

UNIVERSIDADE DO SAGRADO CORAÇÃO

ERICSSON ROBERTO DE BARROS

**DATA MINING: ESTUDO DE TÉCNICAS PARA
ANÁLISE DE DESASTRES NATURAIS**

BAURU
2016

ERICSSON ROBERTO DE BARROS

**DATA MINING: ESTUDO DE TÉCNICAS PARA
ANÁLISE DE DESASTRES NATURAIS**

Trabalho de conclusão de curso apresentado ao Centro de Ciência Exatas e Sociais Aplicadas da Universidade do Sagrado Coração, como parte dos requisitos para obtenção do título de bacharel em Ciência da Computação, sob a orientação do Prof. Me. Patrick Pedreira Silva.

BAURU
2016

Barros, Ericsson Roberto de
B2776d
Data mining: estudo de técnicas para análise de desastres naturais / Ericsson Roberto de Barros. -- 2016.
105f. : il.

Orientador: Prof. Ms. Patrick Pedreira Silva.

Trabalho de Conclusão de Curso (Graduação em Ciência da Computação) – Universidade do Sagrado Coração – Bauru – SP.

1. Desastres naturais. 2. Inteligência artificial. 3. Descoberta de conhecimento. 4. Mineração de dados. 5. Weka. I. Silva, Patrick Pedreira. II. Título.

ERICSSON ROBERTO DE BARROS

**DATA MINING: ESTUDO DE TÉCNICAS PARA ANÁLISE DE
DESASTRES NATURAIS**

Trabalho de Conclusão de Curso apresentado ao Centro de Ciências Exatas e Sociais Aplicadas da Universidade do Sagrado Coração como parte dos requisitos para obtenção do título de Bacharel em Ciência da Computação, sob a orientação do Prof. Me. Patrick Pedreira Silva.

Banca examinadora:

Prof. M.e Patrick Pedreira Silva
Universidade Sagrado Coração

Prof. Dr. Elvio Gilberto da Silva
Universidade Sagrado Coração

Prof. M.e Henrique Pachioni Martins
Universidade Sagrado Coração

Bauru, 29 de novembro de 2016.

AGRADECIMENTOS

Primeiramente agradeço a Deus por ter me abençoado e me dado força em todo este período, para a elaboração deste trabalho. Agradeço a Deus também pelas coisas que aprendi, pelos dias de dificuldades, mas, o senhor sempre esteve ao meu lado, guiando meu coração para a solução dos meus problemas.

Ao meu pai Pedro Roberto de Barros e minha mãe Nilda de Assis Ribeiro de Barros que sempre me incentivaram em toda minha vida acadêmica, ao meu orientador Prof. M.e Patrick Pedreira Silva que tanto me ajudou e incentivou neste trabalho, tendo sempre paciência e confiança ao longo das supervisões das minhas atividades. É um prazer tê-lo como meu orientador.

Quero também agradecer à Universidade do Sagrado Coração com todos seus qualificados profissionais e pessoas acima de tudo.

Agradeço também aos meus colegas de curso, que com certeza, serão ótimos profissionais.

“Seu Trabalho vai preencher uma parte grande da sua vida, e a única maneira de ficar realmente satisfeito é fazer o que você acredita ser um ótimo trabalho. E a única maneira de fazer um excelente trabalho é amar o que você faz.”

(STEVE JOBS)

RESUMO

Atualmente, com o avanço na coleta e armazenamento de dados é comum o acúmulo de grandes informações, muitas vezes não relevantes, sendo necessária a combinação de conhecimentos de diferentes áreas, como a Matemática e a Estatística para a análise desses dados. Esta tarefa pode consumir muito tempo, o que pode inviabilizar sua realização. Por este motivo, o uso de técnicas e ferramentas computacionais é de suma importância para o auxílio na análise de dados, a exemplo da Mineração de Dados. Mineração de Dados (Data Mining) é o processo de exploração de grandes quantidades de dados com o objetivo de encontrar anomalias, padrões e correlações para suportar a tomada de decisões e proporcionar vantagens estratégicas. Com isso, o objetivo desse estudo foi analisar dados sobre desastres naturais e executar algoritmos de Mineração de Dados para gerar algum tipo de conhecimento implícito nos dados. O conhecimento extraído neste trabalho tem como base o estudo de ocorrências referentes a uma base de dados disponibilizados pelo IPMET relatadas entre o período de janeiro de 1986 a janeiro de 2016 contemplando assim 30 anos, relacionada à cidade de Bauru-SP. A partir do estudo realizado foi possível identificar os vários setores problemáticos da cidade de Bauru em relação a desastres naturais, por meio da aplicação de algoritmos de Mineração de Dados implementados pela ferramenta WEKA. Os experimentos apresentam uma importante contribuição em termos de quais aspectos são característicos a um determinado setor da cidade, se as condições estruturais do setor sofrem ou não influência de um determinado fenômeno. Também foi possível notar a importância das etapas do KDD e da Mineração de Dados nesse contexto, pois permitem a realização de estudos detalhados do padrão de fenômenos e ocorrências nos bairros da cidade, além de permitir a avaliação de medidas preventivas para tal problema. A aplicação das técnicas de Mineração de Dados com o objetivo de descobrir novos conhecimentos auxilia no processo de exploração de uma base de dados, permitindo gerar informações úteis e, assim, auxiliar às tomadas de decisões.

Palavras-chave: Desastres naturais. Inteligência Artificial. Descoberta de conhecimento. Mineração de Dados. Weka.

ABSTRACT

Currently, with advances in data collection and storage is common the accumulation of large information, often not relevant, requiring a combination of knowledge from different fields, such as mathematics and statistics for the analysis of such data. This task can be time consuming, which could prevent its realization. For this reason, the use of techniques and computational tools is critical to aid in data analysis, such as the Data Mining. Data Mining (Data Mining) is the process of holding large amounts of data in order to find anomalies, patterns and correlations to support decision-making and provide strategic advantages. Thus, the aim of this study was to analyze data on natural disasters and perform data mining algorithms to generate some kind of implicit knowledge in the data. The knowledge extracted in this paper is based on the study of occurrences relating to a database provided by IPMET reported for the period from January 1986 to January 2016 thus covering 30 years, related to the city of Bauru-SP. From the study it was possible to identify several problematic sectors of the city of Bauru in relation to natural disasters, through the application of Data Mining algorithms implemented by WEKA tool. Experiments have an important contribution in terms of which aspects are characteristic to a certain sector of the city, is the industry's structural conditions suffer or no influence of a particular phenomenon. It was also noted the importance of the steps of the KDD and Data Mining in this context as they allow the realization of standard detailed studies of phenomena and events in the city's neighborhoods, and allows the evaluation of preventive measures for such a problem. The application of data mining techniques in order to discover new knowledge helps in the process of exploitation of a database, allowing generating useful information and thus helping to decision making.

Keywords: Natural disasters. Artificial intelligence. Knowledge discovery. Data Mining. Weka.

LISTA DE ILUSTRAÇÕES

Figura 1 - Fases da Mineração de Dados.....	8
Figura 2 – Técnicas e tarefas da Mineração de Dados.....	9
Figura 3 – Técnicas e tarefas da Mineração de Dados.....	12
Figura 4 – Tipos de chuvas.....	13
Figura 5 - Fases de um raio.....	15
Figura 6 – Tipos de desastres naturais no Brasil (1990-2006).....	17
Figura 7– Distribuição das ocorrências de desastres naturais no Brasil.....	18
Figura 8– Localização Bauru/SP.....	19
Figura 9 – Brasão da cidade de Bauru/SP.....	20
Figura 10– Fachada IPMET Bauru – Radar Banda-S.....	22
Figura 11 - Home Page IPMET.....	24
Figura 12 - Estrutura básica de um SBC.....	29
Figura 13 – SBC e seu Ambiente.....	30
Figura 14 – Software Weka.....	32
Figura 15 – Exemplo de árvore de classificação no Weka.....	33
Figura 16 - Exemplo de regras de associação no Weka.....	34
Figura 17 - Arquivo formato ARFF.....	41
Figura 18 - Etapas da Mineração de Dados.....	42
Figura 19 - Tela inicial Weka.....	43
Figura 20 - Dados carregados no Weka.....	44
Figura 21 - Algoritmo KMeans.....	45
Figura 22 - Configurações do KMeans.....	45
Figura 23 - Número de clusters do KMeans.....	46
Figura 24 - Saída KMeans.....	47
Figura 25 - Kmeans com 4 cluster.....	48
Figura 26 - Saída Kmeans com 4 cluster.....	49
Figura 27 - Passos Mineração - Transformação.....	50
Figura 28 - Base de dados por mês.....	51
Figura 29 - Visualização dos 3 clusters.....	52
Figura 30 - Passos Mineração - Mineração dos dados.....	53
Figura 31 - Saída JRip.....	58
Figura 32 - Regras JRip.....	58
Figura 33 - Classifier algoritmo J48.....	59
Figura 34 - Saída J48.....	60
Figura 35 - Arvore.....	61
Figura 36 - Algoritmo Apriori.....	63
Figura 37 - Saída Apriori.....	64
Figura 38 - Passos Mineração - Conclusão.....	64
Figura 39 - Clima em Bauru.....	66
Figura 40 - Bairros que sofrem com a chuva.....	67
Figura 41 - Setor 1.....	73

Figura 42 - Setor 2	74
Figura 43 - Setor 3	76
Figura 44 - Setor 4	78
Figura 45 - Setor 5	80
Figura 46 - Setor 6	81
Figura 47 - Setor 7	82
Figura 48 - Setor 8	84
Figura 49 - Setor 9	85
Figura 50 - Setor 10	87
Figura 51 - Setor 11	89
Figura 52 - Setor 12	90

LISTA DE TABELAS

Tabela 1 - Algoritmo KMeans	51
Tabela 2 - Classificação das instâncias	54
Tabela 3 - Métricas de Desempenho por classe	55
Tabela 4 - Coeficiente Kappa	56
Tabela 5 - Estatísticas dos erros	56

LISTA DE ABREVIATURAS E SIGLAS

- AM** – Automated Mathematician
- API** - Interface de Programação de Aplicação
- BC** – Base de Conhecimento
- FINEP** – Financiadora de Estudos e Projetos
- GPL** - General Public License
- IA** – Inteligência Artificial
- IPMET** – Instituto de Pesquisas Meteorológicas
- KDD** - knowledge Discovery in Data bases
- MI** – Motor de Inferência
- SBC** – Sistemas Baseado em Conhecimento
- SIPMET** – Sistema Paulista de Meteorologia
- UNESP** – Universidade Estadual Paulista
- WEKA** - Waikato Environment for Knowledge Analysis

SUMÁRIO

1	INTRODUÇÃO	8
2	OBJETIVOS	11
2.1	OBJETIVO GERAL	11
2.2	OBJETIVOS ESPECÍFICOS	11
3	TEORIA E ANÁLISE DE DESASTRES NATURAIS	12
3.1	CHUVA	13
3.2	RAIO	15
3.3	GRANIZO	16
3.4	DESASTRES NATURAIS NO BRASIL	17
3.4.1	PREVENÇÃO	18
3.5	ESTUDO DE CASO: BAURU	19
3.5.1	Economia	20
3.5.2	Clima	20
3.5.3	Ipmet	21
3.5.3.1	<i>Utilização das informações geradas</i>	22
4	INTELIGÊNCIA ARTIFICIAL	24
4.1	HISTÓRICO	25
4.2	APRENDIZADO DE MÁQUINA	26
4.3	AQUISIÇÃO DE CONHECIMENTO	28
4.3.1	Sistemas baseados em conhecimento	28
4.3.2	Processo de aquisição de conhecimento	29
4.3.3	Base de conhecimento	30
4.4	WEKA	31
4.5	ALGORITMOS DE APRENDIZAGEM SUPERVISIONADA	33
4.6	ALGORITMOS DE APRENDIZAGEM NÃO SUPERVISIONADA	34
4.7	LIMPEZA DOS DADOS	35
4.7.1	Valores ausentes	35
4.7.2	Valores fora de padrão	36
4.7.3	Dados inconsistentes	36
5	TRABALHOS CORRELATOS	38
6	METODOLOGIA	39
6.1	FERRAMENTA UTILIZADA - WEKA	39

6.2	BASE DE DADOS	40
6.3	ETAPAS DA MINERAÇÃO	42
6.3.1	Seleção e Pré-processamento	42
6.3.1.1	<i>Cluster – SimpleKMeans (5 clusters)</i>	44
6.3.1.2	<i>Cluster – SimpleKMeans (4 clusters)</i>	47
6.3.2	Transformação dos dados	49
7	RESULTADOS	53
7.1	SELEÇÃO DOS ALGORITMOS	53
7.2	ALGORITMO: JRIP	57
7.3	ALGORITMO: J48	59
7.4	ALGORITMO: REGRAS DE ASSOCIAÇÃO	62
7.5	INTERPRETAÇÃO E AVALIAÇÃO.....	64
8	CONCLUSÃO	68
	REFERÊNCIAS	69
	ANEXOS	72
	ANEXO A	72

1 INTRODUÇÃO

Inúmeras pesquisas têm sido conduzidas para o desenvolvimento de técnicas com objetivo de extrair informações a partir de um grande volume de dados e transformar estas informações em conhecimento útil. Esta área é conhecida na literatura como KDD (*Knowledge Discovery in Data bases*) (FAYYAD et al., 1996).

De acordo com FAYYAD et al., (1996), o KDD é um processo, de várias etapas, não trivial, interativo e iterativo, para identificação de padrões compreensíveis, válidos, novos e potencialmente úteis a partir de grandes conjuntos de dados.

Esta técnica pode identificar padrões e descobrir informações relevantes que auxiliam, por exemplo, nas estratégias de marketing, no comportamento de clientes em relação às compras, entre outras coisas.

A fase de Mineração de Dados é onde realmente se extrai as informações através de algoritmos que executam uma determinada tarefa, conseqüentemente, gerando um padrão a partir de instâncias de uma base de dados.

Figura 1 - Fases da Mineração de Dados



Fonte: Elaborada pelo autor.

Estas fases (Figura 1) são cíclicas e têm como objetivo dar uma direção a todo o estudo em uma base de dados que seja candidata à Mineração. Para garantir o sucesso dessas fases da Mineração de Dados, devem ser seguidas algumas técnicas e tarefas. Na Figura 2 estão citadas as mais conhecidas e utilizadas.

Figura 2 – Técnicas e tarefas da Mineração de Dados

Técnica	Tarefas	Algoritmos
Descoberta de Regras de Associação	Associação	Apriori, AprioriTid, AprioriHybrid, AIS, SETM (Agrawal e Srikant, 1994) e DHP (Chen <i>et al.</i> , 1996).
Árvores de Decisão	Classificação Regressão	CART, CHAID, C5.0, Quest (Two Crows, 1999); ID-3 (Chen <i>et al.</i> , 1996); SLIQ (Metha <i>et al.</i> , 1996); SPRINT (Shafer <i>et al.</i> , 1996).
Raciocínio Baseado em Casos ou MBR	Classificação Segmentação	BIRCH (Zhang <i>et al.</i> , 1996); CLARANS (Chen <i>et al.</i> , 1996); CLIQUE (Agrawal <i>et al.</i> , 1998).
Algoritmos Genéticos	Classificação Segmentação	Algoritmo Genético Simples (Goldberg, 1989); Genitor, CHC (Whitley, 1993); Algoritmo de Hillis (Hillis, 1997); GA-Nuggets (Freitas, 1999); GA-PVMINER (Araújo <i>et al.</i> , 1999).
Redes Neurais Artificiais	Classificação Segmentação	Perceptron, Rede MLP, Redes de Kohonen, Rede Hopfield, Rede BAM, Redes ART, Rede IAC, Rede LVQ, Rede Counterpropagation, Rede RBF, Rede PNN, Rede Time Delay, Neocognitron, Rede BSB (Azevedo, 2000), (Braga <i>et al.</i> , 2000), (Haykin, 2001)

Fonte: Elaborada pelo autor.

Para determinar qual tarefa a ser resolvida, deve-se ter um bom conhecimento do domínio da aplicação e saber o tipo de informação que se quer obter (FAYYAD, 1996).

O uso da Mineração de Dados está se expandindo a cada dia. Hoje existe uma gama de ferramentas no mercado: Weka, Linguagem R, Endeca, Knime, dentre outras.

Os usos dessas ferramentas permitem tratar vários problemas de descoberta de conhecimento com muito mais eficiência já que todo o processo é automatizado. Um dos problemas que podem ser estudados com auxílio da Mineração de Dados é aqueles relacionados aos desastres naturais. Em muitas situações existem diversas bases de dados que guardam informações sobre esses fenômenos e, em alguns casos, pouco se conhece sobre suas causas e sua frequência. Neste contexto esta investigação tem como principal objetivo promover a descoberta de conhecimento em uma base de dados sobre desastres naturais ocorridos na cidade de Bauru-SP, utilizando técnicas de Mineração de Dados.

O processo de Mineração de Dados auxiliará a geração automática de conhecimento sobre a base de dados, permitindo verificar o potencial de utilização

desses métodos na tarefa proposta. Adicionalmente, poderão ser geradas informações relevantes para Instituições Governamentais que poderão ser utilizadas com o intuito de diminuir ou sanar esses problemas, auxiliando a tomada de decisão e planejamentos futuros.

2 OBJETIVOS

Apresenta-se nos tópicos a seguir o objetivo geral e os objetivos específicos da pesquisa.

2.1 OBJETIVO GERAL

Promover a descoberta de conhecimento em uma base de dados sobre desastres naturais, utilizando técnicas de Mineração de Dados.

2.2 OBJETIVOS ESPECÍFICOS

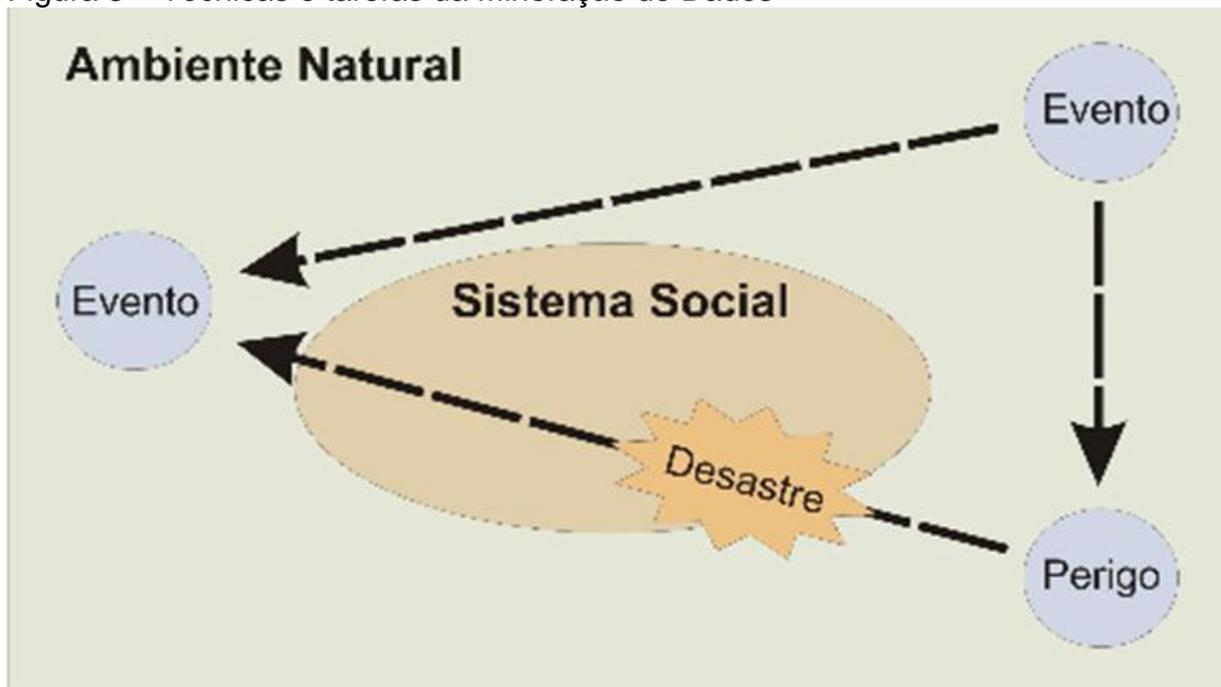
- a) Pesquisar sobre o processo de KDD (*Knowledge Discovery in Data bases*) e Mineração de Dados;
- b) Selecionar e adaptar uma base de dados sobre desastres naturais para exploração;
- c) Definir os algoritmos de Mineração de Dados a serem explorados;
- d) Selecionar uma ferramenta de Mineração de Dados;
- e) Aplicar os algoritmos sobre a base de dados;
- f) Coletar e analisar os resultados, associando-os à geração de conhecimento;

3 TEORIA E ANÁLISE DE DESASTRES NATURAIS

Furacões, inundações, escorregamentos, secas, entre outros, são fenômenos naturais severos, diretamente influenciados pelas características regionais, como topografia, solo, vegetação, condições meteorológicas. Quando estes fenômenos ocorrem em locais onde os seres humanos vivem, os danos e prejuízos causados são considerados como “desastres naturais” (CASTRO, 1998).

Segundo Castro (1998), desastre é definido como o resultado de eventos adversos, naturais ou provocados pelo homem, sobre um ecossistema (vulnerável), causando danos humanos, materiais e/ou ambientais e consequentes prejuízos econômicos e sociais, conforme mostra a Figura 3.

Figura 3 – Técnicas e tarefas da Mineração de Dados



Fonte: Elaborada pelo autor.

