



**UNIVERSIDADE ESTADUAL DA PARAÍBA  
CENTRO DE CIÊNCIAS EXATAS E SOCIAIS APLICADAS  
CAMPUS VII – GOVERNADOR ANTÔNIO MARIZ – CAMPUS VII  
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

**JOSÉ TRAJANO MENDES NETO**

**UTILIZAÇÃO DA CIÊNCIA DE DADOS PARA IDENTIFICAÇÃO DA RELAÇÃO  
ENTRE FATORES CLIMÁTICOS E A INCIDÊNCIA DE CASOS DE DENGUE NO  
MUNICÍPIO DE PATOS PB**

**PATOS – PB**

**2019**

JOSÉ TRAJANO MENDES NETO

**UTILIZAÇÃO DA CIÊNCIA DE DADOS PARA IDENTIFICAÇÃO DA RELAÇÃO  
ENTRE FATORES CLIMÁTICOS E A INCIDÊNCIA DE CASOS DE DENGUE NO  
MUNICÍPIO DE PATOS PB**

Trabalho de Conclusão de Curso apresentado ao Curso de Bacharelado em Ciência da Computação da Universidade Estadual da Paraíba, em cumprimento à exigência para obtenção do grau de bacharelado em Ciência da Computação.

**ORIENTADOR:** MSc. Pablo Roberto Fernandes de Oliveira

PATOS-PB

2019

JOSÉ TRAJANO MENDES NETO

UTILIZAÇÃO DA CIÊNCIA DE DADOS PARA IDENTIFICAÇÃO DA RELAÇÃO  
ENTRE FATORES CLIMÁTICOS E A INCIDÊNCIA DE CASOS DE DENGUE NO  
MUNICÍPIO DE PATOS PB

Trabalho de Conclusão de Curso  
apresentado ao Curso de Bacharelado em  
Ciência da Computação da Universidade  
Estadual da Paraíba, em cumprimento à  
exigência para obtenção do grau de  
bacharelado em Ciência da Computação.

Aprovado em: 18/06/2019.

**BANCA EXAMINADORA**

Pablo Roberto Fernandes de Oliveira  
Prof. Msc. Pablo Roberto Fernandes de Oliveira  
(Orientador)

Wellington Candeia de Araújo  
Prof. Dr. Wellington Candeia de Araújo  
(Examinador)

Amanda Mayara Sobral Rodrigues  
Prof. Amanda Mayara S. Rodrigues  
(Examinadora)

É expressamente proibido 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 do trabalho.

M538u Mendes Neto, Jose Trajano.

Utilização da Ciência de Dados para identificação da relação entre fatores climáticos e a incidência de casos de dengue no município de Patos - PB [manuscrito] / Jose Trajano Mendes Neto. - 2019.

48 p. : il. colorido.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Computação) - Universidade Estadual da Paraíba, Centro de Ciências Exatas e Sociais Aplicadas, 2019.

"Orientação : Prof. Me. Pablo Roberto Fernandes de Oliveira, Coordenação do Curso de Computação - CCEA."

1. Dengue. 2. Ciência de Dados. 3. Análise de Dados. I.

Título

21. ed. CDD 004

## **AGRADECIMENTOS**

Aos meus pais Irlane e Messias, por todo incentivo e amor, que apesar das dificuldades, me ajudaram a realizar meus objetivos.

Aos meus irmãos Matheus, Elizabeth e Jonas, pelo carinho e amizade, que mesmo indiretamente contribuíram com tudo.

Agradeço a minha família, por tudo, pelo apoio, amor e dedicação em todos os momentos.

Aos meus amigos, pelo apoio e amizade durante todo meu percurso acadêmico, em especial Lucas pelo companheirismo, Jakeline, Thais, Camilla, Michelle e Renata.

Agradeço ao meu professor e orientador Pablo Roberto, pelo empenho e dedicação durante todas as etapas do desenvolvimento da minha pesquisa.

Gostaria de deixar meus sinceros agradecimentos aos professores Wellington Candeia, por todas as oportunidades permitidas, ao professor Pablo Suárez, por todo incentivo e dedicação durante todo o curso e a todos os outros professores que por toda minha graduação puderam contribuir com meu desenvolvimento pessoal e acadêmico.

Deixo meu imenso agradecimento a UEPB, por ter me proporcionado cada momento que vivenciei durante todo esse tempo.

*Passamos toda a vida nos preocupando com o futuro. Fazendo planos para o futuro. Tentando prever o futuro. Como se desvendá-lo fosse aliviar o impacto. Mas o futuro está sempre mudando. O futuro é o lar dos nossos medos mais profundos e das nossas maiores esperanças. Mas uma coisa é certa: quando ele finalmente se revela, o futuro nunca é como imaginamos.*

**Meredith Grey**

## RESUMO

A doença dengue é um dos maiores problemas de saúde pública atualmente, que ao longo dos anos tem se tornado foco de pesquisas técnico-científicas em todo mundo. Este trabalho teve como objetivo realizar uma análise exploratória de dados referentes à dengue no município de Patos, a fim de verificar relações entre fatores climáticos e os casos notificados de dengue. Foram utilizados dados de casos de dengue e do clima em Patos, disponibilizados pelas bases do SINAN, INMET e AESA/PB. Foram realizadas análises utilizando as bibliotecas da linguagem de programação Python, utilizadas para visualização e aplicação do algoritmo de regressão linear e predição de dados. Como resultado, verificou-se que os casos de dengue estão relacionados diretamente com as mudanças climáticas na cidade de Patos, tendo maior incidência no período entre março e julho, quando as condições climáticas são favoráveis: diminuição da temperatura, aumento nos níveis de chuva e umidade do ar. Percebeu-se ao final que, apesar da forte relação entre os padrões comportamentais durante o ano, os dados disponíveis não permitiram que fosse implementado um modelo preditivo adequado para prever os casos de dengue em um período futuro no município de Patos.

