



UEPB

**UNIVERSIDADE ESTADUAL DA PARAÍBA
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA E TECNOLOGIA EM SAÚDE**

FERNANDO MEDEIROS FILHO

**REGRAS DE ASSOCIAÇÃO PARA ANÁLISE DE CASOS DE DENGUE EM
MUNICÍPIOS PARAIBANOS**

**CAMPINA GRANDE
2017**

FERNANDO MEDEIROS FILHO

**REGRAS DE ASSOCIAÇÃO PARA ANÁLISE DE CASOS DE DENGUE EM
MUNICÍPIOS PARAIBANOS**

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Ciência e Tecnologia em Saúde da Universidade Estadual da Paraíba, em cumprimento aos requisitos necessários para obtenção do título de Mestre.

Orientador: Prof. Dr. Wellington Candeia de Araújo

**CAMPINA GRANDE
2017**

É expressamente proibida a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano da dissertação.

M488r Medeiros Filho, Fernando.

Regras de associação para análise de casos de dengue em municípios paraibanos [manuscrito] / Fernando Medeiros Filho. - 2017.

70 p. : il. color.

Digitado.

Dissertação (Mestrado Profissional em Ciência e Tecnologia em Saúde) - Universidade Estadual da Paraíba, Pró-Reitoria de Pós-Graduação e Pesquisa, 2017.

"Orientação: Prof. Dr. Wellington Candeia de Araújo, Pró-Reitoria de Pós-Graduação e Pesquisa".

1. Mineração de dados. 2. Epidemiologia. 3. Dengue. I.
Título.

21. ed. CDD 614.4

FERNANDO MEDEIROS FILHO

REGRAS DE ASSOCIAÇÃO PARA ANÁLISE DE CASOS DE DENGUE EM
MUNICÍPIOS PARAIBANOS

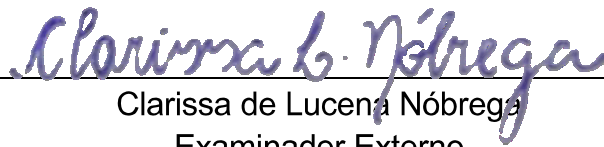
Dissertação de mestrado aprovada como
requisito necessário para obtenção do título
de Mestre no Programa de Pós-Graduação
em Ciência e Tecnologia em Saúde da
Universidade Estadual da Paraíba.

Aprovada em 03 de Julho de 2017

Banca Examinadora:



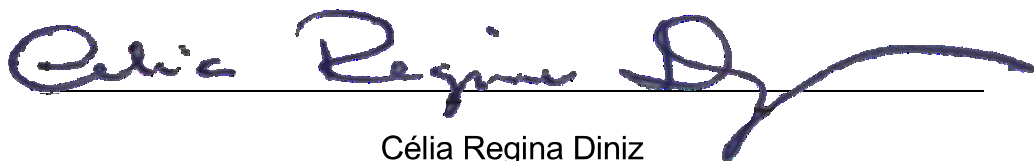
Prof^a Dr. Marcelo Ribeiro da Silva
Examinador Externo



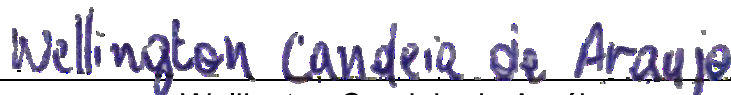
Clarissa de Lucena Nóbrega
Examinador Externo



Prof. Dr. Nadja Maria da Silva Oliveira Brito
Examinador da UEPB



Célia Regina Diniz
Examinador da UEPB



Wellington Candeia de Araújo
Orientador

Campina Grande - PB
2017

AGRADECIMENTOS

Agradeço primeiramente a DEUS, já que Ele colocou pessoas tão especiais a meu lado, sem as quais certamente não teria dado conta!

Agradeço a Universidade Estadual da Paraíba (UEPB) por ter me acolhido desde a graduação e, agora, no mestrado (serei eternamente grato); também sou grato ao Programa de Pós-Graduação em Ciência e Tecnologia em Saúde e ao Núcleo de Tecnologias Estratégicas em Saúde (NUTES) por me formarem uma pessoa melhor em todos os sentidos. Levarei com grande honra o seu nome por onde andar.

Agradeço principalmente a minha família por terem me apoiado e ficarem ao meu lado nas horas que eu mais precisava. Dentre eles posso citar Renilda (minha mãe) Fabiana (minha irmã) e Fernando (meu pai). A vocês expesso o meu maior agradecimento.

Agradeço imensamente a meu orientador, Wellington Candeia de Araújo, que foi a melhor pessoa do mundo para se trabalhar. Sua inteligência, humildade e bom senso são admiráveis.

A Rafael Rodrigues, por ser tão importante na minha vida. Sempre a meu lado, me pondo para cima e me fazendo acreditar que posso mais que imagino. Obrigado por ter feito do meu sonho, o nosso sonho!

À Prefeitura Municipal de São Mamede, meu local de trabalho, que sempre foram flexíveis com minha rotina trabalho/estudo. Entre essas pessoas Chagas Lopes, Umberto Jefferson, Marcelino Elizeu (eterno chefe) e Rafaela Macedo (prima querida).

*“Ao meu avô Pedro Teté, suas histórias sempre me motivaram.
Você permanecerá eternamente minha lembrança e,
principalmente em meu coração.”*

In memoriam

RESUMO

O objetivo deste trabalho é aplicar as técnicas de Mineração de Dados nas informações contidas nos boletins epidemiológicos do portal do Ministério da Saúde que dizem respeito às informações de Dengue de 18 municípios do Estado da Paraíba. Através desse volume de dados foram aplicadas regras de associação por meio de análises com o algoritmo APRIORI. As técnicas foram usadas através do *software* livre WEKA e o algoritmo por ele implementado para gerar as regras de associação. Ao longo do processo foi feita a visualização, o reconhecimento e a organização dos padrões descobertos através de mapas, encontrando relações entre as treze análises que foram realizadas ao longo do trabalho. Elas envolveram as 52 semanas epidemiológicas e a partir disso encontradas regras que são aplicáveis aos programas de combate a Dengue. Os padrões identificados fizeram parte da grande proposta dessa dissertação - Direcionar as ações de combate à Dengue nos municípios analisados. Levando a concluir que ações que envolvem a Vigilância Sanitária e Ambiental dos municípios em questão podem ter um melhor desempenho se as ações forem interiorizadas, segmentadas e adaptadas para a realidade encontrada.

Palavras-chave: Mineração de dados. Epidemiologia. Dengue.

ABSTRACT

The objective of this work is to apply the techniques of Data Mining in the information contained in epidemiological bulletins of the portal of the Ministry of Health that concern the information of Dengue of 18 municipalities of the State of Paraíba. Through this volume of data association rules were applied through analyzes with the APRIORI algorithm. The techniques were used through the free software WEKA and the algorithm implemented by it to generate association rules. Throughout the process the visualization, recognition and organization of the patterns discovered through maps were made, finding relationships among the thirteen analyzes that were carried out at the time of work. They involved the 52 epidemiological weeks and from that found rules that are applicable to programs to combat Dengue. The identified patterns were part of the great proposal of this dissertation - To direct the actions to combat Dengue in the analyzed municipalities. Leading to conclude that actions involving Sanitary and Environmental Surveillance of the municipalities in question can perform better if the actions are internalized, segmented and adapted to the reality found.

Keywords: Data Mining. Epidemiology. Dengue.

LISTA DE ILUSTRAÇÕES

Figura 1: Mapa global da distribuição prevista de <i>Aedes aegypti</i>	15
Figura 2: Incidência de Dengue, CHKV e ZIKV em 2016.....	16
Figura 3: Visão geral das etapas do processo KDD.....	18
Figura 4: Ilustração dos principais termos na MD.....	21
Figura 5: Elementos principais da MD.....	22
Figura 6: Tarefas da Mineração de Dados.....	26
Figura 7: Representação de uma rede neural tipo MLP.....	27
Figura 8: Representação do software WEKA.....	31
Figura 9: Representação da parte “Explorer” do WEKA.....	31
Figura 10: Pseudocódigo do algoritmo APRIORI.....	34
Figura 11: Parâmetros ajustados no WEKA.....	35
Figura 12: Distribuição da amostra selecionada para estudo.....	37
Figura 13: Mesorregiões da Paraíba.....	38
Figura 14: Tela de análise do ATOM com os atributos utilizados.....	40
Figura 15: Municípios com padrões nos casos de Dengue entre as SE 5-8.....	42
Figura 16: Municípios com padrões nos casos de Dengue entre as SE 9-12.....	48
Figura 17: Municípios com padrões nos casos de Dengue entre as SE 13-16.....	51
Figura 18: Municípios com padrões nos casos de Dengue entre as SE 17-20.....	53
Figura 19: Municípios com padrões nos casos de Dengue entre as SE 21-24.....	56
Figura 20: Municípios com padrões nos casos de Dengue entre as SE 25-28.....	58
Figura 21: Municípios com padrões nos casos de Dengue entre as SE 29-32.....	60
Figura 22: Infográfico com a representação das regras geradas.....	64

LISTA DE QUADROS

Quadro 1: Resumo dos elementos básicos da MD	23
Quadro 2: Resumo dos principais autores sobre os métodos de MD	24
Quadro 3: Calendário de notificação para o ano de 2017	39
Quadro 4: Atributos utilizados nas análises	40
Quadro 5: Casos de dengue na 1ª Semana Epidemiológica entre (2007-2016).....	41
Quadro 6: Casos de dengue na 2ª Semana Epidemiológica entre (2007-2016).....	42
Quadro 7: Casos de dengue na 4ª Semana Epidemiológica entre (2007-2016).....	42
Quadro 8: Casos de dengue na 4ª Semana Epidemiológica entre (2007-2016).....	43
Quadro 9: Algumas regras de associação geradas entre as Semanas 5-8	45
Quadro 10: Exemplificação de uma regra de associação	45
Quadro 11: Algumas regras de associação geradas entre as Semanas 9-12	48
Quadro 12: Algumas regras de associação geradas entre as Semanas 13-16	51
Quadro 13: Algumas regras de associação geradas entre as Semanas 17-20	53
Quadro 14: Algumas regras de associação geradas entre as Semanas 21-24	56
Quadro 15: Algumas regras de associação geradas entre as Semanas 25-28	58
Quadro 16: Algumas regras de associação geradas entre as Semanas 29-32	60
Quadro 17: Resumo dos municípios selecionados ao longo das 13 análises.....	62

LISTA DE GRÁFICOS

Gráfico 1: Média Pluviométrica entre os municípios retornados na análise 2	47
Gráfico 2: Média Pluviométrica entre os municípios retornados na análise 3	50
Gráfico 3: Média Pluviométrica entre os municípios retornados na análise 4	52
Gráfico 4: Média Pluviométrica entre os municípios retornados na análise 5	55
Gráfico 5: Média Pluviométrica entre os municípios retornados na análise 6	57
Gráfico 6: Média Pluviométrica entre os municípios retornados na análise 7	59
Gráfico 7: Média Pluviométrica entre os municípios retornados na análise 8	61

SUMÁRIO

1 INTRODUÇÃO	12
2 REFERENCIAL TEÓRICO	14
2.1 Saúde, epidemiologia e Dengue	14
2.2 Aspectos técnicos em Computação	18
2.2.1 Descoberta do Conhecimento em Banco de Dados	18
2.2.2 Mineração de Dados	20
2.2.2.1 Como funciona a Mineração de Dados	22
2.2.2.2 Tarefas de Mineração de dados	23
2.2.2.3 Técnicas de Mineração de dados	26
2.2.2.4 Metodologias de aplicação em Mineração de Dados	27
2.2.2.5 Limitações	28
2.3 Síntese da Sessão	29
3 MATERIAIS E MÉTODOS	30
3.1 Fonte de Dados	30
3.2 Recursos de Software	30
3.3 Transformação dos dados	32
3.4 Regras de associação	32
3.4.1 Algoritmo APRIORI	33
3.5 Seleção dos atributos	34
3.6 Regras de associação no WEKA	35
3.7 Pós-processamento dos Dados	36
3.8 Síntese da Sessão	36
4 RESULTADOS E DISCUSSÕES	37
4.1 Análise 1: Semanas Epidemiológicas 1 a 4	41
4.2 Análise 2: Semanas Epidemiológicas 5 a 8	44
4.3 Análise 3: Semanas Epidemiológicas 9 a 12	48
4.4 Análise 4: Semanas Epidemiológicas 13 a 16	50
4.5 Análise 5: Semanas Epidemiológicas 17 a 20	53
4.6 Análise 6: Semanas Epidemiológicas 21 a 24	55
4.7 Análise 7: Semanas Epidemiológicas 25 a 28	57
4.8 Análise 8: Semanas Epidemiológicas 29 a 32	59
4.9 Análise 9, 10, 11, 12 e 13: Semanas Epidemiológicas 33-36, 37-40, 41-44, 45-48, 49-52	61
4.10 Considerações a respeito dos resultados obtidos	61
4.11 Direcionamento das campanhas de Dengue	62

4.12 Síntese da Sessão	65
5 CONSIDERAÇÕES FINAIS	66
REFERÊNCIAS	67

1 INTRODUÇÃO

Os vírus transmitidos por artrópodes (arbovírus) estão entre as causas mais importantes de doenças infecciosas emergentes ou reemergentes em todo o mundo (CARDOSO et al., 2015). Entre eles, o mosquito *Aedes aegypti* tem representado um grande problema à Saúde Pública através de doenças, como: Dengue, Zika, Chikungunya, entre outras. Controlar esses vírus representa um grande desafio, principalmente pela demanda de ferramentas de monitoramento e controle que visam potencializar a busca ativa e passiva do mosquito (MORRISON et al., 2004; WHO, 1997).

Em 2017, até a Semana Epidemiológica 19, a região Nordeste registrou o maior número de casos prováveis de Dengue (45.431 casos; 31,5%) em relação ao total do país, o que mostra a carência de ferramentas de combate ao mosquito como controle químico, epidemiológico e tecnológico, sendo o último uma necessidade atual e foco deste trabalho.

Tendo em vista o advento das tecnologias, é notório que a informação se tornou um elemento de grande valia no contexto global. Extrair conhecimentos de fontes de dados cria um diferencial na perspectiva estratégica nas mais diversas esferas corporativas. Neste contexto, a utilização dos dados com recursos de computação natural, como algoritmos genéticos (AG) e Mineração de Dados (MD) mostram-se úteis para detectar e implementar mecanismos de prevenção e controle de epidemias (SOUZA; VIANNA; MORAES, 2007). Algumas características como dados abertos e disponíveis na nuvem, além de tecnologias para análise rápida de grandes volumes de dados também contribuem para o aumento destas técnicas em aplicações em saúde pública (KRAEMER, 2015).

Este trabalho tem como hipótese a utilização de técnicas de MD sob uma análise associativa das informações contidas nos boletins epidemiológicos, tendo o intuito de encontrar relações atípicas ou típicas, resumindo os dados em novas formas que podem servir como ferramentas tanto compreensíveis quanto úteis para saúde pública. Tais métodos vêm sendo utilizados em diversas áreas, tendo em vista sua extensa aplicabilidade em resultados válidos, implementáveis e de grande importância na resolução de problemas.

Geralmente as bases de dados que são utilizadas na MD podem ter milhões de registros e milhares de variáveis. É improvável que todas as variáveis sejam independentes e sem uma estrutura de correlação entre elas.

A realização desse estudo é relevante sob diferentes perspectivas, entre as quais se destacam: a gravidade em que se encontram os casos de Dengue no país, o alto grau de variantes das doenças, os grandes investimentos em saúde pública e o caráter social, político, humano, ambiental e econômico envolvidos na realização de atividades que contribuam para a diminuição da doença.

Sendo assim, o objetivo geral deste trabalho é aplicar as técnicas de Mineração de Dados nas informações sobre a Dengue contidas nos boletins epidemiológicos do portal do Ministério da Saúde em 18 municípios paraibanos, aplicando regras de associação por meio de análises do algoritmo APRIORI.

Os objetivos específicos deste trabalho é analisar os boletins epidemiológicos do Ministério da Saúde e seus dados existentes no estado da Paraíba; aplicar o algoritmo APRIORI nos dados obtidos; visualizar, reconhecer e organizar os padrões encontrados nos municípios selecionados para direcionar as ações de combate ao mosquito com base nas cinquenta e duas SE e seus meses correspondentes e direcionar as ações de combate à Dengue nos municípios analisados.

2 REFERENCIAL TEÓRICO

2.1 Saúde Pública, Epidemiologia e Dengue

A Dengue é transmitida pelo mosquito da família *Flaviviridae* e do gênero *Flavivirus* (GUHA-SAPIR; SCHIMMER, 2005). Seus vetores principais são o *Aedes aegypti*, que prefere crescer em áreas urbanas, e o *Aedes albopictus*, que não está tão adaptado aos ambientes domésticos, preferindo o ambiente natural, possuindo morfologia e capacidade proliferativa semelhantes ao *Aedes aegypti*, mas foi responsável apenas por alguns surtos da doença em países do continente asiático (RODHAIN; ROSEN, 1997).

Os episódios do *Aedes aegypti* foram primeiro descritos no Egito em 1762 por Linnaeus, sendo o mosquito presente em praticamente todo o continente americano, no Sudeste da Ásia, e em toda a Índia. Notou-se que o ingresso dessa espécie no Brasil ocorreu no período colonial, entre os séculos XVI e XIX, durante o comércio de escravos. Devido à destruição do *habitat* natural, uma parte da população silvestre sofreu um processo seletivo que beneficiou a dispersão e sobrevivência da espécie nos centros urbanos (ROZEBOOM, 1960).

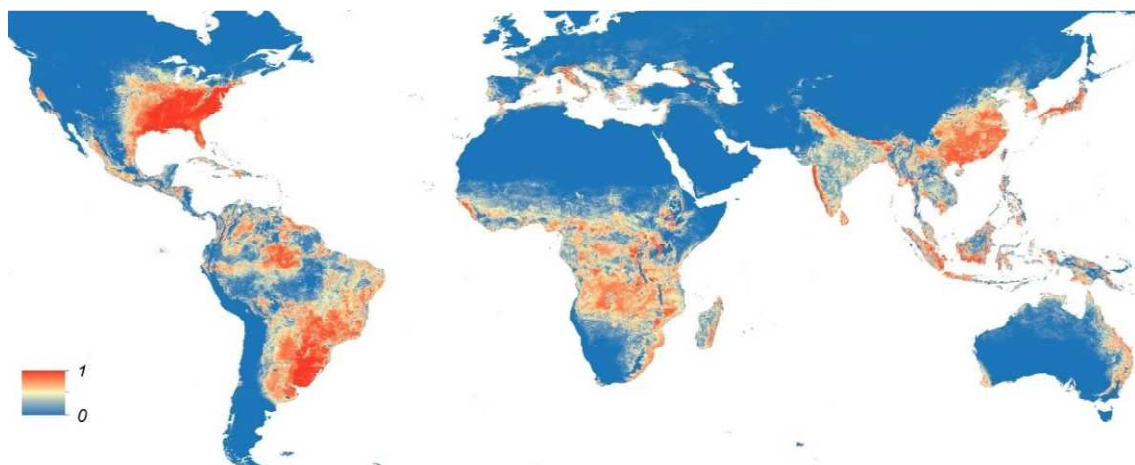
No Brasil, documenta-se que a primeira epidemia clínica e laboratorialmente comprovada ocorreu entre 1981 e 1982 em Boa Vista, Roraima. Ainda nos anos 80 aconteceu epidemia em grandes proporções no Rio de Janeiro que atingiu quase toda a cidade. Nesse ponto em diante, epidemias ocorreram em diversos estados e começaram a ficar constantes (DIAS, 2010).

O habitat do *Aedes aegypti* influencia como potencial área de transmissão da Dengue e outras doenças relacionadas. Elas ocorrem em zonas urbanas e densamente povoadas e/ou, ocasionalmente, encontrados em áreas rurais. A migração das zonas rurais para as áreas urbanas inclui alguns fatores, como, por exemplo, o abastecimento de água e sua contribuição para a proliferação de mosquitos (CARMO, 2015). No Nordeste do Brasil, mais de 75% dos locais de reprodução se devem ao armazenamento de água precário, enquanto que na região Sudeste a maioria dos locais de reprodução ocorrem em outros vícios como o acúmulo de água em jarros, vasos de plantas e calhas do telhado (BRASIL, 2016)

Mesmo diante do grande empenho no combate ao mosquito, ainda tem sido um desafio controlá-lo. As grandes campanhas fazem seu papel informacional, mas a doença ainda mostra dados preocupantes. Os resultados negativos não devem apenas se dar a problemas de conscientização, há uma discrepância entre a infraestrutura das cidades, abastecimento de água e sistema de coleta de lixo, pois esses fatores são cruciais para comprometer a validação do método de combate (GUBLER, 1989)

A Figura 1 mostra onde pode ser encontrado o mosquito *Aedes aegypti*, reforçando a ideia que lugares tropicais são tendenciosos a proliferação do mosquito. O mapa mostra a probabilidade de ocorrência (de 0 azul a 1 vermelho) a uma distância espacial.