Desastre está relacionado com eventos naturais ou através das consequências dos eventos que possam conter riscos causados pelo homem, ocasionando, assim, danos e prejuízos ao ambiente.

Os desastres são normalmente inesperados e súbitos, de uma magnitude e gravidade capaz de produzir danos e prejuízos diversos, resultando em mortos e feridos. Portanto, exigem ações preventivas e restituidoras, que envolvem diversos

setores governamentais e privados, visando uma recuperação que não pode ser alcançada por meio de procedimentos rotineiros (KOBAYAMA, et al. , 2006).

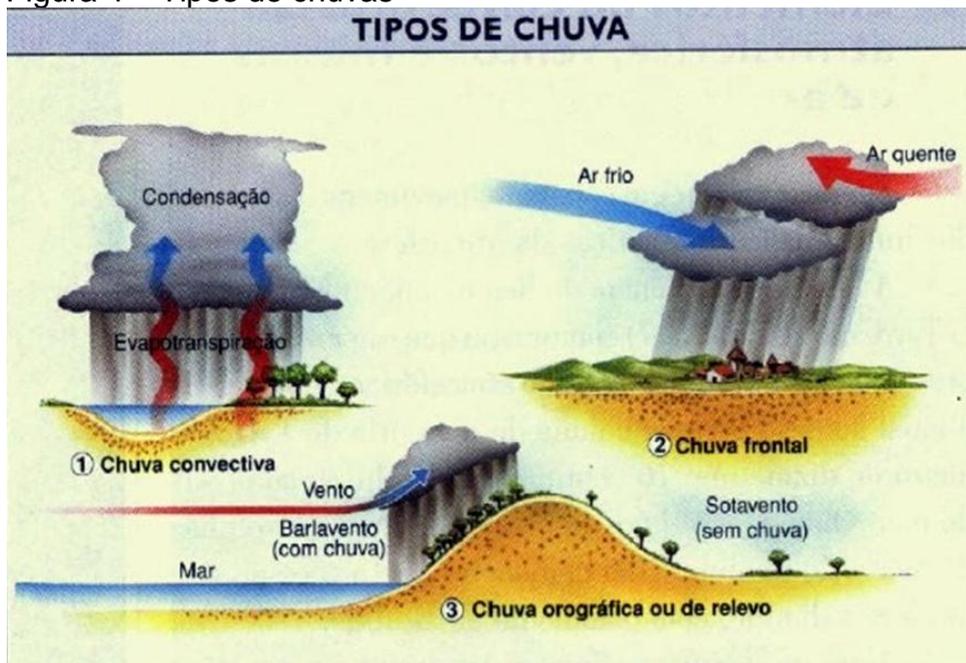
Segundo White (1974), há cinco passos para o estudo de desastres naturais: estimar a área ocupada pelo ser humano nas áreas de perigo, determinar a faixa de ajuste possível contra eventos extremos, examinar como a população percebe os desastres naturais, examinar os processos de seleção de medidas adequadas e estimar os efeitos da política sobre essas medidas. Hewitt (1983) adicionou mais um passo, que seria entender como aspectos socioeconômicos da sociedade contribuem à geração de desastres.

3.1 CHUVA

Segundo Digo (2009), chuva é um fenômeno meteorológico que consiste na precipitação de água no estado líquido sobre a superfície da Terra. A chuva forma-se nas nuvens. Nem todas as chuvas atingem o solo, algumas se evaporam enquanto estão ainda a cair, num fenômeno que recebe o nome de virga e acontece, principalmente, em períodos/locais de ar seco.

São três os tipos de chuvas na natureza, que são apresentados na Figura 4.

Figura 4 – Tipos de chuvas



Fonte: BIBLIOTECA AMBIENTAL. Precipitação pluviométrica - Chuvas. Belo Horizonte MG, 2011.

- a) Chuvas Convectivas – Provocadas por diferenças de locais de aquecimento nas camadas atmosféricas. São chuvas formadas pela ascensão das massas de ar quente da superfície, carregadas de vapor d'água. Ao subir o ar sofre resfriamento provocando a condensação do vapor de água presente e, conseqüentemente, a precipitação. São características deste tipo de precipitação as chuvas de curta duração, alta intensidade, trovoadas, rajadas de ventos e pela sua abrangência em pequenas áreas;
- b) Chuvas Orográficas – Chuvas que são oriundas da passagem de uma massa de ar quente e úmido por uma cadeia de montanha, provocando a ascensão forçada do ar, que gradativamente se esfria provocando a condensação do vapor de água e conseqüentemente a formação de nuvens que se precipitam. Caracterizam-se pela sua longa duração e baixa intensidade e por não apresentarem qualquer tipo de descarga elétrica (MIRANDA, 2011).
- c) Chuvas Frontais – são originárias do deslocamento de frentes frias ou quentes contra frentes contrárias termicamente. Ou seja, são chuvas que ocorrem ao longo da linha de descontinuidade, separando uma massa de ar de características diferentes. Ou seja, quando uma massa de ar quente e úmida, estacionária, ou oriunda do quadrante norte encontra com uma massa de ar frio, oriunda do quadrante sul. A frente fria, mais densa, entra por baixo, levando para cima a massa de ar quente. Quando esta massa de ar quente possui elevada umidade relativa, a chuva é iminente. É uma chuva de menor intensidade, com pingos menores, e de longa duração. É uma chuva de menor intensidade, com pingos menores, e de longa duração. Ocorre por vários dias, apresentando pausas e chuviscos entre fases mais intensas. Na metade sudeste do continente, ocorrer em qualquer época do ano, mas tem maior duração nos meses frios, quando os fenômenos atmosféricos são menos intensos. Pode produzir ventos fortes e grande quantidade de raios. Ocorre em uma imensa área simultaneamente (MIRANDA, 2011).

3.2 RAI0

Raio é o caminho luminoso da descarga atmosférica, às vezes utilizada, mais formalmente, para designar uma descarga para o solo (MENDES; DOMINGUES, 2002).

Um raio começa com pequenas descargas dentro da nuvem. Estas descargas liberam os elétrons que começarão seu caminho de descida em direção ao solo. Esse caminho de descida é tortuoso e truncado em passos de 50 metros, como que buscando o caminho mais fácil. Esta busca de uma conexão com a terra é muito rápida e pouco luminosa para ser visto a olho nu (SABA, 2001).

A Figura 5 mostra uma descarga elétrica liberada por uma nuvem de tempestade que é conhecida como líder escalonado. Quando o líder escalonado está a certa altura do solo, ele atrai outra descarga elétrica de carga oposta, chamada de descarga conectante. A união dessas duas cargas dá origem ao canal do raio, caminho esse que por ser ionizado é altamente condutor. A descarga de retorno, neste mesmo instante, dá-se então origem a uma grande potencia onde muita luz é liberada, temos então um raio (SABA, 2001).

Figura 5 - Fases de um raio



Fonte: EARTH SCIENCE OFFICE. NASA – lightning, 2001.

3.3 GRANIZO

O granizo é a precipitação sólida de partículas de gelo, transparentes ou translúcidas, com formato geralmente arredondados ou irregulares, apresentando diâmetro igual ou superior a 5 milímetros. O granizo é formado em nuvens de grande desenvolvimento vertical, que podem atingir alturas de até 1600 metros, denominadas de cumulonimbus. Essas nuvens são responsáveis por gerarem tempestades com correntes ascendentes e descendentes extremamente intensas, propiciando ambientes favoráveis para ocorrência de precipitação de granizo e chuvas fortes (CASTRO, 2003).

As pedras de granizo que caem do topo das nuvens são cobertas por uma camada de umidade que congela quando são levadas novamente ao topo das nuvens pelas correntes ascendentes, aumentando de tamanho. Esse processo se repete inúmeras vezes no interior das nuvens. Em virtude da união das gotas congeladas, as pedras de granizo crescem rapidamente até tornarem-se muito pesadas para serem suportadas pelas correntes ascendentes, logo precipitam em forma de granizo (MARCELINO et al., 2006).

Segundo Cunha (2001), a precipitação de granizo ocorre durante um tempo menor se comparado à precipitação pluvial, porém associados, em uma mesma tempestade severa. Pois apenas uma parte da nuvem de tempestade possui características e condições adequadas para formar o granizo. Logo, a precipitação de granizo cai durante alguns minutos, enquanto a tempestade continua em forma de precipitação pluvial.

Uma saraivada (chuva de granizo) pode atingir distâncias de até 100 quilômetros, em uma faixa de 16 quilômetros de largura quando as nuvens de tempestade estão em movimento. Pedras de granizo com tamanho elevado podem cair com velocidade de até 200 quilômetros por hora, com massa superior a meio quilograma (CUNHA et al., 2001).

Os locais mais atingidos por granizo no Brasil concentram-se nas regiões Sul, Sudeste e parte meridional do Centro – Oeste, mais especificamente nas áreas de planalto do Paraná, Santa Catarina e Rio Grande do Sul (CASTRO, 2003).

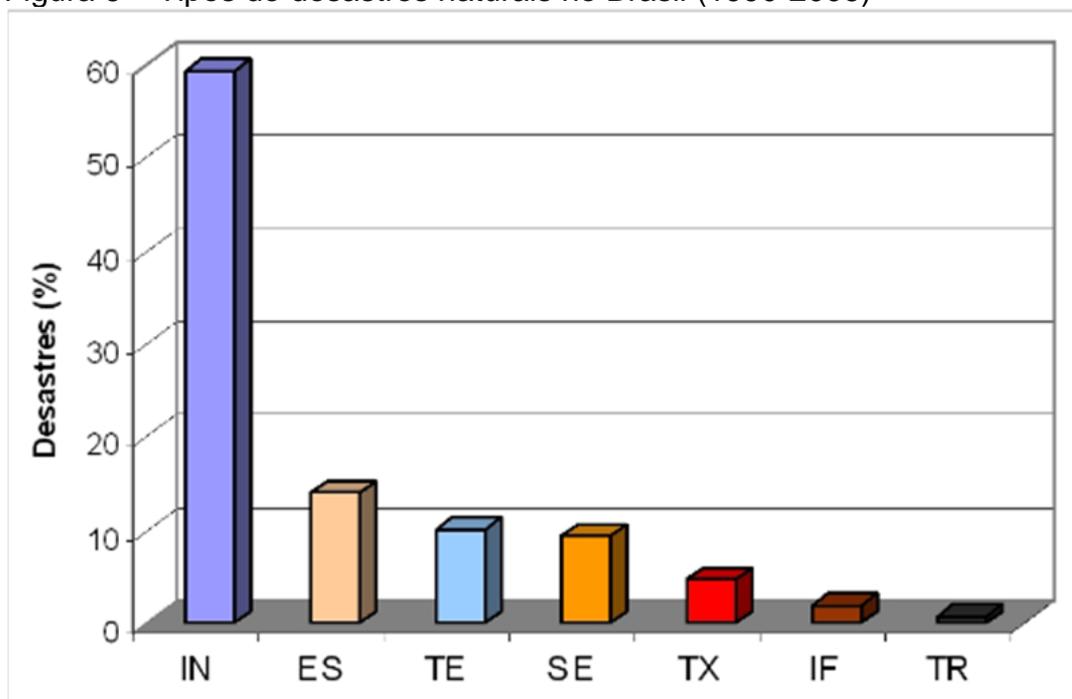
3.4 DESASTRES NATURAIS NO BRASIL

O Brasil sofreu mais de 30 mil desastres naturais entre os anos de 1988 e 2010, o que dá uma média de 1.363 catástrofes por ano, de acordo com o levantamento do Ministério da Integração Nacional (GORZONI, 2010).

As calamidades que acontecem no Brasil como, por exemplo, enchentes e, por consequência, escorregamentos, seca, queimadas, geada, são causadas por um conjunto de fatores não só climáticos: o dedo do homem é o fator decisivo nesse prejuízo. Como, por exemplo, levantar casas em beiras de rios, em encostas de morros ou mesmo construir hidrelétricas sem preocupação com a manutenção adequada. Por isso, pode-se dizer que há tempos as coisas aconteciam por razões geológicas, mas hoje elas se dão em boa parte pelas mãos humanas.

Os tipos de desastres mais frequentes foram as inundações como é demonstrada na última pesquisa levantada pelo Instituto Nacional de Pesquisas Espaciais (INPE) (Figura 6), representadas pelas graduais e bruscas, com 59% dos registros, seguidas pelos escorregamentos (14%).

Figura 6 – Tipos de desastres naturais no Brasil (1990-2006)

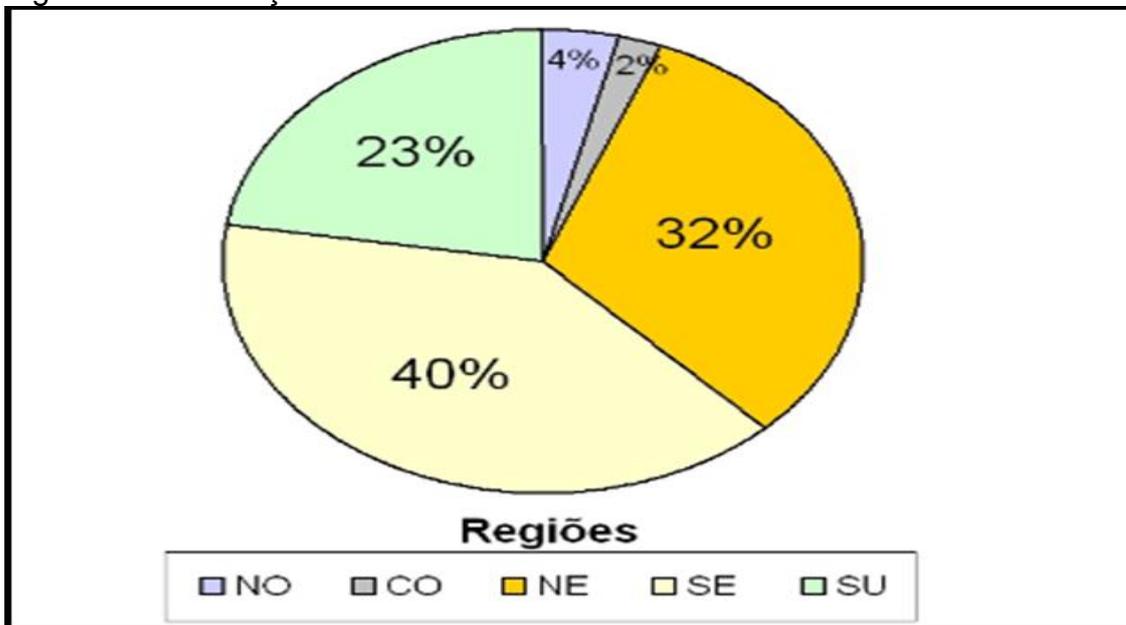


Legenda: IN – Inundação, ES – Escorregamento, TE – Tempestades, SE – Seca, TX – Temperatura Extrema, IF – Incêndio Florestal e TR – Terremoto.

Fonte: MINISTÉRIO DA CIÊNCIA E TECNOLOGIA. Instituto Nacional de Pesquisas Espaciais - Desastres Naturais e Geotecnologias: Conceitos básicos. INPE Santa Maria, 2007.

Com relação à distribuição espacial, mais de 60% dos casos ocorreram nas regiões Sudeste e Sul (Figura 7). No Brasil, essa distribuição está mais associada às características geoambientais do que as socioeconômicas das regiões afetadas. Nessas regiões as instabilidades atmosféricas são frequentes devido à passagem de frentes frias no inverno, da ocorrência de complexos convectivos de mesoescala na primavera e da formação dos sistemas convectivos no verão, que desencadeiam as chuvas intensas e concentradas para essa estação (MONTEIRO; FURTADO, 1995; SANT' ANNA NETO, 1995; SILVA DIAS, 1996; MARCELINO, 2003).

Figura 7– Distribuição das ocorrências de desastres naturais no Brasil



Legenda: NO – Norte, CO – Centro Oeste, NE – Nordeste, SE – Sudeste, SU – Sul.

Fonte: MINISTÉRIO DA CIÊNCIA E TECNOLOGIA. Instituto Nacional de Pesquisas Espaciais - INPE Santa Maria, 2007.

3.4.1 PREVENÇÃO

Os desastres naturais que ocorrem no Brasil, em sua maioria, não podem ser evitados. Porém, as intervenções e atividades humanas podem agravar ou minimizar os impactos causados pelos desastres. O agravamento está diretamente associado ao desrespeito à natureza, através dos desmatamentos, queimadas, assoreamento de rios, acúmulo de lixo, ocupação desordenada, edificações mal construídas, ausência de plano diretor, etc.

Para que ocorra a redução do impacto dos desastres ou a minimização, é preciso que sejam tomadas medidas preventivas que podem ser classificadas estruturais e não estruturais. As medidas estruturais são aquelas de cunho corretivo, como as obras de engenharia. Apesar de minimizar o problema em curto prazo, as medidas estruturais são caras, paliativas e frequentemente ocasionam outros impactos ambientais e geram uma falsa sensação de segurança. As não estruturais, de caráter educativo, apesar dos resultados a médio e longo prazo, são de baixo custo, de fácil implementação, e permitem uma correta percepção do risco. Como exemplo, destacam-se os mapeamentos, as análises de vulnerabilidade, os zoneamentos das áreas de risco e a educação ambiental (NCEM, 1998; JELKOVIC, 2001; ISDR, 2004).

3.5 ESTUDO DE CASO: BAURU

Fundado em 1896, a uma altitude de 526 m, Bauru é um município brasileiro situado no estado de São Paulo (Figura 8) (SANTOS, 2011).

Figura 8– Localização Bauru/SP



Fonte: UNIVERSIDADE ESTADUAL PAULISTA. Faculdade de Ciências – Histórico. Bauru, 2010.

Com um território de 673,5 km², onde vivem 366.992 habitantes, o lema da cidade, presente em seu brasão é a frase “Custos vigilat” (Figura 9), que em português significa “Sentinela alerta” (SANTOS, 2011).

Figura 9 – Brasão da cidade de Bauru/SP



Fonte: VIVENDO BAURU. Sobre Bauru. Bauru, 2011.

3.5.1 Economia

Por ter facilidades de transporte rodoferroviário e uma relativa infertilidade de suas terras, o setor de serviços e comércios são as principais atividades econômicas do município (SANTOS, 2011).

A agricultura é incipiente, focando-se no cultivo de abacaxi e frutas tropicais. A pecuária sempre esteve presente no município e a cana-de-açúcar ganhou espaço nos últimos anos, com a instalação de diversas usinas no interior paulista (SANTOS, 2011).

O setor industrial é fortemente representado por indústrias de transformação, metalomecânica e alimentícias. Nos setores gráfico e alimentício, Bauru possui empresas líderes nacionais de seus setores, com grande volume de exportação e comércio interno (SANTOS, 2011).

3.5.2 Clima

O clima de Bauru é o tropical de altitude, com verão quente e chuvoso e inverno ameno e seco. A temperatura média anual fica em torno dos 22C^o com índice pluviométrico de 1500 mm (SANTOS, 2011).

No município, a mata atlântica predomina na vegetação. A média de temperatura anual de Bauru fica em 22,6 graus Celsius.

O mês mais quente, março, conta com temperatura média de 25,0 °C, sendo a média máxima de 31,0 °C e a mínima de 19,0 °C. E o mês mais frio, julho, com média de 19,0 °C, sendo 26,0 °C e 12,0 °C as médias máxima e mínima, respectivamente. Outono e primavera são estações de transição (Centro de pesquisas meteorológicas e climáticas aplicadas à agricultura, 2012).

3.5.3 Ipmet

O Instituto de Pesquisas Meteorológicas (IPMET) é um instituto de meteorologia localizado em Bauru, no interior do estado de São Paulo, e uma unidade complementar da Universidade Estadual Paulista (UNESP). Foi criado com o intuito de realizar pesquisas meteorológicas voltadas à utilização em previsão do tempo para o estado de São Paulo, além de quantificar e monitorar tempestades e a pluviosidade nessa região, através do uso de um radar meteorológico.

O IPMET (Figura 10) tem sua origem no antigo instituto de pesquisas criado em 1969 pela fundação Educacional de Bauru, que mais tarde iria se tornar a Universidade de Bauru (posteriormente absorvida pela UNESP), com o objetivo de subsidiar as diversas áreas de ensino com ênfase na meteorologia. Em 1972 a instituição passou a ser denominada pelo nome atual de Instituto de pesquisa Meteorológicas. Em 1988, o instituto foi incorporado à UNESP (UNESP. IPMET Faculdade de Ciências - Centro de Meteorologia de Bauru. Bauru, 2014).

Figura 10 – Fachada IPMET Bauru – Radar Banda-S



Fonte: UNIVERSIDADE ESTADUAL PAULISTA. IPMet
Faculdade de Ciências - Centro de Meteorologia de Bauru, 2014.

No ano seguinte, 1989, o IPMET participou da elaboração do projeto do SIPMET, o Sistema Paulista de Meteorologia, cujo objetivo principal era o de prover o Estado de São Paulo com informações meteorológicas confiáveis e passíveis de serem disseminadas aos usuários em tempo hábil para a tomada de providências pelas autoridades competentes, no sentido de proteção à vida, no caso de ocorrências de eventos severos (UNESP. IPMET Faculdade de Ciências - Centro de Meteorologia de Bauru. Bauru, 2014).

Em 1992, o antigo radar, em funcionamento por dezoito anos, é substituído por um equipamento mais moderno, um modelo banda-S, com efeito, Doppler. Esse equipamento, além da quantificação das chuvas, passa a possibilitar a medição de campos de vento e, em consequência, o oferecimento de dados do deslocamento dessas chuvas. A Financiadora de Estudos e Projetos (FINEP), mais uma vez, destinou os recursos necessários ao empreendimento (UNESP. IPMET Faculdade de Ciências - Centro de Meteorologia de Bauru. Bauru, 2014).