**Palavras-Chave:** Dengue. Análise de Dados. Predição. Ciência de Dados.

## **ABSTRACT**

Dengue disease is one of the major public health problems today, which over the years has become the focus of technical-scientific research worldwide. This paper aimed to perform an exploratory analysis of dengue data in the municipality of Patos, in order to verify relationships between climatic factors and reported cases of dengue. Data from dengue cases and the climate in Patos, made available by the SINAN, INMET and AESA/PB databases were used. Analyzes were performed using the Python programming language libraries, used for visualization and application of the linear regression and prediction algorithm. As a result, it was verified that the cases of dengue are directly related to the climatic changes in the Patos city, having a greater incidence in the period between March and July, when the climatic conditions are favorable: decrease of the temperature, increase in the levels of rainfall and air humidity. It was concluded that, despite the strong relationship between behavioral patterns during the year, the available data did not allow the implementation of an adequate predictive model to predict dengue cases in a future period in the municipality of Patos.

**Keywords:** Dengue. Data analysis. Prediction. Data Science.

## LISTA DE GRÁFICOS

Gráfico 1 - Comparação entre as variáveis pluviométricas .....	25
Gráfico 2 - Variação anual da chuva e casos de dengue .....	28
Gráfico 3 - Média mensal de casos de dengue entre 2005 a 2015 .....	29
Gráfico 4 - Média mensal das chuvas entre 2005 a 2015 .....	30
Gráfico 5 - Média mensal da umidade entre 2005 a 2015 .....	30
Gráfico 6 - Média mensal da temperatura entre 2005 a 2015 .....	31
Gráfico 7 - Proporção entre os meses do ano sem casos de dengue .....	32
Gráfico 8 - Proporção entre os meses do ano sem chuvas .....	32

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>9</b>
<b>2</b>	<b>REFERENCIAL TEÓRICO .....</b>	<b>11</b>
2.1	Dengue .....	11
2.2	Ciência de Dados .....	12
2.3	Pré-processamento e normalização dos dados .....	14
2.4	Estratégia de correção de <i>missing values</i> .....	15
2.5	Python .....	16
2.6	Scikit-learn .....	17
2.7	Correlação de variáveis .....	17
2.8	Aprendizado de Máquina .....	18
2.9	Regressão Linear .....	19
2.10	Trabalhos Relacionados .....	21
<b>3</b>	<b>METODOLOGIA.....</b>	<b>22</b>
3.1	Coleta de dados.....	22
3.1.1	Tratamento dos dados.....	23
3.1.2	Dados climáticos: chuvas .....	23
3.2	Tabulação dos dados.....	24
<b>4</b>	<b>ANALISE E DISCUSSÃO DOS RESULTADOS.....</b>	<b>26</b>
4.1	Período de dengue em Patos .....	26
4.2	Análise de correlação de dados .....	32
4.3	Aplicação e resultados da Regressão Linear.....	36
<b>5</b>	<b>CONSIDERAÇÕES FINAIS.....</b>	<b>41</b>

## 1 INTRODUÇÃO

A Organização Mundial da Saúde (OMS) classifica a dengue como uma das principais questões de saúde pública no mundo. Avalia-se que 2,5 bilhões de habitantes, estão sob risco de contrair dengue ocorrendo cerca de 50 milhões de casos anualmente. Do total avaliado, em torno de 550 mil precisam de hospitalização e pelo menos 20 mil morrem em decorrência da doença (BRASIL, 2009).

A dengue tem como seu principal vetor o mosquito *Aedes aegypti*, estando amplamente presente em 4.318 municípios (BRASIL, 2015), este vetor está associado igualmente a transmissão de outras doenças graves como febre de chikungunya, febre por zika virus e febre amarela.

Desde 1908 o pesquisador Antonio Gonçalves Peryassú estudava aspectos referentes sobre o ciclo de vida, hábitos e a biologia do *Aedes aegypti*, baseado em estudos e experimentos realizados no Rio de Janeiro - RJ. Peryassú identificou que existiam relações entre o mosquito, a temperatura e a densidade populacional. Após o primeiro levantamento detalhado, Peryassú identificou que a presença do mosquito *Aedes aegypti* era maior quando a densidade populacional também era maior em certas áreas da cidade. Seus experimentos revelaram também que a queda de temperatura para menos de 20°C interferia no desenvolvimento do mosquito, consequentemente reduzindo extremante os casos (Instituto Oswaldo Cruz, 2017).

Atualmente a dengue vem apresentando comportamento que difere do qual Peryassú identificou. Segundo Câmara et al. (2009) as epidemias de dengue estão mais frequentes nos meses mais quentes do ano, período do clímax reprodutivo do *Aedes aegypti*. A taxa de metabolismo do mosquito aumenta em meses com temperatura mais elevada, tornando assim locais mais quentes propícios a proliferação do mosquito.

Os investimentos e avanços nas pesquisas com vacinas, campanhas educativas, monitoramento e controle desse vetor estão cada vez maiores, no entanto ainda não existe um método eficaz para erradicação, controle e monitoramento do *Aedes aegypti* (SANTOS; SILVA, 2016).

Diante disso, a análise exploratória de dados tem sido utilizada no auxílio à busca de relações entre fatores ambientais e a propagação das doenças, para dessa maneira, detectar e implementar mecanismos de prevenção e controle de epidemias (GONZÁLEZ; RODAS, 2015).

Na pesquisa deste trabalho, pretendeu-se analisar os dados de casos de dengue disponibilizados pelo Sistema de Informação de Agravos de Notificação (SINAN) para identificar relações entre as variáveis climáticas com os casos notificados no período de 2005 a 2015 no município de Patos, Paraíba, utilizando técnicas e recursos computacionais como Aprendizado de Máquina (*Machine Learning*).

Aliada a técnicas de inteligência computacional para análise e predição de dados e extração de conhecimento, especialmente aprendizado de máquina, análises têm se mostrado viáveis em áreas com pouco ou grande volume de dados, em que há a necessidade da interpretação e visualização eficaz e inteligente da informação.

Neste contexto, esta pesquisa tem como objetivo geral analisar os dados referentes aos casos de dengue no município de Patos - Paraíba, aplicando técnicas computacionais para compreender o comportamento da doença no município relacionado aos fatores climáticos. E como objetivos específicos:

- Observar relações entre variáveis meteorológicas e os casos de dengue no município.
- Verificar a possibilidade de prever casos de dengue através de variáveis climatológicas.
- Descrever as informações obtidas por meio da extração de conhecimento.
- Identificar a possibilidade de infestação do mosquito e da doença dengue em um determinado período.