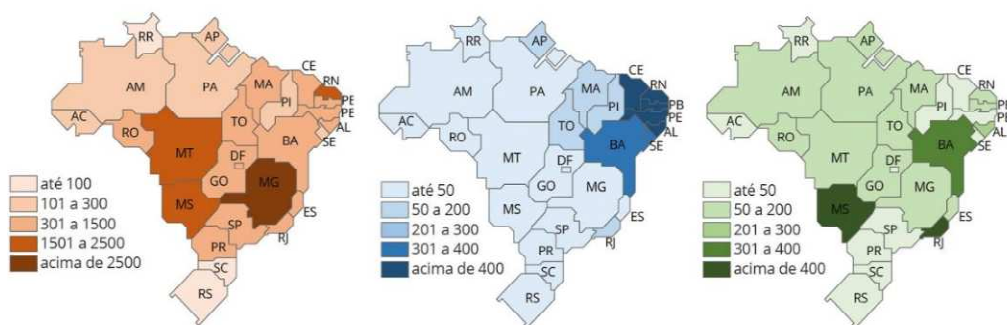
Figura 1: Mapa global da distribuição prevista de *Aedes aegypti*



Fonte: Kraemer (2015)

A quantidade de casos de doenças epidemiológicas reflete, na maioria das vezes, o viés de serviços de saúde, bem como a experiência da população com a doença. Essa bilateralidade nem sempre é ocasionada pelos investimentos financeiros de combate à epidemia, pois a complexidade dos problemas vai desde localização, clima, hábitos, vigilância, economia, desenvolvimento social, entre outros fatores. A Figura 2 mostra os dados de Dengue (cor laranja), Chikungunya (cor azul) e Zika (cor verde), eles representam a incidência por 100 mil habitantes em 2016.

Figura 2: Incidência de Dengue, Chikungunya e Zika em 2016



Fonte: Dantas; Lenharo (2017)

Há algum tempo a população brasileira enfrenta muitos casos de Dengue e outras doenças relacionadas ao mosquito transmissor já supracitado. O problema já é conhecido e bastante estudado no Brasil, tendo em vista os diversos surtos ocasionados. Mas então por que tal crise ainda permanece? Essa é a uma das perguntas que nos fazem pensar a real característica do impasse encontrado, todavia é imprescindível o caráter epidêmico da Dengue e sua intensa carência por serviços de urgência e emergência, bem como as longas filas nos atendimentos.

Um estudo realizado em Recife-PE demonstrou associação entre a soroprevalência e o nível socioeconômico dos bairros, sendo maior no estrato amostral formado pelos bairros mais pobres. Neste estudo, não sair do bairro para trabalhar ou estudar foi um fator de risco no estrato mais pobre. Nos estratos intermediário e alto, o número de pessoas por cômodo associou-se à maior prevalência, e ainda, no estrato alto, morar em casa (em comparação a morar em apartamento) foi também um fator de risco para a soroprevalência (BRAGA et al., 2010).

Segundo Breilh (1991) entender a relação entre o aspecto social e o biológico de uma epidemia é tão grande que rompe com a ideia de que há uma separação entre essas duas instâncias como a que existiria entre duas partes diferentes do mundo.

Laurell (1977 apud NUNES, 1983) divide o processo de saúde e doença em três níveis: singular, particular e universal. O aspecto biológico está nos dois primeiros níveis, contudo o social encontra-se no terceiro. O autor representou o primeiro como enfermidade relacionada ao processo biológico do indivíduo e o segundo como enfermidade que pode ser entendida como o resultado da

influência recíproca dos fatores biológicos (hospedeiro, agente e ambiente). Já no terceiro nível diz que a saúde-enfermidade se apresenta como um fenômeno social e coletivo.

Buscar entender o significado da Dengue é corroborar os pensamentos de Bertolli Junior (1993) e Breilh (1991), os quais dizem que as doenças devem ser tratadas como objetos sociais e que a abordagem biológica não consegue entender por sua totalidade. Eles dizem que um estudo biológico que é complementado pelas ciências sociais por meio de um delineamento histórico do processo saúde e doença revela que fenômenos fisiopatológicos variam em conformidade com ambientes sociais e ecológicos específicos. Contextualizando os autores, o aspecto tecnológico, aliado aos dois já citados faz com que o processo saúde e doença se torne mais amplo, acessível e completo, tendo em vista suas grandes possibilidades, o que se retrata no estudo.

Outro ponto que merece destaque nesse texto é como a informação em saúde pode influenciar a incidência de casos da doença tratada nesse trabalho. Os meios de comunicação são vários e vão de um panfleto entregue a uma pessoa na rua até o compartilhamento nas redes sociais.

Partindo do pressuposto que a democratização das informações em saúde é umas das propostas da promoção em saúde no Brasil, é necessário compreender dois pontos importantes que são a quantidade de informação *versus* a sensação de dever cumprido através da redução dos casos. Nisso, faz-se necessário que tais informações sejam unificadas e direcionadas bilateralmente entre os atores sociais. A comunicação e a educação não devem mais ser consideradas processos unidirecionais e sim, processos de circulação de significados sociais entrelaçados (RANGEL, 2008).

O Ministério das Comunicações lançou em novembro de 2016 um aplicativo que objetiva trazer informações que ajudem a combater o mosquito o *Aedes aegypti*. A ferramenta “Desafio Aedes” trata-se de um jogo que exige dos usuários o cumprimento de tarefas que previnem a proliferação do mosquito (BRASIL, 2016). Diante disso, a comunicação e educação em saúde fazem com que cada pessoa se torne interlocutor de informações úteis para a comunidade, multiplicando as informações. Vale salientar que as equipes de Saúde são fundamentais para diagnosticar e fazer o tratamento adequado.

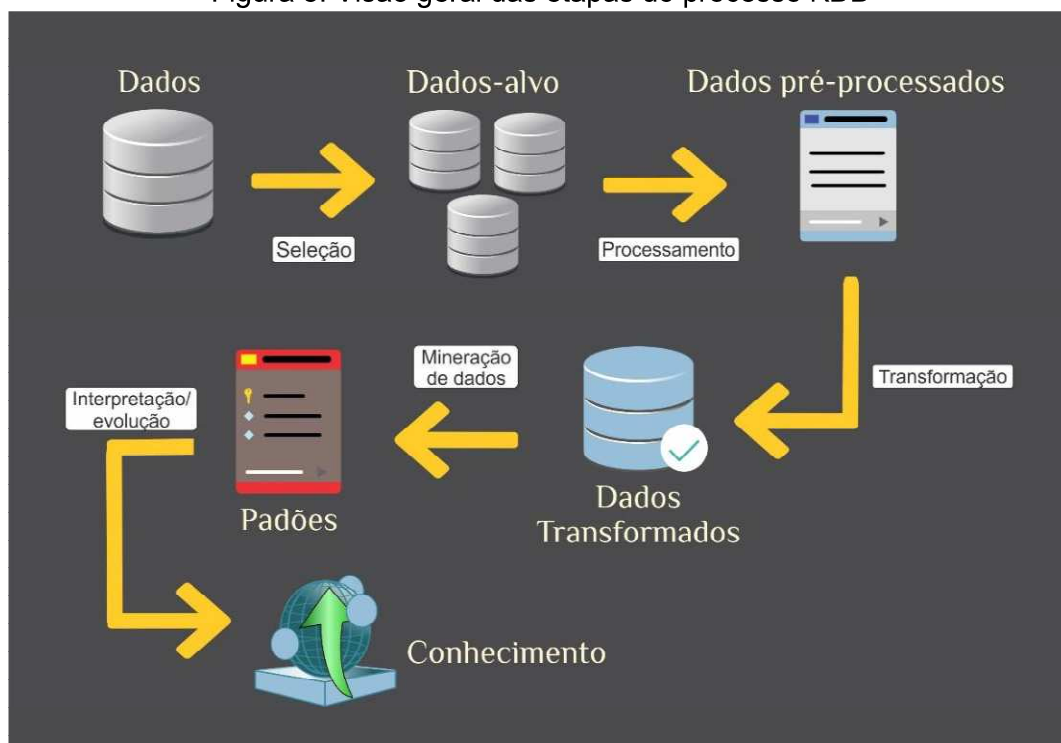
2.2 Aspectos Técnicos em Computação

2.2.1 Descoberta de Conhecimento em Banco de Dados

A descoberta de conhecimento em Banco de Dados ou KDD (*Knowledge-Discovery in Databases*) é um processo não trivial que tem como foco a identificação de padrões válidos, novos, potencialmente úteis e compreensíveis. Este processo foi formalizado no final da década de 80, elucidando o grande processo na busca por conhecimento e as grandes aplicações na área de Mineração de Dados (FAYYAD, 1996).

Essa descoberta de conhecimento é um processo que visa extrair informações e conseqüentemente conhecimento por meio de uma manipulação dos dados. As etapas do processo KDD estão ilustradas na Figura 3:

Figura 3: Visão geral das etapas do processo KDD



Fonte: Fayyad; Piatestky; Smyth (1996) - Adaptado

Os autores consideram os dados como um conjunto de fatores e o padrão como uma expressão em algum idioma que descreve um subconjunto dos dados ou um modelo aplicável ao mesmo. Portanto, a extração de um padrão também designa o ajuste de um modelo aos dados; estrutura de pesquisa a partir de

dados; ou, em geral, fazer qualquer descrição de alto nível de um conjunto de dados (FAYYAD, 1996).

Em linhas gerais, tem-se na seleção os dados que são escolhidos como um subconjunto de interesse, ou dispostos em um subconjunto de amostragem. As demais etapas podem ser resumidas em:

- O **Pré-processamento** é o momento em que são utilizadas ferramentas para preparação dos dados visando deixá-los em um formato mais apropriado para as próximas etapas. Aqui são tratadas distorções, ausência de dados ou até mesmo uma reorganização das informações; na transformação os dados são processados e disponibilizados em uma forma diferente do original, mas com suas propriedades mantidas, visando normalizar e/ou reduzir sua dimensão;
- **MD** propriamente dita é a extração dos dados já citados com auxílio de métodos computacionais (algoritmos, softwares) com o objetivo de revelar padrões, tendências etc;
- **Interpretação ou evolução dos dados**, etapa na qual há o julgamento dos dados pós-minerados para, se necessário, refazer ou alterar os processos anteriores.

Vale salientar que os padrões descobertos devem ser válidos em novos dados com algum tipo de certeza. Visa-se que os padrões sejam novos (pelo menos para o sistema e de preferência para o usuário) e utilizáveis, ou seja, extrair alguma vantagem para o usuário ou tarefa. Finalmente, os padrões devem ser compreensíveis logo após o processamento (FAYYAD; PIATETSKY; SMYTH, 1996).

Goldberg e Deb (1991) fazem uma organização em três grupos das atividades realizadas na KDD, são eles:

1. Tarefas relacionadas ao desenvolvimento tecnológico, que engloba todas as iniciativas de concepção e desenvolvimento de recursos de suporte para serem utilizados na busca por conhecimentos novos em grandes bases de dados;
2. As atividades de execução de processos de KDD que dizem respeito à busca efetiva de conhecimento em bases de dados. As

ferramentas produzidas pelas atividades de desenvolvimento tecnológico são utilizadas na execução de KDD;

3. Atividades que envolvem a aplicabilidade de resultados de processos de KDD. Elas visam obter modelos de conhecimento úteis a partir de grandes bases de dados, e se relacionam a aplicação dos resultados no contexto em que foi realizado o processo de KDD.

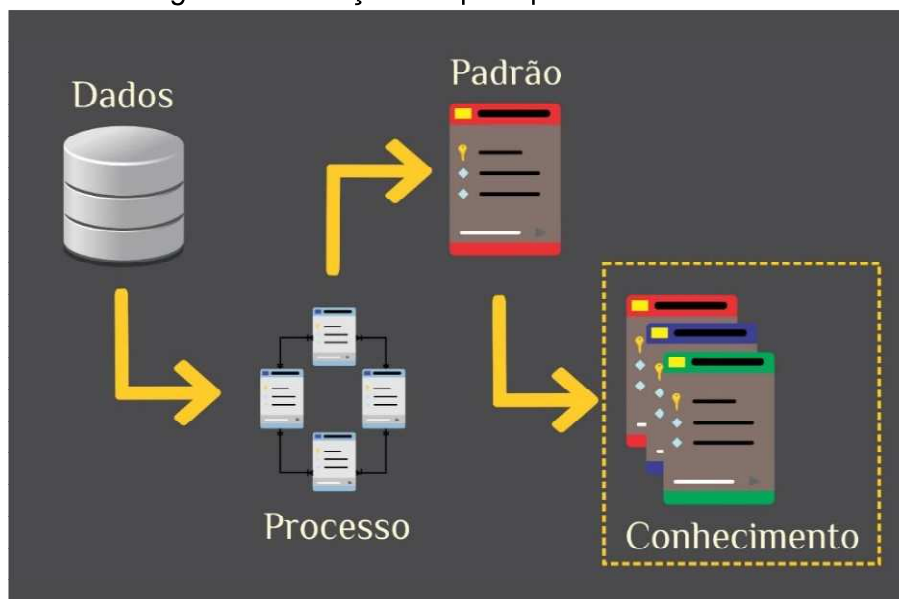
A maioria dos trabalhos relacionados a KDD tem se concentrado na MD. No entanto, as outras etapas são tão importantes (e provavelmente mais) para a aplicação bem-sucedida do KDD na prática. Entretanto este trabalho tem enfoque na aplicabilidade das técnicas de MD sob a pesquisa elencada.

2.2.2 Mineração de Dados

A MD tem o objetivo de retirar de informações preditivas ou descritivas escondidas de grandes bancos de dados a partir de aplicações utilizando algoritmos para extrair, identificar padrões, estruturas, tendências e revelar novidades que sejam úteis. As ferramentas de MD identificam as possibilidades de correlações que existem nas bases de dados. Isso acontece por meio de técnicas para exploração de dados através do desenvolvimento de aplicações que venham a extrair, dos bancos de dados informações críticas, com o objetivo de subsidiar plenamente o processo decisório de uma organização (QUINLAN, 1992).

Todavia, para que se tenha um conhecimento sobre o que é MD, faz-se necessário uma conceituação dos principais termos existentes e Witten e Frank (2005) elencam os seguintes: **Dados**, conjunto de fatos presentes em um repositório. Ex: Valores dos campos de um registro de vendas em uma base qualquer; **Padrões**, denota alguma abstração de um subconjunto dos dados em alguma linguagem descritiva de conceitos; **Processo**, a extração do conhecimento envolve várias etapas como preparação dos dados, busca por padrões e avaliação do conhecimento; **Conhecimento**, definido em função do seu domínio de aplicação, utilidade, originalidade e compreensão. A Figura 4 ilustra tais parâmetros:

Figura 4: Ilustração dos principais termos na MD



Fonte: Próprio Autor

Trabalhar com MD é ter consciência do grande volume dados que pode ser encontrado, bem como o seu resultado. Essas informações, muitas vezes, desconhecidas pelos responsáveis da base de dados, permitem uma maior agilidade na administração do novo conjunto de informações adquiridas.

Um exemplo básico da aplicabilidade é uma empresa. Ao fazer uso das técnicas de MD ela será capaz de criar pré-requisitos para entender o comportamento do consumidor, identificando suas afinidades entre as escolhas de produtos e serviços, prevendo hábitos de compras e analisando comportamentos habituais para detecção de fraudes (PINTO, 2005).

A grande motivação de se trabalhar com MD é analisar as grandes quantidades de dados que estão sendo acumulados em todo o mundo. Kass-Hout e Alhinnawi (2013) afirmam que ao redor do mundo as pessoas e organizações compartilham diariamente um grande volume de informações sobre saúde ou questões ligadas a ela

Todavia, há uma grande diferença entre os dados armazenados e o conhecimento que poderia ser construído a partir deles e, a MD, pode ajudar sobre esse conhecimento aprofundado, a partir de técnicas de análise.

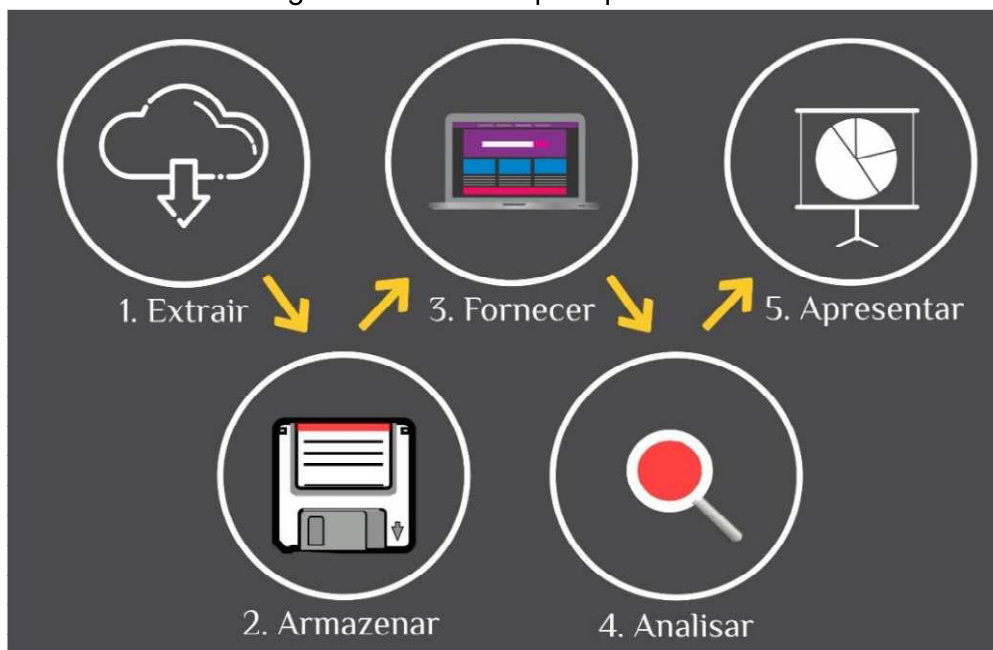
2.2.2.1 Como funciona a Mineração de Dados

Basicamente, um software de MD analisa associações e padrões em dados de operação armazenados com base em consultas do usuário. Há vários tipos de software no mercado, eles geralmente envolvem: estatística, aprendizagem de máquina e redes neurais.

Outro termo importante são padrões sequenciais, ou seja, os dados são explorados para antecipar padrões de comportamento e tendências. Por exemplo, o mosquito da Dengue trafega sobre uma região e as semanas epidemiológicas mostram afinidade por determinadas datas e situações, clima e outras variantes.

KAUR, GOYAL e LU (2001) citam cinco elementos principais para a MD: **Extrair**, transformar e carregar dados de transações no sistema de *data warehouse*; **Armazenar** e gerenciar os dados em um sistema de banco de dados multidimensional; **Fornecer** acesso a dados para analistas de negócios e profissionais de tecnologia da informação; **Analisar** os dados pelo software aplicativo; **Apresentar** os dados em um formato útil, como um gráfico ou tabela.

Figura 5: Elementos principais da MD



Fonte: Próprio Autor

Além dos cinco elementos básicos para a MD, os autores citam seis níveis de análise:

Quadro 1: Resumo dos elementos básicos da MD

Redes neurais artificiais	Modelos preditivos não-lineares que aprendem através do treinamento e se assemelham a redes neurais biológicas na estrutura.
Algoritmos genéticos	Técnicas de otimização que utilizam processos como combinação genética, mutação e seleção natural em um desenho baseado nos conceitos de evolução natural.
Árvores de Decisão	Estruturas em forma de árvore que representam conjuntos de decisões. Eles fornecem um conjunto de regras que você pode aplicar a um novo conjunto de dados (não classificado) para prever quais registros terão um determinado resultado.
Método de vizinho mais próximo	Uma técnica que classifica cada registro em um conjunto de dados com base em uma combinação das classes do(s) registro(s) k mais semelhante(s) a ele em um conjunto de dados históricos.
Indução de regras	A extração de regras <i>if-then</i> úteis a partir de dados baseados em significância estatística.
Visualização de dados	A interpretação visual de relações complexas em dados multidimensionais. As ferramentas gráficas são usadas para ilustrar relacionamentos de dados.

Fonte: Próprio Autor

2.2.2.2 Tarefas de Mineração de Dados

Fayyad (1996) elenca vários métodos para se extrair conhecimento em uma base dados. Conquanto, é importante saber os dois objetivos básicos da MD que, na prática, visam a predição e descrição. O primeiro envolve a utilização de algumas variáveis ou campos no banco de dados para prever valores desconhecidos ou futuros de outras variáveis de interesse, já a descrição tem o objetivo de encontrar padrões interpretáveis por humanos descrevendo os dados. Ainda que os limites entre previsão e descrição não sejam nítidos (alguns dos modelos preditivos podem ser descritivos, na medida em que são

compreensíveis, e vice-versa), a distinção é útil para a compreensão do objetivo geral de descoberta. A importância relativa da predição e da descrição para aplicações específicas de MD pode variar consideravelmente. As metas de predição e descrição podem ser obtidas usando uma variedade de métodos de mineração de dados específicos.