3.5.3.1 *Utilização das informações geradas*

Os benefícios adquiridos pelas informações meteorológicas atingem praticamente todas as áreas da atividade humana, dentre as mais importantes podemos citar:

- a) Defesas civis

- b) Agricultura
- c) Transporte (aéreo, naval ou rodoviário)
- d) Segurança pública
- e) Construção civil
- f) Turismo
- g) Geração e distribuição de energia
- h) Recursos hídricos
- i) Imprensa
- j) Saúde
- k) Seguradoras
- l) População em geral

A parceria com as defesas civis para acesso a imagens de radar em tempo real, e da emissão de alertas quando da aproximação de tempestades severas, por exemplo, é de fundamental importância para a tomada de medidas visando a salvaguarda de vidas, com a retirada de populações de áreas de risco e proibição de acesso a áreas sujeitas a escorregamentos ou alagamentos (UNESP. IPMET Faculdade de Ciências - Centro de Meteorologia de Bauru. Bauru, 2014).

A população em geral também se beneficia com essas informações já que pode contar com a emissão de boletins atualizados a cada hora, ou menos, no caso da ocorrência de chuvas na área de alcance dos radares. Esses boletins são disponibilizados na homepage da instituição (Figura 11).

Figura 11 - Home Page IPMET

The image shows the home page of the IPMET website. At the top, there is the UNESP logo and the name of the university, 'UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO" Câmpus de Bauru'. Below this, the page is titled 'IPMet - Faculdade de Ciências'. On the left, there is a vertical navigation menu with various options such as 'Instituição', 'Cadastro', 'Radar Cidades', 'Imagem do Radar', 'Radar PPI GIS', 'Imagem do Satélite', 'Boletim do Tempo', 'Alertas no Twitter', 'Previsão Numérica', 'Estação Meteorológica', 'Chuva Acumulada', 'Chuva Diária', 'Banco de Dados', 'Cooperação IPMet e CPTEC', 'Serviços', 'Observador Voluntário', 'Estações do Ano', 'Pesquisas/Projetos', 'Saiba Mais', 'Sítios Relacionados', 'Fale Conosco', and 'Como Chegar'. The main content area is divided into several sections. The top section features a weather radar image of São Paulo state with the headline 'Boletim do Tempo' and 'Chuvvas isoladas no estado de São Paulo.' Below this, there is a section titled 'IPMet conta com nova supervisão' with the names of the supervisors, 'Assumem: Rodolfo Langhi e Jeferson P. Machado'. To the right of this, there is a section titled 'Chuva acumulada em fevereiro chega a 351,3mm de chuva' with the text 'É o mais chuvoso dos últimos dezesseis anos'. Below these sections, there is a search bar with the text 'Buscar Todas' and 'Pesquisar'. The right sidebar contains a login form with fields for 'Usuário:' and 'Senha:', and buttons for 'Entrar' and 'Limpar'. Below the login form, there is contact information for the 'Estação Meteorológica Bauru' and a section titled 'Previsão do Tempo e Suporte Meteorológico' with phone numbers. At the bottom of the page, there are several promotional banners, including one for 'LAUDOS METEOROLÓGICOS' and another for 'OPERAÇÃO VERÃO' with the text 'Dicas da Defesa Civil' and 'Deseja agendar uma visita?'.

Fonte: UNIVERSIDADE ESTADUAL PAULISTA. IPMet Faculdade de Ciências – Centro de Meteorologia de Bauru. Bauru, 2016.

A seguir será demonstrada a tecnologia utilizada para a descoberta de conhecimento e padrões de acordo com a base de dados adquiridas do IPMET. Primeiramente com um breve histórico e técnicas necessárias para assim poder obter o resultado esperado com esses dados.

4 INTELIGÊNCIA ARTIFICIAL

A Inteligência Artificial (IA) pode ser definida como o conjunto de problemas e metodologias estudadas por seus pesquisadores, sendo que um sistema inteligente seria aquele que tem a capacidade de raciocinar, planejar, resolver problemas, armazenar conhecimento, comunicar-se através de uma linguagem e aprender.

A Inteligência Artificial não tem uma definição simples, até o momento, simplesmente é definida como a coleção de problemas e metodologias estudada pelos pesquisadores de Inteligência Artificial. Pode parecer sem sentido essa definição, mas ela reforça um argumento importante: a Inteligência Artificial, como

toda ciência, é um empreendimento humano e talvez seja mais bem entendida nesse contexto (LUGER et al., 2004).

Para Alan Turing (1950), o pensamento é uma atividade interior muito especial, sendo impossível descrever cientificamente. Com isso, Turing chegou à conclusão que para um computador ser considerado inteligente, ele deveria primeiramente ser submetido a um teste prático, onde um interrogador humano, depois de propor algumas perguntas por escrito, não conseguir descobrir se as respostas escritas vêm de uma pessoa ou não (NAVEGA, 2002).

Contudo, nenhum computador conseguiu passar no teste de Turing, apesar do seu apelo intuitivo, é vulnerável a várias críticas justificáveis. Uma das mais importantes se refere ao seu viés direcionado para as tarefas de solução de problemas puramente simbólicos. Ele não testa as habilidades necessárias para a percepção ou para a destreza manual, muito embora elas sejam componentes importantes da inteligência humana (LUGER et al., 2004).

4.1 HISTÓRICO

Embora nos séculos XVIII, XIX e no início do século XX a formulação da ciência e da matemática tenha criado o pré-requisito intelectual para o estudo da Inteligência Artificial, somente com a introdução do computador digital no século XX é que a IA se tornou uma disciplina cientificamente viável. No final dos anos 1940, os computadores eletrônicos digitais demonstraram seu potencial para disponibilizar memória e poder de processamento requerido pelos programas inteligentes. Com isso, era possível agora, programar sistemas de raciocínio formal em um computador e testar empiricamente a sua aptidão para exibir inteligência (LUGER et al., 2004).

A partir da década de 1970, houve uma maior disseminação do uso de técnicas de computação baseadas em IA para a solução de problemas reais. Eram criados programas de computador que através de aquisição de conhecimento de especialistas de um dado domínio, por exemplo, do domínio da medicina, eram então codificados, quase sempre por regras lógicas. Esses programas eram conhecidos como Sistemas Especialistas ou Sistemas Baseados em Conhecimento (FACELI et al., 2011).

Eram realizadas entrevistas com os especialistas para a aquisição do conhecimento, e assim descobrir que regras eles utilizam quando da tomada de decisão. Muitas vezes, faltava pouca cooperação por parte do especialista, por causa do seu receio de ser dispensado após repassar o conhecimento solicitado (FACELI et al., 2011).

Com a crescente complexidade dos problemas a serem tratados computacionalmente, e do elevado volume de dados gerados por diferentes setores, tornou-se óbvio a necessidade de ferramentas computacionais mais sofisticadas, que fossem mais autônomas, reduzindo a necessidade de intervenção humana e dependência de especialistas. Mas para isso, essas técnicas deveriam ser capazes de criar por si próprias, a partir da experiência passada, uma hipótese ou função capaz de resolver o problema que se deseja tratar. A esse processo de indução de uma hipótese a partir da experiência passada dá-se o nome Aprendizado de Máquina (FACELI et al., 2011).

Frequentemente, é esquecida que as ferramentas que são criadas para os diversos propósitos tendem a moldar a concepção que as pessoas têm do mundo a partir de sua estrutura e suas limitações. Embora aparentemente restritiva, essa interação é um aspecto essencial da evolução do conhecimento humano: uma ferramenta (e as teorias científicas são, no final das contas, apenas ferramentas) é desenvolvida para solucionar um problema em particular. Conforme essa ferramenta é usada e refinada, ela própria parece sugerir outras aplicações, levando a outras questões e, por fim, ao desenvolvimento de novas ferramentas (LUGER et al., 2004).

4.2 APRENDIZADO DE MÁQUINA

Existem várias definições de aprendizado de máquina na literatura. A capacidade de melhorar o desempenho na realização de alguma tarefa por meio da experiência, apresentada em Mitchell (1997), é uma delas.

Em aprendizado de máquina, computadores são programados para aprender com a experiência passada. Para isso, empregam um princípio de inferência denominado indução, no qual se obtêm conclusões genéricas a partir de um

conjunto particular de exemplos. Assim, algoritmos de aprendizado de máquina aprendem a induzir uma função ou hipótese capaz de resolver um problema a partir de dados que representam instancias do problema a ser resolvido (FACELI et al., 2011).

Embora o aprendizado de máquina seja uma área difícil, há diversos programas que sugerem que isso não é impossível, ao contrário, auxiliam e nos permite interpretar dados em busca de resultados satisfatórios.

Um programa pioneiro é o AM (Automated Mathematician – matemático automatizado), criado para descobrir leis matemáticas (LENAT, 1977). A partir de conceitos e axiomas da teoria dos conjuntos inicialmente fornecidos, o AM foi capaz de induzir conceitos matemáticos importantes como cardinalidade, aritmética inteira e muitos resultados da teoria dos números. O AM formulou novos teoremas a partir da modificação da sua base de conhecimento atual e usou heurísticas para encontrar o “melhor” teorema entre uma série de teoremas possíveis (LUGER et al., 2004).

Mais recentemente, Cotton (2000) concebeu um programa que inventa automaticamente sequências “interessantes” de inteiros.

Existem várias aplicações bem-sucedidas de técnicas de aprendizado de máquinas na solução de problemas reais, dentre as quais podem ser citadas:

- a) Reconhecimento de palavras faladas;
- b) Predição de taxas de cura de pacientes com diferentes doenças;
- c) Detecção do uso fraudulento de cartões de crédito;
- d) Condução de automóveis de forma autônoma em rodovias;
- e) Ferramentas que jogam gamão e xadrez de forma semelhante a campeões;
- f) Diagnóstico de câncer por meio de análise de dados de expressão gênica.

O desenvolvimento de algoritmos cada vez mais eficazes e eficientes e a elevada capacidade dos recursos computacionais têm favorecido a expansão da área de aprendizado de máquina.

Outras motivações para as pesquisas em aprendizado de máquina incluem a possibilidade de aumentar a compreensão de como se dá o aprendizado nos seres vivos. Além disso, algumas tarefas são naturalmente mais bem definidas por meio de exemplos.

4.3 AQUISIÇÃO DE CONHECIMENTO

Existe uma estimativa de que a cada 20 meses dobra-se a quantidade de dados armazenada nos bancos de dados do mundo (WITTEN et al., 2011). Esses dados são gerados por transações financeiras, monitoramento ambiental, obtenção de dados clínicos e genéticos, captura de imagens, navegação na internet, apenas para citar algumas atividades. Com o aumento crescente da quantidade de dados gerada, tem aumentado muito a distância entre a quantidade de dados existentes e a porção desses dados que é analisada e compreendida (FACELI et al., 2011).

A partir da década de 1980 empresas começaram a perceber o grande crescimento de dados armazenados nos computadores, e que não tinham utilidades para elas, assim, teve início ao estudo da Mineração de Dados (data mining). Era necessária uma forma automática de avaliação de todos os dados pertinentes, para poder classificá-los e, conseqüentemente, descartar o que não fosse necessário armazenar, economizando espaço de armazenamento, e conseqüentemente, dinheiro (AFONSO; GUEDES; MAGALHÃES, 2010).

4.3.1 Sistemas baseados em conhecimento

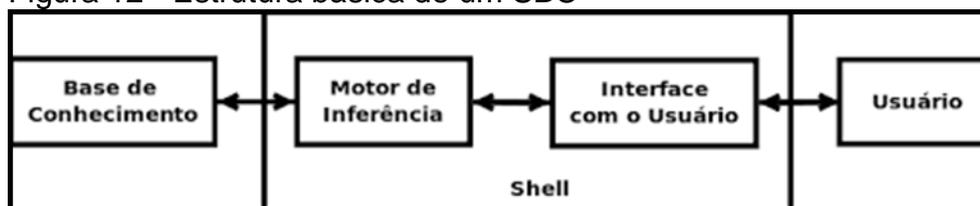
O desenvolvimento de Sistemas Baseados em Conhecimento (SBC) é uma das áreas de Inteligência Artificial que tem recebido grande investimento em pesquisas, e tem tido uma razoável projeção em diversos campos de conhecimento (NICOLETTI, et al. , 1994).

Os SBC são programas de computador que usam uma base de conhecimento para resolver problemas.

Segundo Nicoletti (1994), SBC podem ser definidos como programas de computador que usam conhecimento representado explicitamente, para resolver problemas. Assim conhecimento e processo de resolução de problemas são pontos centrais no desenvolvimento de um SBC.

O desenvolvimento de um SBC é dividido em três módulos principais, como mostra a Figura 12, onde:

Figura 12 - Estrutura básica de um SBC



Fonte: Nicoletti (1994, p.7).

- 1) Base de Conhecimento (BC): Contém o conhecimento específico do domínio da aplicação, o que inclui:
 - I. Fatos sobre o domínio;
 - II. Regras que descrevem relações no domínio;
 - III. Métodos e heurísticas para a resolução de problemas no domínio.
- 2) Motor de Inferência (MI): Mecanismo responsável pelo processamento do conhecimento da BC usando, para isso, alguma linha de raciocínio. Implementa as estratégias de inferência e controle. Quando o conhecimento está expresso como regras, as estratégias de controle usadas pelo MI normalmente são:
 - I. Backward Chaining ou encadeamento para trás;
 - II. Forward Chaining ou encadeamento para frente.

Quando MI usa a estratégia Backward Chaining, ele pesquisa uma lista de hipóteses e procura reunir evidências para viabilizar a conclusão da validade de alguma(s) dela(s).

- 3) Interface com o Usuário: Módulo responsável pela comunicação entre o usuário e o sistema. Deve fornecer também, justificativa e explicações referentes às conclusões obtidas baseadas na BC, bem como do raciocínio utilizado (NICOLETTI, et al. , 1994).

O conjunto composto pelo Motor de Inferência e Interface com o Usuário é, usualmente, chamado de shell do SBC (MONARD, 1991).

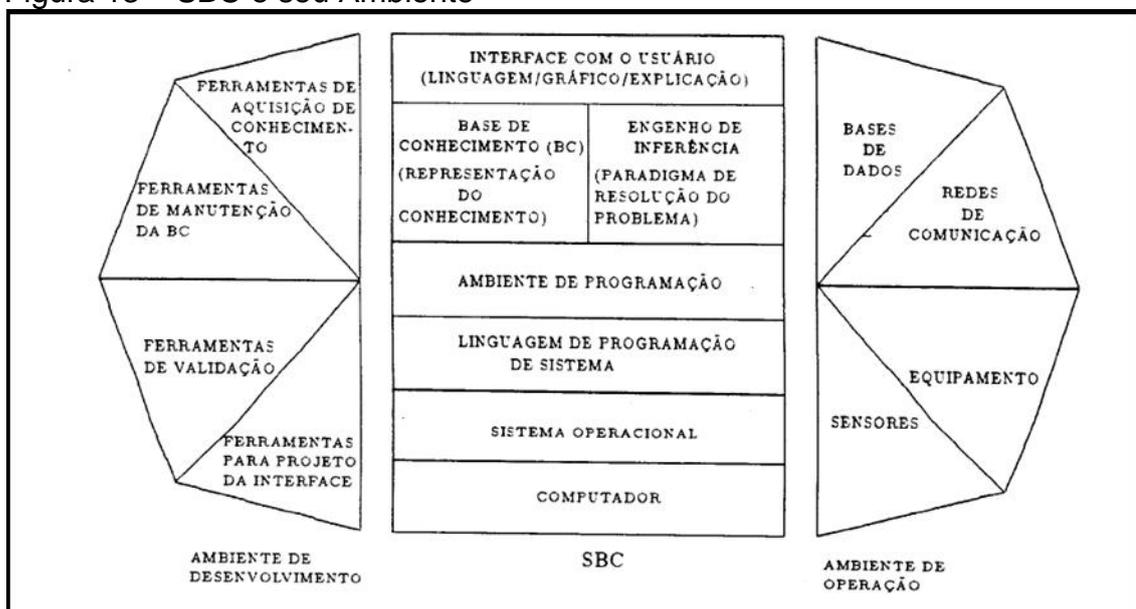
4.3.2 Processo de aquisição de conhecimento

A tarefa de aquisição de conhecimento, na grande maioria das abordagens existentes, consiste basicamente na transferência do conhecimento (informações e

formas de condução do raciocínio) do especialista humano à base de conhecimento do SBC (Figura 13) envolvendo, essencialmente, extração, análise e interpretação do conhecimento que o especialista humano usa quando resolve problemas (NICOLETTI, et al., 1994).

É importante frisar bem que o processo de extração do conhecimento do especialista envolve um imenso questionamento que, em certos casos, pode afetar a sua própria percepção de como elabora o raciocínio.

Figura 13 – SBC e seu Ambiente



Fonte: Nicoletti (1994, p. 9).

4.3.3 Base de conhecimento

A informação usada pelo sistema de aprendizado necessita ser representada de alguma forma. A escolha da representação é crítica na determinação da solução.

Um sistema representacional deve ser sempre analisado com relação a várias características, descritas a seguir (NICOLETTI, et al. , 1994).

- i. Expressividade de representação: Em qualquer sistema de IA é importante o uso de uma representação na qual o conhecimento possa ser facilmente expresso;
- ii. Facilidade de inferência: O custo computacional na realização de inferências deve ser cuidadosamente considerado. Um tipo de

inferência necessário em sistemas de aprendizado é a da comparação de duas descrições para determinar a sua equivalência.

- iii. Modificabilidade da base de Conhecimento: Devido à sua natureza, um sistema de aprendizado deve estar sempre modificando partes de sua Base de Conhecimento. Consequentemente, a maioria dos sistemas de aprendizado emprega representações que facilitam a tarefa de inserção de conhecimento na Base.
- iv. Extensibilidade da representação: Trata-se da construção de sistemas representacionais que permitam meta conhecimento como parte da Base de Conhecimento de forma a permitir que o sistema examine e altere sua própria representação, adicionando termos e estruturas representacionais.

Concluindo, sabe-se que um sistema não aprende a partir do nada, ou seja, sem um conhecimento anterior adquirido. O sistema de aprendizado deve usar algum conhecimento para entender a informação fornecida pelo ambiente, para elaborar situações e hipóteses, para testá-las e refiná-las. Portanto, talvez seja mais apropriado abordar um sistema de aprendizado como um sistema que amplia e aperfeiçoa o conhecimento que já possui.

4.4 WEKA

Ao longo dos últimos anos, o crescente emprego de processos de Mineração de Dados pelas empresas motivou o surgimento de algumas dezenas de ferramentas comerciais e livres para este fim. Um dos sistemas que acabou alcançando grande destaque chama-se Weka (Waikato Environment for Knowledge Analysis) e foi desenvolvido por uma universidade da Nova Zelândia (The University of Waikato) (PERBONI, 2013).

O Weka é um software desenvolvido em Java, formado por um conjunto de implementações de algoritmos de diversas técnicas de Mineração de Dados com as especificações GPL (General Public License) como pode ser visto na Figura 14, sendo utilizada por estudantes e professores de universidade, embora ela também seja muito utilizada por profissionais que desejam aprender os conceitos básicos de

Mineração de Dados. Por ter sua interface gráfica, é possível conduzir processos de Mineração de Dados de forma simples, realizando avaliações dos resultados obtidos e a comparação de algoritmos (DAMASCENO, 2013).

Figura 14 – Software Weka



Fonte: Elaborada pelo autor.

O Weka possui uma GUI (Graphical User Interface) que provê um ponto inicial para o usuário poder explorar as suas ferramentas e aplicações.

A GUI do Weka é composta por quatro botões, como mostra a Figura 14.

Os botões são usados para iniciar as seguintes aplicações:

- i. Explorer – Um ambiente para exploração dos dados no Weka.
- ii. Experimenter – Um ambiente de testes estatísticos e experimentos de desempenho entre os esquemas.
- iii. KnowledgeFlow – Um ambiente que suporta praticamente as mesmas funcionalidades do ambiente explorer, com a diferença de ter uma interface drag-and-drop.
- iv. SimpleCLI – Uma interface de linha de comando que fornece um ambiente para executar os comandos Weka.

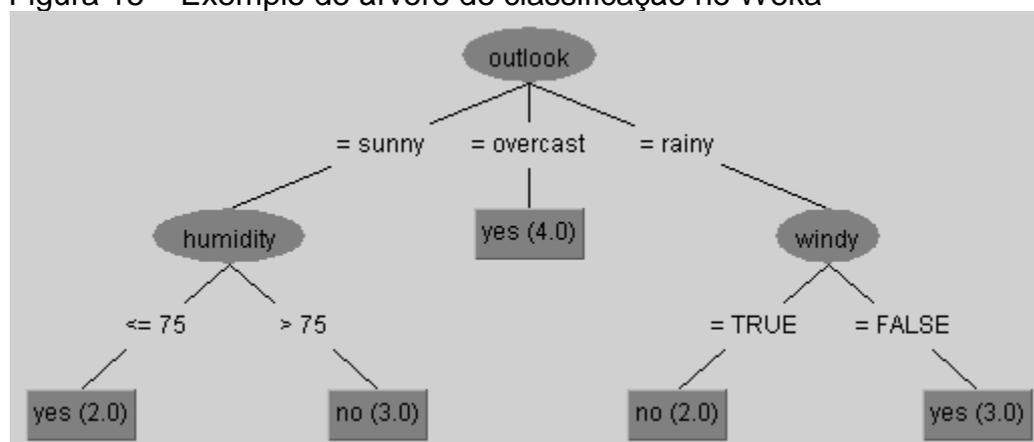
Para este trabalho foi utilizado apenas o ambiente Explorer, que será detalhado na seção referente à metodologia.