Devido à dimensão territorial, o Brasil possui diversas variações climáticas ao longo do país. O Brasil é considerado um país tropical, onde prevalecem climas quentes e úmidos, tornando-o mais suscetível a endemias decorrentes de mosquitos, em particular o *Aedes Aegypti*, vetor transmissor da dengue. Sendo assim, este trabalho se justifica pela importância de se verificar o comportamento da dengue na cidade de Patos, considerando os fatores climáticos da região, a fim de analisar se os casos de dengue estão relacionados aos fatores relativos à dinâmica do vetor da doença.

## 2 REFERENCIAL TEÓRICO

Neste capítulo serão apresentados os conceitos utilizados para o desenvolvimento desta pesquisa. Também serão apresentados conceitos fundamentais para a melhor compreensão sobre análise de dados e aprendizado de máquina.

### 2.1 Dengue

Causada por um vírus de genoma RNA, do gênero *Flavivirus*, do qual são conhecidos quatro sorotipos (DENV-1, DENV-2, DENV-3 e DENV-4) a dengue é uma doença infecciosa que pode causar desde infecções assintomáticas até formas mais graves que podem levar a morte. Conforme relatado pelo Ministério da Saúde através do Plano de Contingência Nacional, no Brasil o padrão epidemiológico da doença tem variado no decorrer dos anos. Inicialmente, casos de dengue clássica ocorriam, particularmente, em adultos jovens. No entanto, entre 2007 e 2009, observou-se aumento das formas graves, atingindo, principalmente, crianças (BRASIL, 2015).

A dengue demonstra possuir um comportamento sazonal no Brasil, ocorrendo de maneira mais frequente, entre os meses de outubro e maio. Assim sendo, o monitoramento de indicadores entomo-epidemiológicos pode detectar antecipadamente vulnerabilidade de um determinado local para ocorrência da doença, bem como alterações no padrão de comportamento da dengue. Desta forma, é recomendado que, a partir do mês de outubro, o monitoramento seja intensificado (BRASIL, 2015).

Segundo o Ministério da Saúde a dengue é uma das doenças de notificação compulsória, devendo todo caso suspeito ou confirmado ser notificado ao Serviço de Vigilância Epidemiológica, por meio do SINAN nas fichas de notificação e investigação (BRASIL, 2008).

O vírus da dengue é capaz de alterar seu potencial epidêmico e suas apresentações clínicas a medida que se move entre as populações, fazendo com que as apresentações epidemiológicas das infecções se mostrem de modo bastante variado.

Os padrões da proliferação de dengue dependem da interação entre alguns dos fatores relacionados, como o movimento do vetor, raio de dispersão, taxa de

reprodução e longevidade, clima (chuvas, umidade, temperatura), saneamento e entre outros. Alguns padrões de comportamento podem se repetir em locais com grandes densidades populacionais e com índices elevados de infestação.

O período sazonal das infecções pelos vírus da dengue é perceptível, na maioria dos estados. A incidência eleva-se significativamente nos primeiros meses do ano, alcançando maior destaque em março a maio, seguida de uma redução considerável dos casos a partir de junho. Este padrão epidemiológico da dengue, que nem sempre é observado em outros países, tem sido justificado pelo aumento na densidade populacional do *Aedes aegypti*, em virtude do aumento da temperatura e umidade, registradas durante o verão e outono no Brasil (TEIXEIRA et. al, 1999).

## 2.2 Ciência de Dados

A Ciência de Dados incorpora vários elementos e baseia-se em técnicas e teorias derivadas de múltiplos campos em engenharia e ciências, sendo assim profundamente ligada diversas disciplinas, deste modo viabilizando uma nova área altamente interdisciplinar. Dessa forma, associado a aplicação interdisciplinar, a ciência de dados surge como componente cada vez mais importante nas diversas áreas, tais como saúde, financeira, astronomia, bioinformática, mobilidade urbana, segurança digital, comunicação e entre outros (PORTO e ZIVANI, 2014). A Ciência de dados permite aplicar técnicas com finalidade de descoberta de padrões como também de descoberta de conhecimento através de bases de dados.

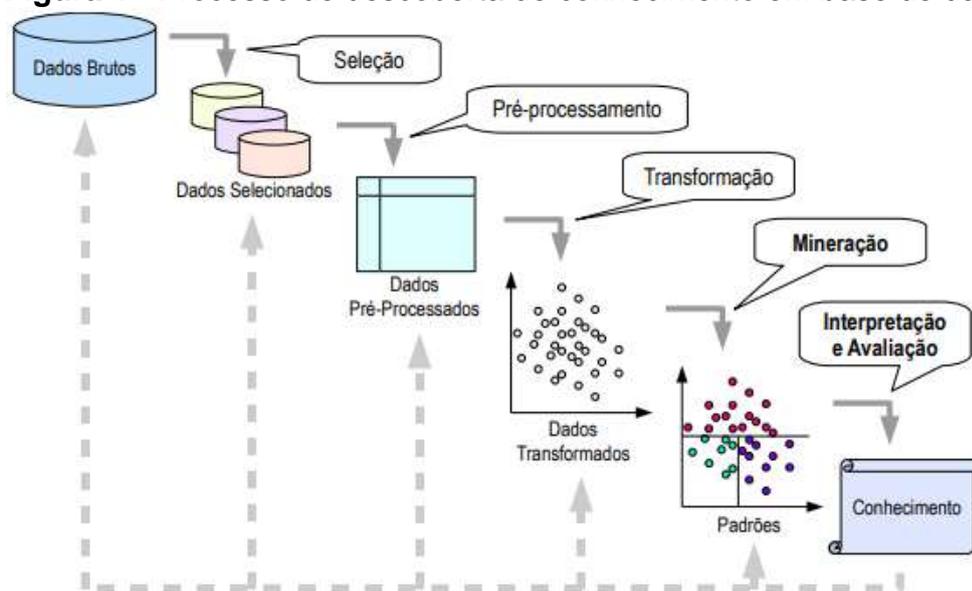
Mannila (1996) cita que, *Knowledge Discovery in Databases* (KDD) é um processo que frequentemente é chamado de mineração de dados. Esse método visa descobrir informações úteis de grandes coleções de dados. O conhecimento adquirido por esse processo de descoberta pode gerar regras que descrevem algumas propriedades dos dados, padrões frequentes e agrupar objetos no banco de dados.