Muitos autores têm tratado os métodos de MD de forma diferenciada. O Quadro 2 resume os principais métodos:

Quadro 2: Resumo dos principais autores sobre os métodos de MD

Autor	Tarefas
(Fayyad, 1996)	Classificação, regressão, agrupamento (<i>clustering</i>), sumarização, dependência de modelos, detecção de mudança e desvio.
(Adriaans; Zantinge, 1996)	Descoberta do conhecimento e Descrição.
(Berry; Linoff, 1997)	Classificação, estimação, predição, afinidade em grupos, agrupamentos (<i>clustering</i>) e descrição .
(Westphal; Blaxton, 1998)	Classificação, estimação, segmentação e descrição
(Elmasri; Navathe, 1999)	Previsão, identificação, classificação e otimização
(Thuraisingham, 1999)	Classificação, detecção de frequência, análise de dependência dos dados e análise de desvio
(Mena, 1999)	Predição, classificação, agrupamento (<i>clustering</i>), segmentação, associação, visualização e otimização.
(Weis; Indurkha, 1999)	Predição, detecção de desvio, segmentação, agrupamento (<i>clustering</i>), análise de ligações e regras de associação, sumarização e visualização.

Fonte: Próprio Autor

Um dos métodos mais clássicos é o apresentado por Fayyad (1996), no trabalho intitulado "*From Data Mining to Knowledge Discovery in Databases*" ou "Da MD à descoberta de conhecimento em bancos de dados". Esse trabalho foi a base do texto aqui apresentado.

Weiss e Kulikowski (1994) e Hand (1981) dizem que a **classificação** dados é uma forma de mapear uma função aprendida. Um exemplo disso são os métodos KDD que classificam tendências nos mercados financeiros, além da

identificação automatizada de objetos de interesse em bases de dados de grandes imagens (FAYYAD,1996).

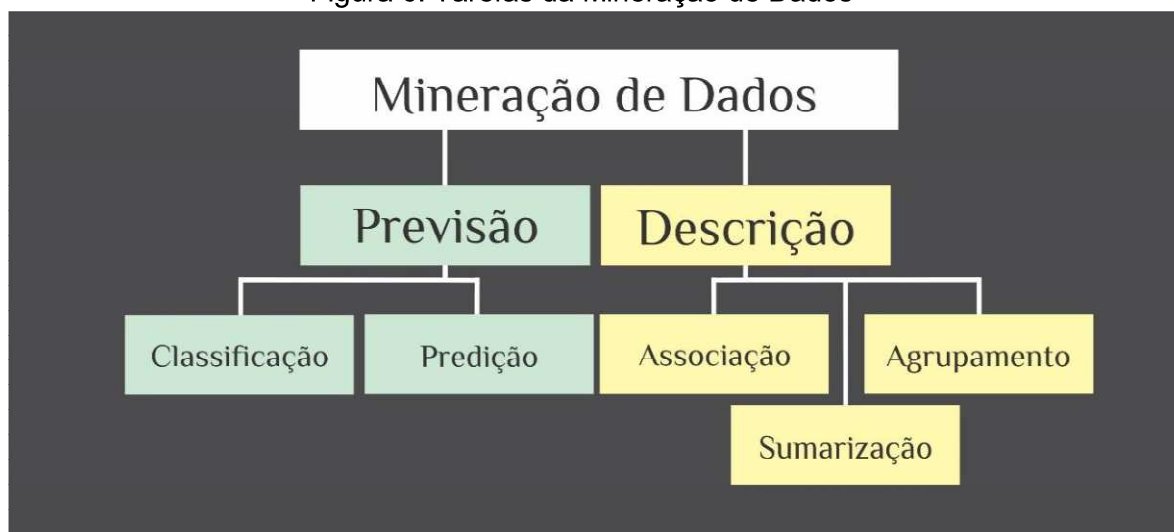
A **regressão** é aprender uma função que mapeia um item de dados para uma variável de predição de valor real. As aplicações de regressão são muitas, por exemplo, predizendo a quantidade de casos de Dengue presentes em uma região dadas as medições epidemiológicas; estimando a probabilidade de um paciente ter a doença dados os resultados de um conjunto de testes diagnósticos.

Jain e Dubes (1988) dizem que **agrupamento** de dados é uma função descritiva em comum na qual se almeja identificar um conjunto finito de categorias ou *clusters* para descrição. As categorias podem ser mutuamente exclusivas e exaustivas ou consistir em uma representação mais rica, como categorias hierárquicas ou sobrepostas. Exemplos de aplicações de *clustering* em um contexto de descoberta de conhecimento incluem a descoberta de subpopulações homogêneas para consumidores em bancos de dados de *marketing* (CHEESEMAN, 1990).

Sumarização busca identificar e apresentar em um formato compreensível, as principais características dos dados em um conjunto de dados (DIAS; 2001, 2002). Exemplo: identificar o perfil das pessoas que adquirem uma determinada doença. Métodos mais sofisticados envolvem a derivação de regras sumárias (AGRAWAL; PSAILA, 1995), técnicas de visualização multivariada e a descoberta de relações funcionais entre variáveis (ZEMBOWICZ; ZYTKOW, 1996). As técnicas de resumo são frequentemente aplicadas à análise de dados exploratórios interativos e à geração automatizada de relatórios.

A **dependência** verifica a existência de um modelo que descreva a dependência entre variáveis. Os modelos de dependência existem em dois níveis: (1) o nível estrutural do modelo especifica (frequentemente em forma gráfica) quais variáveis são localmente dependentes umas das outras e (2) o nível quantitativo do modelo especifica os pontos fortes das dependências usando alguns valores numéricos. A Figura 6 resume esquematicamente as tarefas citadas anteriormente:

Figura 6: Tarefas da Mineração de Dados



Fonte: Próprio autor

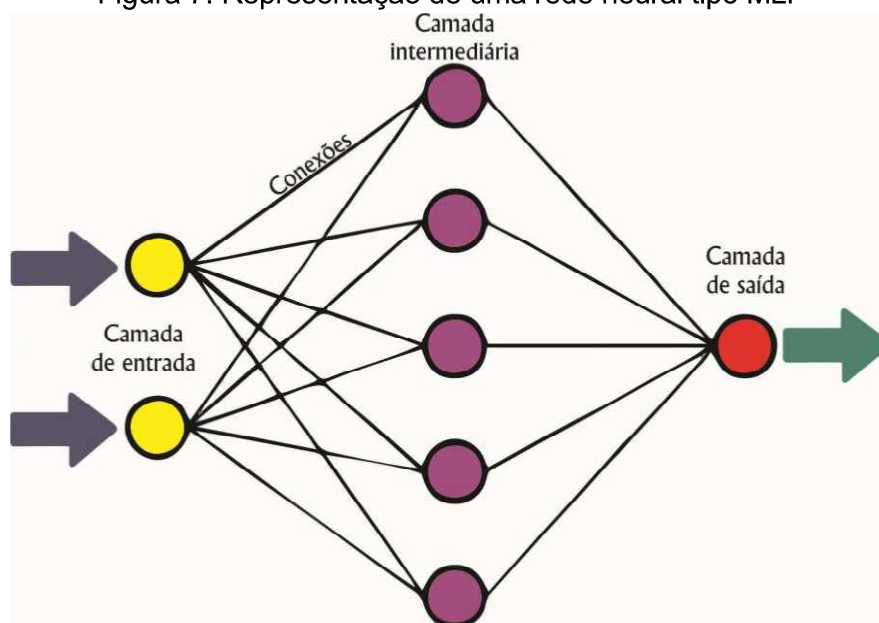
2.2.2.3 Técnicas de mineração de dados

Na MD as tarefas são desenvolvidas por diferentes técnicas usadas para diferentes propósitos (HARRISON, 1998). Logo abaixo estão elencadas algumas técnicas de MD que são mais utilizadas:

Descoberta de Regras de Associação: foi inicialmente apresentada por Agrawal (1995) e identifica conjuntos de itens que ocorrem simultaneamente e frequentemente em banco de dados. A técnica consiste em uma correlação estatística entre os itens de dados (GOEBEL; GRUENWALD, 1999). Entre os diversos algoritmos utilizados na técnica, temos: DHP (*Direct Hashing and Pruning*), APRIORI, Partition, DIC (*Dynamic Itemset Counting*), Eclat, Maxclique e Cumalte.

Redes Neurais: Criam um conjunto de dados de teste, permitindo que a rede neural possa aprender padrões baseados em resultados conhecidos e, posteriormente, definir a rede neural em grandes quantidades de dados. Na Figura 5 é possível ver uma representação de uma rede neural. Alguns exemplos de algoritmos para esta técnica são: Perceptron, Rede MLP, Rede Hopfield, Rede BAM, Redes ART, Rede IAC, Rede LVQ, Rede Counterpropagation, Rede RBF, Rede PNN, Rede Time Delay, Neocognitron, Rede BSB (HARRISON, 1998). A Figura 7 ilustra uma Rede Neural.

Figura 7: Representação de uma rede neural tipo MLP



Fonte: Próprio autor

Algoritmos Genéticos: estão relacionados a problemas de otimização. Algoritmos que implementam esta técnica, como: Algoritmo Genético Simples, CHC, Algoritmo de Hillis, GA-Nuggets, GA-PVMINER, NSGA II (HARRISON, 1998).

Árvores de decisão: ajudam com a classificação e geração de informações que são muito descritivas, dando auxílio para o entendimento dos dados. Ou seja, é um fluxograma representado por uma árvore na qual cada nó interno significa um teste em um atributo, cada ramo representa o resultado do teste e cada folha a distribuição dos registros. Os algoritmos que implementam esta técnica são: CART, CHAID, C5.0, Quest, ID-3, SLIQ, SPRINT (HARRISON, 1998; GOEBEL, GRUENWALD, 1999).

2.2.2.4 Metodologias de Aplicação em Mineração

Há diversas abordagens para representar como o usuário irá fazer a MD e obter seus resultados. Basicamente existem três tipos de abordagens *top-down*, a *bottom-up* e a híbrida. A primeira também é chamada de teste de hipótese, pois parte de uma ideia pré-concebida e espera-se validá-la ou contestá-la. Na segunda abordagem, *bottom-up* ou busca de conhecimento o explorador de dados parte em busca do desconhecido, ou seja, o que ainda não

é de conhecimento prévio. A híbrida é uma combinação das duas (GOLDSCHMIDT, 2015).

Durante a utilização de uma dessas abordagens acima elencadas, o minerador de dados terá que usar a busca de conhecimento na forma direta ou indireta. A direta ou também conhecida como supervisionada usa um valor que será prognosticado, uma classe que atribuirá registros ou um relacionamento a ser explorado. Em linhas gerais, nesse ponto, existe uma ideia vaga do que se procura. Todavia, para se aplicar esse tipo de busca é necessário identificar as fontes dos dados selecionados, preparar esses dados para análise, construir e treinar e avaliar o modelo computacional (BRAGA; LUDERMIR; CARVALHO, 2000)

Na busca de conhecimento indireta ou não-supervisionada há tudo bem definido e já tem-se noção do que será encontrado. A metodologia é semelhante ao direto, entretanto ainda serão geradas novas hipóteses para testes depois de avaliado o modelo computacional.

2.2.2.5 Limitações

No mercado profissional e na internet há diversas ferramentas poderosas que podem minerar desde os dados mais organizados ou simples até os mais complexos. Todavia, os programas não fazem todo o trabalho e duas grandes limitações são os tipos de dados e quem os manipulam.

A Mineração de Dados ajuda a revelar os padrões existentes nos dados bem como seus relacionamentos, porém ela não entrega ao usuário final o significado desses padrões. A análise final baseada na metodologia usada deve ser feita pela pessoa responsável pelos dados, ou seja, a legitimidade do que foi descoberto só é real se comparada com as situações cotidianas do contexto vivido.

Outro ponto que deve ser citado é a confiabilidade dos dados encontrados. É claro que um bom profissional irá se certificar que seu material de estudo é confiável e oficial. Entretanto, mesmo dados oficiais são sujeitos a erros e equívocos.

Outra limitação da Mineração de Dados é que, embora possa identificar conexões entre comportamentos e/ou variáveis, ela não identifica necessariamente um relacionamento causal. Por exemplo, um aplicativo pode

identificar que um padrão de comportamento, como a propensão para comprar bilhetes de avião pouco antes do voo está programado para partir, está relacionado a características como renda, nível de educação e uso da Internet. No entanto, isso não indica necessariamente que o comportamento de compra do ticket é causado por uma ou mais dessas variáveis. De fato, o comportamento do indivíduo pode ser afetado por algumas variáveis adicionais como a ocupação (a necessidade de fazer viagens com aviso prévio), o status familiar (um familiar doente que precisa de cuidados) ou um *hobby* (KAUR; GOYAL; LU, 2001).

2.3 Síntese da Seção

Nessa seção foi realizada uma síntese dos autores estudados durante a revisão bibliográfica, bem como o estabelecimento de conceitos que foram cruciais para a elaboração da metodologia e dos resultados do trabalho. Some-se a isso a interdisciplinaridade entre Saúde Pública, Epidemiologia e Dengue com a Mineração de Dados, dando suporte para que a tecnologia seja mais um elemento de combate a doenças, como é o caso daquelas causadas pelo *Aedes aegypti*.

3 MATERIAL E MÉTODO

3.1 Fonte de Dados

Este trabalho utilizou como fonte de dados o Sistema de Notificação de Agravos de Notificação (SINAN) do Ministério da Saúde, Brasil. Os dados de Dengue compreendem os anos de 2007 a 2016, totalizando 10 anos de informações de 18 municípios do estado Paraíba que fazem parte da faixa populacional de 20.000 a 40.000 habitantes, são eles: Pombal (32.110 hab.), Esperança (31.098 hab.), São Bento (30.879 hab.), Monteiro (30.852 hab.), Catolé do Rocha (28.759 hab.), Alagoa Grande (28.479 hab.), Pedras de Fogo (27.032 hab.), Solânea (26.693 hab.), Lagoa Seca (25.900 hab.), Itabaiana (24.481 hab.), Areia (23.829 hab.), Itaporanga (23.192 hab.), Rio Tinto (22.976 hab.), Bananeiras (21.851 hab.), Conde (21.400 hab.), Princesa Isabel (21.283 hab.), Mari (21.176 hab.) e Caaporã (20.362 hab.). O mapa dos municípios refere-se ao censo de 2010, provenientes da aplicação *WebCart* do Instituto Brasileiro de Geografia e Estatística (IBGE).

3.2 Recursos de software

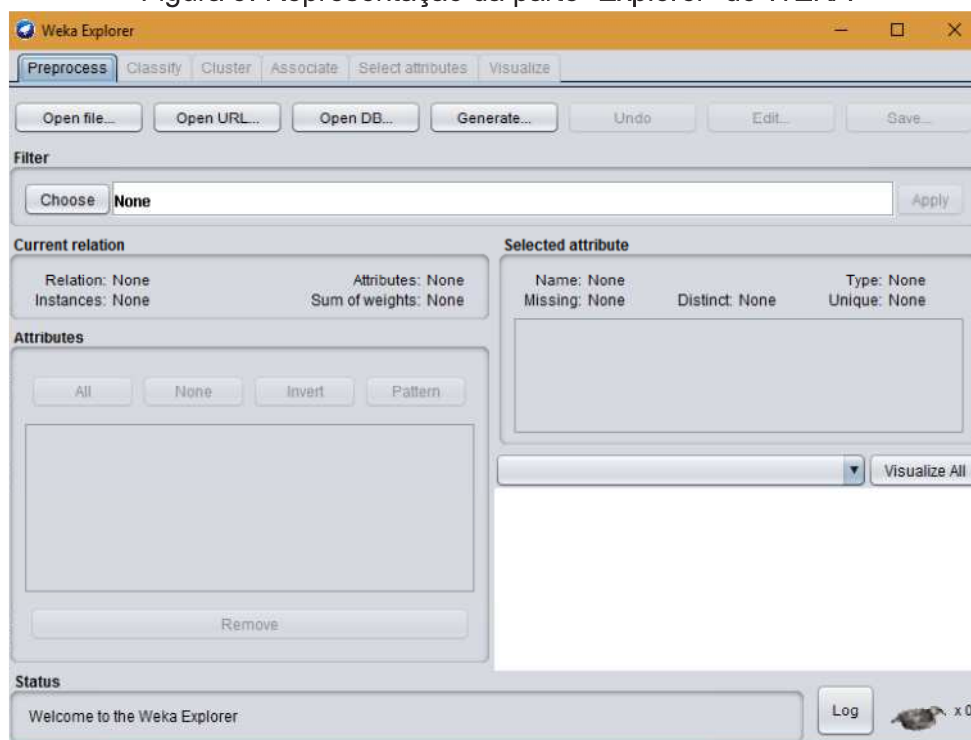
Foi utilizado o *open-source* WEKA, uma alternativa para análise do conhecimento de fonte aberta emitida sob a GPU (*General Public License*), que inclui um conjunto de ferramentas para completar muitas tarefas de MD, além de métodos e modelos apresentando vários *hands-on*, juntamente com arquivos de entrada disponíveis a partir do livro no site www.dataminingconsultant.com. A versão utilizada é a 3.8.0 © 1999-2016 criada pela University of Waikato na Nova Zelândia, conforme ilustra a Figura 8 e Figura 9.

Figura 8: Representação do software WEKA



Fonte: próprio autor

Figura 9: Representação da parte “Explorer” do WEKA



Fonte: Próprio autor

O Weka fornece uma interface uniforme com vários tipos de algoritmos de aprendizagem diferentes, além de reunir métodos de pré e pós-processamento (Figura 7), avaliando o resultado de planos de aprendizagem em qualquer conjunto de dados.

3.3 Transformação dos dados

Para que uma MD seja bem-sucedida é necessário muito mais do que selecionar um algoritmo de aprendizado e executá-lo sobre os dados. Nos dados desta pesquisa existem esquemas de aprendizagem nos vários parâmetros existentes. Esta fase irá melhorar de forma acentuada a escolha adequada de valores de parâmetros, o que dependerá dos dados disponíveis.

Há vários esquemas de aprendizado e vários valores de parâmetros que podem ser usados com os dados. Todavia, para evitar o problema de superalimentação, priorizou-se trabalhar com semanas epidemiológicas que posteriormente são aliados com dados baseados em geolocalização (IBGE) e dados pluviométricos da Agência Executiva de Gestão de Águas do Estado da Paraíba (AESA).

3.4 Regras de associação

Regras de associação são um método que visa descobrir relações interessantes, padrões frequentes, associações, correlações ou estruturas entre conjuntos de itens entre variáveis em um banco de dados. Não obstante, seu objetivo é encontrar essas regras, do tipo X (antecessor) e Y (consequente). Conforme Agrawal (1993) um modelo matemático na qual as regras de associação deveriam ter como base um **suporte** e uma **confiança** mínimos especificados pelo analista, onde:

- I é um conjunto de itens de venda descritos por $I = \{i_1, i_2, \dots, i_m\}$;
- T representa uma transação de venda tal que $T \subseteq I$;
- TID representa o chave que identifica a transação;
- D representa um conjunto de transações de vendas;
- X e Y conjuntos de itens de venda contidos em uma transação $X \subseteq T$ e $Y \subseteq T$;
- Uma regra de associação é uma implicação de $X \Rightarrow Y$ onde $X \subset I$, $Y \subset I$, e $X \cap Y = \phi$;
- s é o Suporte de uma alguma regra $X \Rightarrow Y$ é dado através do total de transações contido no subconjunto de transações que contém $X \cup Y$ sobre o conjunto de transações D . O suporte é descrito pela seguinte fórmula:

$$\text{Suporte} = \frac{\text{N}^\circ \text{ de registros da tabela que contêm o conjunto}}{\text{N}^\circ \text{ total de registros da tabela}} \quad (1)$$

c é a Confiança de uma determinada regra $X \Rightarrow Y$ é dada através do total de registros do subconjunto $X \cup Y$ sobre o total de registros do subconjunto X. A confiança é descrita pela seguinte fórmula:

$$\text{Confiança} = \frac{\text{N}^\circ \text{ de registros da tabela que contêm todos os itens da regra}}{\text{N}^\circ \text{ de registros da tabela que contêm o antecedente da regra}} \quad (2)$$

3.4.1 O algoritmo APRIORI

O APRIORI foi utilizado para este trabalho por ser considerado um dos principais algoritmos para extração de Regras de Associação. Ele originou o *Software Intelligent Miner*, sendo inicialmente proposto por uma equipe de pesquisa QUEST da IBM. Ele faz recursivas buscas no Banco de Dados à procura dos conjuntos frequentes que satisfazem um suporte mínimo estabelecido.

Os procedimentos de extração de regras de associação são decompostos em duas etapas separadas, primeiro, o suporte mínimo é aplicado para encontrar todos os itens frequentes em um banco de dados. Em segundo lugar, esses conjuntos de itens frequentes e a restrição de confiança mínima são usados para formar regras (AGRAWAL; IMIELINSKI; SWAMI, 1993).

Todavia, enquanto o segundo passo é direto, o primeiro passo precisa de mais atenção. Encontrar todos os itens frequentes em um banco de dados é uma tarefa difícil, pois envolve pesquisar todos os itens possíveis (combinações de itens).