4.5 ALGORITMOS DE APRENDIZAGEM SUPERVISIONADA

Nos tipos de algoritmos que possuem esta denominação, a aprendizagem do modelo é supervisionada, ou seja, é fornecida uma classe a qual cada registro do treinamento pertence. São preditivos esses algoritmos, pois suas tarefas de mineração executam inferências nos dados com o objetivo de fornecer previsões ou tendências, com isso obter informações não disponíveis a partir dos dados disponíveis. Os tipos de algoritmos usados são (CARDOSO, 2007):

- a. Classificação: é possível determinar o valor de um atributo através dos valores de um subconjunto dos demais atributos da base de dados. Um exemplo desse algoritmo seria num conjunto de dados comerciais, onde se deseja descobrir qual o perfil dos clientes que consomem cosméticos importados. Com o algoritmo de classificação pode-se inferir, ou seja, prever que clientes do sexo feminino, com renda superior a R\$ 1.500,00 e com idade acima de 30 anos compram cosméticos importados. Neste caso, o atributo “compra cosmético importado” é denominado classe, pois é o atributo alvo da classificação cujos possíveis valores, neste caso, são “sim” ou “não”. As formas mais comuns de representação de conhecimento dos algoritmos de classificação são regras e árvores. Os algoritmos Id3, C45, J48, ADTTree, UserClassifier, PredictionNode, Splitter, ClassifierTree, M5Prime, por exemplo, geram como resultado árvores de classificação (Figura 15), enquanto que outros como Prism, Part, OneR geram regras de classificação.

Figura 15 – Exemplo de árvore de classificação no Weka



Fonte: Waikato (2004).

- b. Seleção de Atributos: existem atributos que têm um peso maior ou até determinante nas tarefas de Mineração de Dados. Por exemplo, no caso de um cliente, a sua renda com certeza é um atributo determinante nos seus hábitos de consumo. Com algoritmos de seleção de atributos é possível determinar os atributos de fato relevantes para a mineração dos dados, separando-os dos atributos irrelevantes, como por exemplo, nome do cliente (que neste caso não influencia seus hábitos de consumo). O Weka disponibiliza vários algoritmos para esta categoria de mineração, dentre eles InformationGain, PrincipalComponents e ConsistencyEval.

4.6 ALGORITMOS DE APRENDIZAGEM NÃO SUPERVISIONADA

Nestes algoritmos o rótulo da classe de cada amostra do treinamento não é conhecido e o número de conjunto de classes a ser treinado não pode ser conhecido. Além disso, são também descritivos, pois descrevem de forma concisa os dados disponíveis, fornecendo características das propriedades gerais dos dados minerados. Os tipos de algoritmos usados são (CARDOSO, 2007):

- a) Associação: quando uma classe de uma tarefa de mineração não é determinada como no caso da classificação, uma boa opção é o algoritmo de associação Apriori do Weka. O próprio algoritmo elege os atributos determinantes (lado esquerdo da regra) e os atributos resultantes (lado direito) na tarefa revelando associações entre valores dos atributos, tendo o algoritmo sua ênfase no compromisso entre precisão e cobertura (Figura 16).

Figura 16 - Exemplo de regras de associação no Weka

```
Best rules found:
1. outlook=overcast 4 ==> play=yes 4    conf:(1)
2. temperature=cool 4 ==> humidity=normal 4    conf:(1)
3. humidity=normal windy=FALSE 4 ==> play=yes 4    conf:(1)
4. outlook=sunny play=no 3 ==> humidity=high 3    conf:(1)
5. outlook=sunny humidity=high 3 ==> play=no 3    conf:(1)
6. outlook=rainy play=yes 3 ==> windy=FALSE 3    conf:(1)
7. outlook=rainy windy=FALSE 3 ==> play=yes 3    conf:(1)
8. temperature=cool play=yes 3 ==> humidity=normal 3    conf:(1)
9. outlook=sunny temperature=hot 2 ==> humidity=high 2    conf:(1)
10. temperature=hot play=no 2 ==> outlook=sunny 2    conf:(1)
```

Fonte: Waikato (2004).

- b) Clustering: em algumas situações, torna-se necessário verificar como as instâncias de uma determinada base de dados se agrupam devido a características intrínsecas de seus atributos, sem que seja definida uma classe para a tarefa. A partir da definição de uma métrica de similaridade para cada atributo e uma função de combinação destas métricas em uma métrica global, os atributos são agrupados com base no princípio da maximização da similaridade intraclasse e da minimização da similaridade interclasse. Weka possui os algoritmos Cobweb, Simple Kmeans e em para tarefas que demandam a descoberta de padrões de agrupamento nos dados. Como exemplo pode utilizar algoritmos de clustering para identificar subgrupos homogêneos de clientes de uma determinada loja.

4.7 LIMPEZA DOS DADOS

No mundo real, bases de dados estão suscetíveis a erros quando dados são inseridos. Erros como ausência de dados, dados fora dos padrões e inconsistências são bastante comuns. O trabalho de limpeza dos dados dá-se no preenchimento de dados ausentes, padronização dos dados fora de padrões e correção dos inconsistentes.

A seguir serão mostrados alguns métodos de limpeza que podem ser aplicados no início da etapa de KDD.

4.7.1 Valores ausentes

Existem alguns métodos que podem ser utilizados para trabalhar com valores que, por algum motivo não estão dispostos nas bases tais como (HAN; KAMBER, 2001):

- I. Ignorar o registro – Pouco funcional, pois apenas na condição da tupla conter vários valores ausentes é que se torna funcional.

- II. Valor constante para preencher dados – Pode mascarar os resultados dependendo da variância do atributo. Não muito recomendável.
- III. Valor médio da mesma classe a qual a tupla pertence – Utilizado quando o valor do atributo é numérico e quando seu significado é passível de atribuição a um valor médio.
- IV. Valores estatísticos para preencher dados – Utilizando técnicas de regressão ou ferramentas de inferências, tais como um formalismo bayesiano ou indução por arvores de decisão.

4.7.2 Valores fora de padrão

São erros encontrados em atributos que possam sofrer desvios acentuados em seus valores ou cadastrados de forma incorreta, fugindo assim, do padrão dos outros registros. A seguir algumas técnicas são explicadas para correção desses valores (HAN; KAMBER, 2001; CORTES, 2002).

- I. Binning – Essa técnica consiste em ordenar valores e agrupa-los. Após o agrupamento aplicar uma técnica como média aritmética, mediana, valor limite, para o ajuste dos valores em cada grupo. Após isso substituir os valores pelos calculados em cada grupo.
- II. Agrupamento – Valores fora do padrão podem ser detectados quando dados similares são dispostos em clusters ou grupos. Os valores fora destes grupos podem ser considerados fora do padrão e, assim, excluídos.
- III. Inspeção humana e computador – Às vezes o próprio programador pode fazer algumas medições, identificar os valores e excluí-los.
- IV. Regressão – Dados podem ser ajustados por funções de regressão. Estas funções ajustam duas variáveis, num gráfico, desde que uma possa ser predita por outra.

4.7.3 Dados inconsistentes

Existem várias maneiras de retirar inconsistências num banco de dados. Algumas delas podem ser retiradas manualmente através de referências externas.

Estes erros são mais comuns quando um usuário dá entrada incorretamente num dado. Outras formas de retirar dados inconsistentes num banco de dados é utilizar ferramentas de engenharia de conhecimento. Inconsistências podem ser causadas por integrações de dados, dados de banco de dados diferentes, ou ainda de bases diferentes. Há a possibilidade também de ocorrer redundância de dados. Dados com nomes parecidos pertencentes a um mesmo atributo (LOUZADA; DINIZ, 2000).

5 TRABALHOS CORRELATOS

Existe uma grande variedade de trabalhos acadêmicos e pesquisas acerca da descoberta de conhecimento utilizando bases de dados específicas. São variadas as informações que podem ser encontradas e de como podem ser implementados os diversos algoritmos para a fase de treinamentos e testes. Os resultados identificam padrões que antes não podiam ser detectados de forma manual, podendo ser utilizados em diversas áreas do conhecimento humano.

Dentre os trabalhos pode-se citar: “Identificação de padrões na ocorrência de emergências e desastres associados a níveis de precipitação” de autoria de ESTÉBANEZ (2012), onde se utiliza técnicas de Mineração de Dados para a identificação de padrões na ocorrência de emergências e desastres com base em registros de precipitação para cinco cidades do Equador. Os resultados obtidos foram utilizados em conjunto com previsões de precipitação de prazos imediatos e curto, a fim de fornecer informação sobre quais eventos poderiam acontecer com a precipitação prevista nesses períodos.

Outro trabalho similar seria “Técnicas de Mineração de Dados no estudo de desastres naturais e planejamento urbano” de autoria de SOUZA (2004), onde são utilizadas técnicas computacionais para a previsão de desastres naturais e para o estudo do impacto causado nos corpos hídricos decorrentes do crescimento urbano.

6 METODOLOGIA

Pesquisas realizadas com propósitos acadêmicos tendem, no primeiro momento, a assumir o caráter de pesquisa exploratória, pois é pouco provável que o pesquisador tenha uma definição clara do que irá investigar. Segundo Gil (2010), as pesquisas exploratórias têm por finalidade proporcionar maior familiaridade com o problema, com o intuito de torna-lo mais explícito ou a construir hipóteses.

Diante disso, esse projeto é uma pesquisa exploratória, pois o objetivo é a aplicação de algoritmos de Mineração de Dados a fim de encontrar padrões, que possam indicar algum tipo de conhecimento implícito sobre uma base de dados, no caso, dados de desastres naturais ocorridos na cidade de Bauru no interior do estado de São Paulo.

O projeto foi desenvolvido em duas etapas: na primeira foram realizadas pesquisas, estudos e desenvolvimento do embasamento teórico e na segunda etapa foram feitos os testes com a ferramenta Weka com o objetivo de gerar conhecimento sobre os dados estudados.

Para o desenvolvimento do embasamento teórico e a descrição das atividades, houve a necessidade de realizar pesquisas sobre:

- i. Conceituação de Inteligência Artificial;
- ii. Conceitos e aplicações de aprendizagem de Máquina;
- iii. Fases do processo de Descoberta de Conhecimento em Base de Dados;
- iv. Conceitos, arquitetura e tarefas de Data Mining;
- v. Características e funcionamento da ferramenta de Mineração de Dados Weka;
- vi. Algoritmos de classificação da ferramenta Weka;
- vii. Testes com os algoritmos de classificação e associação da ferramenta Weka.

6.1 FERRAMENTA UTILIZADA - WEKA

Weka é uma ferramenta, de código aberto, desenvolvida em JAVA, que usa algoritmos de aprendizagem de máquina para tarefas de Mineração de Dados. Os

algoritmos podem ser aplicados diretamente ao conjunto de dados ou chamados por um código JAVA. Weka possui ferramentas para pré-processamento, classificação, regressão, agrupamento, regras de associação e visualização (LAROSE, 2005). A escolha dessa ferramenta deveu-se a:

- I. Ser um software livre e possuir código aberto.
- II. Por ser implementado na linguagem Java, que tem como sua principal característica a sua portabilidade, sendo possível utiliza-lo em diferentes sistemas operacionais.
- III. Ter vários métodos úteis para Mineração de Dados tais como: Métodos de classificação, métodos para predição numérica, métodos de agrupamento e métodos de associação.
- IV. Possuir uma interface gráfica amigável.

6.2 BASE DE DADOS

A base de dados para a descoberta de conhecimento foi obtida através do Website do Instituto de pesquisa Meteorológica de Bauru (IPMET). O IPMET possui uma base de dados de desastres naturais cujo download pode ser realizado gratuitamente. A busca dos desastres naturais pode ser filtrada por períodos de datas que estão entre 01/01/1980 até a data corrente, por tipo de fenômeno como, por exemplo, granizo, chuvas, raios, enchentes, entre outros, podem ser feitas também pelo tipo de danos causados, como danos em veículos, destelhamento, feridos, desabrigados, dentre outros. Existe a possibilidade de filtrar por cidades que estão situadas nos estados de São Paulo e Paraná.

Depois de realizado o download do arquivo foi preciso adaptá-lo para o formato a ser utilizado para a mineração. A princípio o arquivo está em formato texto (txt), para ser utilizado na ferramenta Weka, houve a necessidade de convertê-lo para o formato ARFF.

O formato ARFF é utilizado como padrão para estruturar as bases de dados manipuladas pela Weka. Ele corresponde a um arquivo texto contendo um conjunto de registros, procedido por um pequeno cabeçalho e por fim de um conjunto de dados iniciados pela palavra “@data” como se pode observar na Figura 17.

Figura 17 - Arquivo formato ARFF

```

64
65
66
67 @RELATION iris
68
69 @ATTRIBUTE sepalength→REAL
70 @ATTRIBUTE sepalwidth→REAL
71 @ATTRIBUTE petallength→REAL
72 @ATTRIBUTE petalwidth→REAL
73 @ATTRIBUTE class→{Iris-setosa,Iris-versicolor,Iris-virginica}
74
75 @DATA
76 5.1,3.5,1.4,0.2,Iris-setosa
77 4.9,3.0,1.4,0.2,Iris-setosa
78 4.7,3.2,1.3,0.2,Iris-setosa
79 4.6,3.1,1.5,0.2,Iris-setosa
80 5.0,3.6,1.4,0.2,Iris-setosa
81 5.4,3.9,1.7,0.4,Iris-setosa
82 4.6,3.4,1.4,0.3,Iris-setosa
83 5.0,3.4,1.5,0.2,Iris-setosa
84 4.4,2.9,1.4,0.2,Iris-setosa
85 4.9,3.1,1.5,0.1,Iris-setosa
86 5.4,3.7,1.5,0.2,Iris-setosa
87 4.8,3.4,1.6,0.2,Iris-setosa
88 4.8,3.0,1.4,0.1,Iris-setosa
89 4.3,3.0,1.1,0.1,Iris-setosa
90 5.8,4.0,1.2,0.2,Iris-setosa
91 5.7,4.4,1.5,0.4,Iris-setosa
92 5.4,3.9,1.3,0.4,Iris-setosa

```

Fonte: Elaborada pelo autor.

A declaração da Relação, ou seja, o nome do arquivo que será usado no Weka é definido pelo parâmetro @relation seguido pelo seu nome. As declarações dos atributos são feitas através de uma sequência de linhas @attribute. A ordem da declaração indica a posição de cada atributo na seção DATA. A Weka suporta atributos numéricos e categóricos.

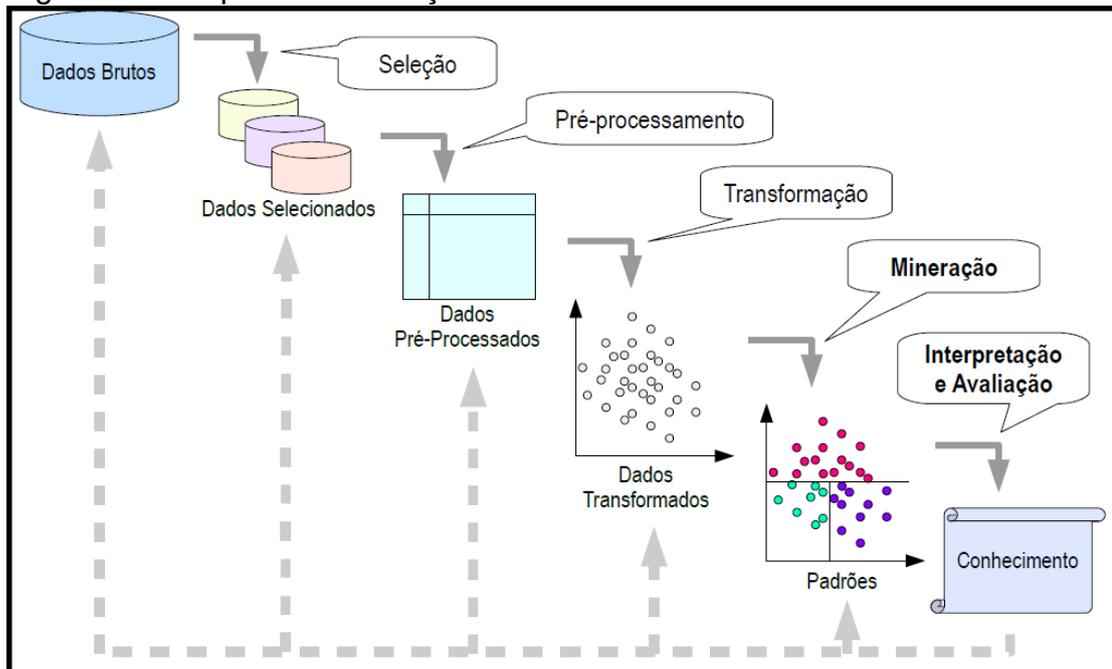
Atributos numéricos devem ser indicados através das palavras numeric ou real.

Para os atributos categóricos é preciso fornecer uma lista, indicando todos os valores do atributo. Ex.: {Iris-setosa, Iris-versicolor, Iris-virginica...}. Também são suportados os tipos date e string.

6.3 ETAPAS DA MINERAÇÃO

Para realizar a Mineração de Dados foi preciso seguir algumas etapas prévias, como se pode observar na Figura 18. As primeiras delas, a seleção e pré-processamento, são descritas a seguir.

Figura 18 - Etapas da Mineração de Dados



Fonte: Elaborada pelo autor.

6.3.1 Seleção e Pré-processamento

O primeiro passo para a realização da mineração foi disponibilizar os dados para a seleção e realizar seu pré-processamento, onde o objetivo foi melhorar a qualidade, tratando-os para evitar possíveis distorções na extração do conhecimento. Também foram feitas transformações nessa base, com relação à forma como os dados estão representados a fim de facilitar e retirar possíveis barreiras para a etapa de Mineração de Dados. Foi selecionada a base de dados disponibilizada pelo site do IPMet e importada para o Weka.

Ao iniciar o Weka são mostradas algumas opções para realizar a mineração, foi escolhida a opção “Explorer”, como se pode ver na Figura 19.

Figura 19 - Tela inicial Weka



Fonte: Elaborada pelo autor.

A opção “Explorer” possui algumas ferramentas úteis como:

- I. Preprocess: Escolhe e modifica os dados utilizados
- II. Classify: Treina e testa sistemas de aprendizagem que classificam ou realiza regressão.
- III. Cluster: Análise de clusters.
- IV. Permite aprender regras de associação para os dados.
- V. Select attributes: Seleciona os atributos mais relevantes nos dados.
- VI. Visualize: Gráfico 2D interativo dos dados.

A base de dados foi carregada e, como está ilustrado na Figura 20, no item “current relation”, está o nome da relação, o número de instâncias e atributos.

Figura 20 - Dados carregados no Weka

The screenshot shows the Weka Explorer interface. The 'Current relation' panel is highlighted with a red box and shows the following information:

- Relation: desastres_naturais
- Instances: 2217
- Attributes: 6
- Sum of weights: 2217

The 'Selected attribute' panel shows a table with the following data:

No.	Label	Count	Weight
1	janeiro-86	2	2.0
2	fevereiro-86	1	1.0
3	agosto-86	1	1.0
4	dezembro-86	1	1.0
5	janeiro-87	8	8.0
6	fevereiro-87	5	5.0
7	maio-87	7	7.0
8	setembro-87	4	4.0
9	novembro-87	3	3.0

The 'Attributes' panel shows a list of attributes with checkboxes:

- dia
- período
- fenomeno
- endereco
- ocorrencia_danos
- vitimas

The 'Status' panel at the bottom shows 'OK'.

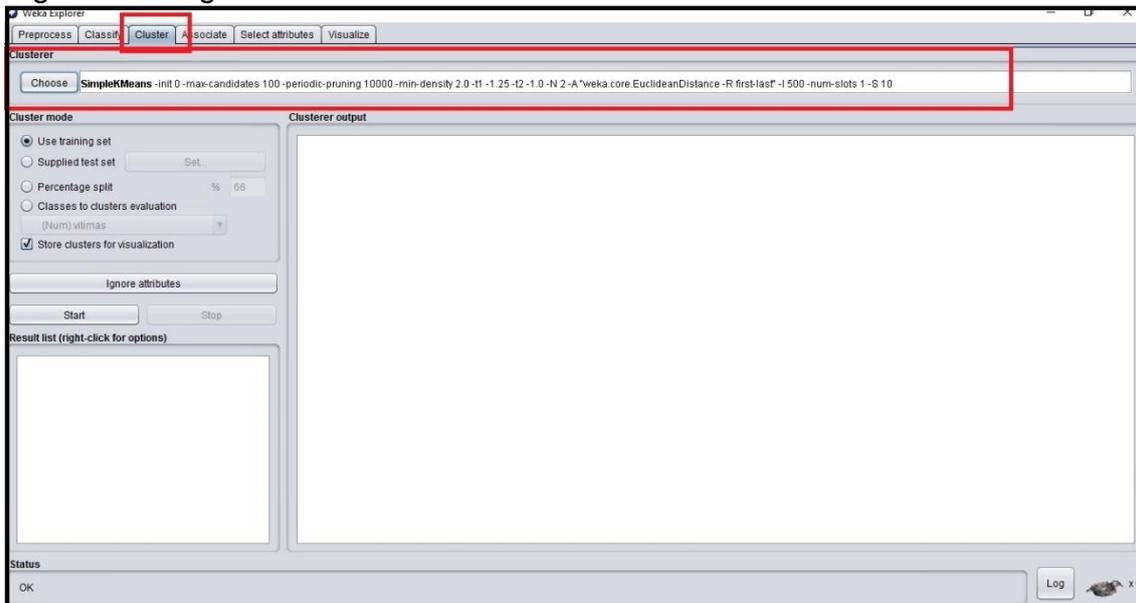
Fonte: Elaborada pelo autor.