Mineração de Dados (*Data Mining*) é uma das fases do processo chamado Descoberta de Conhecimento em Bancos de Dados (*Knowledge Discovery in Databases*). O processo de descoberta de conhecimento em bancos de dados é definido como o processo não-trivial de identificação de padrões válidos, novos, potencialmente úteis e compreensíveis a partir de dados (SANTOS, 2009).

Santos (2009) destaca o processo de descoberta de conhecimentos em bancos de dados ilustrado na Figura 1. Pode-se enumerar os passos para o processo

de descoberta de conhecimento com a lista a seguir, correspondendo às etapas deste processo.

**Figura 1** - Processo de descoberta de conhecimento em base de dados



Fonte: Santos (2009)

- 1 Compreender o domínio, do conhecimento prévio relevante e dos objetivos do usuário final do processo;
- 2 Criar o conjunto de dados para a utilização no processo de KDD através da seleção dos dados ou atributos relevantes; (**seleção**)
- 3 Pré-processar os dados, limpeza e remoção de ruído e desvios, estratégia de correção de dados desconhecidos ou faltantes, normalização, extração de atributos numéricos de documentos e logs, etc; (**pré-processamento**)
- 4 Reduzir e reprojeter os dados, isto pode ser feito através da seleção de atributos relevantes para representar melhor os dados sem perda de precisão, sempre dependendo do objetivo a ser atingido; (**transformação**)
- 5 Definir o método de mineração considerando o objetivo do processo (classificação, regressão, agrupamento, etc.);
- 6 Estabelecer o(s) algoritmo(s) de mineração de dados baseado no objetivo geral e na estrutura imposta aos dados. A decisão do(s) algoritmo(s) envolve a escolha de modelos, parâmetros e formas de execução;
- 7 Minerar os dados em busca de padrões de interesse usando algoritmos e dados selecionados; (**mineração**)

- 8 Interpretar os resultados e avaliar os padrões, regras, etc. encontrados pelo processo de mineração;
- 9 Avaliar os conhecimentos adquiridos, documentação e elaboração de relatórios. (**interpretação e avaliação**)

Usuários de um sistema KDD devem ter uma sólida compreensão do domínio dos dados, com o propósito de selecionar subconjuntos corretos de dados, classes adequadas de padrões e critérios apropriados para o interesse desses padrões. Assim, os sistemas KDD devem ser classificados como ferramentas interativas e não como sistemas de análises automáticas. Descobrir o conhecimento dos dados deve, portanto, ser visto como um processo que inclui várias etapas sendo delas: compreender o domínio, preparar o conjunto de dados, extrair padrões (data mining), pós-processamento de padrões descobertos e colocar os resultados em uso (MANNILA, 1996).

Segundo Maimon e Rokach (2010), mineração de dados (*Data Mining*) é um termo criado para descrever o processo de filtrar e avaliar grandes bancos de dados para encontrar padrões e relacionamentos. Atualmente o processo de mineração visa a modelagem e análise baseada em evidências, e seu uso é aplicado em diversas áreas de conhecimento como medicina, finanças, segurança e marketing.

Han e Kamber (2001, p. 5) relatam que, “a mineração de dados deveria ter sido mais apropriadamente chamada de ‘mineração de conhecimento a partir de dados’.”

A análise das características presentes em um conjunto de dados possibilita a descoberta de padrões e tendências comportamentais que podem fornecer informações sobre os processos que os gerou. Muitas informações podem ser obtidas por simples estatísticas, outras delas através de técnicas de visualização de dados. As técnicas de exploração de dados auxiliam na seleção de técnicas posteriores como o pré-processamento e aprendizado. Entretanto, a maior parte dos padrões e conhecimento são adquiridos com o uso de técnicas mais sofisticadas como o Aprendizado de Máquina (HAN; KAMBER, 2001).

### **2.3 Pré-processamento e normalização dos dados**

A fase de pré-processamento dos dados é constantemente tida como sendo uma fase em que abrange uma grande quantidade de conhecimento do domínio.

Muitos pesquisadores citam que os dados coletados diretamente de banco de dados são de má qualidade, isto é, possuem informações incorretas e imprecisas, além dos chamados *missing values* (valores desconhecidos). Apesar dos algoritmos utilizados na fase de mineração de dados tenham sido desenvolvidos para manipular os dados em tais situações, é esperado que esses algoritmos produzam resultados mais precisos e confiáveis caso a maioria dos problemas presentes nos dados sejam corrigidos ou removidos (BATISTA, 2003).

Desta maneira, o pré-processamento de dados se torna um processo semi-automático, pois entende-se que essa fase necessita da capacidade do analista de dados em identificar os problemas que estão presentes nos dados, como também a natureza desses problemas com o objetivo de utilizar os métodos mais apropriados para solucionar cada um dos problemas (BATISTA, 2003).

Batista (2003) destaca que, na fase de normalização dos dados os valores dos atributos e variáveis são transformados de seus intervalos originais para um intervalo específico, como, por exemplo,  $[-1,1]$  e  $[0,1]$ . Essas transformações são especialmente valiosas para métodos que calculam distâncias entre variáveis. O método do *k-nearest neighbors algorithm* (algoritmo k-vizinhos mais próximos) tende a dar mais importância para atributos que apresentam um intervalo maior entre seus valores.

## 2.4 Estratégia de correção de *missing values*

Na continuação das fases de pré-processamento, a qualidade dos dados é a preocupação central no Aprendizado de Máquina, uma vez que na maioria dos algoritmos de aprendizado dependem do conhecimento extraído estritamente a partir dos dados, a qualidade do conhecimento adquirido é determinada pela qualidade dos dados de entrada fornecidos (BATISTA, 2003).

Como citado na seção anterior, um dos problemas no Aprendizado de Máquina é a presença de valores desconhecidos ou ausentes. Esses valores consistem na não medição dos valores de alguma variável, seja por defeitos em equipamentos, recusa por partes de entrevistados ou até mesmo do agente de coleta não registrar os dados.

Embora a frequência da ocorrência de valores desconhecidos no conjunto de dados seja alta, muitos analistas tratam esses valores de forma bastante simples.