O pseudocódigo abaixo (Figura 10) exemplifica como funciona o algoritmo onde L_k corresponde ao conjunto de todos os conjuntos-de-k-itens frequentes e C_k é o conjunto de todos os conjuntos-de-k-itens candidatos, ou seja, potencialmente frequentes, na passagem k. O algoritmo gera um conjunto de itens frequentes, a partir disso ele realiza mais três operações: primeiro ele gera o conjunto de itens candidatos baseado no conjunto de itens frequentes anterior $k-$, logo após ele percorre os dados e o suporte para cada candidato em C_k , e, por último, gera um conjunto de itens frequentes L_k (AGRAWAL; SRIKANT, 1994).

Figura 10: Pseudocódigo do Algoritmo APRIORI

```

Procedure Apriori

  L1 = {Conjunto de itens frequentes};

  for (k=2; Lk-1 ≠ ∅; k++) do
    Ck = apriori_gen(Lk-1);
    forall transações t ∈ D do Ct
      = subset(Ck, t);
      t.count++;
      forall candidatos c ∈ Ct do
        c.count++;
      od
    od
    Lk = {c ∈ Ck | c.count/t.count ≥ suporte_mínimo};
  od
  Resposta = ∪k Lk;
  Fim

```

Fonte: AGRAWAL; SRIKANT (1993)

No WEKA ele já está implementado em sua suíte de algoritmos, possibilitando mais confiabilidade em sua utilização, pois ele traz um ambiente com um formato padronizado (".arff").

3.5 Seleção dos atributos

O trabalho foi dividido em 13 análises que representam quatro semanas epidemiológicas (cada), totalizando 52 (semanas), referentes aos 18 municípios selecionados da Paraíba. Cada análise contém 18 atributos (referente aos municípios) e os dados das semanas epidemiológicas em questão analisadas entre os anos de 2007 e 2016, totalizando 10 anos de dados. O formato das análises foi escolhido com base no estudo referencial feito para a criação deste trabalho. Além da revisão foram feitos vários testes baseados em tentativas e erros - levando em consideração o *hardware* utilizado para as análises e da leitura das informações após a análise dos dados.

Os atributos selecionados seguem o seguinte padrão do WEKA:

1. Os nomes dos municípios não podem ter espaços, então nomes compostos foram juntados;
2. Por padrão deve existir a palavra @ATTRIBUTE que representa um atributo;

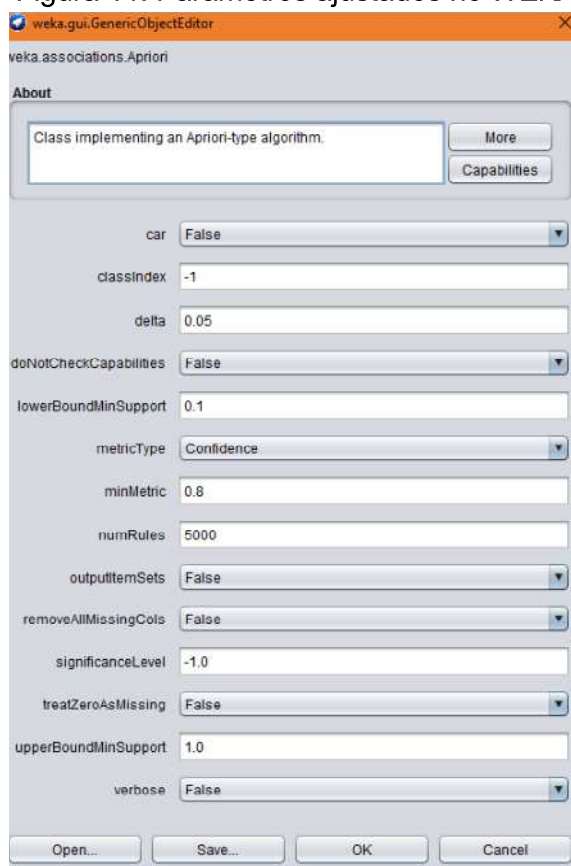
- Entre aspas estão o formato que os dados estarão, no caso “S” para representar que teve caso (s) de dengue na referida semana, e “N” para indicar que não houve.

Para aplicar esta estrutura foi utilizado o editor de código aberto ATOM, um programa gratuito e disponível na *internet*.

3.6 Regras de associação no WEKA

Foram feitas 13 análises independentes que representaram as 52 semanas epidemiológicas contidas no ano. De cada análise foram obtidas 5.000 linhas de regras que foram analisadas, escolhidas e que compõem o conjunto de regras deste trabalho. As análises foram ajustadas para uma taxa de confiabilidade de 80% (métrica utilizada para gerar mais regras), conforme as regras de cada análise. Cada conjunto de regras está analisado conforme sua semana epidemiológica, mostrando que há padrões nos dados. A Figura 11 representa tais parâmetros da análise.

Figura 11: Parâmetros ajustados no WEKA



Fonte: Próprio Autor

3.7 Pós-processamento dos Dados

Após encontradas as regras de associação, analisaram-se os dados para que um conhecimento seja gerado. Contudo, as informações encontradas precisaram de um método visual para ser apresentadas e, nesta dissertação, foram feitos mapas para tal representação. Nesta etapa os atributos com grande similaridade e, em alguns casos, foram submetidos a uma comparação analítica com outras fontes já citadas nesta metodologia. Buscou-se observar nos casos onde não houve regras de associação aplicáveis uma possível causa, nesses casos foram utilizadas outras fontes para comparação.

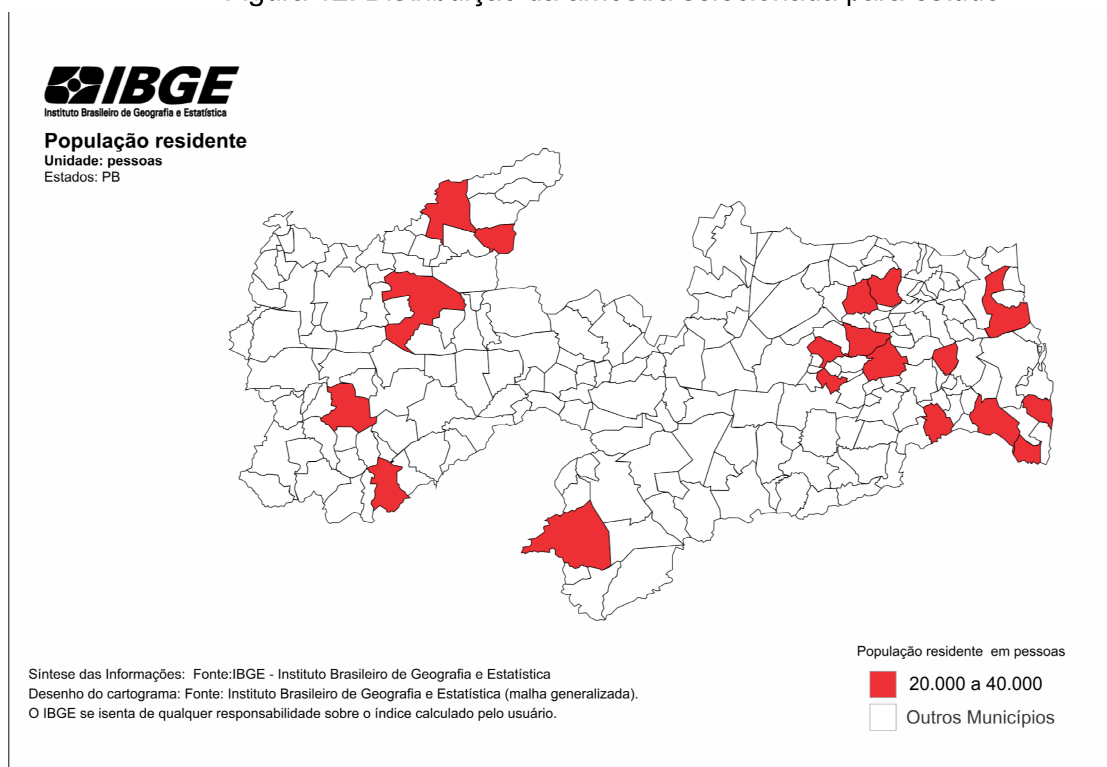
3.8 Síntese da Seção

Nesta seção foi apresentada a metodologia do trabalho, ou seja, o molde em que esta pesquisa foi construída. Em síntese buscou-se utilizar fontes de dados oficiais que deram o respaldo ao trabalho, bem como o processo de descoberta do conhecimento e suas fases principais: Pré-processamento, Mineração de Dados e Pós-processamento.

4 RESULTADOS E DISCUSSÃO

O trabalho em questão utilizou 18 municípios do estado Paraíba que fazem parte da faixa populacional de 20.000 a 40.000 habitantes, são eles: Pombal, Esperança, São Bento, Monteiro, Catolé do Rocha, Alagoa Grande, Pedras de Fogo, Solânea, Lagoa Seca, Itabaiana, Areia, Itaporanga, Rio Tinto, Bananeiras, Conde, Princesa Isabel, Mari e Caaporã. Figura 12 ilustra a amostra selecionada.

Figura 12: Distribuição da amostra selecionada para estudo

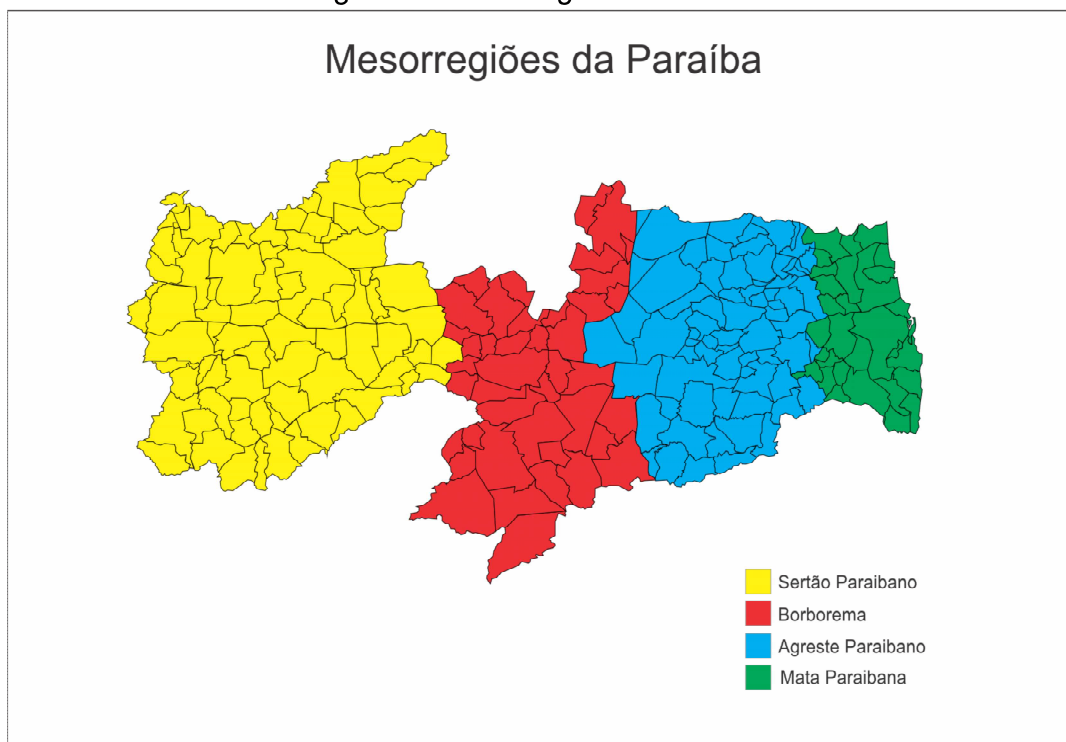


Fonte: Próprio Autor

A faixa populacional escolhida está disposta em todas as quatro mesorregiões conforme ilustra a Figura 13 e que compõem os 18 municípios paraibanos. É importante ressaltar a escolha dessa amostra, pois conforme os testes realizados pelo próprio autor aliado aos dados de **casos prováveis** (casos notificados, excluindo-se os casos descartados) do SINAN, observou-se que muitos municípios apresentavam pouquíssimos ou nenhum caso notificado para o mês em questão. Portanto a escolha da faixa populacional foi importante para que as análises através do algoritmo APRIORI fossem feitas da forma mais

confiável possível, técnica que é utilizada na fase do pós-processamento de dados.

Figura 13: Mesorregiões da Paraíba



Fonte: Próprio autor

Foram 13 análises que representaram os 10 anos de dados de dengue (2007-2016). Cada análise processa, através de regras de associação, quatro semanas epidemiológicas de cada ano a fim de encontrar relações entre os casos de dengue e a amostra selecionada, ou seja, municípios em desenvolvimento populacional semelhantes.

Vale ressaltar que quatro semanas não representam necessariamente a contagem de meses convencional (janeiro, fevereiro, março etc), e sim o calendário do SINAN para o ano, conforme ilustra o Quadro 3. Apenas há uma pequena modificação na estrutura das semanas nos anos bissextos, mas que não alterou a análise dos resultados.

Quadro 3: Calendário de notificação para o ano de 2017

Semana	Início	Término	Semana	Início	Término
1	01/01/2017	07/01/2017	27	02/07/2017	08/07/2017
2	08/01/2017	14/01/2017	28	09/07/2017	15/07/2017
3	15/01/2017	21/01/2017	29	16/07/2017	22/07/2017
4	22/01/2017	28/01/2017	30	23/07/2017	29/07/2017
5	29/01/2017	04/02/2017	31	30/07/2017	05/08/2017
6	05/02/2017	11/02/2017	32	06/08/2017	12/08/2017
7	12/02/2017	18/02/2017	33	13/08/2017	19/08/2017
8	19/02/2017	25/02/2017	34	20/08/2017	26/08/2017
9	26/02/2017	04/03/2017	35	27/08/2017	02/09/2017
10	05/03/2017	11/03/2017	36	03/09/2017	09/09/2017
11	12/03/2017	18/03/2017	37	10/09/2017	16/09/2017
12	19/03/2017	25/03/2017	38	17/09/2017	23/09/2017
13	26/03/2017	01/04/2017	39	24/09/2017	30/09/2017
14	02/04/2017	08/04/2017	40	01/10/2017	07/10/2017
15	09/04/2017	15/04/2017	41	08/10/2017	14/10/2017
16	16/04/2017	22/04/2017	42	15/10/2017	21/10/2017
17	23/04/2017	29/04/2017	43	22/10/2017	28/10/2017
18	30/04/2017	06/05/2017	44	29/10/2017	04/11/2017
19	07/05/2017	13/05/2017	45	05/11/2017	11/11/2017
20	14/05/2017	20/05/2017	46	12/11/2017	18/11/2017
21	21/05/2017	27/05/2017	47	19/11/2017	25/11/2017
22	28/05/2017	03/06/2017	48	26/11/2017	02/12/2017
23	04/06/2017	10/06/2017	49	03/12/2017	09/12/2017
24	11/06/2017	17/06/2017	50	10/12/2017	16/12/2017
25	18/06/2017	24/06/2017	51	17/12/2017	23/12/2017
26	25/06/2017	01/07/2017	52	24/12/2017	30/12/2017

Fonte: SINAN (2017)

Depois de delimitados os municípios que seriam trabalhados e a estrutura das análises, iniciou-se o pré-processamento dos dados. Primeiro foi feito o tratamento nos dados, utilizando uma planilha eletrônica gerada no *site* do SINAN. Nesta etapa subdividiram-se em sub-planilhas; a limpeza de dados foi feita para selecionar apenas os municípios desejados, já que a planilha do SINAN continha todos os municípios brasileiros; a transformação dos dados foi realizada com o intuito de adequar as informações para o formato WEKA já mencionado nesse texto. Nesta etapa os valores dos casos (numéricos) foram transformados em um valor lógico (verdadeiro ou falso; 0 ou 1), sendo atribuído

@ATTRIBUTE RioTinto {S,N}	Rio Tinto
@ATTRIBUTE Bananeiras {S,N}	Bananeiras
@ATTRIBUTE Conde {S,N}	Conde
@ATTRIBUTE Princesaisabel {S,N}	Princesa Isabel
@ATTRIBUTE Mori {S,N}	Mori
@ATTRIBUTE Caapora {S,N}	Caaporã

Fonte: Próprio autor

4.1 Análise 1: Semanas Epidemiológicas 1 a 4

A primeira análise foi composta pelas quatro primeiras semanas epidemiológicas dos municípios selecionados. Todavia, o algoritmo não encontrou nenhuma regra aceitável para futuras considerações. Isso se justifica devido ao pouco número de casos nas quatro primeiras semanas nos 10 anos, conforme ilustram os Quadros 5, 6, 7 e 8 a seguir.

Quadro 5: Casos de dengue na 1ª Semana Epidemiológica entre (2007-2016)

	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Alagoa Grande	0	1	0	0	2	0	0	0	0	0
Areia	0	1	0	0	0	0	0	0	0	0
Bananeiras	0	0	0	0	0	0	0	0	0	0
Caaporã	0	1	0	0	0	0	0	0	0	0
Catolé do Rocha	0	4	0	4	0	0	1	0	0	1
Conde	0	0	0	0	0	1	0	0	0	1
Esperança	3	1	0	0	2	0	0	0	0	0
Itabaiana	0	0	0	0	0	1	0	0	1	1
Itaporanga	0	0	0	0	1	0	0	2	0	16
Lagoa Seca	0	0	0	0	0	0	2	0	0	0
Mari	0	0	0	0	0	0	0	0	0	0
Monteiro	1	3	0	0	14	0	0	11	1	351
Pedras de Fogo	0	0	0	0	3	0	0	0	0	0
Pombal	0	1	0	0	0	0	0	0	0	0
Princesa Isabel	0	2	0	0	0	1	0	0	0	4
Rio Tinto	0	0	0	0	0	0	0	0	2	0
São Bento	0	2	0	0	0	1	0	0	2	3
Solânea	0	0	0	0	0	0	0	0	0	0

Fonte: SINAN (2017)

Quadro 6: Casos de dengue na 2ª Semana Epidemiológica entre (2007-2016)

	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Alagoa Grande	1	0	0	0	0	0	0	0	0	1
Areia	0	0	0	0	0	0	0	0	0	0
Bananeiras	1	0	0	0	0	0	0	0	0	0
Caaporã	0	0	0	0	0	0	0	0	0	0
Catolé do Rocha	0	4	0	5	0	1	1	2	1	1
Conde	0	0	0	0	1	0	0	1	0	0
Esperança	1	1	0	0	1	0	2	0	0	0
Itabaiana	0	0	0	0	0	0	0	0	0	0
Itaporanga	0	0	0	0	2	0	0	0	2	45
Lagoa Seca	0	0	0	0	1	0	1	0	0	0
Mari	0	0	0	0	0	0	0	0	0	1
Monteiro	1	5	0	1	11	1	0	11	6	174
Pedras de Fogo	0	0	0	0	1	0	0	0	0	1
Pombal	0	0	0	0	1	0	0	0	0	0
Princesa Isabel	0	0	0	0	0	0	1	0	0	10
Rio Tinto	0	0	0	0	0	0	0	0	0	0
São Bento	0	2	0	0	0	0	0	0	0	21
Solânea	2	0	0	0	0	0	0	0	0	0

Fonte: SINAN

Quadro 7: Casos de dengue na 3ª Semana Epidemiológica entre (2007-2016)

	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Alagoa Grande	1	1	0	0	0	0	0	0	2	1
Areia	0	0	0	0	0	0	0	0	0	0
Bananeiras	0	0	0	0	0	0	0	0	0	2
Caaporã	1	0	0	0	1	0	0	0	0	3
Catolé do Rocha	0	3	0	6	0	1	0	1	0	3
Conde	0	1	0	0	0	0	0	0	0	0
Esperança	4	1	0	0	1	0	0	0	0	0
Itabaiana	0	0	0	0	0	0	1	0	0	3
Itaporanga	0	0	0	0	1	0	4	2	0	156
Lagoa Seca	0	2	0	0	1	0	0	1	0	0
Mari	2	0	0	0	0	0	0	0	0	0
Monteiro	2	10	1	2	9	0	0	7	2	139
Pedras de Fogo	0	0	0	0	1	0	0	0	0	0
Pombal	0	1	0	0	4	0	0	0	0	0
Princesa Isabel	0	0	0	0	0	0	0	0	0	17
Rio Tinto	0	0	0	0	0	0	0	0	0	0
São Bento	0	4	0	0	0	1	0	0	0	47
Solânea	1	0	0	0	0	0	0	0	0	0

Fonte: SINAN

Quadro 8: Casos de dengue na 4ª Semana Epidemiológica entre (2007-2016)

	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Alagoa Grande	0	1	0	0	0	0	0	0	0	0
Areia	0	0	0	0	1	0	0	0	0	0
Bananeiras	1	0	0	0	0	0	0	0	1	0
Caaporã	1	1	0	0	1	0	0	0	0	7
Catolé do Rocha	0	4	0	16	2	2	2	2	0	4
Conde	0	1	0	0	0	0	0	0	0	0
Esperança	2	1	0	0	4	0	0	0	0	6
Itabaiana	0	0	0	0	0	1	0	0	0	2
Itaporanga	0	0	0	0	1	0	2	0	6	233
Lagoa Seca	0	0	0	0	2	0	0	1	0	0
Mari	0	0	0	0	0	0	0	0	0	0
Monteiro	5	2	4	1	20	4	0	6	8	142
Pedras de Fogo	0	1	0	0	2	0	0	0	0	2
Pombal	0	0	0	0	6	0	0	0	1	1
Princesa Isabel	0	0	0	0	0	0	0	0	0	34
Rio Tinto	2	0	0	0	0	0	0	1	0	0
São Bento	0	22	0	0	0	0	0	0	0	57
Solânea	0	0	0	0	0	0	0	0	0	0

Fonte: SINAN

Nas primeiras semanas não foram encontradas regras de associação, mas é importante observar uma grande diferença nos dados de 2016 dos municípios de Itaporanga (450 casos), Monteiro (806 casos) e São Bento (128 casos), ambos em janeiro de 2016. Uma possível justificativa para o elevado número de casos é a grande preocupação da população e dos órgãos de vigilância quanto aos agravos de Dengue, CHKV e ZIKV, ainda associados aos casos de Microcefalia. O ano foi incomum, ou seja não há uma associação com base nos anos anteriores. Ainda, através de uma abordagem empírica, foi analisado o índice pluviométrico e foi observado que entre dezembro de 2015 e janeiro de 2016 foi registrado um aumento expressivo no volume de chuvas comparado com o mesmo período do ano anterior, dezembro de 2014 e janeiro de 2015. Em Itaporanga o volume de chuvas subiu de 30,4 mm para 165,2, Monteiro de 20,6mm para 322,3mm e São Bento de 7mm para 187,7mm, todos no período já citado, respectivamente.