Como os dados não possuem nenhuma classe de saída de interesse para poder utilizar o aprendizado supervisionado foram realizados vários testes a fim de criar clusters (classes) automaticamente. Para isso, foi utilizado o algoritmo de clusterização denominado SimpleKMeans, descrito a seguir. A criação das classes visou permitir a aplicação de algoritmos de classificação.

6.3.1.1 Cluster – SimpleKMeans (5 clusters)

Os dados foram carregados e, por meio da guia Cluster foi, então, escolhido o algoritmo SimpleKMeans (ao clicar no botão Choose), como é demonstrado na Figura 21.

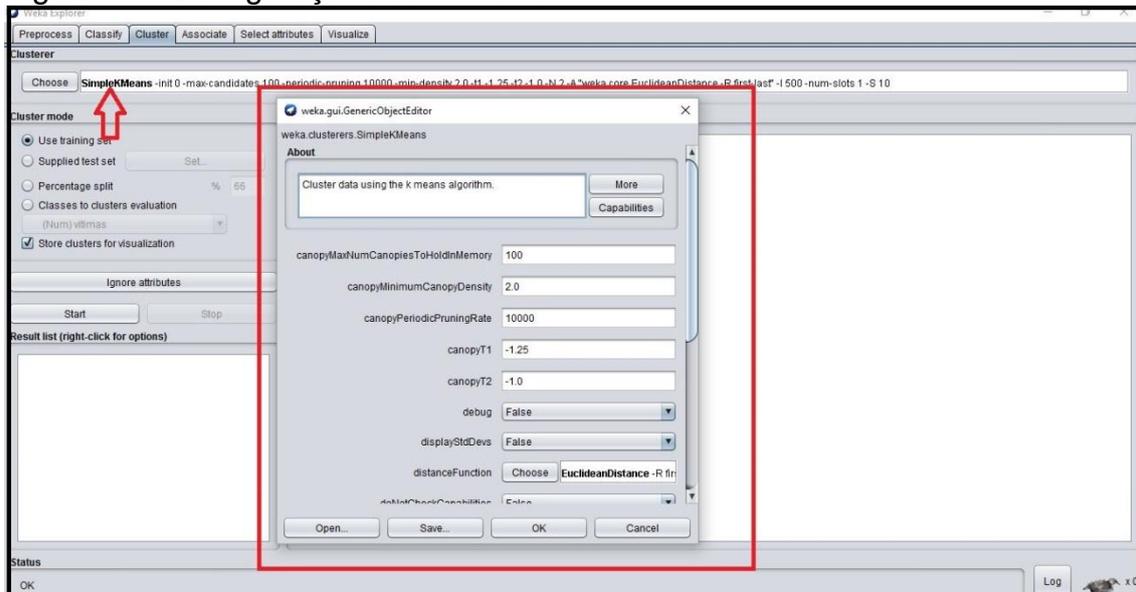
Figura 21 - Algoritmo KMeans



Fonte: Elaborada pelo autor.

É possível definir alguns atributos para o uso desse algoritmo, clicando sobre a palavra SimpleKMeans. Neste caso, é aberta uma janela com algumas configurações como se pode ver na Figura 22.

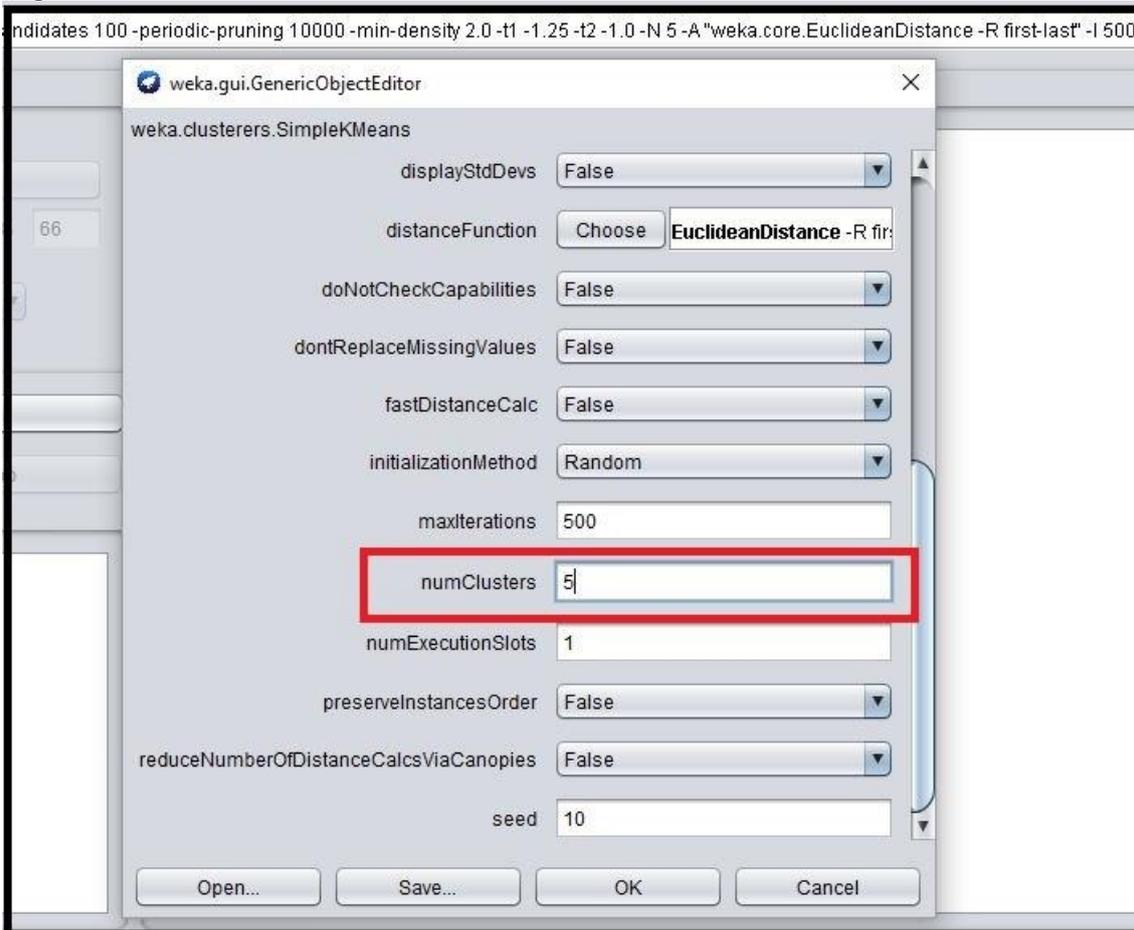
Figura 22 - Configurações do KMeans



Fonte: Elaborada pelo autor.

O único atributo do algoritmo que foi modificado foi o campo numClusters, que indica quantos clusters deseja-se criar. Foi alterado o valor padrão de 2 para 5, assim como mostra a Figura 23. O número de 5 classes foi escolhido aleatoriamente para poder verificar a disposição dos dados em cluster.

Figura 23 - Número de clusters do KMeans



Fonte: Elaborada pelo autor.

Após executar o algoritmo foi dado como saída um agrupamento dos dados divididos inicialmente em 5 classes, como mostra na Figura 24.

Figura 24 - Saída KMeans

```

Attribute          Full Data          Cluster#
                   (2217.0)          0          1          2          3          4
                   (2217.0)          (451.0)          (83.0)          (742.0)          (416.0)          (525.0)
-----
dia                fevereiro-02      fevereiro-01      janeiro-11      fevereiro-02      fevereiro-95      janeiro-94
periodo            sem_informacao    Noturno sem_informacao    Vespertino sem_informacao sem_informacao
fenomeno           2.7501            2.7228            4.1205          2.3356            3.1298            2.8419
endereco           setor_4            setor_12          setor_11        setor_12          setor_4            setor_5
ocorrencia_danos   32.8372           31.949            43.0723         31.7291           30.6442           35.2857
vitas              4.4118            0.8404            13.5181         -0.7358           3.9736            13.6629

Time taken to build model (full training data) : 0.08 seconds

=== Model and evaluation on training set ===

Clustered Instances
0      451 ( 20%)
1       83 (  4%)
2      742 ( 33%)
3      416 ( 19%)
4      525 ( 24%)

```

Fonte: Elaborada pelo autor.

A saída mostra como cada cluster se une. Os números são o valor médio de todos no cluster. Cada cluster mostra um tipo de comportamento, do qual foi possível observar alguns pontos:

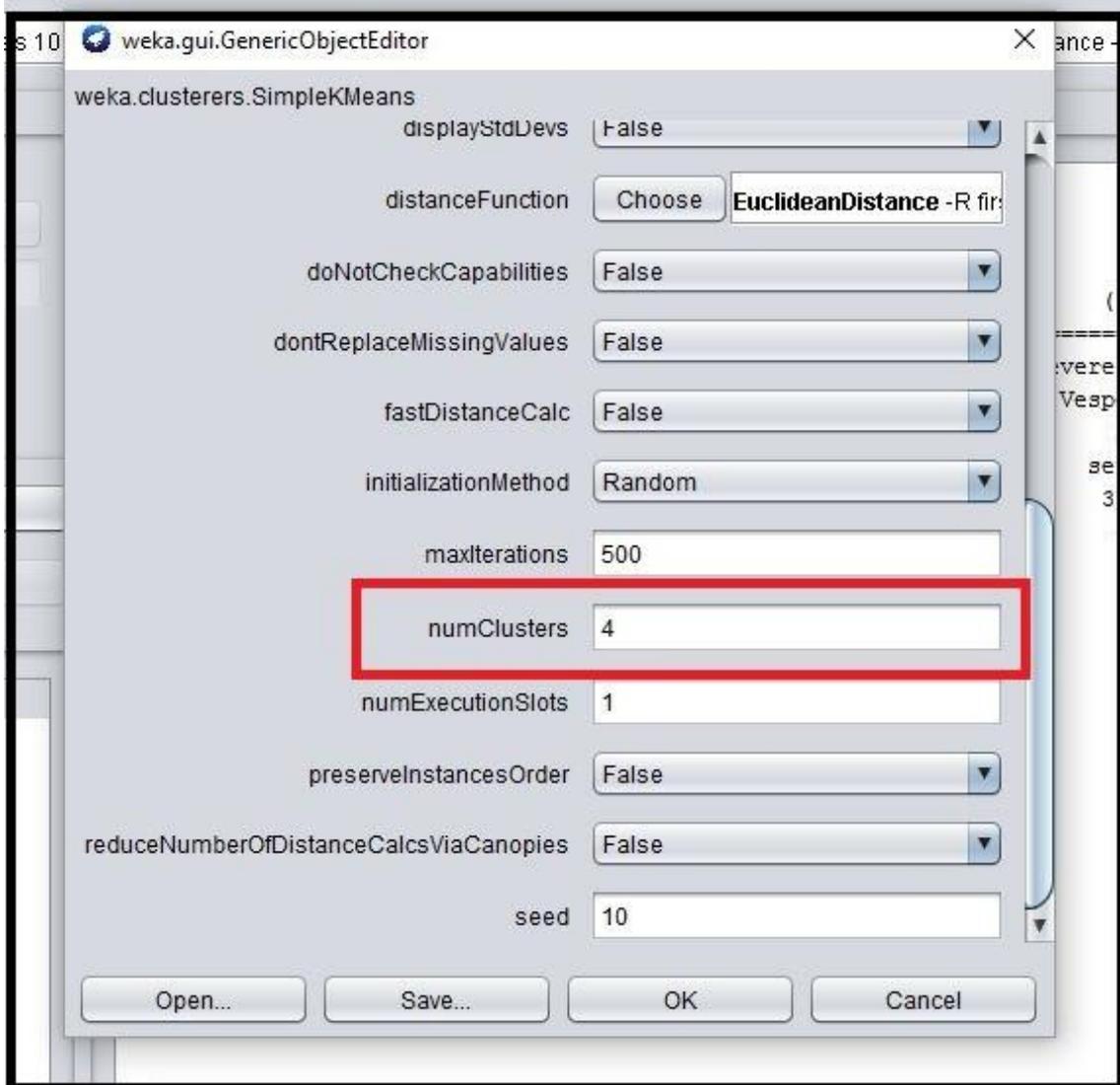
- O cluster 1 mesmo tendo um elevado número médio de ocorrência (43.0723) e vítimas (13.5181) teve o menor número de instâncias com uma porcentagem de 4%, totalizando 83, se tornando “irrelevante” para tirar alguma conclusão.
- O cluster 2 obteve o maior número de instâncias, observa-se que este cluster tem como característica ter o período de fevereiro de 2002 com maiores ocorrências de danos.
- Logo em seguida o cluster 0 e o cluster 3 seguem praticamente empatados com 20% e 19% respectivamente do total de instâncias.

6.3.1.2 Cluster – SimpleKMeans (4 clusters)

Como o cluster 1 teve um número pequeno de dados (apenas 4%), optou-se por diminuir o número de clusters para ser observado como ficaria a distribuição dos agrupamentos.

Foram realizados os mesmos passos anteriormente indicados, mudando apenas o número de clusters (grupos) de 5 para 4, como mostra na Figura 25.

Figura 25 - Kmeans com 4 cluster



Fonte: Elaborada pelo autor.

Após executar o algoritmo foi dado como saída um agrupamento dos atributos divididos em 4 classes, como mostra na Figura 26.

Figura 26 - Saída Kmeans com 4 cluster

```

Final cluster centroids:
Attribute          Full Data          Cluster#
                   (2217.0)          0              1              2              3
=====
dia                fevereiro-02      fevereiro-01   janeiro-94     fevereiro-02   fevereiro-95
periodo            sem_informacao   Noturno sem_informacao   Vespertino sem_informacao
fenomeno           2.7501           2.7323        2.8513         2.3395         3.3479
endereco           setor_4          setor_12       setor_5        setor_12       setor_4
ocorrencia_danos  32.8372         31.8822       36.315        31.7936       31.1429
vitimas           4.4118          0.7773        13.9558       -0.739         4.8111

Time taken to build model (full training data) : 0.08 seconds

=== Model and evaluation on training set ===

Clustered Instances
0      467 ( 21%)
1      565 ( 25%)
2      751 ( 34%)
3      434 ( 20%)

```

Fonte: Elaborada pelo autor.

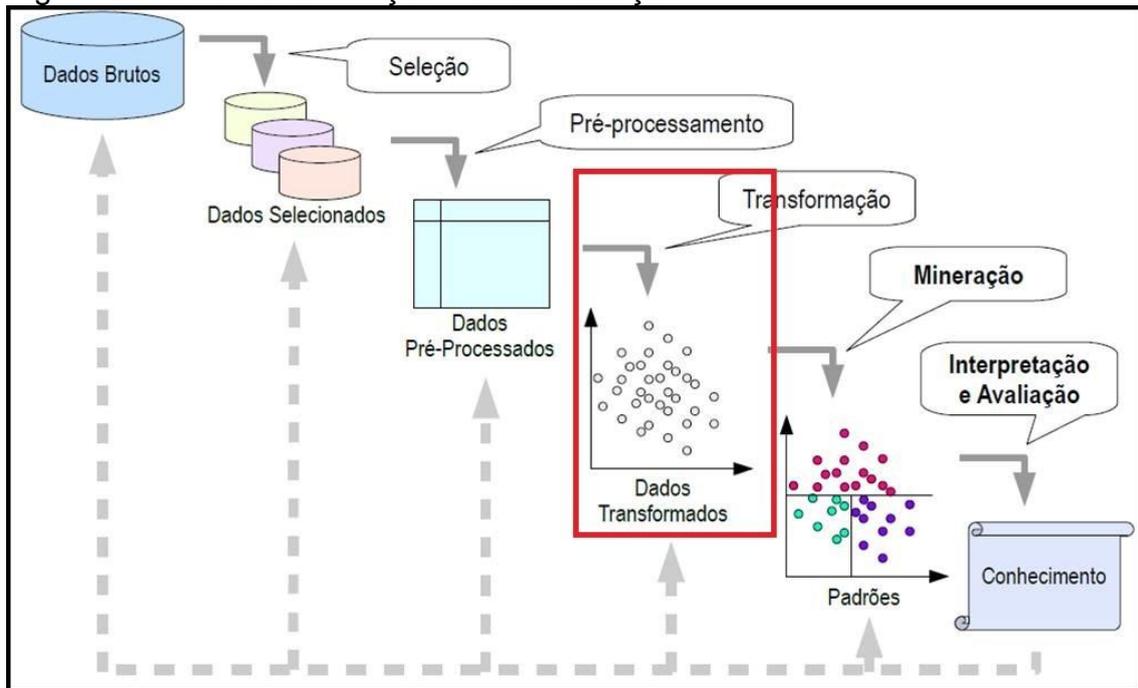
Com 4 clusters os dados ficaram melhor distribuídos do que quando foram considerados cinco grupos. Considerando ainda a etapa de pré-processamento foram removidos alguns atributos.

Essa transformação foi feita na etapa posterior, descrita a seguir.

6.3.2 Transformação dos dados

O próximo passo foi a transformação dos dados (Figura 27) para uma melhor visualização dos mesmos.

Figura 27 - Passos Mineração - Transformação



Fonte: Elaborada pelo autor.

Com 4 clusters os resultados (números de instâncias em cada cluster) foram bem distribuídos, entretanto, o atributo dia ainda não estava adequado (muitos valores distintos) para realizar a mineração, por isso, esse atributo foi transformado para meses (de janeiro até dezembro), diminuindo, assim, os valores dos seus domínios.

Como os algoritmos de aprendizagem não supervisionada, a exemplo da regra de associação, necessitam de apenas valores nominais, os atributos (fenômeno, ocorrência e vítimas) que tinham associados valores numéricos foram transformados e passaram a conter valores nominais, conforme ilustra a Figura 28.

Figura 28 - Base de dados por mês

```

1 @RELATION desastres_naturais
2
3 @attribute mes {janeiro, fevereiro, marco, abril, maio, junho, julho, agosto, setembro, outubro, novembro, dezembro}
4 @attribute periodo {sem_informacao, matutino, vespertino, noturno}
5 @attribute fenomeno {granizo, chuvas_fortes, ventos_fortes, raio, chuvas_moderadas}
6 @attribute endereco {setor_1, setor_2, setor_3, setor_4, setor_5, setor_6, setor_7, setor_8, setor_9, setor_10, setor_11, setor_12}
7 @attribute ocorrencia_danos {desabrigados, desalojados, feridos, vitimas_fatais, queda_barreira, queda_de_arvores, transbordamentos, danos_pontes,
8 inundacoes_graduais, enchentes, alagamentos, deslizamento_terra, queda_muro, desabamentos, congestionamento, danos_veiculos, destelhamento,
9 corte_fornecimento_energia_agua, destruicao_plantacao, erosao_buraco, queda_poste, pessoa_arrastada_enxurrada, desbarracamento_margem_rio, acidente_carro,
10 danificacao_pavimentacao, rompimento_rede_agua_esgoto, danos_causados_raios, enxurradas, inundacoes_bruscas }
11 @attribute vitimas {sim, nao}
12
13 @data
14 janeiro,sem_informacao,chuvas_fortes,setor_12,queda_de_arvores,nao
15 janeiro,sem_informacao,chuvas_fortes,setor_12,desabamentos,nao
16 fevereiro,sem_informacao,chuvas_fortes,setor_5,erosao_buraco,nao
17 agosto,sem_informacao,chuvas_fortes,setor_12,inundacoes_graduais,nao
18 dezembro,sem_informacao,chuvas_fortes,setor_1,erosao_buraco,nao
19 janeiro,vespertino,chuvas_fortes,setor_12,inundacoes_graduais,nao
20 janeiro,vespertino,chuvas_fortes,setor_12,inundacoes_graduais,nao
21 janeiro,vespertino,chuvas_fortes,setor_12,inundacoes_graduais,nao
22 janeiro,sem_informacao,chuvas_fortes,setor_4,inundacoes_graduais,nao
23 janeiro,sem_informacao,chuvas_fortes,setor_4,inundacoes_graduais,nao
24 janeiro,vespertino,chuvas_fortes,setor_3,erosao_buraco,nao
25 janeiro,sem_informacao,chuvas_fortes,setor_3,erosao_buraco,nao
26 janeiro,sem_informacao,chuvas_fortes,setor_1,erosao_buraco,nao
27 fevereiro,vespertino,chuvas_fortes,setor_1,inundacoes_graduais,nao
28 fevereiro,vespertino,chuvas_fortes,setor_4,inundacoes_graduais,nao
29 fevereiro,vespertino,chuvas_fortes,setor_4,inundacoes_graduais,nao
30 fevereiro,vespertino,chuvas_fortes,setor_12,inundacoes_graduais,nao

```

Fonte: Elaborada pelo autor.

Com as alterações feitas, optou-se por reduzir o número de clusters para apenas 3 (um menor número de grupos/classes poderia favorecer as tarefas de classificação e extração de conhecimento). Na Tabela 1 pode-se observar um comparativo do algoritmo KMeans para criação automática de classes (clusters), mostrando as porcentagens de cada cluster.