Entretanto, esses tratamentos de dados desconhecidos devem ser cuidadosamente pensados, do contrário, distorções podem ser introduzidas no conhecimento adquirido (BATISTA, 2003).

Aliada a técnicas de manipulação de dados, faz-se necessário utilizar softwares a fim de auxiliar no desenvolvimento da análise de dados como também de melhorar a visualização dos resultados extraídos. Com isso, a linguagem de programação *Python*, associada às suas bibliotecas, tem-se tornado cada vez mais utilizada para este propósito.

## 2.5 Python

Segundo Millman e Aivazis (2011) durante os últimos anos, o *Python* (linguagem de programação de alto nível, interpretada, orientada a objetos e de tipagem forte) tornou-se sem dúvida, o padrão para pesquisa científica exploratória, orientada à computação e interativa.

Com a evolução do *Python*, uma comunidade relativamente pequena, mas comprometida, de engenheiros e cientistas que usavam o *Python* para necessidades de computação lentamente migrou para a utilização do módulo Numérico (baseado em matrizes). Esta comunidade continuou com o aperfeiçoamento do módulo numérico do *Python* e começou a desenvolver e compartilhar os pacotes adicionais para a computação científica. Embora o *Python* possuísse uma sintaxe expressiva e uma grande coleção de tipos de dados integrados (*strings*, listas, dicionários, tuplas), ficou claro que, para dispor de uma estrutura que era necessário para computação científica, o *Python* precisava fornecer algum tipo de *array* para computação numérica. Com isso, o *Python* em seus releases de atualizações foi ganhando novos recursos e sintaxes, em especial voltados para comunidade científica (MILLMAN; AIVAZIS, 2011).

Apesar do *Python* não ter sido projetado inicialmente para atender as necessidades específicas computacionais da comunidade científica, ele rapidamente atraiu interesse de engenheiros e cientistas. Nos seus primórdios em meados dos anos 90 o *Python* era utilizado para a computação numérica avançada e só se tornou popular nos últimos anos quando incorporado à computação algébrica na qual tratava do desenvolvimento de operações simbólicas que envolviam fatoração de derivadas, integrais e polinômios, como também cálculo de matrizes (MILLMAN; AIVAZIS, 2011).

A utilização do *Python* para análise de dados, computação interativa e exploratória e visualização de dados é inevitavelmente comparada com as muitas outras linguagens de programação e ferramentas de uso geral, como R e MATLAB. Ao longo dos últimos anos, as bibliotecas (o pandas principalmente) de manipulação de dados contaram com um forte e melhorado suporte dos desenvolvedores, com intuito de aprimorar o seu uso e abranger novos usuários o que tornou o *Python* uma forte alternativa para tarefas de manipulação de dados. Assim, a linguagem de programação *Python* ganhou destaque tanto em programação de propósito geral quanto como uma linguagem única para criação de aplicações focadas em computação científica e análise de dados (MCKINNEY, 2012).

## 2.6 Scikit-learn

Na linguagem de programação *Python* a biblioteca do *Scikit-learn* possui as técnicas de aprendizado de máquina mais utilizadas com os algoritmos disponíveis. O *Scikit-learn* diferente das outras ferramentas para aprendizado de máquina é uma ferramenta simples e eficiente para análise de dados, acessível e capaz de ser reutilizada em diversos contextos (PEDREGOSA, 2011).

Diante aos algoritmos disponibilizados pelo *Scikit-learn* estão presentes os algoritmos: classificação, regressão, agrupamento, redução da dimensionalidade, seleção de modelo e pré-processamento. Com isso, será avaliada a viabilidade de cada um identificando qual se adequa melhor aos resultados esperados da pesquisa.

## 2.7 Correlação de variáveis

O termo coeficiente de correlação é popularmente encontrado na área de Estatística, segundo Garson (2009) correlação é uma medida de associação bivariada (isto é, tamanho ou força do efeito) da relação entre duas variáveis. A correlação de variáveis é medida no intervalo de -1 e +1, no qual 0 caracteriza-se como uma relação aleatória, 1 uma relação linear positiva perfeita e -1 como uma relação linear negativa perfeita. Ou seja, pontos próximos ao 0 indicam uma relação desprezível, e indica uma forte relação quando se aproxima dos extremos -1 e +1.

A correlação auxilia na análise de dados expondo o quanto uma variável pode estar dependente de outra ou o inverso. Como exemplo, podemos citar a relação

positiva entre altura e peso, no qual geralmente pessoas mais altas (acima da média) tendem também a estar acima da média de peso, como também pessoas mais baixas tendem a ter o peso mais baixo em relação a média. Deste modo, podemos ver que a correlação é positiva quando a associação entre  $x$  e  $y$  é positiva e vice-versa.

A utilização da correlação pode variar mediante as variáveis alvos de análise, isto é, conforme o tipo de variável utilizada faz-se necessário aplicar as técnicas mais adequadas para o estudo. Dentre elas destaca-se o método de Pearson para análise linear do coeficiente de correlação.

O coeficiente de correlação de Pearson pode variar entre  $-1$  e  $+1$ , os valores quando negativos indicam uma correlação do tipo inversa, ou seja, quando  $x$  aumenta  $y$  diminui. Deste modo, os valores positivos do coeficiente ocorrem quando  $x$  e  $y$  variam no mesmo sentido, isto é, quando  $x$  cresce  $y$  também cresce ou quando  $x$  diminui  $y$  diminui também. Logo, o coeficiente de correlação de Pearson avalia a relação linear entre duas variáveis contínuas (SANTOS, 2007).

## **2.8 Aprendizado de Máquina**

Monard e Baranauskas (2003) definem Aprendizado de Máquina como uma área de Inteligência Artificial (IA) com objetivo de desenvolver técnicas computacionais sobre o aprendizado e o desenvolvimento de sistemas capazes de obter conhecimento de forma automática.

Batista (2003) destaca que existem diversas abordagens de aprendizado que são utilizadas por um sistema computacional para esse processo de aprendizado tais como, o aprendizado por hábito, por instrução, por dedução, por indução e por analogia. O aprendizado por indução é um dos mais úteis pois possibilita a obtenção de novos conhecimentos a partir de exemplos. Porém, o aprendizado por indução é também um dos mais desafiadores, pois o conhecimento adquirido pode ultrapassar os limites das premissas, e não existem garantias de que esse conhecimento gerado seja verdadeiro.