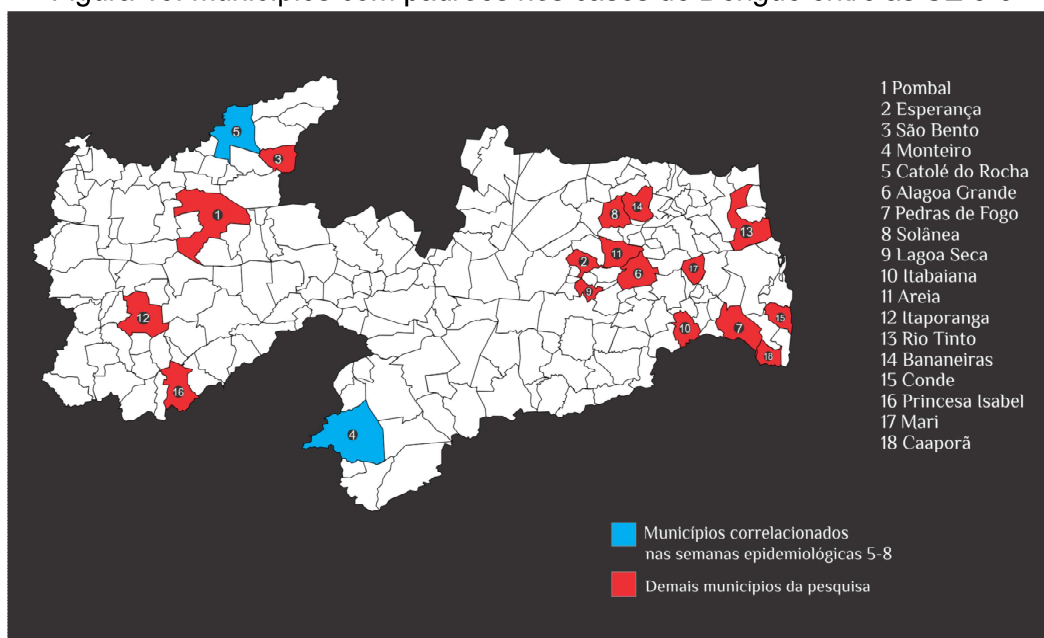
Os dados pluviométricos da AESA permitem concluir que independentemente da estação do ano e as médias anuais (nas semanas especificadas), as campanhas de monitoramento e informação devem observar

o aumento do índice pluviométrico fornecido pelas agências reguladoras. Não obstante, as campanhas nacionais que geralmente seguem estações do ano bem definidas, como o verão, na qual os casos geralmente aumentam nas regiões sul/sudeste, devem ser interiorizadas para atender as necessidades locais. Os dados foram comparados através de uma busca ativa no sítio eletrônico: <http://www.aesa.pb.gov.br/>.

4.2 Análise 2: Semanas Epidemiológicas 5 a 8

Entre as semanas 5 a 8, a análise das regras de associação foi observado um padrão entre os municípios de Catolé do Rocha e Monteiro (Figura 15), (Quadro 9), levando a concluir que existe uma forte associação de casos durante os 10 anos de notificações nos municípios citados. A análise da regra foi vista até uma taxa de confiabilidade de 80% entre 5000 linhas de regras geradas pelo WEKA. Foram sintetizadas 15 regras no Quadro 8, que traz informações que vão de 91% a 95% de confiabilidade, trazendo regras válidas dentro da faixa selecionada.

Figura 15: Municípios com padrões nos casos de Dengue entre as SE 5-8



Fonte: Próprio autor

Quadro 9: Algumas regras de associação geradas entre as Semanas 5-8

1136. Monteiro=S CatoledoRocha=S RioTinto=N 20 ==> Solanea=N 19 <conf:(0.95)> lift:(1.09) lev:(0.04) [1] conv:(1.25)
1137. Monteiro=S CatoledoRocha=S Solanea=N 20 ==> RioTinto=N 19 <conf:(0.95)> lift:(1.15) lev:(0.06) [2] conv:(1.75)
1140. CatoledoRocha=S PedrasdeFogo=N Solanea=N 20 ==> Areia=N 19 <conf:(0.95)> lift:(1.27) lev:(0.1) [3] conv:(2.5)
1144. CatoledoRocha=S Solanea=N Princesalsabel=N 20 ==> RioTinto=N 19 <conf:(0.95)> lift:(1.15) lev:(0.06) [2] conv:(1.75)
1145. CatoledoRocha=S Conde=N Caapora=N 20 ==> Solanea=N 19 <conf:(0.95)> lift:(1.09) lev:(0.04) [1] conv:(1.25)
1146. CatoledoRocha=S LagoaSeca=N Areia=N 20 ==> RioTinto=N 19 <conf:(0.95)> lift:(1.15) lev:(0.06) [2] conv:(1.75)
1147. CatoledoRocha=S Itabaiana=N Areia=N 20 ==> RioTinto=N 19 <conf:(0.95)> lift:(1.15) lev:(0.06) [2] conv:(1.75)
1148. CatoledoRocha=S Itabaiana=N Areia=N 20 ==> Bananeiras=N 19 <conf:(0.95)> lift:(1.27) lev:(0.1) [3] conv:(2.5)
1149. CatoledoRocha=S Bananeiras=N Mori=N 20 ==> Areia=N 19 <conf:(0.95)> lift:(1.27) lev:(0.1) [3] conv:(2.5)
1150. CatoledoRocha=S Conde=N Caapora=N 20 ==> Areia=N 19 <conf:(0.95)> lift:(1.27) lev:(0.1) [3] conv:(2.5)
1151. CatoledoRocha=S Bananeiras=N Mori=N 20 ==> RioTinto=N 19 <conf:(0.95)> lift:(1.15) lev:(0.06) [2] conv:(1.75)
2033. Monteiro=S CatoledoRocha=S 22 ==> Solanea=N 20 <conf:(0.91)> lift:(1.04) lev:(0.02) [0] conv:(0.92)
2034. Monteiro=S CatoledoRocha=S 22 ==> RioTinto=N 20 <conf:(0.91)> lift:(1.1) lev:(0.05) [1] conv:(1.28)
2035. CatoledoRocha=S Itabaiana=N 22 ==> Areia=N 20 <conf:(0.91)> lift:(1.21) lev:(0.09) [3] conv:(1.83)
2036. CatoledoRocha=S Bananeiras=N 22 ==> Mori=N 20 <conf:(0.91)> lift:(1.07) lev:(0.03) [1] conv:(1.1)

Fonte: Próprio Autor

No Quadro 9 é possível analisar algumas regras geradas no WEKA sobre os municípios entre a 5ª e a 8ª semanas epidemiológicas. Na ocasião, é visto que a não associação com outros municípios e a correlação com os já descritos - Monteiro e Catolé do Rocha. A leitura das regras é feita da seguinte forma, como apresentado no Quadro 10:

Quadro 10: Exemplificação de uma regra de associação

2034. Monteiro=S CatoledoRocha=S 22 ==> RioTinto=N 20 <conf:(0.91)> lift:(1.1) lev:(0.05) [1] conv:(1.28)
--

Fonte: Próprio Autor

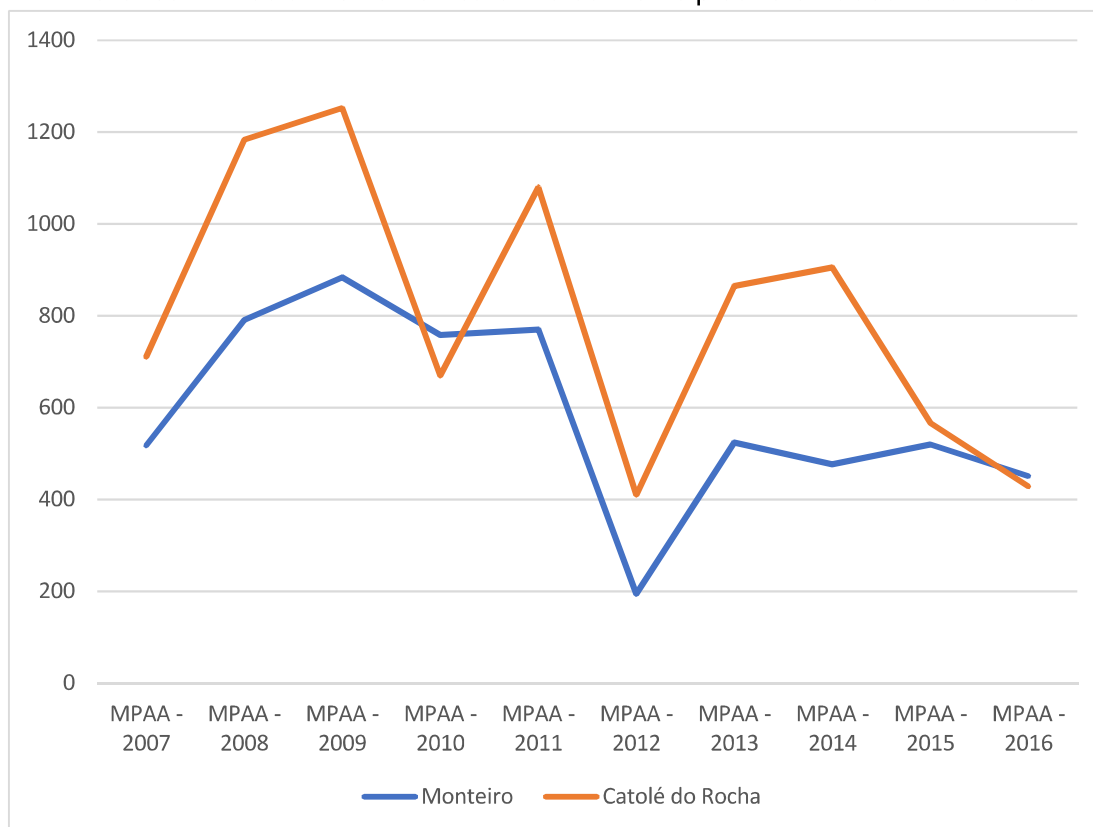
- O “2034” representa a linha que a regra se encontra;
- “Monteiro=S” diz que o atributo “Monteiro” teve como associação o “S” que no caso representa que “SIM” para os casos de Dengue na semana;
- O “22” representa o número de instancias que a regras cobre;
- O “20” mostra o número de instancias preditas corretamente;
- “Conf” é o ajuste de confiança, ou seja, o percentual de ocorrência da regra;
- Os demais parâmetros como “lift”, “lev” e “conv” dependem dos resultados das instâncias, cada um tem uma fórmula definida, e são calculados automaticamente durante as iterações do algoritmo.

No caso a leitura é feita da seguinte forma: “Se houve Dengue em Monteiro então também houve em Catolé do Rocha”. Tanto parâmetros positivos como negativos podem representar uma regra de associação, cabe ao analista julgar quais são aplicáveis ou não.

Buscando uma correção com outras fontes de dados, experimentos foram feitos para descobrir se os atributos (municípios) retornados seguiam outros padrões observáveis. Portanto, extraiu-se aos dados das Médias Pluviométricas Anuais Acumuladas (MPAA) em 10 anos de dados (2007-2016). As informações pluviométricas são da Agência Executiva de Gestão de Águas do Estado da Paraíba (AESA).

O Gráfico 1 diz respeito as médias anuais dos municípios que foram correlacionados nesta análise 2, buscando entender mais profundamente suas características e peculiaridades. O eixo x representa a MPAA durante os anos que se seguem (2007 a 2016) e o eixo y a quantidade em milímetros de chuvas para os anos correspondentes.

Gráfico 1: Média Pluviométrica entre os municípios retornados na análise 2



Fonte: Próprio Autor

No caso acima é possível concluir que, além de existir um padrão nos casos, há uma simetria nos picos para os municípios em questão (Monteiro e Catolé do Rocha). Ou seja, praticamente todas vezes que o índice pluviométrico sobe ou decai em Monteiro, também se eleva ou decai em Catolé do Rocha, o que pode ser configurado um padrão no pico pluviométrico nesses municípios.

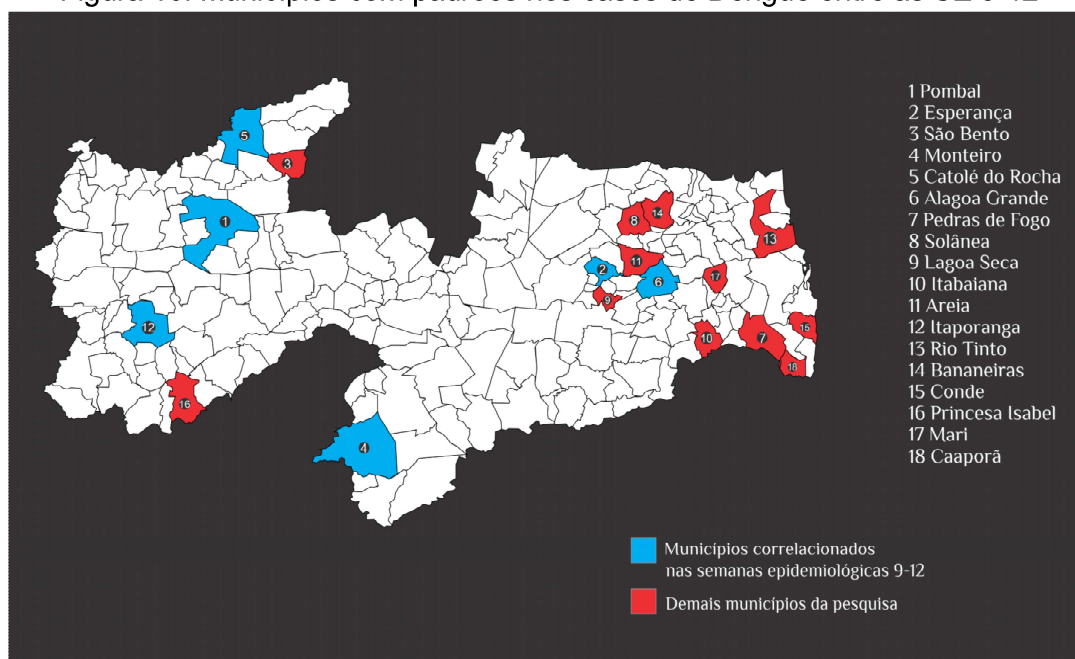
Portanto, esta análise estabelece pontos interessantes a respeito dos dados retornados, levando a concluir que:

- Catolé do Rocha e Monteiro vêm seguindo um padrão no número de casos durante os 10 anos;
- Catolé do Rocha e Monteiro estão intimamente relacionados, também, aos índices pluviométricos, tendo picos semelhantes e confirmando a relação entre eles;
- Entre os 18 municípios analisados, nas Semanas Epidemiológicas 5 a 8, uma forte estrutura correlacional aponta a necessidade de ações interiorizadas nesse período do ano.

4.3 Análise 3: Semanas Epidemiológicas 9 a 12

Entre a nona e décima segunda (9-12) Semanas Epidemiológicas foram identificados padrões nos municípios de Pombal, Monteiro, Catolé do Rocha, Esperança, Itaporanga e Alagoa Grande (Figura 16). A análise foi feita até uma taxa de confiabilidade de 80% tendo sido analisadas 5000 linhas de regras de associação geradas pelo WEKA. Tais regras de associação propõem que os municípios citados têm uma relação e casos de dengue entre aproximadamente os dias 26/02 a 25/03, mostrando que, caso sejam feitas campanhas de combate à Dengue segmentadas, o grau de sucesso das mesmas será maior. Algumas regras de associação podem ser vistas no Quadro 11.

Figura 16: Municípios com padrões nos casos de Dengue entre as SE 9-12



Fonte: Próprio Autor

Quadro 11: Algumas regras de associação geradas entre as Semanas 9-12

82. Pombal=S 16 ==> Esperanca=S 16 <conf:(1)> lift:(1.48) lev:(0.13) [5] conv:(5.2)
83. Pombal=S 16 ==> CatoledoRocha=S 16 <conf:(1)> lift:(1.25) lev:(0.08) [3] conv:(3.2)
85. Pombal=S CatoledoRocha=S 16 ==> Esperanca=S 16 <conf:(1)> lift:(1.48) lev:(0.13) [5] conv:(5.2)
87. Pombal=S 16 ==> Esperanca=S CatoledoRocha=S 16 <conf:(1)> lift:(1.74) lev:(0.17) [6] conv:(6.8)
90. AlagoaGrande=S Solanea=N 16 ==> CatoledoRocha=S 16 <conf:(1)> lift:(1.25) lev:(0.08) [3] conv:(3.2)
99. Monteiro=S AlagoaGrande=S Itaporanga=S 16 ==> Esperanca=S 16 <conf:(1)> lift:(1.48) lev:(0.13) [5] conv:(5.2)

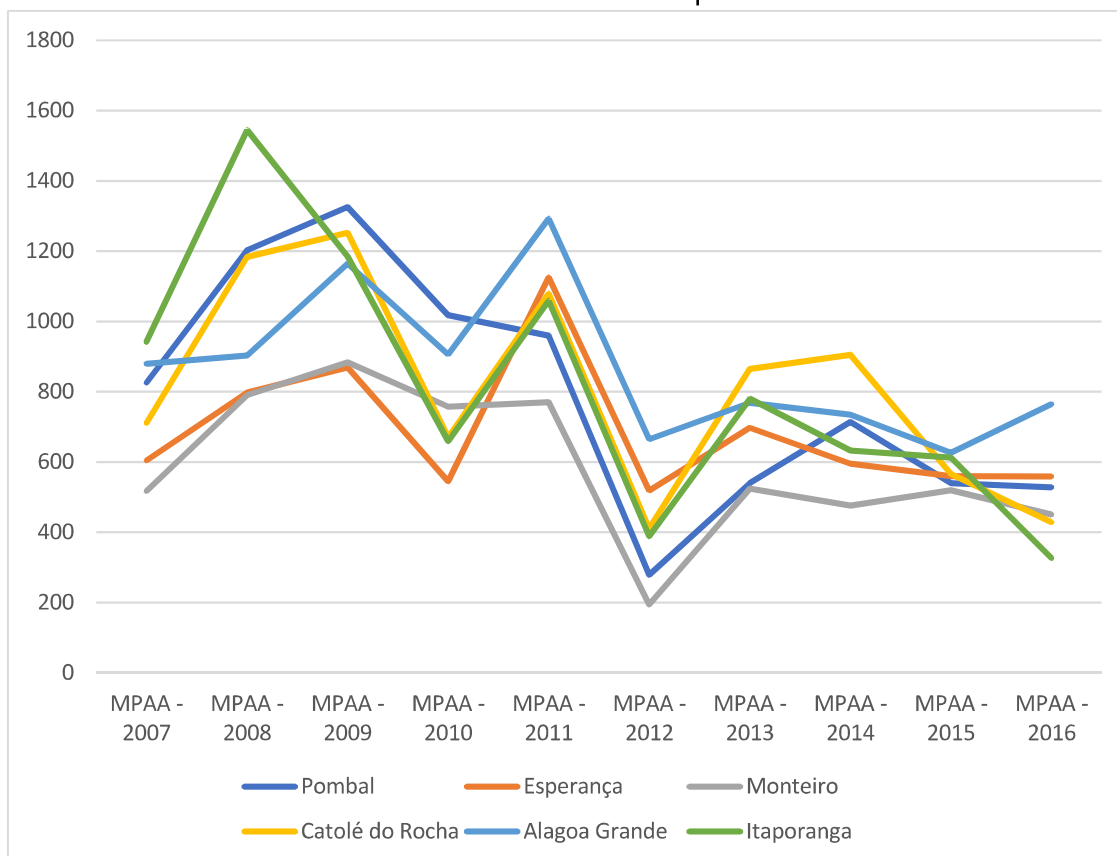
101. Monteiro=S AlagoaGrande=S Itaporanga=S 16 ==> CatoledoRocha=S 16 <conf:(1)> lift:(1.25) lev:(0.08) [3] conv:(3.2)
128. Monteiro=S CatoledoRocha=S AlagoaGrande=S Itaporanga=S 16 ==> Esperanca=S 16 <conf:(1)> lift:(1.48) lev:(0.13) [5] conv:(5.2)
129. Esperanca=S Monteiro=S AlagoaGrande=S Itaporanga=S 16 ==> CatoledoRocha=S 16 <conf:(1)> lift:(1.25) lev:(0.08) [3] conv:(3.2)
130. Monteiro=S AlagoaGrande=S Itaporanga=S 16 ==> Esperanca=S CatoledoRocha=S 16 <conf:(1)> lift:(1.74) lev:(0.17) [6] conv:(6.8)
160. Pombal=S Itaporanga=S 15 ==> Esperanca=S 15 <conf:(1)> lift:(1.48) lev:(0.12) [4] conv:(4.88)
161. Pombal=S Itaporanga=S 15 ==> CatoledoRocha=S 15 <conf:(1)> lift:(1.25) lev:(0.07) [2] conv:(3)
174. Pombal=S CatoledoRocha=S Itaporanga=S 15 ==> Esperanca=S 15 <conf:(1)> lift:(1.48) lev:(0.12) [4] conv:(4.88)
175. Pombal=S Esperanca=S Itaporanga=S 15 ==> CatoledoRocha=S 15 <conf:(1)> lift:(1.25) lev:(0.07) [2] conv:(3)
176. Pombal=S Itaporanga=S 15 ==> Esperanca=S CatoledoRocha=S 15 <conf:(1)> lift:(1.74) lev:(0.16) [6] conv:(6.38)

Fonte: Próprio Autor

Nesta análise, também foram vistas semelhanças em relação a elevação e diminuição na MPAA (Gráfico 2) ao longo dos anos, o que reforça a ideia que as regras são válidas, apenas com algumas diferenças, como, por exemplo, entre Alagoa Grande que se elevou no ano de 2016 em relação aos demais que decaíram. Em geral existe uma semelhança que permite concluir que:

- Os municípios de da Figura 16 seguem um padrão no número de casos ao longo dos 10 anos, reforçando a ideia de ações de combate à Dengue de forma segmentada e interiorizada, analisando as peculiaridades;
- As MPAA reforçam a ideia que existe relação intrínseca na análise e o pico no índice de chuvas com o padrão nas regras de associação encontradas.