Tabela 1 - Algoritmo KMeans

Comparação - Algoritmo Cluster			
	Kmeans		
	5 Cluster	4 Cluster	3 Cluster
Cluster 0	20,35%	21,30%	38,42%
Cluster 1	4,12%	25,12%	37,15%
Cluster 2	33,50%	34,30%	24,43%
Cluster 3	19,85%	19,28%	
Cluster 4	22,18%		

5 Cluster: Na etapa de pré-processamento
Na etapa de pré-processamento

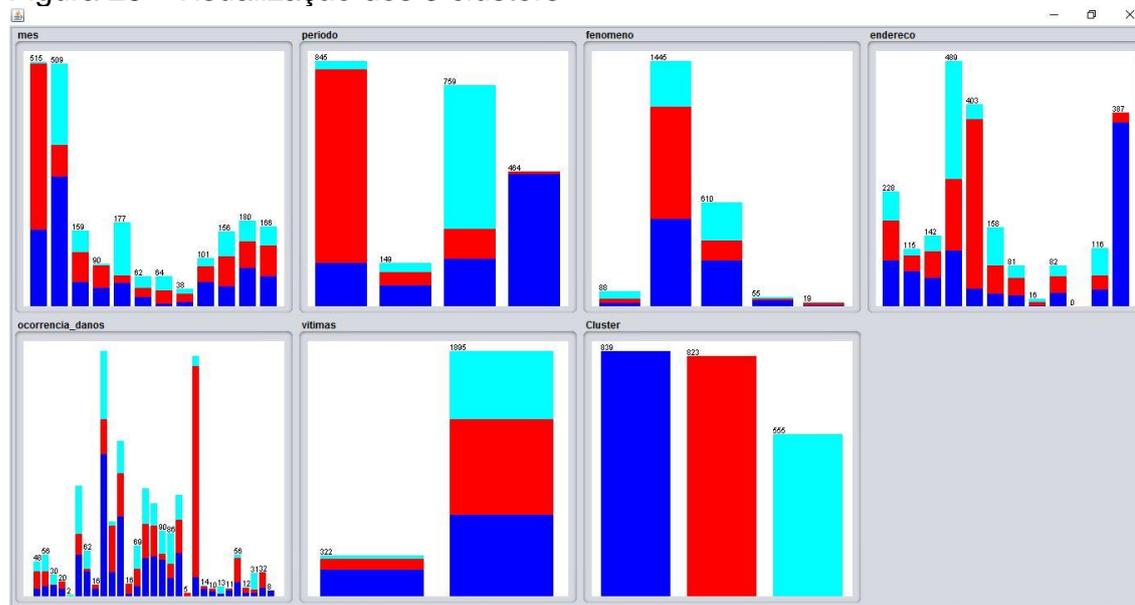
4 Cluster: para melhorar a saída dos dados,
eliminando a classe (Cluster 1 - 4%)
irrelevante.

3 Cluster: Após o Processo de transformação.

Fonte: Elaborada pelo autor.

A Figura 29 mostra a distribuição dos dados em 3 clusters.

Figura 29 - Visualização dos 3 clusters



Fonte: Elaborada pelo autor

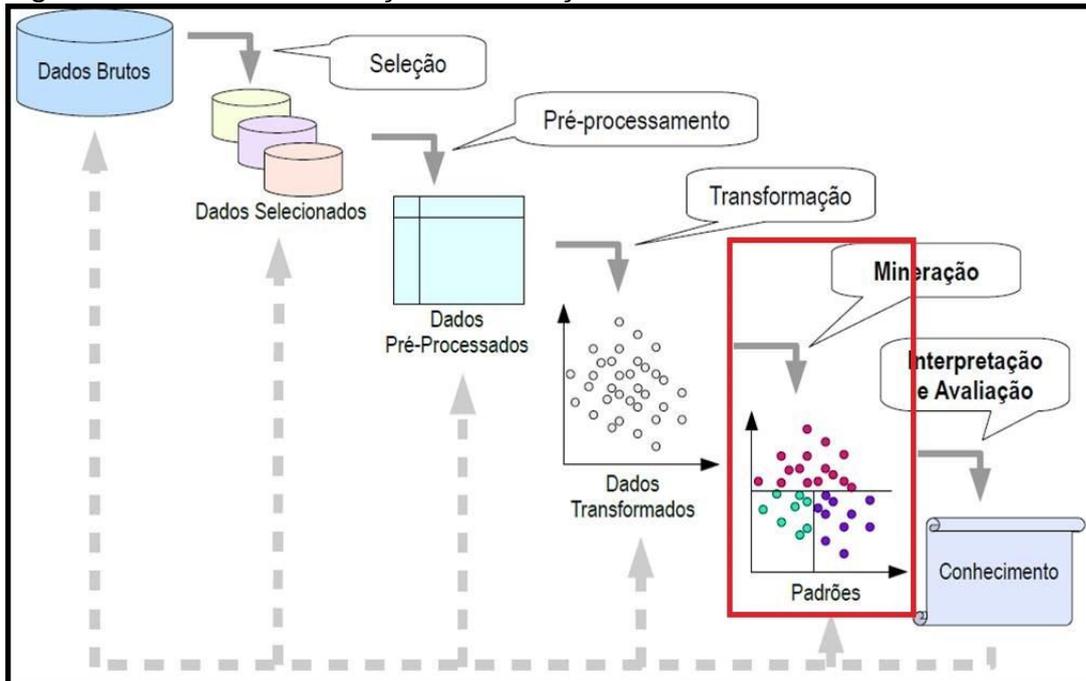
Os clusters estão divididos por cores onde, o cluster 0, o 1 e o cluster 2 são o azul, o vermelho e o azul claro, respectivamente.

- O cluster 0 é formado por dados referentes em maior parte ao setor 12 (Figura 52 no anexo) onde está localizado o centro da cidade e onde são constatados diversos casos de inundações nas suas principais avenidas e isso ficou claro pois, no gráfico os maiores incidentes são de inundações e alagamentos todos ocorridos no começo do ano.
- O cluster 1 pode ser chamado de mais precário de Bauru, pois sua localização está centrada nos bairros menos abastados como é o caso dos setores 4 e 5 (Núcleo Fortunato Rocha Lima e Favela São Manoel, por exemplo) – Figuras 44 e 45 - e os casos mais recorrentes são de chuvas fortes o que, conseqüentemente, gera erosões e buracos, acarretando em maior número de vítimas.
- O cluster 2 poderia ser visto como um subconjunto do cluster 1, já que sua maior localidade está no setor 4 (Figura 44, nos anexos) e suas ocorrências na maior parte vão ao encontro com as mesmas ocorrências do cluster 1 como inundações, quedas de árvores e desabamentos, por exemplo.

7 RESULTADOS

O próximo passo foi a mineração dos dados em si (Figura 30) cujo objetivo foi a escolha dos algoritmos a serem utilizados na extração do conhecimento, conforme descrito a seguir.

Figura 30 - Passos Mineração - Mineração dos dados



Fonte: Elaborada pelo autor

7.1 SELEÇÃO DOS ALGORITMOS

Dentre os algoritmos de Aprendizado Supervisionado foram selecionados JRIP e o J48. Estes algoritmos foram escolhidos devido ao vasto uso em tarefas de Mineração de Dados e geração de conhecimento, quantidade de material disponibilizado assim como o guia de estudos desses algoritmos fornecido no site do WEKA.

Características dos algoritmos testados:

- a) JRIP: O JRIP ou Ripper (Repeated Incremental Pruning to Produce Error Reduction), foi proposto por William W. Cohen. É um algoritmo baseado em regras de decisão, tais como SE e ENTÃO. Este algoritmo

basicamente se divide em duas fases: a primeira fase gera um conjunto de regras para a comparação e a segunda etapa otimiza o conjunto de regras inicial para diminuir erros e tornar o processo mais seletivo, sendo estes dois estágios repetidos inúmeras vezes.

- b) OneR: É um algoritmo de classificação que obtém as regras de árvore de decisão parciais construídas usando o J48. Este algoritmo realiza testes em cada atributo do seu conjunto de dados, criando regras para construção da árvore.
- c) J48: É um algoritmo baseado em árvore de decisão. Segundo Pereira (2002) árvore é uma coleção finita de 0 ou mais nodos. Caso a árvore obtenha zero nodo então se diz que ela é nula, do contrário ela apresenta um nodo raiz, e demais nodos denominados sub-árvores. O número de sub-árvores indica o grau. Quando um determinado nodo não possui sub-árvores, este nodo é chamado de folha.
- d) Randon Tree: O algoritmo Random Tree considera apenas alguns atributos escolhidos aleatoriamente para cada nó da árvore, não realiza poda e também não permite a estimativa de probabilidade de classes.

Após a execução desses algoritmos foi possível realizar algumas comparações que serão demonstradas a seguir. Todos foram testados com relação à classificação das instâncias nos 3 clusters definidos anteriormente (seção 6.3.2).

Tabela 2 - Classificação das instâncias

ALGORITMO	INSTÂNCIAS CORRETAMENTE CLASSIFICADAS	% CORRETAS	INSTÂNCIAS ERRONEAMENTE CLASSIFICADAS	% INCORRETAS	TOTAL DE INSTÂNCIAS
J48	2084	94,10%	133	5,60%	2217
RandomTree	2025	91,33%	133	8,67%	2217
Jrip	2072	93,45%	145	6,54%	2217
OneR	1969	88,81%	248	11,18%	2217

Fonte: Elaborada pelo autor

Na Tabela 2 pode-se notar o percentual de instâncias classificadas corretamente e incorretamente, mostrando uma pequena variação entre os algoritmos J48, RandomTree e JRIP, observa-se que o algoritmo OneR teve uma

alta taxa de classificações incorretas superior aos demais, o que mostra que seu desempenho geral foi inferior.

Tabela 3 - Métricas de Desempenho por classe

ALGORITMO	MÉTRICAS DE DESEMPENHO POR CLASSE					CLASSE
	TAXA TP	TAXA FP	PRECISÃO	COBERTURA	F-MEASURE	
J48	0,47	0,01	0,84	0,47	0,61	SIM
	0,98	0,52	0,91	0,98	0,95	NÃO
RamdonTree	0,72	0,05	0,69	0,7	0,71	SIM
	0,94	0,29	0,95	0,94	0,94	NÃO
Jrip	0,73	0,02	0,81	0,7	0,75	SIM
	0,97	0,29	0,95	0,97	0,96	NÃO
OneR	0,32	0,01	0,76	0,32	0,46	SIM
	0,98	0,57	0,89	0,98	0,93	NÃO

Fonte: Elaborada pelo autor

Legenda:

- a) TAXA TP = Taxa de Verdadeiro positivo são as instâncias classificadas positivamente.
- b) TAXA FP = Taxa de Falso positivo são as instâncias em que o modelo que foi criado prevê que ela deve ser positiva, mas, ao invés disso, o valor real é negativo.
- c) PRECISÃO = Corresponde à razão entre o número de instâncias classificadas corretamente em uma classe e o número de instâncias classificadas como aquela classe (Tanto corretamente quanto incorretamente).
- d) COBERTURA = Corresponde à razão entre o número de instâncias classificadas corretamente em uma classe e o número total de instâncias da classe.
- e) F-MEASURE = Média harmônica entre a precisão e a cobertura.

Na Tabela 3 tem-se o desempenho de cada algoritmo em relação à saída (sim ou não) para o atributo vítimas. Pode-se notar em relação a sua média harmônica (F-Measure) que os algoritmos não têm muita variação, não podendo escolher qual o algoritmo mais apropriado.

Na Tabela 4 é mostrado mais um argumento para a escolha do algoritmo que é o coeficiente Kappa.

O Kappa corresponde ao resultado da comparação entre a classificação feita por cada classificador e o conjunto de treinamento, que representa o classificador perfeito.

Tabela 4 - Coeficiente Kappa

COEFICIENTE KAPPA		
ALGORITMO	KAPPA	INTERPRETAÇÃO
J48	0,56	Concordância quase imperfeita
RamdonTree	0,65	Concordância quase imperfeita
Jrip	0,72	Concordância quase perfeita
OneR	0,41	Concordância quase imperfeita

Fonte: Elaborada pelo autor

Nesse coeficiente Kappa, quanto mais próximo ao valor 1 a concordância se torna mais adequada. Com isso, o algoritmo JRIP se torna o mais adequado com um valor de 0,7211 e o algoritmo OneR o menos adequado com um valor de 0,4094.

Pode-se observar ainda os coeficientes de erro para cada algoritmo. A Tabela 5 mostra que o algoritmo JRIP possui a menor taxa de erro com um valor de 68,2146%.

Tabela 5 - Estatísticas dos erros

ESTATÍSTICAS DOS MODELOS DE CLASSIFICAÇÃO				
ALGORITMO	ERRO ABSOLUTO MÉDIO	RAIZ DO ERRO QUADRÁTICO MÉDIO	ERRO ABSOLUTO TOTAL NORMALIZADO	RAIZ DO ERRO QUADRÁTICO RELATIVO
J48	0,15	0,27	59,47%	78,82%
RamdonTree	0,09	0,26	38,75%	76,00%
Jrip	0,11	0,24	43,66%	68,21%
OneR	0,12	0,33	45,00%	94,92%

Fonte: Elaborada pelo autor

Na aprendizagem supervisionada foram utilizados os algoritmos JRIP (Classificação-regra) e J48 (Classificação-árvore de decisão).

Além da aplicação dos algoritmos de Mineração de Dados, o Weka oferece os Algoritmos de Fragmentação, que tratam da forma que o programa vai interagir com

a Base de Dados. Os dois utilizados são: Cross-Validation (Folds) e Percentage Split. Na Cross-Validation o algoritmo realiza um laço de repetição de iterações, onde os folds são os números de pares e subconjuntos treinamento-teste fornecidos na entrada. Os resultados dos testes geram dados estatísticos, sendo finalizado e gerando as informações. O número de Folds altera o resultado final da aplicação do Algoritmo de Mineração de Dados, dessa forma para se atingir um bom resultado deve-se testar crescentemente ou decrescentemente o número dos mesmos. Na Percentage Split cria-se um subconjunto de treinamento com $i\%$ do tamanho da base de dados fornecida, sendo i a percentagem dada. Basicamente então, trata-se da redução do tamanho da Base de Dados em uma percentagem do seu total, visando à melhoria das estatísticas.

7.2 ALGORITMO: JRIP

Como mostrado na Figura 31 foi selecionado o algoritmo JRip e este obteve como saída um percentual de 93,4% de instâncias classificadas corretamente (com relação à existência de vítimas).

Figura 31 - Saída JRip

```

Classifier output

Time taken to build model: 0.24 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      2072           93.4596 %
Incorrectly Classified Instances    145            6.5404 %
Kappa statistic                    0.7211
Mean absolute error                 0.1085
Root mean squared error            0.2404
Relative absolute error             43.6623 %
Root relative squared error        68.2146 %
Total Number of Instances          2217

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                0,708   0,027   0,817     0,708   0,759     0,724   0,851    0,727    sim
                0,973   0,292   0,951     0,973   0,962     0,724   0,851    0,954    nao
Weighted Avg.   0,935   0,253   0,932     0,935   0,933     0,724   0,851    0,921

=== Confusion Matrix ===

  a  b  <-- classified as
228 94 |  a = sim
 51 1844 |  b = nao

```

Fonte: Elaborada pelo autor

A Figura 32 demonstra que os dados induziram 12 regras de classificação.

Figura 32 - Regras JRip

```

=== Classifier model (full training set) ===

JRIP rules:
=====

(mes = fevereiro) and (periodo = noturno) and (endereço = setor_2) => vitimas=sim (44.0/0.0)
(mes = fevereiro) and (periodo = noturno) and (endereço = setor_3) => vitimas=sim (25.0/3.0)
(mes = fevereiro) and (periodo = noturno) and (fenômeno = ventos_fortes) => vitimas=sim (35.0/5.0)
(mes = fevereiro) and (Cluster = cluster0) and (endereço = setor_4) => vitimas=sim (48.0/15.0)
(ocorrendia_danos = feridos) => vitimas=sim (27.0/0.0)
(ocorrendia_danos = desabrigados) and (periodo = sem_informacao) => vitimas=sim (33.0/5.0)
(periodo = sem_informacao) and (mes = fevereiro) and (ocorrendia_danos = desalojados) => vitimas=sim (7.0/0.0)
(periodo = sem_informacao) and (mes = janeiro) and (endereço = setor_5) and (fenômeno = ventos_fortes) => vitimas=sim (24.0/9.0)
(mes = abril) and (fenômeno = ventos_fortes) and (periodo = noturno) => vitimas=sim (16.0/0.0)
(periodo = sem_informacao) and (mes = fevereiro) and (ocorrendia_danos = inundacoes_graduais) and (endereço = setor_1) => vitimas=sim (5.0/1.0)
(periodo = sem_informacao) and (ocorrendia_danos = vitimas_fatais) => vitimas=sim (7.0/1.0)
=> vitimas=nao (1946.0/90.0)

Number of Rules : 12

```

Fonte: Elaborada pelo autor

A análise das regras geradas permite extrair conhecimento dos dados da base, pode-se notar que as maiores ocorrências de vítimas acontecem em:

- a) Começo do ano: As maiores ocorrências estão entre janeiro e fevereiro ocorrendo vítimas no setor 2 (Figura 42, nos anexos) que é um setor de área mais antiga onde existe um assentamento precário que é a favela da vila Santista e no setor 4 (Figura 44, nos anexos) onde faltam obras de infraestrutura nos diversos conjuntos habitacionais existentes por lá como por exemplo, no núcleo Fortunato Rocha Lima.
- b) Outro dado interessante é quando a ocorrência são ventos fortes e com pessoas sendo desalojadas sempre ocorrem vítimas.
- c) O setor 1 (Figura 41, nos anexos) é um local com um padrão habitacional médio/alto, com grande oferta de comércio e serviços, ocorrem vítimas quando acontecem inundações graduais também prevalecendo no início do ano.

7.3 ALGORITMO: J48

Como ilustrado na Figura 33 foi selecionado na opção Classifier o algoritmo J48.

Figura 33 - Classifier algoritmo J48



Fonte: Elaborada pelo autor.

Como saída o J48 trouxe uma precisão de classificação de 91% como mostra a Figura 34.

Figura 34 - Saída J48

```

Classifier output

Time taken to build model: 0.05 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      2020           91.1141 %
Incorrectly Classified Instances    197            8.8859 %
Kappa statistic                    0.5613
Mean absolute error                 0.1478
Root mean squared error             0.2777
Relative absolute error             59.4703 %
Root relative squared error         78.8242 %
Total Number of Instances          2217

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC   ROC Area  PRC Area  Class
0,472  0,014  0,849   0,472  0,607   0,592  0,817   0,620   sim
0,986  0,528  0,917   0,986  0,950   0,592  0,817   0,953   nao
Weighted Avg.  0,911  0,453  0,907   0,911  0,900   0,592  0,817   0,905

=== Confusion Matrix ===

  a  b  <-- classified as
152 170 |  a = sim
 27 1868 |  b = nao

```

Fonte: Elaborada pelo autor

Com isso, foi possível verificar a classificação dos dados por meio de uma árvore de decisão induzida (Figura 35).

Figura 35 - Árvore

```

período = matutino: nao (149.0/2.0)
período = vespertino: nao (759.0/46.0)
período = noturno
| mes = janeiro: nao (93.0/1.0)
| mes = fevereiro
| | endereço = setor_1
| | | fenômeno = granizo: sim (0.0)
| | | fenômeno = chuvas_fortes: nao (6.0)
| | | fenômeno = ventos_fortes: sim (10.0)
| | | fenômeno = raio: sim (0.0)
| | | fenômeno = chuvas_moderadas: sim (0.0)
| | endereço = setor_2: sim (44.0)
| | endereço = setor_3
| | | fenômeno = granizo: sim (0.0)
| | | fenômeno = chuvas_fortes: sim (22.0)
| | | fenômeno = ventos_fortes: nao (3.0)
| | | fenômeno = raio: sim (0.0)
| | | fenômeno = chuvas_moderadas: sim (0.0)
| | endereço = setor_4: sim (48.0/15.0)
| | endereço = setor_5: nao (6.0)
| | endereço = setor_6: nao (1.0)
| | endereço = setor_7: sim (0.0)
| | endereço = setor_8: sim (0.0)
| | endereço = setor_9: sim (0.0)
| | endereço = setor_10: sim (0.0)
| | endereço = setor_11: nao (1.0)
| | endereço = setor_12
| | | fenômeno = granizo: sim (0.0)
| | | fenômeno = chuvas_fortes: nao (8.0)
| | | fenômeno = ventos_fortes: sim (20.0)
| | | fenômeno = raio: sim (0.0)
| | | fenômeno = chuvas_moderadas: sim (0.0)
| mes = março: nao (26.0/1.0)

```

Fonte: Elaborada pelo autor.

A análise do algoritmo J48 (Figura 35) foi feita utilizando como classe de saída o atributo vítimas, tendo assim como alguns fatos importantes:

- a) Não há ocorrência de vítimas quando o período é matutino ou vespertino, mas existem vários casos onde não é informado em qual período foi ocorrido (sem_informação), supondo então que podem ter ocorridos em qualquer período.
- b) Analisando o período “sem_informação”, pode-se tirar algumas conclusões como, sempre existem vítimas quando há desabrigados e quando tiverem desalojados, ocorrem vítimas quando o mês é fevereiro.
- c) Quando o período é noturno e o mês é fevereiro é possível observar as ocorrências distribuídas por setores (bairros) da cidade, como por exemplo, nos setores mais precários da cidade onde faltam obras de

infraestrutura como no setor 4 (Figura 44, nos anexos) e também nos bairros mais antigos da cidade como no setor 2 (Figura 42, nos anexos) existem vítimas.