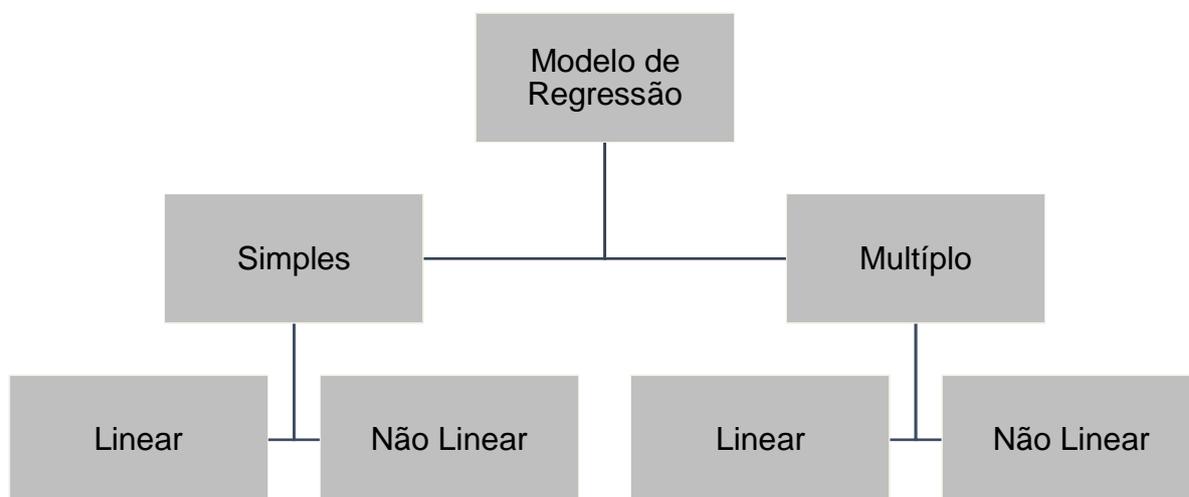
O aprendizado indutivo divide-se em supervisionado e não-supervisionado. No aprendizado supervisionado é fornecido ao algoritmo de aprendizado, um conjunto ou amostras de exemplos de treinamento para os quais o rótulo da classe associada é conhecido. Em geral, cada amostra é descrita por um vetor de valores de características, e o rótulo da classe associada (MONARD e BARANAUSKAS, 2003).

Já no aprendizado não-supervisionado, o indutor analisa os exemplos fornecidos e tenta determinar se alguns deles podem ser agrupados de alguma maneira, formando agrupamentos ou *clusters* (CHEESEMAN e STUTZ, 1990).

## 2.9 Regressão Linear

Santos (2007) explica que a utilização de qual modelo de regressão linear depende de quais variáveis propõe-se estudar, de acordo com a Figura 2 os modelos de regressão se classificam em dois, o simples e o múltiplo. O modelo de regressão simples é utilizado quando existe somente uma variável dependente, já o múltiplo há duas ou mais variáveis para investigação.

**Figura 2** - Modelos de Regressão



**Fonte:** Santos (2009)

Na estatística umas das maiores preocupações ao analisar dados, é a fase de criação de modelos que determinem estruturas do fenômeno em observação. O modelo de regressão é um dos métodos estatísticos mais comum para investigar a relação entre variáveis (SANTOS, 2007).

Regressão é o processo de prever um valor a partir de um conjunto de dados semelhante a à técnica de classificação. Isto é, baseado em uma coleção de dados de entrada, é possível prever novos valores a partir desses dados. O método de classificação para previsão de dados difere de forma sucinta da regressão quando a variável prevista é discreta, enquanto na regressão é contínua. Quando se fala em em

regressão, geralmente é dito como regressão linear, então os termos regressão e regressão linear são usados de maneira conjunta. Regressão linear significa que se pode somar as entradas multiplicadas por algumas constantes para obter uma saída (HARRINGTON, 2012).

Na regressão mede-se relação entre as variáveis, no qual se destacam variáveis de resposta que medem um resultado de um estudo e variáveis explicativas que explicam ou influenciam mudanças em uma variável de resposta. Pode-se definir regressão linear como uma linha reta, que descreve como uma variável de resposta  $y$  muda quando uma variável explicativa  $x$  também muda (MOORE et al., 2003). Costuma-se usar uma linha de regressão para prever o valor de  $y$  para um determinado valor de  $x$ . Por exemplo, prever a temperatura ou tempo ideal para colheita.

A regressão linear apresenta como pontos positivos a capacidade de fácil interpretação dos resultados e ser computacionalmente simples, mas por outro lado a regressão modela dados não lineares. O método de regressão linear é adequado para o uso tanto de valores numéricos quanto nominais (HARRINGTON, 2012).

A presença ou ausência de relação linear pode ser investigada sob dois pontos de vista, quando quantifica a força da relação utiliza-se a correlação e explicitando a forma dessa relação a regressão (SANTOS, 2007).

Para a abordagem geral da regressão linear seguem-se algumas etapas para sua utilização:

1. **Coleta:** qualquer método de coleta.
2. **Preparação:** necessita-se de valores numéricos para regressão linear. Os valores nominais devem ser mapeados para valores binários.
3. **Análise:** visualização em gráficos 2D.
4. **Treinamento:** Encontrar os pesos de regressão.
5. **Teste:** Medir o  $R^2$  (coeficiente de determinação, medida de ajustamento de um modelo estatístico), ou correlação do valor e dos dados previstos, para medir o sucesso dos modelos.
6. **Utilização:** Com a regressão, consegue-se prever um valor numérico para um número de entradas.

Após o desenvolvimento das etapas a avaliação do modelo é uma etapa posterior para algumas avaliações devem ser realizadas para se ter alguma idéia da

eficácia e adequação do modelo. Normalmente dentro das técnicas utilizadas para avaliar a eficácia, utiliza-se inicialmente o coeficiente de correlação, mas de forma mais específica a medida de adequação mais indicada é o coeficiente de determinação do modelo. De maneira geral este coeficiente indica quanto o modelo foi eficaz em explicar os dados coletados (SANTOS, 2007).