Gráfico 2: Média Pluviométrica entre os municípios retornados na análise 3

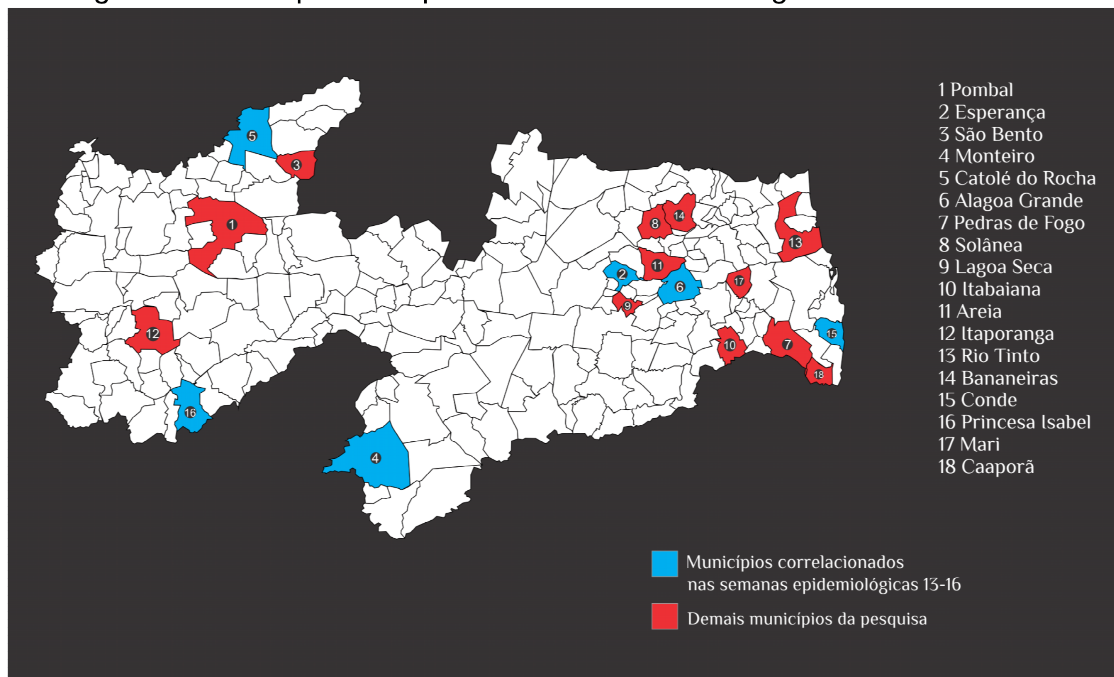


Fonte: Próprio Autor

4.4 Análise 4: Semanas Epidemiológicas 13 a 16

Entre a décima terceira e décima sexta Semanas Epidemiológicas foram identificadas regras de associação entre os municípios de Esperança, Princesa Isabel, Alagoa Grande, Catolé do Rocha, Conde e Monteiro. Na análise foi avaliada uma taxa de confiabilidade de até 80% e a analisadas de 5000 linhas de regras. A Figura 17 ilustra os municípios com associação e o Quadro 12 algumas regras encontradas. Na respectiva análise é possível observar que os municípios citados nessa seção estão diretamente associados e, assim como foi observado, ações de combate a Dengue focadas no período de associação serão mais efetivas.

Figura 17: Municípios com padrões nos casos de Dengue entre as SE 13-16



Fonte: Próprio Autor

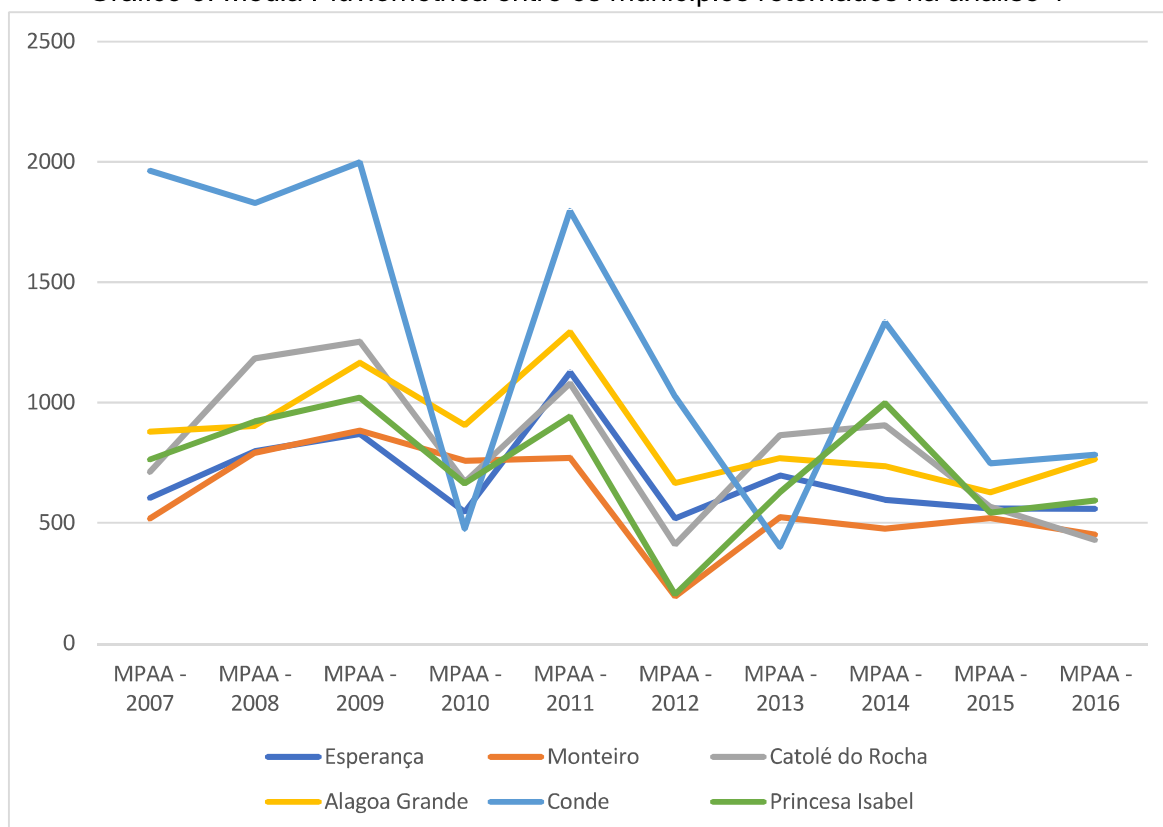
Quadro 12: Algumas regras de associação geradas entre as Semanas 13-16

1. Esperança=S Princesalsabel=S 22 ==> AlagoaGrande=S 22 <conf:(1)> lift:(1.48) lev:(0.18) [7] conv:(7.15)
2. Esperanca=S 25 ==> AlagoaGrande=S 24 <conf:(0.96)> lift:(1.42) lev:(0.18) [7] conv:(4.06)
3. Esperanca=S Monteiro=S 22 ==> AlagoaGrande=S 21 <conf:(0.95)> lift:(1.41) lev:(0.15) [6] conv:(3.58)
5. CatoledoRocha=S Princesalsabel=S 21 ==> AlagoaGrande=S 20 <conf:(0.95)> lift:(1.41) lev:(0.15) [5] conv:(3.41)
7. Conde=S 24 ==> CatoledoRocha=S 22 <conf:(0.92)> lift:(1.08) lev:(0.04) [1] conv:(1.2)
8. AlagoaGrande=S Princesalsabel=S 24 ==> Esperanca=S 22 <conf:(0.92)> lift:(1.47) lev:(0.18) [7] conv:(3)
22. Esperanca=S 25 ==> AlagoaGrande=S Princesalsabel=S 22 <conf:(0.88)> lift:(1.47) lev:(0.18) [7] conv:(2.5)
29. CatoledoRocha=S AlagoaGrande=S 23 ==> Princesalsabel=S 20 <conf:(0.87)> lift:(1.34) lev:(0.13) [5] conv:(2.01)
37. Esperanca=S 25 ==> Monteiro=S AlagoaGrande=S 21 <conf:(0.84)> lift:(1.4) lev:(0.15) [6] conv:(2)
38. CatoledoRocha=S Solanea=N 25 ==> Monteiro=S 21 <conf:(0.84)> lift:(1.02) lev:(0.01) [0] conv:(0.88)
42. Conde=S 24 ==> AlagoaGrande=S 20 <conf:(0.83)> lift:(1.23) lev:(0.09) [3] conv:(1.56)
43. Esperanca=S AlagoaGrande=S 24 ==> CatoledoRocha=S 20 <conf:(0.83)> lift:(0.98) lev:(-0.01) [0] conv:(0.72)
44. Monteiro=S AlagoaGrande=S 24 ==> CatoledoRocha=S 20 <conf:(0.83)> lift:(0.98) lev:(-0.01) [0] conv:(0.72)
47. CatoledoRocha=S 34 ==> Monteiro=S 28 <conf:(0.82)> lift:(1) lev:(-0) [0] conv:(0.85)
49. AlagoaGrande=S 27 ==> Esperanca=S Princesalsabel=S 22 <conf:(0.81)> lift:(1.48) lev:(0.18) [7] conv:(2.02)

Fonte: Próprio Autor

O Quadro 12 mostra que a regra na linha 5, por exemplo, Catolé do Rocha, Princesa Isabel e Alagoa Grande tiveram casos de dengue nas respectivas semanas epidemiológicas sob uma taxa de confiabilidade de 95%. Essa regra em conjunto com as outras que foram analisadas mostram a correlação entre esses e os outros municípios retornados.

Gráfico 3: Média Pluviométrica entre os municípios retornados na análise 4



Fonte: Próprio Autor

O Gráfico 3 apresenta a MPAA durante os 10 anos desses municípios da análise em questão, mostrando a semelhança existente entre o aumento e diminuição das chuvas nos municípios.

Diante do exposto, os resultados permitem concluir que:

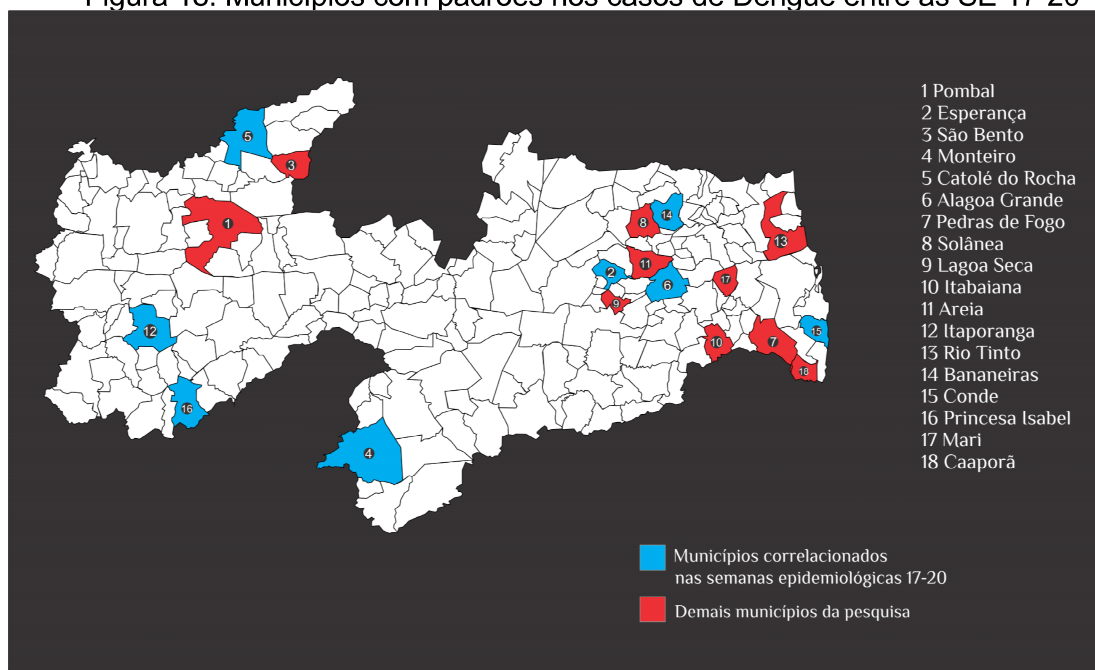
- Esperança, Princesa Isabel, Alagoa Grande, Catolé do Rocha, Conde e Monteiro tiveram uma padronização nos casos nas semanas e durante os 10 anos de análise para as 13^a a 16^a semanas epidemiológicas;
- Ações que envolvem a Vigilância Sanitária e Ambiental dos municípios em questão poderiam ter um melhor desempenho se as ações forem interiorizadas;

- O quadro pluviométrico fortalece a ideia que na análise, traz aumentos e quedas padronizadas, mostrando sua relação intrínseca, apresentando apenas uma diferença no município de Conde na qual os picos foram fora do comum.

4.5 Análise 5: Semanas Epidemiológicas 17 a 20

Entre a décima sétima e vigésima SE foram identificadas regras de associação entre os municípios de Bananeiras, Princesa Isabel, Esperança, Monteiro, Alagoa Grande, Conde, Itaporanga, Catolé do Rocha. A análise foi realizada a uma taxa de confiabilidade de até 80% e a análise de 5000 regras de associação. A Figura 18 e o Quadro 13 ilustram tais representações.

Figura 18: Municípios com padrões nos casos de Dengue entre as SE 17-20



Fonte: Próprio Autor

Quadro 13 Algumas regras de associação geradas entre as Semanas 17-20

15. AlagoaGrande=S Princesalsabel=S 25 ==> Esperanca=S 23 <conf:(0.92)> lift:(1.42) lev:(0.17) [6] conv:(2.92)
16. Esperanca=S Princesalsabel=S 25 ==> AlagoaGrande=S 23 <conf:(0.92)> lift:(1.19) lev:(0.09) [3] conv:(1.88)
20. CatoledoRocha=S AlagoaGrande=S Princesalsabel=S 23 ==> Esperanca=S 21 <conf:(0.91)> lift:(1.4) lev:(0.15) [6] conv:(2.68)
21. Esperanca=S AlagoaGrande=S Princesalsabel=S 23 ==> CatoledoRocha=S 21 <conf:(0.91)> lift:(1.14) lev:(0.07) [2] conv:(1.53)

22. Esperanca=S CatoledoRocha=S 23 ==> AlagoaGrande=S Princesalsabel=S 21 <conf:(0.91)> lift:(1.46) lev:(0.17) [6] conv:(2.88)
24. AlagoaGrande=S RioTinto=N 22 ==> CatoledoRocha=S 20 <conf:(0.91)> lift:(1.14) lev:(0.06) [2] conv:(1.47)
25. Conde=S 22 ==> CatoledoRocha=S Itaporanga=S 20 <conf:(0.91)> lift:(1.58) lev:(0.18) [7] conv:(3.12)
38. CatoledoRocha=S Itaporanga=S 23 ==> Conde=S 20 <conf:(0.87)> lift:(1.58) lev:(0.18) [7] conv:(2.59)
41. Esperanca=S 26 ==> CatoledoRocha=S AlagoaGrande=S 22 <conf:(0.85)> lift:(1.21) lev:(0.1) [3] conv:(1.56)
42. CatoledoRocha=S Princesalsabel=S 26 ==> Esperanca=S 22 <conf:(0.85)> lift:(1.3) lev:(0.13) [5] conv:(1.82)
43. Esperanca=S 26 ==> CatoledoRocha=S Princesalsabel=S 22 <conf:(0.85)> lift:(1.3) lev:(0.13) [5] conv:(1.82)
47. Monteiro=S CatoledoRocha=S 25 ==> AlagoaGrande=S 21 <conf:(0.84)> lift:(1.08) lev:(0.04) [1] conv:(1.13)
48. Monteiro=S Princesalsabel=S 25 ==> CatoledoRocha=S 21 <conf:(0.84)> lift:(1.05) lev:(0.03) [1] conv:(1)
49. Monteiro=S CatoledoRocha=S 25 ==> Princesalsabel=S 21 <conf:(0.84)> lift:(1.12) lev:(0.06) [2] conv:(1.25)
50. AlagoaGrande=S Princesalsabel=S 25 ==> Esperanca=S CatoledoRocha=S 21 <conf:(0.84)> lift:(1.46) lev:(0.17) [6] conv:(2.13)

Fonte: Próprio Autor

Esta análise foi a que mais teve municípios envolvidos, totalizando oito em relação a amostra principal estudada. Aqui é possível observar, por exemplo a regra de número 50 (Quadro 13), que mostra uma relação entre quatro municípios paraibanos (Alagoa Grande, Princesa Isabel, Esperança e Catolé do Rocha), apresentando que esta associação ocorre em 84% dos casos. A regra 21 do mesmo quadro diz que em 92% dos casos houve relação entre os casos em Esperança, Alagoa Grande, Princesa Isabel e Catolé do Rocha.

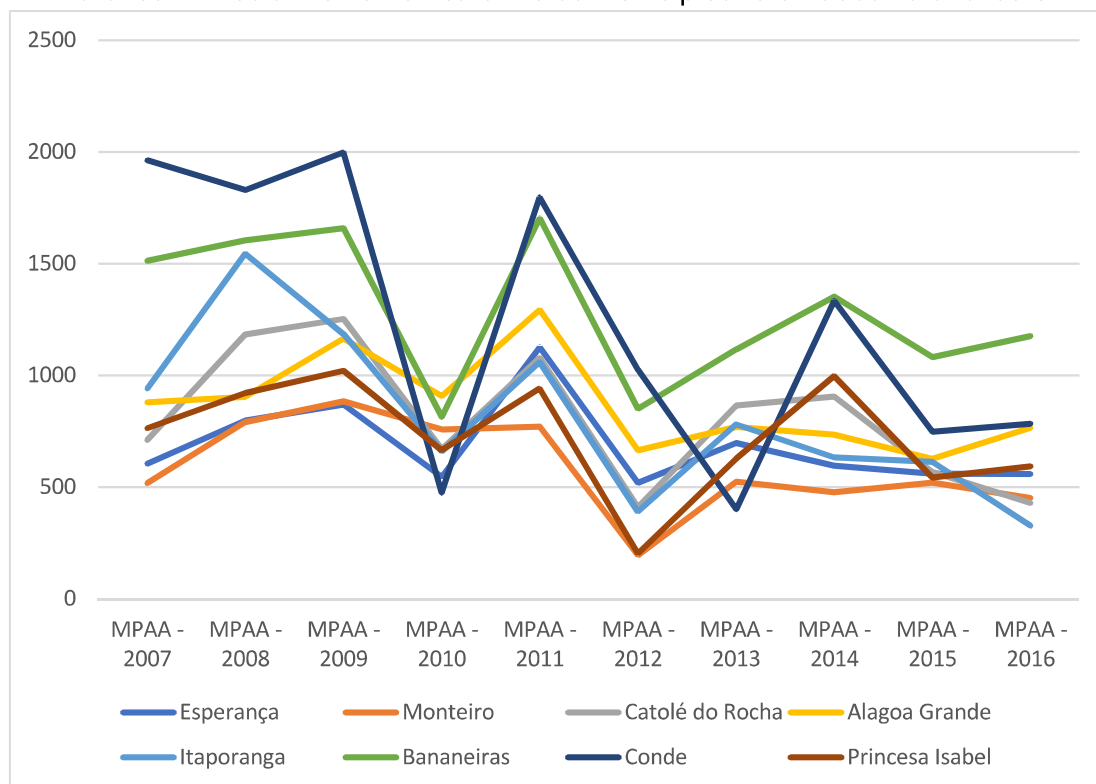
Seguindo a mesma linha de análise, o Gráfico 4 representa outro dado importante, a MPAA diante dos municípios que tiveram regras de associação fortemente relacionadas. Neste caso, quando comparados os atributos usados nesta análise, foi possível observar uma padronização semelhante entre o aumento e a diminuição.