- d) É possível observar também que a ocorrência de vítimas sempre está relacionada a chuvas fortes e moderada, raios, granizo e ventos fortes.

Como foi possível notar, os algoritmos JRip e J48 têm resultados parecidos mostrando que é possível a extração de conhecimentos nesses dados.

7.4 ALGORITMO: REGRAS DE ASSOCIAÇÃO

A regra de associação é muito útil para encontrar tendências que possam ser usadas para entender e explorar padrões de comportamento dos dados.

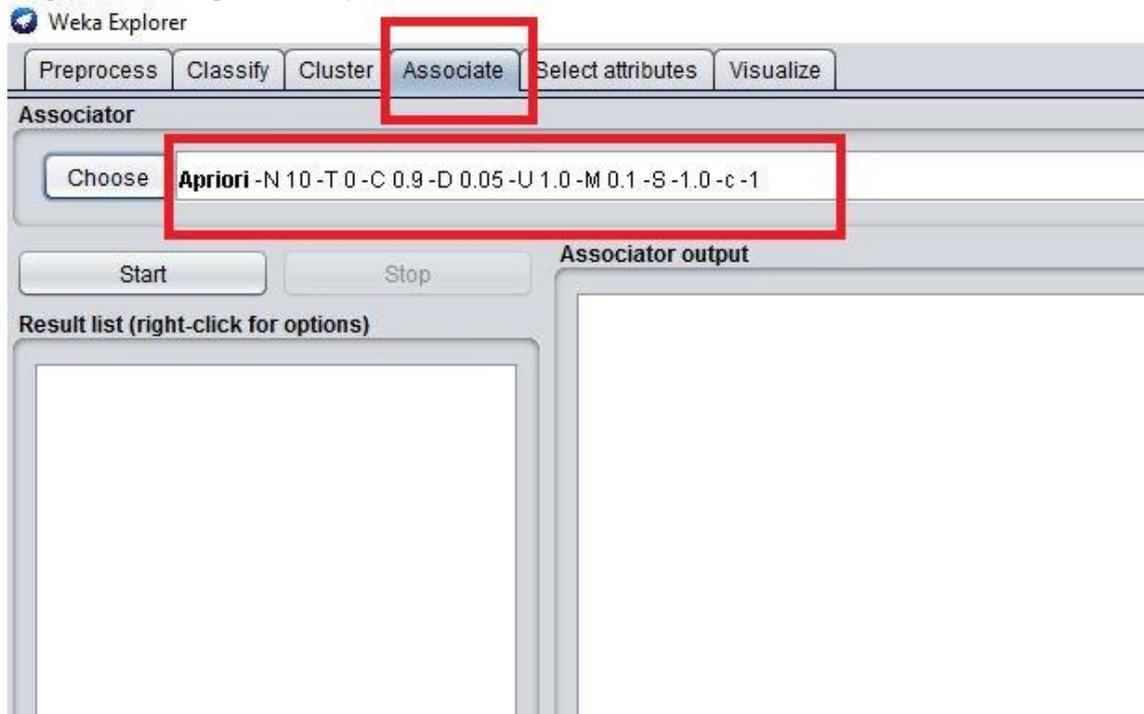
Uma regra de associação é um relacionamento na forma $A \Rightarrow B$, onde A e B são conjunto de itens e $A \cap B = \emptyset$. Existem alguns fatores e índices estatísticos para observar:

- a) Fator de suporte “Sup”: É definido como a razão do número de tuplas que satisfaçam tanto A quanto B sobre o número total de tuplas, isto é, representa a porcentagem de transações da base de dados que contêm os itens de A e B, indicando sua relevância.
- b) Fator de Confiança “Conf”: É definido como a razão do número de tuplas que satisfaçam tanto A quanto B sobre o número de tuplas que satisfazem A, isto é, representa dentre as transações que possuem os itens de A, a porcentagem de transações que possuem também os itens de B, indicando a validade da regra.
- c) Grau de interesse “lift”: Indica o quanto mais frequente torna-se o B, quando A ocorre, isto é, $Lift = Conf(A \cup B) / Sup(B)$.

A tarefa de descoberta de regras de associação consiste na extração de todas as regras com Sup e Conf maior ou igual ao Sup e Conf mínimos (chamados SupMín e ConfMín) especificados pelo usuário.

Foi utilizado o algoritmo Apriori da regra de associação como mostra a Figura 36.

Figura 36 - Algoritmo Apriori



Fonte: Elaborada pelo autor.

Como demonstra na Figura 37, existem algumas regras do algoritmo de associação que merecem destaque:

- a) De acordo com as regras 2, 3, 4, 5 e 10 o cluster 1 é onde ocorrem o maior número de ocorrências de chuvas fortes ocasionando erosões e buracos, assim como demonstrado no algoritmo de classificação por cluster onde, foram constatados que o cluster 1 é o mais “precário” de Bauru.
- b) O cluster 0, onde está localizado o centro da cidade, é onde ocorrem várias inundações ocasionadas por chuvas fortes nas principais avenidas da cidade como mostra na regra 7.
- c) As regras 1 e 9 indicam várias ocorrências de chuvas fortes no cluster 2, porém sem vítimas.

Figura 37 - Saída Apriori

```

Associator output
Minimum support: 0.1 (222 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 18
Size of set of large itemsets L(2): 42
Size of set of large itemsets L(3): 31
Size of set of large itemsets L(4): 9
Size of set of large itemsets L(5): 1

Best rules found:

1. periodo=vespertino fenomeno=chuvas_fortes Cluster=cluster2 237 ==> vitimas=nao 235 <conf:(0.99)> lift:(1.16) lev:(0.01) [32] conv:(11.47)
2. periodo=sem_informacao fenomeno=chuvas_fortes ocorrencia_danos=erosao_buraco 257 ==> Cluster=cluster1 254 <conf:(0.99)> lift:(2.66) lev:(0.07) [158] conv:(40.4)
3. periodo=sem_informacao fenomeno=chuvas_fortes ocorrencia_danos=erosao_buraco vitimas=nao 238 ==> Cluster=cluster1 235 <conf:(0.99)> lift:(2.66) lev:(0.07) [146]
4. periodo=sem_informacao ocorrencia_danos=erosao_buraco 280 ==> Cluster=cluster1 276 <conf:(0.99)> lift:(2.66) lev:(0.08) [172] conv:(35.21)
5. periodo=sem_informacao ocorrencia_danos=erosao_buraco vitimas=nao 260 ==> Cluster=cluster1 256 <conf:(0.98)> lift:(2.65) lev:(0.07) [159] conv:(32.7)
6. periodo=noturno 464 ==> Cluster=cluster0 456 <conf:(0.98)> lift:(2.6) lev:(0.13) [280] conv:(32.04)
7. periodo=noturno fenomeno=chuvas_fortes 278 ==> Cluster=cluster0 271 <conf:(0.97)> lift:(2.58) lev:(0.07) [165] conv:(21.6)
8. periodo=noturno vitimas=nao 311 ==> Cluster=cluster0 303 <conf:(0.97)> lift:(2.57) lev:(0.08) [185] conv:(21.48)
9. fenomeno=chuvas_fortes Cluster=cluster2 271 ==> vitimas=nao 263 <conf:(0.97)> lift:(1.14) lev:(0.01) [31] conv:(4.37)
10. periodo=sem_informacao endereco=setor_5 290 ==> Cluster=cluster1 280 <conf:(0.97)> lift:(2.6) lev:(0.08) [172] conv:(16.58)

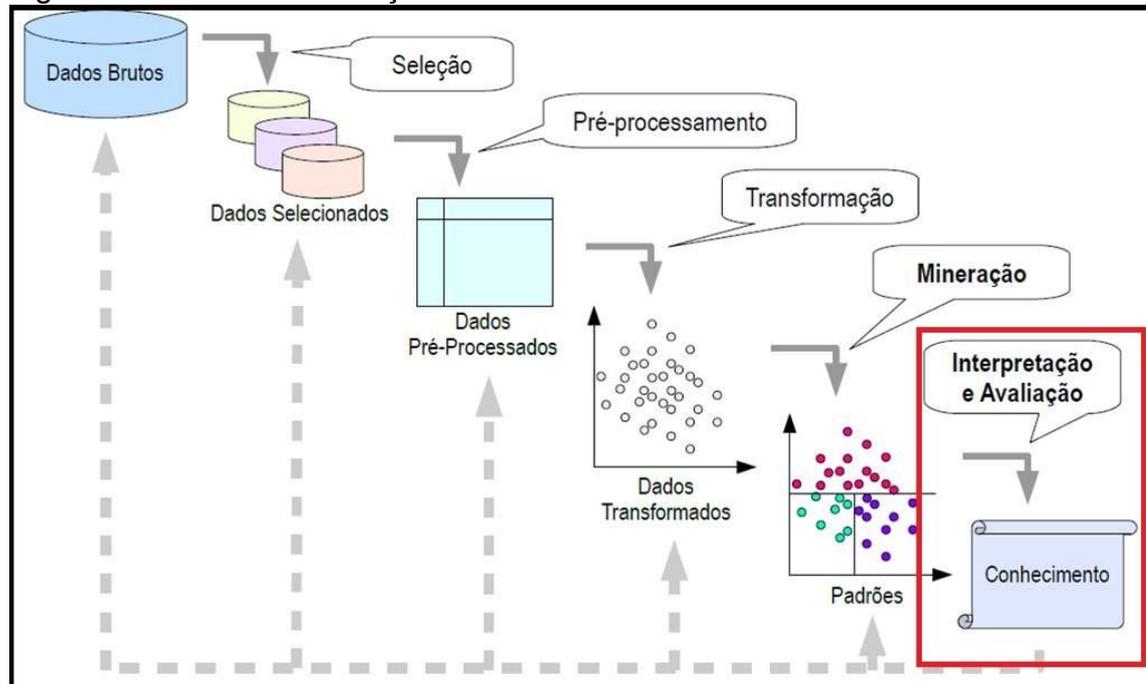
```

Fonte: Elaborada pelo autor.

7.5 INTERPRETAÇÃO E AVALIAÇÃO

O último passo foi a interpretação e avaliação dos resultados obtidos (Figura 38) onde o objetivo foi analisar os resultados a fim de julgar o modelo obtido da fase anterior.

Figura 38 - Passos Mineração - Conclusão



Fonte: Elaborada pelo autor.

A partir dos estudos realizados, foi possível identificar os vários setores problemáticos da cidade de Bauru em relação a desastres naturais, por meio da aplicação de algoritmos de Mineração de Dados implementados pelo WEKA. Os experimentos apresentam uma importante contribuição em termos de quais aspectos são característicos a um determinado setor da cidade, se as condições estruturais do setor influenciam ou não para um determinado fenômeno. Também foi possível notar a importância das etapas do KDD e da Mineração de Dados nesse contexto, pois permitem a realização de estudos detalhados do padrão de fenômenos e ocorrências nos bairros da cidade, além de permitir a avaliação de medidas preventivas para tal problema.

A aplicação das técnicas de Mineração de Dados com o objetivo de descobrir novos conhecimentos auxilia no processo de exploração de uma base de dados, permitindo gerar informações úteis e assim auxiliar às tomadas de decisões. Diante disso, este trabalho apresentou uma proposta de estudo de padrões de desastres naturais na cidade de Bauru.

Foi analisada uma base de dados disponibilizados pelo IPMET, onde foram feitas algumas transformações nos dados, por não estarem padronizadas ou serem irrelevantes, causou muito retrabalho e perda de tempo. O ideal é que o pesquisador faça uma análise prévia da base de dados a ser trabalhada, antes da elaboração de um cronograma ou planejamento de trabalho.

Após o processo de transformação dos dados, foram identificadas algumas tarefas de data mining que poderiam ser aplicadas, como técnicas de Associação e Classificação como os algoritmos Apriori, JRip, J48, utilizando a ferramenta de Mineração de Dados WEKA.

O conhecimento extraído neste trabalho tem como base o estudo de ocorrências relatadas entre o período de janeiro de 1986 a janeiro de 2016 contemplando assim 30 anos.

De acordo com os resultados obtidos, foi possível notar que o fenômeno ocorrido predominante em Bauru é a chuva forte com inundações e ocorrência de erosões e buracos nos setores mais precários da cidade como nos bairros Fortunato Rocha Lima e Nova Esperança localizada no setor 4 (Figura 44, nos anexos), que é um setor de ocupação antiga e Madureira e Parque São Geraldo no setor 5 (Figura 45, nos anexos). Com isso, foi possível observar que o problema de erosões e

buracos tanto reclamados pela população de Bauru nos locais mais precários, vem de longa data e sem esperança de melhorias.

Mesmo não tendo vítimas, as ocorrências se dão sempre no começo do ano, ou seja, em janeiro e fevereiro, que são os meses que tendem a ter altos níveis de chuvas como mostra na Figura 38. No cluster 0 foram identificadas várias ocorrências no setor 12 (Figura 52, nos anexos). É nessas regiões que ocorrem diversas inundações nas principais avenidas quando ocorrem chuvas fortes como mostra na Figura 39, e isso foi comprovado com a mineração desses dados que a principio não tinham nada em comum, eram apenas dados e fatos acumulados e agrupados em um arquivo de texto.

Figura 39 - Clima em Bauru

Clima dos Municípios Paulistas				
Bauru				
Latitude: 22g 11m		Longitude: 49g 2m		Altitude: 530 metros
Classificação Climática de Koeppen: Aw				
MÊS	TEMPERATURA DO AR (C)			CHUVA (mm)
	mínima	média	máxima	
JAN	19.0	25.0	30.0	226.6
FEV	19.0	25.0	30.0	196.1
MAR	19.0	25.0	31.0	143.3
ABR	17.0	23.0	30.0	76.2
MAI	14.0	20.0	27.0	62.3
JUN	13.0	20.0	26.0	52.4
JUL	12.0	19.0	26.0	35.4
AGO	13.0	21.0	29.0	25.5
SET	15.0	22.0	30.0	62.7
OUT	16.0	23.0	30.0	122.8
NOV	17.0	24.0	30.0	129.0
DEZ	18.0	24.0	30.0	198.7
Ano	16.0	22.6	29.1	1331.0
Min	12.0	19.0	26.0	25.5
Max	19.0	25.0	31.0	226.6

Fonte: CENTRO DE PESQUISAS METEOROLÓGICAS E CLIMÁTICAS APLICADAS À AGRICULTURA (CEPAGRI). Campinas, 2016.

Na Figura 40 são demonstrados que bairros próximos a córregos e erosões são os que mais têm problemas causados pela chuva. Na mineração foi comprovada isso, como por exemplo, na Avenida Nações Unidas que está localizada no setor 12 (Figura 52, nos anexos) que é onde ocorrem problemas de alagamento por ser uma área próxima a córregos. Nos setores mais precários de Bauru como os setores 4 e 5 (ANEXO A) que possuem falta de infraestrutura, ruas de terra e residências precárias como, por exemplo, o Parque Jaraguá e Vila São Manoel são onde ocorrem várias erosões e inundações causadas por chuvas fortes.

Figura 40 - Bairros que sofrem com a chuva



Fonte: (FERRARI, 2011).

8 CONCLUSÃO

Com o grande incentivo no mundo à estudos e pesquisas relacionadas a questões ambientais, motivados pela necessidade de informações para tomadas de decisões, tem-se obtido grande volume de dados que não possuem nenhum tratamento ou conhecimento, tornando-se assim, a análise humana uma atividade penosa e métodos tradicionais de recuperação de dados, mesmo que sejam sofisticados, não são eficazes para a obtenção de conhecimentos ocultos em massas de dados.

Por este motivo, o uso de técnicas e ferramentas computacionais é de suma importância para o auxílio na análise de dados, a exemplo da Mineração de Dados. Mineração de Dados (Data Mining) é o processo de exploração de grandes quantidades de dados com o objetivo de encontrar anomalias, padrões e correlações para suportar a tomada de decisões e proporcionar vantagens estratégicas. Com isso, o objetivo desse estudo foi analisar dados sobre desastres naturais e executar algoritmos de Mineração de Dados para gerar algum tipo de conhecimento implícito nos dados. A aplicação das técnicas de Mineração de Dados com o objetivo de descobrir novos conhecimentos auxilia no processo de exploração de uma base de dados, permitindo gerar informações úteis e, assim, subsidiar as tomadas de decisões.

Foi possível notar, através dos algoritmos JRip e Apriori, regras que foram extraídas dos dados de desastres naturais úteis e talvez algumas já esperadas como, por exemplo, bairros antigos que possuem falta de infra estrutura são os que estão mais sujeitos às ocorrências de vítimas por inundações e sempre ocorrendo no período entre janeiro a março, que corresponde período de chuvas em Bauru.

Como trabalhos futuros, podem ser aplicados outros algoritmos de mineração como, por exemplo, redes neurais para poder, assim, analisar novas informações e com isso prever resultados. Podem ser feitas também, análises mais aprofundadas da base de dados, buscando novas formas de trabalha-la para poder aplicar novas técnicas de mineração.

REFERÊNCIAS

AFONSO, D.; GUEDES, R.; MAGALHÃES, L. H. **Mineração de Opiniões de Usuários na busca de conhecimento**. Revista das Faculdades Integradas Vianna Júnior. V.1. 2010.

BURTON, I.; KATES, R. W.; WHITE, G. F. **The environment as hazard**. New York: Oxford Univ. Press, 1978. 240 p.

CARDOSO, T. E.; **Desenvolvimento de um algoritmo Híbrido para Mineração de Dados na Área de Vendas**, Santa Cruz do Sul, Dezembro 2007.

CASTRO, A. L. C. **Manual de desastres: desastres naturais**. Brasília: MIN, 2003. 175 p.

Centro de Pesquisas Meteorológicas e Climáticas Aplicadas à Agricultura (CEPAGRI). Disponível em:< http://www.cpa.unicamp.br/outras-informacoes/clima_muni_068.html> Acesso em: 05 mar. 2016.

CORTES, S. D. C.; PORCARO, R. M.; LIFSCHITZ, S.; **Mineração de Dados – Funcionalidades, Técnicas e Abordagens**, Rio de Janeiro, 2002.

CUNHA, G. R. SCHEEREN, P. L e SILVA, M. S. **Granizo e cereais de inverno no Rio Grande do Sul**. Passo Fundo: Embrapa Trigo, 2001.

DAMASCENO, Marcelo; **Introdução a Mineração de Dados Utilizando o WEKA**. Disponível em:< <http://connepi.ifal.edu.br/ocs/anais/conteudo/anais/files/conferences/1/schedConfs/1/papers/258/public/258-4653-1-PB.pdf>> Acesso em: 27 fev. 2016.

ESTÉBANEZ, B. M. K. Universidade Federal do Rio de Janeiro (UFRJ). Rio de Janeiro, 2012.

FACELI, Katia; LORENA, A. C.; GAMA, João; CARVALHO, A. C. P. L. D. de; **Inteligência Artificial: Uma Abordagem de aprendizado de máquina**. Rio de Janeiro, 2011.

FAYAD, U.; **The KDD process for extracting useful knowledge from volumes of data**. Communications of the ACM, p. 27-34, 1996.

Ferrari, W, **Chuva castiga bairros sem infraestrutura e joga na lama a tranquilidade de seus moradores**, 2011. Disponível< <http://www.jcnet.com.br/Bairros/2011/11/chuva-castiga-bairros-sem-infraestrutura-e-joga-na-lama-a-tranquilidade-de-seus-moradores.html>>acesso em 04/10/2016.

GIL. A. C.; **Como Elaborar Projetos de Pesquisa**. 5.ed. São Paulo: Atlas, 2010.

GORZONI, Priscila, **Os nossos desastres naturais**. Disponível em:<
<http://geografia.uol.com.br/geografia/mapas-demografia/31/artigo178161-4.asp>>
Acesso em: 05 mar. 2016.

HAN, J.; KAMBER, M.; **Data Mining – Concepts and Techniques**. [S.l.]: Morgan kaufmann Publishers, 2001.

HEWITT, K. **The idea of calamity in a technocratic age**. In: HEWITT, K. (Ed.) Interpretations of calamity: from the view point of ecology. London: Allen and Unwin, 1983. p.3-32.

LAROSE, D. T.; **Discovering Knowledge in Data, An Introduction to Data**. [S. l.]: John Wiley & Sons, 2005.

LENAT, D. B., **AM: An artificial intelligence approach to discovery in mathematics as heuristic search**, Stanford University, 1977.

LOUZADA N. F.; DINIZ, C. A. R.; **Data Mining – Uma Introdução**, São Paulo, 2000.

LUGER, G. F.; **Inteligência Artificial: Estruturas e Estratégias para a Solução de Problemas Complexos**. 4ª ed. Porto Alegre: Bookman, 2004. 774 p.

MARCELINO, E. V.; NUNES, L. H.; KOBAYAMA, M. **Banco de dados de desastres naturais: análise de dados globais e regionais**. **Caminhos de Geografia**, v.6, n.19, p.130-149, 2006.

MENDES JR., O; DOMINGUES, M. O. **Introdução à Eletrodinâmica Atmosférica**. Revista Brasileira de Ensino de Física, volume 24, nº1, março.2002.

MIRANDA, **Biboca Ambiental**. Disponível em:<<http://bibocaambiental.blogspot.com.br/2011/07/precipitacao-pluviometrica-chuvas.html>> Acesso em 06 mar. 2016.

MONTEIRO, M. A; FURTADO, S. **O clima do trecho Florianópolis – Porto Alegre: Uma abordagem dinâmica**. **Geosul**, v.10, n. 19/20, p.117-133, 1995.

NAVEGA, Sergio, **Princípios Essenciais do Data Mining**. São Paulo: Anais do Infoimagem 2002, Cenadem, Novembro. Disponível em:<
<http://www.inteliwise.com/reports/i2002.pdf> /> Acesso em: 27 fev. 2016.