## **2.10 Trabalhos Relacionados**

Viana e Ignotti (2003) em seu trabalho intitulado “A ocorrência da dengue e variações meteorológicas no Brasil: revisão sistemática” relatam que, a dengue está fortemente associada a variáveis meteorológicas. A variação sazonal da temperatura e da pluviosidade interferem na dinâmica do vetor e na incidência da doença em todo o país, independentemente do compartimento climático. Na pesquisa, ainda é ressaltado o período do segundo ao quarto mês do ano como de maior associação com o clima. Logo, há dificuldades no estabelecimento de padrões sazonais da dengue em relação a variáveis meteorológicas no Brasil.

Santos e Silva (2016) utilizam em sua pesquisa, um método de inteligência computacional, o algoritmo de vaga-lume que é uma técnica baseada em inteligência por enxames para o controle do mosquito *Aedes aegypti* como forma de combater e reduzir os casos de doenças relacionadas a este mosquito.

Halstead (2008) destaca que, a distribuição atual dos vírus da dengue nos países tropicais é resultado do efeito da temperatura e das chuvas. Com isso os casos de dengue aumentam durante as estações quentes e chuvosas, tornando a chuva não somente um fator único para o aumento dos casos de dengue.

### 3 METODOLOGIA

Este trabalho pode ser classificado como uma pesquisa explicativa que segundo Gil (2008, p. 28) “são aquelas pesquisas que têm como preocupação central identificar os fatores que determinam ou que contribuem para a ocorrência dos fenômenos. ” O trabalho compreende também uma pesquisa quantitativa no qual “recorre à linguagem matemática para descrever as causas de um fenômeno, as relações entre variáveis, etc.” (FONSECA, 2002, p. 20).

A partir do domínio da pesquisa, pretende-se utilizar uma metodologia também de pesquisa aplicada, que segundo Gerhardt e Silveira (2009) “objetiva gerar conhecimentos para aplicação prática, dirigidos à solução de problemas específicos. Envolve verdades e interesses locais. ”

Neste estudo se faz necessário da mesma forma o uso de uma metodologia que consiste de pesquisas utilizando-se referências bibliográficas detalhadas e o levantamento das problemáticas da dengue. Com esse levantamento, a pesquisa busca identificar relações entre variáveis climáticas e a incidência de casos de dengue.

#### 3.1 Coleta de dados

O principal foco desta pesquisa foi analisar dados referente às notificações de casos de dengue como também, dados climatológicos na cidade de Patos. Dito isto, foi necessário a criação de uma coleção de dados (*dataset*) que possuísse variáveis climáticas e os casos de dengue notificados durante um período de tempo.

Para a captura dos dados, usou-se como base os órgãos públicos que são responsáveis por esse armazenamento dos dados e sua divulgação. Os dados referentes ao clima foram obtidos através do INMET (Instituto Nacional de Meteorologia), por meio do BDMEP (Banco de Dados Meteorológicos para Ensino e Pesquisa) foi possível selecionar inicialmente diversas variáveis climáticas no período mensal de 2005 a 2015. Outra variável referente especificamente às chuvas acumuladas foi obtida no banco de dados da AESA/PB.

Com base nos dados disponibilizados pelo SINAN através dos Boletins Epidemiológicos emitidos pela Secretaria de Vigilância em Saúde, com caráter técnico-científico e divulgado com periodicidade mensal e semanal, a fim de monitorar

e investigar doenças específicas sazonais no Brasil, foi possível utilizar os dados de casos notificados de dengue para a tabulação no *dataset* criado.

No Boletim Epidemiológico são publicadas descrições sobre o monitoramento de eventos e doenças com potencial para desencadear emergência de Saúde Pública no Brasil, através de análises da situação epidemiológica de doenças e agravos com relatos de investigação de surtos, e para este estudo foram selecionados os boletins referente especificamente à dengue.

O SINAN é alimentado pela notificação e investigação de casos de doenças e agravos constantes da lista nacional de doenças de notificação compulsória. A entrada de dados é realizada a partir de instrumentos de coleta padronizados, como a Ficha Individual de Notificação (FIN) e a Ficha Individual de Investigação (FII) (BRASIL, 2008a).

Os dados obtidos por meio das bases de dados citadas anteriormente, foram selecionados, tratados para implementação e experimentos dos algoritmos de aprendizado de máquina como também técnicas de mineração de dados.

Após o processo de seleção e tratamento, os dados foram avaliados e visualizados, representando-os da melhor maneira o conhecimento gerado, auxiliando para as próximas etapas de análise dos dados.

Para as técnicas de aprendizado de máquina foram utilizados os algoritmos disponíveis na biblioteca do *Scikit-learn* para *Python*.

### 3.1.1 Tratamento dos dados

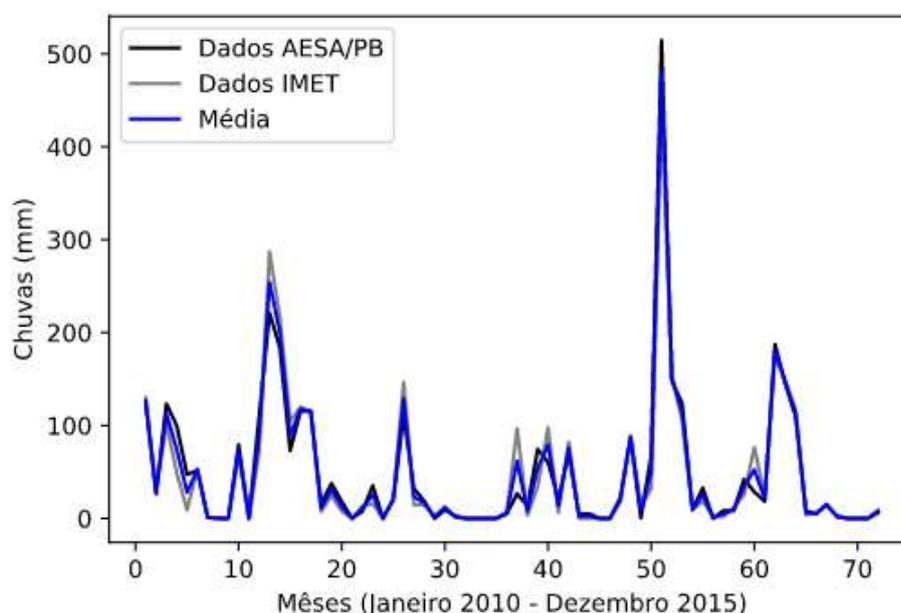
Os dados foram tratados em conjunto com a etapa de criação da base dados utilizada, verificando a presença de atributos e valores nulos, como também a formatação dos valores presentes nos dados capturados. Dessa forma, os dados puderam ser avaliados de maneira inicial, evitando erros de processamento na etapa de análise como também na aplicação dos algoritmos de aprendizado de máquina.