Em geral, as conclusões são semelhantes as outras análises:

- Campanhas de Vigilância Sanitária e Ambiental dos municípios em questão poderiam ter um melhor desempenho seu papel se as ações forem interiorizadas, atendendo as especificidades relacionadas;

- A MPAA é mais uma forma de relacionar os resultados obtidos nas regras de associação, mostrando que existe uma relação entre os padrões nos casos e as taxas pluviométricas, mesmo que em regiões tão diferentes.

Gráfico 4: Média Pluviométrica entre os municípios retornados na análise 5



Fonte: Próprio Autor

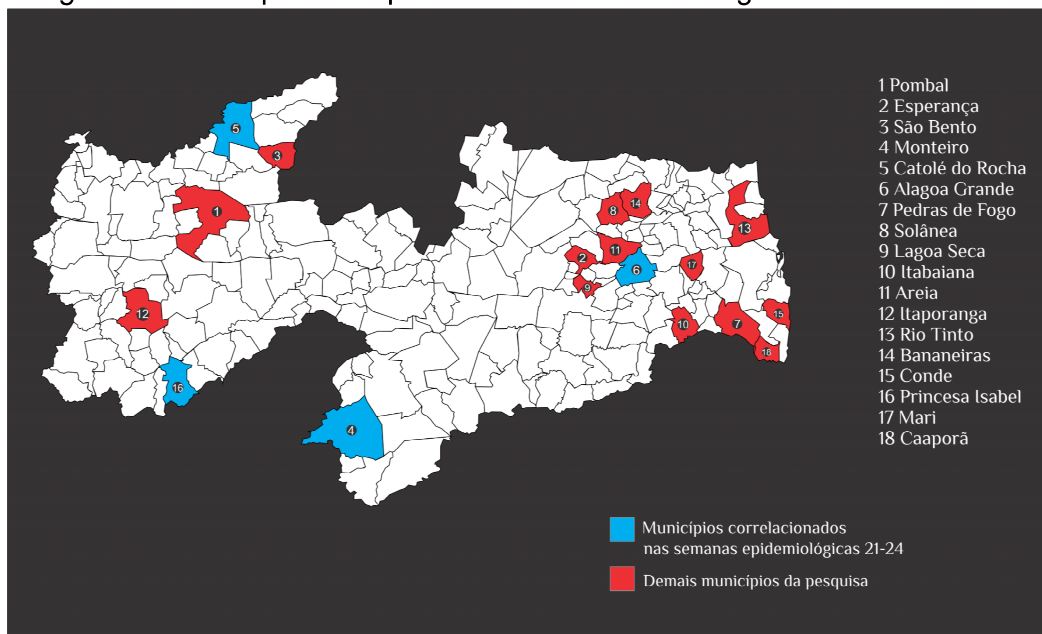
4.6 Análise 6: Semanas Epidemiológicas 21 a 24

Entre a vigésima primeira e vigésima quarta Semanas Epidemiológicas foram identificadas regras de associação entre os municípios de Monteiro, Catolé do Rocha, Alagoa Grande e Princesa Isabel (Figura 19). O Quadro 14 apresenta algumas regras geradas pelo WEKA. É possível observar que, além da relação entre os municípios já citados nessa sessão, o número de municípios com associações começa a decrescer. Tal situação ocorre devido ao número de casos que também começa a ficar menor. Todavia, mesmo com um número baixo, ainda assim, existem relações entre atributos.

Outro ponto importante nestas regras é a representação que existe quando acontece a não associação, como por exemplo, na regra 53 (Quadro 14, que diz que não existe associação nas semanas entre Itabaiana, Rio tinto e

Caaporã, municípios que também não foram elencados em nenhuma análise deste trabalho.

Figura 19: Municípios com padrões nos casos de Dengue entre as SE 21-24



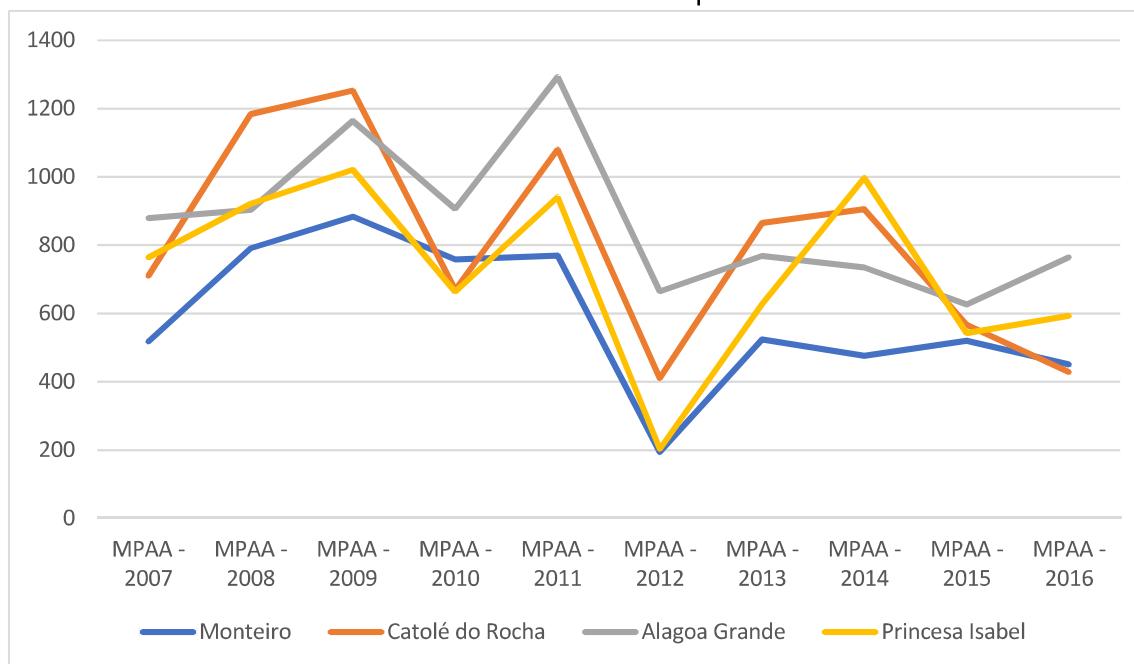
Fonte: Próprio Autor

Quadro 14: Algumas regras de associação geradas entre as Semanas 21-24

34. CatoledoRocha=S AlagoaGrande=S 25 ==> Itabaiana=N 24 <conf:(0.96)> lift:(0.98) lev:(-0.01) [0] conv:(0.31)
40. AlagoaGrande=S Solanea=N 23 ==> Itabaiana=N 22 <conf:(0.96)> lift:(0.98) lev:(-0.01) [0] conv:(0.29)
41. PedrasdeFogo=N Caapora=N 23 ==> Itabaiana=N 22 <conf:(0.96)> lift:(0.98) lev:(-0.01) [0] conv:(0.29)
42. Mori=N 25 ==> Monteiro=S 22 <conf:(0.88)> lift:(1.1) lev:(0.05) [2] conv:(1.25)
43. Mori=N 25 ==> Caapora=N 22 <conf:(0.88)> lift:(1.14) lev:(0.07) [2] conv:(1.41)
52. Itabaiana=N Bananeiras=N 29 ==> Caapora=N 25 <conf:(0.86)> lift:(1.11) lev:(0.06) [2] conv:(1.31)
53. RioTinto=N 28 ==> Caapora=N 24 <conf:(0.86)> lift:(1.11) lev:(0.06) [2] conv:(1.26)
54. Itabaiana=N RioTinto=N 28 ==> Caapora=N 24 <conf:(0.86)> lift:(1.11) lev:(0.06) [2] conv:(1.26)
55. RioTinto=N 28 ==> Itabaiana=N Caapora=N 24 <conf:(0.86)> lift:(1.14) lev:(0.08) [3] conv:(1.4)
56. PedrasdeFogo=N 27 ==> Caapora=N 23 <conf:(0.85)> lift:(1.1) lev:(0.05) [2] conv:(1.22)
57. Itabaiana=N 39 ==> CatoledoRocha=S 33 <conf:(0.85)> lift:(1) lev:(-0) [0] conv:(0.84)
58. Princesalsabel=S 26 ==> Monteiro=S 22 <conf:(0.85)> lift:(1.06) lev:(0.03) [1] conv:(1.04)
59. Princesalsabel=S 26 ==> CatoledoRocha=S 22 <conf:(0.85)> lift:(1) lev:(-0) [0] conv:(0.78)
60. Princesalsabel=S 26 ==> Solanea=N 22 <conf:(0.85)> lift:(1.13) lev:(0.06) [2] conv:(1.3)

Fonte: Próprio Autor

Gráfico 5: Média Pluviométrica entre os municípios retornados na análise 6



Fonte: Próprio Autor

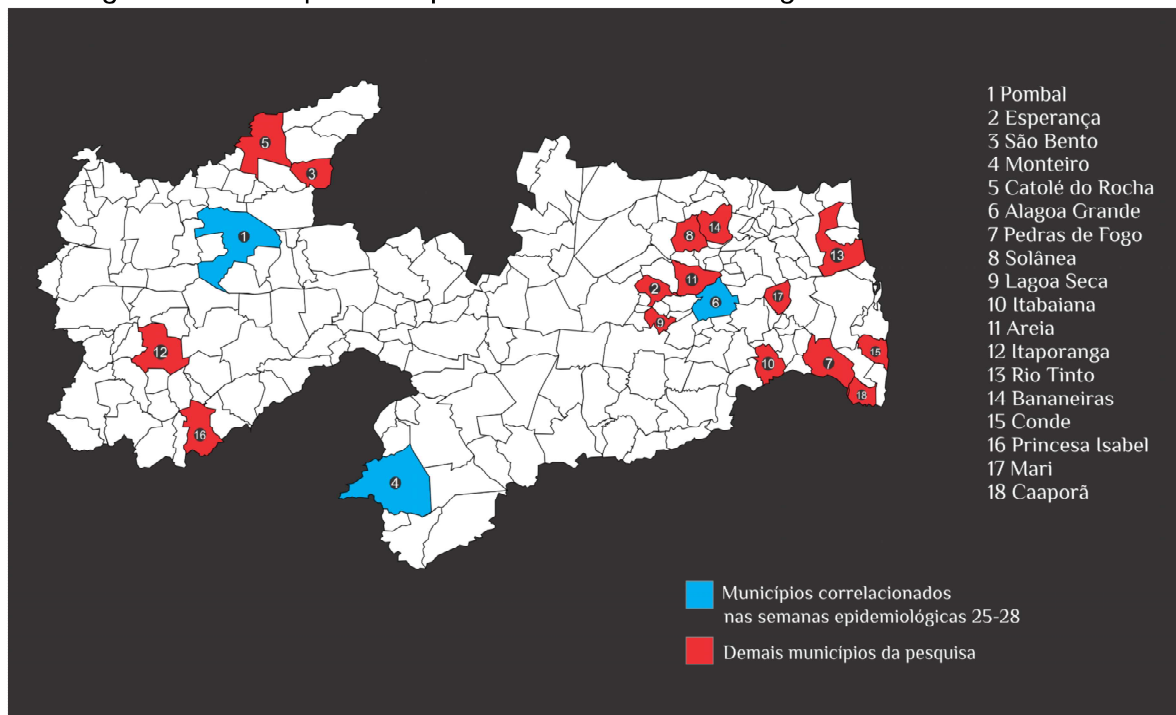
Nesta análise, também foram vistas semelhanças em relação a elevação e diminuição na média pluviométrica (Gráfico 5) ao longo dos anos, o que reforça a ideia que as regras são válidas.

- Os municípios de da Figura 14 seguem um padrão no número de casos ao longo dos 10 anos, reforçando a ideia de ações de combate à Dengue de forma segmentada e interiorizada, analisando as peculiaridades;
- As médias anuais reforçam a ideia que existe relação intrínseca na análise e o pico no índice de chuvas com o padrão nas regras de associação encontradas.

4.7 Análise 7: Semanas Epidemiológicas 25 a 28

Entre a vigésima quinta e vigésima oitava SE foram identificadas regras de associação entre os municípios de Monteiro, Alagoa Grande e Catolé do Rocha. A Figura 20 e o Quadro 15 representam tal afirmação.

Figura 20: Municípios com padrões nos casos de Dengue entre as SE 25-28



Fonte: Próprio Autor

Quadro 15: Algumas regras de associação geradas entre as Semanas 25-28

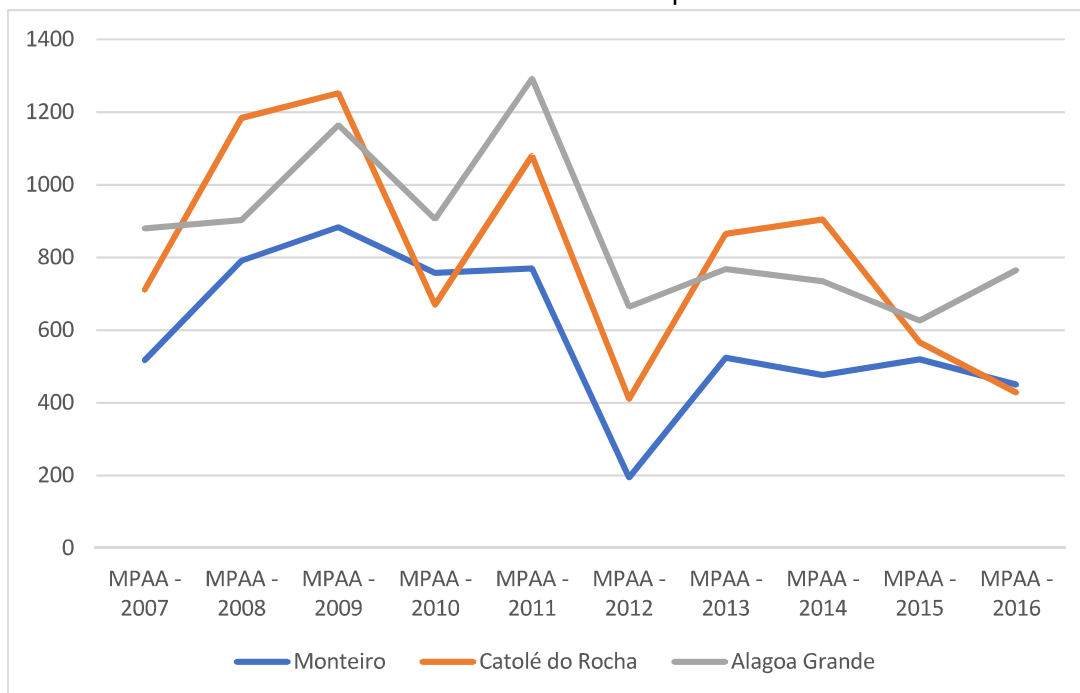
158. Monteiro=S AlagoaGrande=S 18 ==> PedrasdeFogo=N 18 <conf:(1)> lift:(1.03) lev:(0.01) [0] conv:(0.45)
188. Monteiro=S CatoledoRocha=S Itabaiana=N 18 ==> PedrasdeFogo=N 18 <conf:(1)> lift:(1.03) lev:(0.01) [0] conv:(0.45)
465. Monteiro=S CatoledoRocha=S 19 ==> Itabaiana=N 18 <conf:(0.95)> lift:(1.02) lev:(0.01) [0] conv:(0.71)
502. Monteiro=S CatoledoRocha=S PedrasdeFogo=N 19 ==> Itabaiana=N 18 <conf:(0.95)> lift:(1.02) lev:(0.01) [0] conv:(0.71)
503. Monteiro=S CatoledoRocha=S 19 ==> PedrasdeFogo=N Itabaiana=N 18 <conf:(0.95)> lift:(1.05) lev:(0.02) [0] conv:(0.95)

Fonte: Próprio Autor

A análise em questão representa aproximadamente entre a segunda quinzena de julho e a primeira quinzena de agosto e neste período só houveram associação entre três municípios, que estão dispostos em 3 mesorregiões diferentes. A regra 503 do Quadro 15 mostra a similaridade entre os casos de Dengue de Monteiro e Catolé do Rocha, mostrando quando tiveram casos nessas cidades não tiveram em Pedras de Fogo e Itabaiana.

O Gráfico 6 mostra um resumo da MPAA nos municípios correlacionados nessa análise, apontando que há, também, uma similaridade entre o aumento e queda no número de chuvas em questão.

Gráfico 6: Média Pluviométrica entre os municípios retornados na análise 7



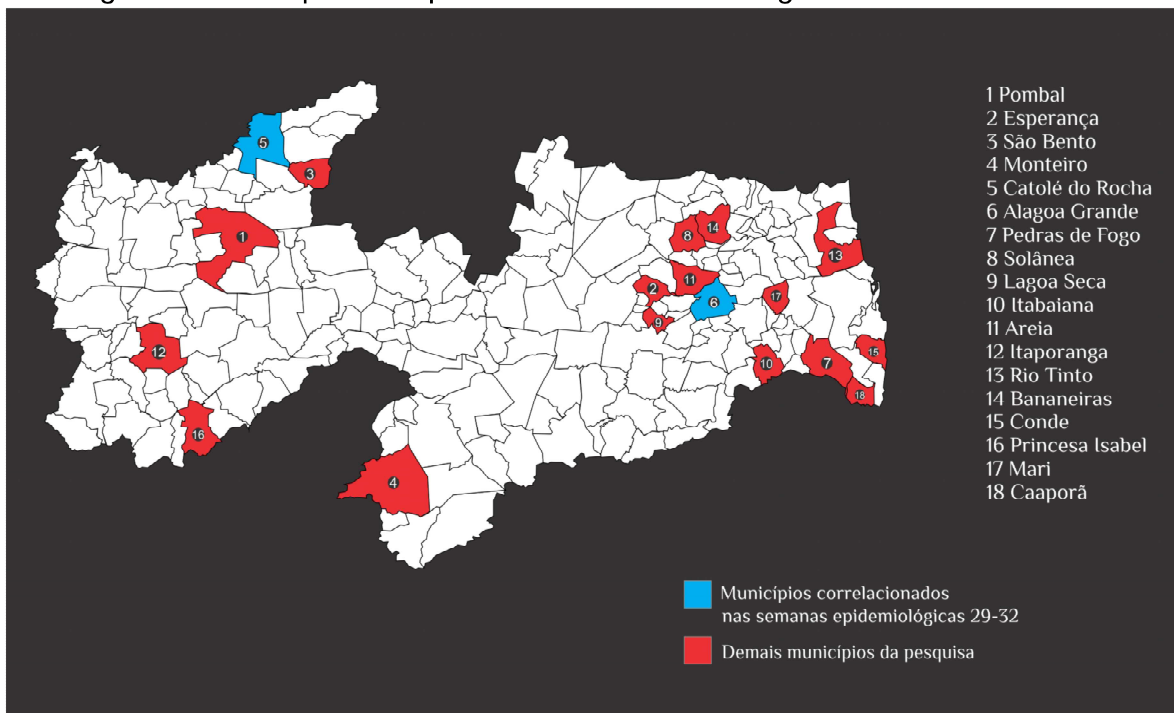
Fonte: Próprio Autor

4.8 Análise 8: Semanas Epidemiológicas 29 a 32

Entre a vigésima quinta e vigésima oitava SE foram identificadas regras de associação entre os municípios de Alagoa Grande e Catolé do Rocha, mostrando que aproximadamente entre as datas correspondentes do calendário epidemiológico, os municípios citados estão com a ocorrência de casos semelhante e seguindo um padrão nesses dez anos. A Figura 21 e o Quadro 16 representam tal afirmação.

Essa análise em questão foi a última que encontrou regras de associação aplicáveis no estudo em questão, mostrando que apenas dois municípios tiveram relação nos 10 anos de dados analisados. A regra de associação 1968 do Quadro 15 mostra que sempre que houve Dengue em Alagoa Grande também teve em Catolé do Rocha, sob uma taxa de confiabilidade de 83%. O Gráfico 7 representa que a MPAA indica uma tendência que pode estar relacionada com os demais casos.