NICOLETTI, M.C. **Limiting the Background Knowledge in Inductive Logic Programming**: Notas do ICMSC-USP, Série Computação, Nº 5, Janeiro 1994.

PERBONI, Marcos, **Mineração de Dados na prática com Weka API**. Pontifícia Universidade Católica de São Paulo (PUC-SP). Disponível em:<
<https://marcosvperboni.wordpress.com/2013/02/15/mineracao-de-dados-na-pratica-com-weka-api/>> Acesso em: 27 fev. 2016.

PEREIRA, Silvio do Lago, **Estrutura de Dados Fundamentais Conceitos e aplicações**, 2002

SABA, M. M. F. **A Física das Tempestades e dos Raios**. Física na Escola, São Paulo, volume 2, nº1, 2001.

SANT' ANNA NETO, J. L. **As chuvas no Estado de São Paulo**: Contribuição ao estudo da variabilidade e tendência da pluviosidade na perspectiva da análise geográfica. São Paulo: USP/FFLCH, 1995.

SANTOS, Bruno, **Sobre Bauru**. Disponível em:<
<http://www.vivendobauru.com.br/sobre-bauru-2/>> Acesso em: 05 mar. 2016.

SILVA DIAS, M. A. F. Complexos convectivos de mesoescala. **Climanálise**. Cachoeira Paulista: MCT/INPE/CPTEC, edição especial, p.173-182, 1996.

SILVA, M. P. D. S.; **Mineração de Dados** – Conceitos, Aplicações e Experimentos com Weka, Mossoró, 2005.

SOUZA. T. F. Faculdades Integradas Espírito-Santenses (FAESA). Vitória, Espírito Santo, 2004.

UNIVERSIDADE ESTADUAL PAULISTA, **IPMet - Centro de Meteorologia de Bauru**. Disponível em:<
http://www.ipmet.unesp.br/index2.php?menu_esq1=&abre=ipmet_html/institucional/IPMet_Historia.htm> Acesso em: 05 mar. 2016.

WITTEN, I. H.; FRANK, E.; HALL, M. A. **Data Mining: Practical machine learning tools and techniques**. 3ª. ed. San Francisco: Morgan Kaufmann, 2011.

ANEXOS

ANEXO A

Setor 1 Figura 41

Bairros:

Vila Ascensão, Vila Santa Izabel, Vila Mesquita, Vila Régis, Vila Noemy, Vila América, Vila Santa Tereza (parte), Jardim Nasralla (parte), Vila Marinazan, Vila Samaritana, Jardim Estoril, Vila Ester, Vila Silvia, Vila Santa Clara, Vila Nova Santa Clara, Vila Gorizia, Vila Frutuoso Dias, Residencial Centreville, Jardim Estoril 2, Jardim Estoril 3, Jardim Estoril 4, Jardim Estoril 5, Vila Zillo, Vila Mariana, Calil Rahal, Jardim Paulista, Jardim Amalia, Vila Riachuelo, Vila Leme da Silva, Vila Guedes de Azevedo, Vila Aeroporto de Bauru, Jardim Aeroporto, Jardim América, Núcleo Residencial do Jardim América, Jardim Europa, Mutirão Carmem Carrijo Coube, Jardim Yolanda, Residencial Ilha de Capri, Vila Aviação, Vila Aviação B, Parque Residencial Paineiras, Parque Residencial Samambaia, Residencial Villagio, Residencial Villagio 2, Residencial Villagio 3, Residencial Spazio Verde, Jardim Mary, Jardim Marabá, Residencial Village Campo Novo, Vila Serrão, Residencial Tivoli, Residencial Tivoli 2, Parque das Nações, Residencial Villa Lobos, Jardim Imperial, Parque Panorama, Chácaras Cardoso, Residencial Lago Sul.

Figura 41 - Setor 1



Fonte: Elaborada pelo autor.

Caracterização:

Setor com ocupação mista, com grande oferta de comércio e serviços, padrão habitacional médio/alto, com tendência à verticalização concentrada em áreas específicas, de alto valor imobiliário, ocasionando congestionamento das redes de infraestrutura e sistema viário. Percebe-se nas áreas mais próximas à área central uma tendência de alteração de uso com esvaziamento habitacional. Nas áreas já urbanizadas apresenta problemas de drenagem devido à deficiência das redes de águas pluviais. Nas áreas ainda não ocupadas, que sofrem uma grande pressão imobiliária, existe a preocupação com as taxas de impermeabilização, tendo em vista tratar-se de área situada a montante do canal da Av. Nuno de Assis, com capacidade de vazão limitada. Possui 5 assentamentos precários, sendo prevista a regularização fundiária das favelas V.Zillo e Ilha de Capri pelo Programa Cidade Legal e reassentamento das favelas do J. Europa, J. Yolanda e Parque das Nações na gleba da União que está sendo transferida para o Município. Em função do preço da terra, os investimentos imobiliários lançados nos últimos anos estão voltados para as classes de alto poder aquisitivo, com loteamentos fechados ou edifícios

residenciais de padrão médio/alto, reforçando a segregação social caracterizada por esse setor.

Setor 2 Figura 42
Bairros:

Vila D´Aro (parte), Jardim Central, Jardim Noroeste, Vila Independência, Vila B. Prates, Vila Santa Ignez, Parque São Joaquim, Vila Razuk, Vila Tentor, Jardim Terra Branca, Vila São Francisco, Jardim Eugênia, Vila Santista, Jardim Jandira, Jardim Solange, Residencial Parque Granja Cecília B, Jardim Shangri-Lá, Condomínio Residencial Jardins do Sul.

Figura 42 - Setor 2



Fonte: Elaborada pelo autor.

Caracterização:

A área mais antiga, já consolidada, apresenta predominância residencial de padrão médio/baixo, mas também abriga 2 empreendimentos fechados de padrão alto. Recebeu nos últimos anos muitos empreendimentos verticais populares (Plano PAR e PMCMV), contribuindo com a implantação de infraestrutura na região e impulsionando o mercado imobiliário há muito estagnado. As glebas vazias sofrem pressão imobiliária por empreendimentos de alto padrão, em especial loteamentos fechados, devido à proximidade com o Setor 1. Possui um assentamento precário, favela da Vila Santista, incluída no Programa Cidade Legal para fins de regularização fundiária.

Setor 3 Figura 43**Bairros:**

Residencial Parque Granja Cecília A, Residencial Monte Verde, Jardim Ouro Verde, Jardim Vitória, Jardim Ferraz, Vila Ipiranga, Vila Popular, Vila 9 de Julho, Residencial Parque dos Sabiás, Vila São João do Ipiranga, Vila Paulista, Vila Nova Paulista, Jardim Gaivota, Vila Nipônica, Vila Nova Nipônica, Água do Sobrado, Jardim Jussara, Núcleo Habitacional Joaquim Guilherme de Oliveira, Residencial Parque das Andorinhas, Residencial Parque dos Sabiás, Parque Viaduto, Bosque da Saúde, Vila Nova Celina, Vila Jardim Celina, Parque São João, Vila Alto Alegre, Vila Rocha, Vila Paraíso, Vila Souto, Vila D'Áro (parte), Vila Falcão (parte), Jardim central (parte), Parque Fortaleza, Jardim Ana Lúcia, Vila Carvalho, Área Cemitério São Benedito, Jardim Esplanada, Vila Maria, Prol. Vila Maria, Vila Giunta, Jardim Brasília, Jardim Aracy, Jardim Dalila, Vila Pelegrina, Jardim Faria, Residencial Dr. Manoel Lopes, Vila Dante Alighieri, Jardim de Allah, Quinta Ranieri.

Figura 43 - Setor 3



Fonte: Elaborada pelo autor.

Caracterização:

Caracteriza-se por edificações de padrão médio e conjuntos habitacionais de interesse social. Em meados da década de 90 recebeu grande incremento residencial com a construção dos condomínios denominados Sabiás e Andorinhas e, mais recentemente, dois loteamentos fechados de padrão médio/alto. Possui dois assentamentos irregulares, Favela do J. Vitória e Cutuba, incluídos no Programa Cidade Legal para fins de regularização fundiária.

Foram identificados diversos casos de irregularidade fundiária na Vila Celina, Nova Celina e Parque Viaduto que necessitam de assessoria para ingresso de Ação de Usucapião.

Os investimentos em habitação acarretaram um aumento de densidade no setor não acompanhado por investimentos em sistema viário, acarretando pontos de

estrangulamento e congestionamento, em especial na Praça Primaz Chujiro Otake, intersecção das principais vias coletoras dessa região.

Setor 4 Figura 44

Bairros:

Chácaras Cornélias, Parque Real, Vila Industrial, Vila Santa Terezinha, Vila Pacífico, Vila Pacífico 2, Vila Bela, Vila Martha, Vila Falcão (parte), Jardim Bela Vista (parte), Vila São João da Bela Vista, Jardim Vitória, Vila Quaggio, Jardim Maravilha, Jardim Marise, Vila Cordeiro, Jardim Elydia, Vila Bechelli (parte), Jardim Gerson França, Residencial Vanessa, Parque Sergipe, Vila São Manoel, Jardim Vânia Maria, Chácaras Rodrigueiro, Parque Santa Edwirges, Núcleo 9 de Julho, Parque Jaraguá, Núcleo Fortunato Rocha Lima, Jardim Eldorado, Jardim Andorfato, Jardim Nova Esperança, Jardim Nova Esperança 2, Fundação Casa Popular, Vila São Sebastião, Jardim Rosa Branca, Jardim Prudência, Vila Santa Filomena, Jardim da Grama, Parque Jandaia, Conjunto Habitacional Presidente Eurico Gaspar Dutra, Vila Presidente Eurico Gaspar Dutra, Parque Santa Cândida, Núcleo Leão 13, Parque Val de Palmas, Núcleo Habitacional Vereador Edson Francisco da Silva.

Figura 44 - Setor 4



Fonte: Elaborada pelo autor.

Caracterização:

Setor de ocupação antiga (a Vila Falcão foi o primeiro núcleo urbano fora da área do Patrimônio) apresenta, ainda, muitas habitações em madeira ou em estado precário nas áreas próximas à ferrovia. É cortado por diversos cursos d'água e duas linhas férreas, o que dificulta a interligação entre os bairros. A predominância é de ocupação residencial de padrão médio/baixo, sem as obras de infraestrutura (redes de drenagem e pavimentação), gerando uma demanda ainda não totalmente atendida pelo poder público. Possui muitos conjuntos habitacionais de interesse social e um grande mutirão, Fortunato Rocha Lima, implantado em 1994/1996 destinado ao desfavelamento, com 536 unidades, foco de muitos problemas sociais. Possui sete assentamentos precários, identificadas no PDP para fins de regularização fundiária ou reassentamento, de acordo com suas características. Para reassentamento da favela do Parque Real, considerada de risco, estão sendo construídas 34 unidades habitacionais com recursos do FNHIS, já em fase de conclusão. Para as favelas do J. Marise e J. Gerson França, a Prefeitura já dispõe

de recursos do FNHIS para construção de UH para reassentamento de parte dos moradores e regularização fundiária dos demais. O mesmo deverá ocorrer nas favelas do Parque Jaraguá e Andorfato por meio do Programa MCMV e Cidade Legal. As favelas São Manoel e Santa Filomena ainda dependem de estudos específicos.

Setor 5 Figura 45
Bairros:

Bairro Santa Fé, Parque Primavera, Parque Roosevelt, Jardim Coral, Núcleo Habitacional Alto Alegre, Jardim Petrópolis, Jardim Progresso, Vila Cidade Jardim, Vila Nova Marajoara, Vila Marajoara, Vila Santa Rosa, Vila São José, Vila Bom Jesus, Residencial Francisco Lemos de Almeida, Jardim Imperatriz, Parque União, Vila Lemos, Vila Bechelli (parte), Vila Gonçalves, Jardim Bela Vista (parte), Vila Camargo, Vila Seabra, Vila Formosa, Vila Vergueiro, Jardim Santana, Bairro Madureira, Jardim Araruna, Parque Vista Alegre, Parque Floresta, Vila Gimenes, Vila Formosa, Parque Alto Sumaré, Jardim Hojas, Parque Residencial do Castelo, Jardim Godoy, Jardim Maria Angélica, Jardim Fonte do Castelo, Parque King, Jardim Jose Kallil, Jardim Nossa Senhora de Lourdes, Jardim Jacyra, Parque Novo São Geraldo, Jardim Maria Célia, Jardim Mainichi, Parque São Geraldo, Parque São Cristóvão, Residencial Santa Cecília, Vila Garcia, Jardim Marília, Jardim Nova Marília, Jardim TV.

Figura 45 - Setor 5



Fonte: Elaborada pelo autor.

Caracterização:

Trata-se de um setor com ocupação antiga, já consolidada com habitações de padrão médio, que apresenta grande número de casas de madeira e habitações em estado precário que necessitam de melhorias. No entanto, também possui grandes glebas (terreno não urbanizado) ainda não parceladas nas áreas mais periféricas em virtude da dificuldade de acesso.

Setor 6 Figura 46

Bairros:

Residencial Vitória Régia, Residencial Nova Bauru, Parque Residencial Pousada da Esperança, Vila São Paulo, Jardim Helena, Núcleo Habitacional

Vanuire, Núcleo Habitacional Edison Bastos Gasparini, Parque City, Residencial Parque Colina Verde.

Figura 46 - Setor 6



Fonte: Elaborada pelo autor.

Caracterização:

Com limites bem definidos em virtude das barreiras físicas das rodovias estaduais SP 300 e SP 321, esse setor fica isolado da malha urbana principal. A ocupação é predominantemente residencial de padrão baixo, com pouca oferta de comércio, serviço e indústria, o que força a população a grandes deslocamentos na busca por postos de trabalho gerando conflitos entre o trânsito urbano (em especial de pedestres e ciclistas) e o trânsito rodoviário, muito intenso nesse trecho da área

urbana. Possui um assentamento precário, Pousada da Esperança, com previsão de regularização fundiária pelo Programa Cidade Legal.

Setor 7 Figura 47

Bairros:

Jardim Ivone, Chácara das Flores, Chácaras Gigo, Núcleo Habitacional Nobuji Nagasawa, Parque das Perdizes, Residencial Nova Flórida, Jardim Flórida, Novo Jardim Pagani, Jardim Silvestre, Parque dos Eucaliptos, Núcleo Habitacional Eldorado, Parque Residencial Jardim Araruna, Vila Conceição, Vila Santa Luzia, Vila Nova Santa Luzia, Núcleo Habitacional Beija Flor.

Figura 47 - Setor 7



Fonte: Elaborada pelo autor.

Caracterização:

Setor ocupado em grande parte por habitações de interesse social, também sofre com o isolamento causado pela barreira das rodovias e poucas opções de interligação entre os bairros. Apresenta pequenos pólos de comércio e serviços e está próximo de uma indústria de grande porte (frigorífico) que oferece opção de emprego.

Setor 8 Figura 48**Bairros:**

Jardim Benícia, Jardim Jamil, Jardim Mendonça, Jardim Chapadão, Parque Rossi, Núcleo Habitacional Mary Dota, Chácaras São João, Conjunto Habitacional Isaura Pitta Garms, Parque Giansante, Loteamento Mário Luiz Rodrigues do Prado, Quinta da Bela Olinda.

Figura 48 - Setor 8



Fonte: Elaborada pelo autor.

Caracterização:

Esse setor apresenta características bem distintas. Na parte norte, possui densidade muito baixa em virtude da dimensão dos lotes da Quinta da Bela Olinda e da falta de infraestrutura no bairro, que não estimula sua ocupação. O Loteamento Mário Luiz Rodrigues do Prado (lotes urbanizados), empreendimento executado pela Prefeitura no início da década de 90, esteve embargado em função de demandas judiciais. Devido ao tempo decorrido, a infraestrutura instalada encontra-se comprometida, necessitando de grandes investimentos para sua recuperação. Parte das glebas vazias pertence ao poder público e COHAB, mas sem possibilidade de utilização para empreendimentos residenciais. A densidade maior ocorre nos bairros

situados ao sul, parte do Núcleo Mary Dota e Isaura Pitta Garms, além de loteamentos populares.

Setor 9 Figura 49
Bairros:

Núcleo Habitacional Engenheiro Octavio Rasi, Distrito Industrial Domingos Biancardi, Parque Paulista, Chácaras Betânia, Ferradura Mirim, O Ferradura, Jardim Nova Bauru, Parque Júlio Nóbrega, Jardim Country Club, Parque Bauru, Parque Bauru Mirim, Núcleo Habitacional Pastor Arlindo Lopes Viana, Núcleo Habitacional Luiz Edmundo Coube, Vila Tecnológica Engenheiro José Queda, Núcleo Habitacional José Regino, Bairro dos Tangarás, Parque Industrial Manchester, Parque Santa Terezinha, Distrito Industrial 2, Parque Baurulândia, Vila Aimorés, Jardim São Judas, Parque Ecológico, UNESP, Jardim Alto Bauru, Parque Santa Rita.

Figura 49 - Setor 9



Fonte: Elaborada pelo autor.

Caracterização:

Com características populares, o Setor 9 apresenta deficiência de infraestrutura de drenagem que compromete, ainda mais, a acessibilidade aos lotes, em especial aos Bairros Tangarás e

Manchester, isolados da malha urbana pela barreira da rodovia. Possuem três assentamentos irregulares, Parque Santa Terezinha, Vila Aimoré e Ferradura (o maior da cidade, com aproximadamente 1200 famílias), todos incluídos no Programa Cidade Legais para fins de regularização fundiária.

Setor 10 Figura 50**Bairros:**

Jardim Guadalajara, Horto Florestal, Mutirão Primavera, Núcleo Habitacional Jardim Redentor, Jardim Cecap, Núcleo Habitacional Bom Samaritano, Jardim Rosas do Sul, Vila Monlevade, Vila Cardia, Residencial do Bosque, Jardim Cruzeiro do Sul, Vila Carmem, Jardim Carvalho, Jardim Carolina, Residencial Moriah, Parque Paulistano, Parque São Jorge, Jardim Marambá, Vila Galvão (parte), Vila Engler, Parque Residencial das Camélias, Parque Residencial Flamboyant, Jardim Alvorada, Parque Hipódromo, Jardim Olímpico, Jardim da Orquídea, Núcleo Habitacional Presidente Ernesto Geisel, Jardim Auri Verde, Jardim do Contorno, Jardim Samburá, Residencial Odete, Jardim Colonial, Residencial Sauípe, Jardim Nicéia, Jardim Santos Dumont, Vila Aviação B.

Figura 50 - Setor 10



Fonte: Elaborada pelo autor.

Caracterização:

Setor com ocupação predominantemente popular, contendo alguns conjuntos habitacionais de grande porte (Redentor e Geisel). Recebeu, a partir do final da década de 80, empreendimentos verticais de padrão médio (Camélias, Flamboyant, J. dos Duques, Vila Verde e Vila Grená, Campo Limpo, Campo Belo, entre outros) que mudaram a configuração do setor, atraindo comércio e serviços para atender a nova demanda. Possui dois assentamentos precários, Jardim Nicéia e Jardim Olímpico, com proposta de regularização fundiária por meio do Programa Cidade Legal. Tem grande potencial de crescimento populacional em função da existência de glebas vazias passíveis de ocupação. No entanto, algumas dessas áreas, ocupadas por vegetação de cerrado, tem restrições à ocupação. Apresenta algumas

dificuldades de acesso e mobilidade pela deficiência do sistema viário, que poderá ser amenizado com a implantação de obra viária ao longo do Córrego da Água Comprida.

Setor 11 Figura 51**Bairros:**

Vila Antártica, Chácara das Flores, Vila das Flores, Vila Vieira, Jardim Avenida, Vila Cardia, Vila Bonfim, Vila Santos Pinto, Bairro Alto Higienópolis, Vila Santo Antônio, Vila Yara, Vila Santa Lúcia, Vila Brunhari, Jardim Brasil, Jardim Deluiggy, Vila Galvão (parte), Vila Maracy, Vila Nova Cidade Universitária, Vila Nova Nise, Vila Nise, Vila Perroca, Vila Altinópolis (parte), Vila Santa Tereza (parte), Jardim Pagani, Jardim Nasralla, Vila Ferraz, Vila Reis, Vila Guedes de Azevedo, Jardim Dona Sarah, FOB-USP, Vila Cidade Universitária, Jardim Panorama, Jardim Planalto, Jardim Infante Dom Henrique, Vila Regina.

Figura 51 - Setor 11



Fonte: Elaborada pelo autor.

Caracterização:

É o setor mais densamente ocupado e estruturado em termos de infraestrutura instalada, oferta de comércio e serviços. O crescimento populacional tem se dado pela intensa verticalização, em especial na região do Bauru Shopping, criando problemas de abastecimento e de circulação viária. A ocupação residencial é de padrão médio/alto, exceto na região da Vila Cardia e Vila Antártica, bairros mais antigos, com habitações de padrão mais modesto, que deverão sofrer impacto com a implantação de empreendimentos comerciais de grande porte na vizinhança.

Setor 12 Figura 52
Bairros:

Área Central.

Figura 52 - Setor 12



Fonte: Elaborada pelo autor.

Caracterização:

Setor delimitado pelo leito ferroviário e principais avenidas da cidade, é servido de completa infraestrutura e caracteriza-se como área predominantemente comercial e de serviços, com forte tendência de esvaziamento habitacional. A verticalização ocorre de maneira esparsa, sem comprometimento de salubridade (insolação e ventilação) como ocorre em outros setores da cidade.