Figura 21: Municípios com padrões nos casos de Dengue entre as SE 29-32



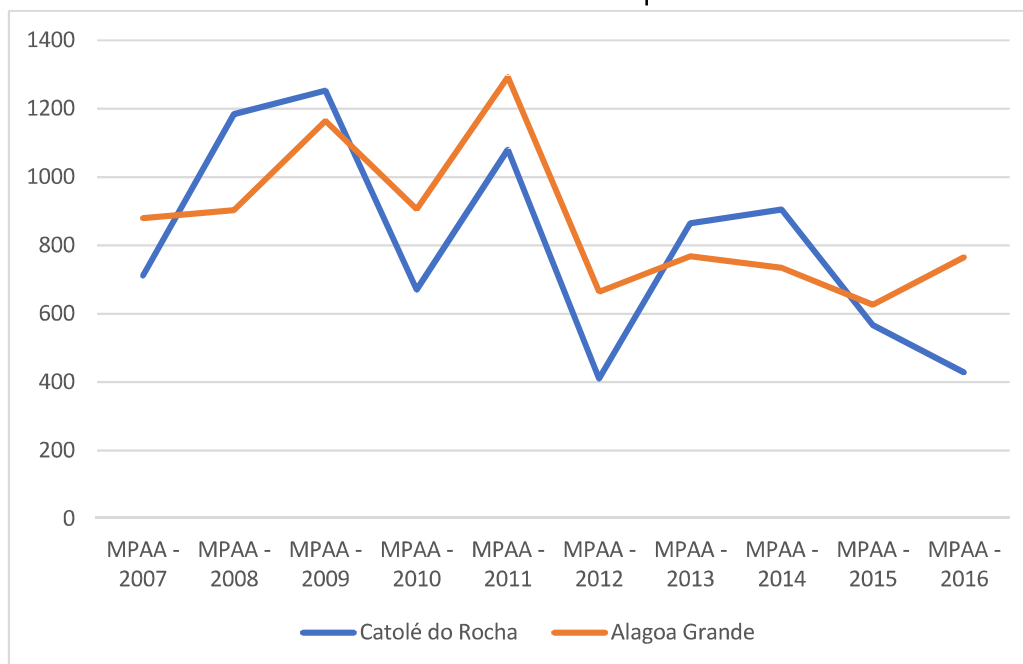
Fonte: Próprio Autor

Quadro 16: Algumas regras de associação geradas entre as SE 29-32

367. CatoledoRocha=S AlagoaGrande=S 20 ==> Areia=N 19 <conf:(0.95)> lift:(1.15) lev:(0.06) [2] conv:(1.75)
1063. CatoledoRocha=S AlagoaGrande=S 20 ==> Itabaiana=N 18 <conf:(0.9)> lift:(0.97) lev:(-0.01) [0] conv:(0.5)
1480. AlagoaGrande=S Areia=N 22 ==> CatoledoRocha=S 19 <conf:(0.86)> lift:(1.33) lev:(0.12) [4] conv:(1.93)
1968. AlagoaGrande=S 24 ==> CatoledoRocha=S 20 <conf:(0.83)> lift:(1.28) lev:(0.11) [4] conv:(1.68)

Fonte: Próprio Autor

Gráfico 7: Média Pluviométrica entre os municípios retornados na análise 8



Fonte: Próprio Autor

4.9 Análise 9, 10, 11, 12 e 13: Semanas Epidemiológicas 33-36, 37-40, 41-44, 45-48 e 49-52

Nas referidas SE não foram encontradas regras de associação aplicáveis nas cidades analisadas. Simulou-se em uma taxa de confiabilidade de 80% e analisaram-se 5000 regras retornadas pelo WEKA. A falta de resultados aceitáveis se dá pelo pouco número de casos nas SE supracitadas, o que leva à conclusão que, segundo as regras de associação analisadas, não há padrões que possam ser usados para análises futuras.

4.10 Considerações a respeito dos resultados obtidos

- As regras de associação são aplicáveis aos casos analisados;
- Os municípios seguem um padrão nos casos de dengue ao longo dos anos;
- Poucos casos de dengue dificultam a mineração de dados nas regras de associação;
- Cada fase do processo de MD é fundamental para o resultado final, principalmente aquelas que dependem mais fortemente do analista (pré-processamento e pós-processamento);

- Foram encontradas mais regras de associação entre as estações do outono e inverno, o que diferencia das campanhas de dengue que geralmente são focadas no verão.
- Os dados das MPAA serviram como panorama para entender uma relação entre o aumento e diminuição no número de chuvas com os padrões encontrados nas análises no WEKA.
- No Quadro 17 existe uma síntese dos municípios e suas análises, mostrando os municípios com relação e os não existem nenhum padrão nos casos ao longo dos anos analisados.

Quadro 17: Resumo dos municípios selecionados ao longo das 13 análises

Municípios	Análise 1	Análise 2	Análise 3	Análise 4	Análise 5	Análise 6	Análise 7	Análise 8	Análise 9	Análise 10	Análise 11	Análise 12	Análise 13
Pombal			X										
Esperança			X	X	X								
São Bento													
Monteiro		X	X	X	X	X	X						
Catolé do Rocha		X	X	X	X	X	X	X					
Alagoa Grande			X	X	X	X	X	X					
Pedras de Fogo													
Solânea													
Lagoa Seca													
Itabaiana													
Areia													
Itaporanga			X		X								
Rio Tinto													
Bananeiras					X								
Conde				X	X								
Princesa Isabel				X	X	X							
Mari													
Caaporã													

Fonte: Próprio Autor

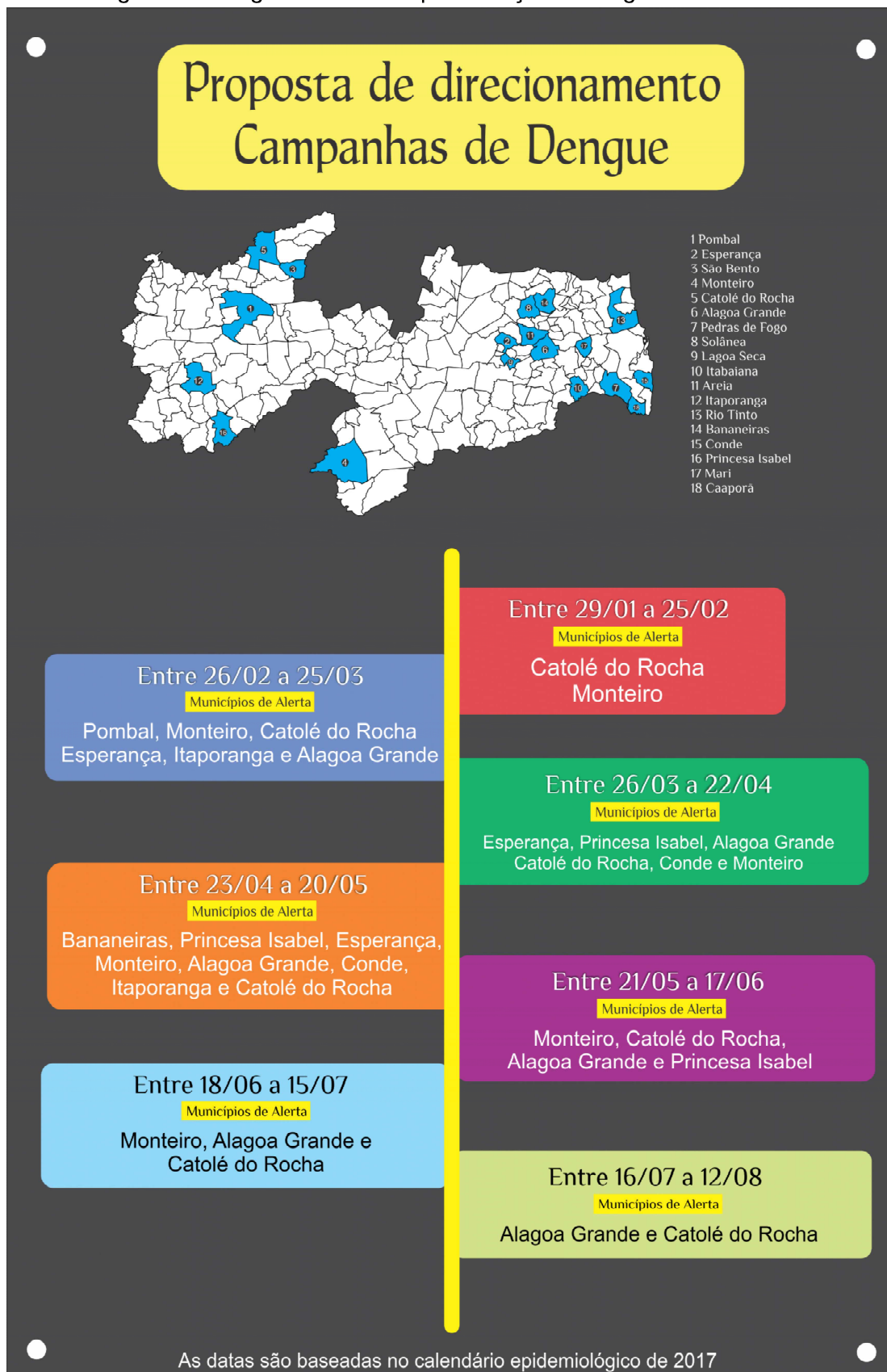
4.11 Sugestão de direcionamento das campanhas de dengue

É fato que o combate ao mosquito *Aedes aegypti* deve ser feito durante todo o ano, mas os sistemas de vigilância usam períodos específicos para intensificarem as campanhas de combate ao mosquito causador da Dengue. A proposta final deste trabalho é criar um melhor direcionamento para que a

informação em saúde chegue mais próximo da realidade e assim tentar diminuir o número de casos.

Através de futuras parcerias entre os municípios alvo desta pesquisa e de suas respectivas associações, um melhor direcionamento de promoção em saúde poderá ser feito. Para ilustrar melhor os resultados encontrados nesta dissertação a Figura 22 apresenta um infográfico das informações encontradas através das regras durante os 10 anos de casos de Dengue nos respectivos municípios.

Figura 22: Infográfico com a representação das regras encontradas



Fonte: Próprio Autor

4.12 Síntese da Seção

Essa seção apresentou os resultados do trabalho em questão, mostrando as relações que houve ao longo da pesquisa e suas particularidades, através de mapas, quadros e gráficos – dividindo as análises de treze e elencando seus resultados conforme a amostra selecionada para cada item. Ao final foi apresentado um infográfico mostrando uma sugestão de datas na qual as campanhas poderiam tomar como base.

5 CONSIDERAÇÕES FINAIS

A aplicabilidade de técnicas de MD são alternativas eficazes para ajudar no combate à Dengue. As regras de associação são eficientes e a ferramenta utilizada é importante gerar regras eficazes, além da análise do autor sobre as regras geradas.

Foram 13 análises que envolveram as 52 semanas epidemiológicas e a partir disso encontradas regras que são aplicáveis aos programas de combate à Dengue, mostrando que a hipótese deste trabalho é executável. Os padrões que têm sido identificados fizeram parte da grande proposta dessa dissertação - Direcionar as ações de combate à Dengue nos municípios analisados. Ações que envolvem a Vigilância Sanitária e Ambiental dos municípios em questão podem ter um melhor desempenho se as ações forem interiorizadas, segmentadas e adaptadas para a realidade encontrada nessa pesquisa.

As taxas pluviométricas ao longo dos anos foram diretrizes importantes para uma concepção análoga aos resultados obtidos, o que é interessante notar sua simetria gráfica ao longo anos e, que propiciaram uma diretriz ao aumento da população do *Aedes aegypti* e ao desenvolvimento do vírus da dengue. Estes dados usados como observação gráfica/visual servirão para estudos posteriores a este trabalho.

Ao final das análises foi criado um infográfico de direcionamento das campanhas, a fim de colaborar para o combate dos casos de Dengue. Trabalhos futuros ficarão responsáveis por refazer os experimentos adotando uma base de dados maior e abrangendo mais municípios, além de analisar outros tipos métodos implementados no WEKA.

Desta forma, as regras de associação podem ser usadas como um aliado às atividades de vigilância entomológica, proporcionando um conhecimento para avaliações de casos ocorridos e sobretudo informações que contribuam para previsões e ações preventivas no combate a proliferação da doença.

REFERÊNCIAS

- AGRAWAL, R; IMIELINSKI, T; SWAMI, A. Mining Association Rules between Sets of Items in Large Databases. *ACM SIGMOD Record*, New York, v. 22, n. 2, p. 207-216, 1993.
- AGRAWAL, R; SRIKANT, R. Fast Algorithms for Mining Association Rules. *Proceedings*, Santiago, 1994.
- AGRAWAL, R; PSAILA, G. Active Data Mining. *American Association for Artificial Intelligence*, Menlo Park, p. 3–8, 1995.
- AGRAWAL, R. et al. Fast Discovery of Association Rules. *AAAI Press*, Menlo Park, p. 307–328, 1996.
- BERTOLLI JUNIOR, C. A história social da tuberculose e do tuberculoso: 1900-1950. 1993. Tese (Livre – Docência). Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo.
- BOECHAT, B. Ministério da Saúde divulga principais criadouros de dengue por região. *Extra*, Rio de Janeiro, 19 mar. 2014. Disponível em: <<https://extra.globo.com/noticias/saude-e-ciencia/dengue/ministerio-da-saude-divulga-principais-criadouros-de-dengue-por-regiao-11921245.html>>. Acesso em: 30 nov. 2016.
- BRAGA, C. et al. Seroprevalence and risk factors for dengue infection in socio-economically distinct areas of Recife. *Acta Tropical*, [S.L.], v. 113, n. 3, p. 234-240, mar./jan. 2015.
- BRAGA, P. A.; LUDERMIR, T. P.; CARVALHO, A. C. P. L. F. Redes neurais artificiais: teoria e aplicações. [S.L.]: LTC, 2000.
- BRASIL, Ministério da Saúde. Levantamento de Índice Rápido para o *Aedes aegypti*. Brasília, nov. 2015. Disponível em: <<http://portalsaude.saude.gov.br/images/pdf/2015/novembro/24/apresentacao-LIRAA-boletim-Dengue.pdf>>. Acesso em: 19 nov. 2016.
- _____, Ministério da Saúde. Aplicativo para celular ajudará a combater o *Aedes aegypti*. Brasília, 2016. Disponível em: <http://www.brasil.gov.br/saude/2016/02/aplicativo-para-celular-ajudara-a-combater-o-aedes-aegypti>. Acessado em: 1 dez. 2016.
- _____, Ministério da Saúde. Monitoramento dos casos de dengue, febre de chikungunya e febre pelo vírus Zika até a Semana Epidemiológica 15, 2017. Brasília, 2017. Disponível em: <<http://portalquivos.saude.gov.br/images/pdf/2017/maio/05/Monitoramento-dos-casos-de-dengue-febre-de-chikungunya-e-febre-pelo-virus-Zika-ate-a-Semana-Epidemiologica.pdf>>. Acessado em: 26 mai. 2017.

BREILH, J. Epidemiologia: economia, politica e saude. 1 ed. São Paulo: UNESP/HUCITEC, 1991.

CARDOSO, B. F. et al. Detection of Oropouche virus segment S in patients and in *Culex quinquefasciatus* in the state of Mato Grosso, Brazil. Mem Inst Oswaldo Cruz, [S.L], v. 110, n. 6, p. 745-54, 2015.

CARMO, Roberto Luiz. Dengue e chikungunya: estudos da relação entre população, ambiente e saúde. 1 ed. Campinas: UNICAMP, 2015.

CHEESEMAN, P. On Finding the Most Probable Model. In Computational Models of Scientific Discovery and Theory Formation, San Francisco, p. 73–95, 1990.

DANTAS, C; LENHARO M. Febre amarela, dengue, zika e chikungunya: entenda as doenças do Aedes que afetam o Brasil. G1. 2017. Disponível em: <http://g1.globo.com/bemestar/noticia/febre-amarela-dengue-zika-e-chikungunya-entenda-as-doencas-do-aedes-que-afetam-o-brasil.html>. Acessado em: 25 fev. 2017.

DIAS, M. M. Um modelo de formalização do processo de sistema de descoberta de Conhecimento em banco de dados. 2001. Tese. Pós Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Santa Catarina.

DIAS, M M. Parâmetros na escolha de técnicas e ferramentas de mineração de dados. Acta Scientiarum, Maringá, v. 24, p. 1715-1725, 2002.

DIAS, Larissa Barbiero De Almeida. Dengue: transmissão, aspectos aspectos clínicos, diagnóstico e tratamento.. Fmrp (USP), Ribeirão Preto, v. 43, n. 2, 2010. Disponível em: <http://revista.fmrp.usp.br/2010/vol43n2/Simp6_Dengue.pdf>. Acesso em: 24 mai. 2017.

FAYYAD, U. Advances in knowledge discovery and data mining. 1 ed. Menlo Park: AAAI Press, 1996.

FAYYAD, U; PIATETSKY, S. G; SMYTH, P. He KDD process for extracting useful knowledge from volumes of data. Communications of the ACM, [S.L], v. 39, p. 27-35, 1996.

GOEBEL, M; GRUENWALD, D. L. A Survey of Data Mining and Knowledge Discovery Software Tools. ACM SIGKDD Explorations, New York, v. 1, n. 1, p. 20-33, 1999.

GOLDBERG, D. E.; DEB, K. A comparative analysis of selection schemes used in genetic algorithms: foundations of genetic algorithms (foga). 1 ed. São Francisco: Morgan Kaufmann Publishers, 1991.

- GOLDSCHMIDT, Ronaldo. Data mining : conceitos, técnicas, algoritmos, orientações e aplicações. 2 ed. Rio de Janeiro : Elsevier, 2015.
- GUBLER, D. Aedes aegypti and Aedes aegypti-Borne Disease Control in the 1990s: Top Down or Bottom Up. *The American Journal of Tropical Medicine and Hygiene*, [S.L], v. 40, n. 6, p. 571-578, 2015.
- GUHA-SAPIR, D; SCHIMMER, B. Dengue fever: new paradigms for a changing epidemiology. *Emerg Themes Epidemiol*, [S.L], v. 2, n. 1, 2005.
- HARRISON, T H. Intranet data warehouse. São Paulo: Berkeley Brasil, 1998.
- JAIN, A.; DUBES, R. Algorithms for clustering data. 1 ed. Englewood Cliffs: Prentice-Hall, 1988.
- KASS-HOUT, T; ALHINNAWI, H. Social media in public health. *British Medical Bulletin*, [S.L], v. 108, n. 1, p. 5-24, 2013.
- KAUR, P; GOYAL, M; LU, J. Data Mining driven agentes for predicting online auction's end price. *Symposium on Computational Intelligence and Data Mining (CIDM)*, Paris, p. 141-147, 2001.
- KRAEMER, M. M. G. The global compendium of Aedes aegypti and Ae. albopictus occurrence. *Scientific Data*, [S.L], v. 2, jul. 2015.
- LAROSE, D. Discovering knowledge in data: An introduction to data mining. Hoboken: Wiley, 2005.
- LAURELL, A. Algunos problemas teóricos y conceptuales de la epidemiologia social. *Revista Centroamericana de Ciencias de La Salud*, [S.L], v. 6, n. 3, p. 79-87, 1977.
- MORRISON, A. et al. Evaluation of a Sampling Methodology for Rapid Assessment of Aedes aegypti Infestation Levels in Iquitos, Peru. *Journal of Medical Entomology*, Peru, v. 41, n. 3, p. 502-510, 2004.
- NUNES, E. Medicina social. 1 ed. São Paulo: Global Editora, 1983.
- PINTO J. Interfase de visibilização de informações para o sistema integrado de potocolos eletrônicos. 2005. Tese. Ciências da Saúde, UFPR, Paraná.
- QUINLAN, J. C4.5: programs for machine learning. 1 ed. San Mateo: Kaufmann, 1992.
- RANGEL, S. M. L. Dengue: educação, comunicação e mobilização na perspectiva do controle: propostas inovadoras. *Saúde e Sociedade*, São Paulo, v. 12, n. 25, p. 433-441, 2008.
- RODHAIN, F; ROSEN, L. Mosquito vectors and dengue virusvector relationships. CAB International, New York, p. 45-60, 1997.

ROZEBOOM, L. *Aedes aegypti* (L.): The yellow fever mosquito. 35 ed. Londres: The Quarterly Review of Biology, 1960.

SOUZA, I.; VIANNA, R.; , R Moraes. Modelagem da incidência do dengue na Paraíba, Brasil, por modelos de defasagem distribuída.. *Cadernos de Saúde Pública.*, [S.L], v. 23, n. 11, p. 2623-2630, 2007.

WEISS, S; KULIKOWSKI, C. *Computer systems that learn*. 1 ed. San Mateo: Kaufmann, 1994.

WHO, World Health Organization. *Dengue hemorrhagic fever: diagnosis, treatment, prevention and control*. Geneva, 1997. Disponível em: <http://www.who.int/sorry/en/>. Acessado em: 15 jan. 2017.

_____, World Health Organization. *The Weekly Epidemiological Record*. Geneva, 2016. Disponível em: <http://www.who.int/wer/en/>. Acessado em: 17 jan. 2017.

WITTEN, I; FRANK, E. *Data mining*. 1 ed. Amsterdam: Morgan Kaufman, 2005.

HAND, D. J. *Discrimination and classification*. Chichester: Wiley, 1981.

ZEMBOWICZ, R; ZYTKOW, J. *From Contingency Tables to Various Forms of Knowledge in Databases*. AAAI Press, Menlo Park, p. 329–351, 1996.