### 3.1.2 Dados climáticos: chuvas

Os dados referentes às chuvas no município de Patos podem ser encontrados nos bancos de dados da AESA/PB e do INMET. Uma vez que os dados são

disponibilizados por fontes distintas, fez-se necessário analisá-los a fim de verificar a compatibilidade das informações contidas nesses bancos. Para validar os dados, foi selecionado uma amostra do conjunto de dados com o objetivo de visualizá-los graficamente. Foram escolhidos para amostra, os dados referentes ao intervalo que compreende os meses de janeiro de 2010 a dezembro de 2015, somando um total de 6 anos ou 72 meses. O Gráfico 1 mostra a variação das chuvas nas duas bases de dados e também foi adicionado uma variável que contém a média dos valores entre as bases, com isso, pôde-se perceber que, a diferença entre as bases é mínima. Logo, para análise dos dados, foi escolhido a da AESA/PB.

**Gráfico 1** - Comparação entre as variáveis pluviométricas



**Fonte:** Elaborado pelo autor, 2019.

### 3.2 Tabulação dos dados

Após a seleção das bases de dados utilizadas, os dados foram tabulados e organizados em uma planilha eletrônica contendo 133 linhas referentes aos meses de cada ano e 9 colunas referentes às variáveis analisadas, logo após a tabulação foi necessário identificar os valores nulos ou inexistentes afim de tratá-los evitando-se erros futuros.

Em seguida, os dados foram convertidos para o formato CSV (*Comma-separated values*) devido ser o formato mais comum de importação e exportação para

planilhas e bancos de dados usados em *Python* como exibido na Tabela 1 os valores iniciais do arquivo.

**Tabela 1** – Amostra dos dados utilizados

	Data	Evapo	Insol	Precip	TempMaxM	TempMinM	Umid	Casos	ChuvAc
0	31/01/2005	62,8	273,2	18,9	36,0	24,5	62,0	1	27,8
1	28/02/2005	96,0	243,5	114,5	35,1	23,7	61,4	3	68,3
2	31/03/2005	111,3	207,0	307,3	35,1	23,8	65,2	29	384,3
3	30/04/2005	86,3	265,0	97,7	33,6	23,6	68,1	99	47,5
4	31/05/2005	63,0	217,2	28,2	32,6	23,0	70,1	128	20,7

**Fonte:** Elaborado pelo autor, 2019.

O quadro 1 descreve os tipos de dados referentes as variáveis utilizadas, a fim de especificar melhor para a análise posterior.

**Quadro 1** – Tipos de variáveis presentes no conjunto de dados

VARIÁVEL	UNIDADE	LIMITE MÍNIMO	LIMITE MÁXIMO
MÊS	data (dd/mm/aaaa)	1	12
INSOLAÇÃO TOTAL	hs (horas)	0	334
TEMPERATURA MÍNIMA	°C (graus celsius)	17	24
TEMPERATURA MÁXIMA	°C (graus celsius)	31	37
UMIDADE RELATIVA DO AR	% (porcentagem)	44	86
EVAPOTRANSPIRAÇÃO REAL	mm (milímetros)	0	149
PRECIPITAÇÃO TOTAL	mm (milímetros)	0	509
CHUVAS ACUMULADAS	mm (milímetros)	0	641
CASOS DE DENGUE	Inteiro	0	286

**Fonte:** Elaborado pelo autor, 2019.

## 4 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Neste capítulo serão discutidos os resultados da análise dos dados da dengue e fatores climáticos, apresentando a possibilidade de predição e as relações entre os dados observados, por meio dos processos de descoberta de conhecimento em base de dados<sup>1</sup>.

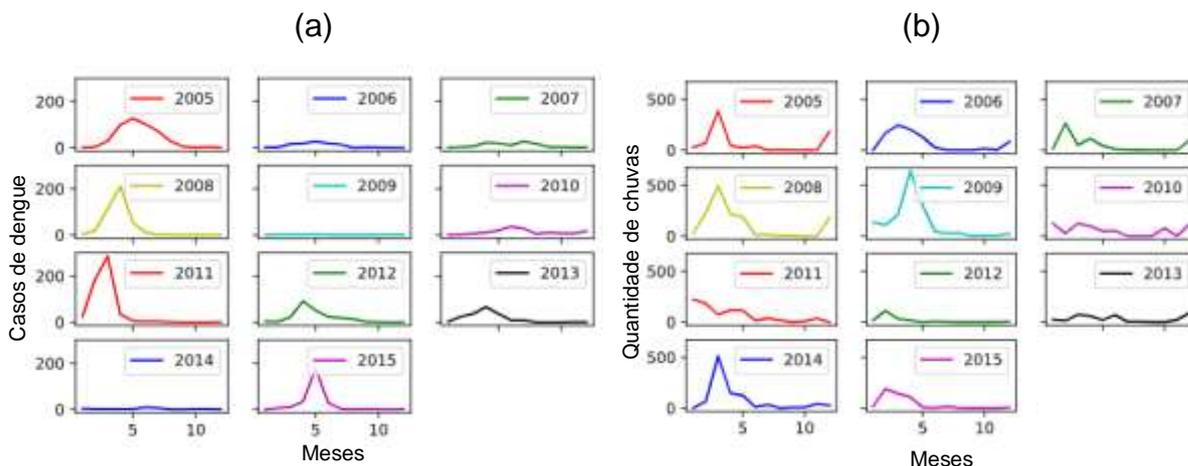
### 4.1 Período de dengue em Patos

Com relação ao período epidemiológico da dengue em Patos PB, verificou-se que, quando comparado ao período epidemiológico do Brasil, o período em que a epidemia se deu na cidade de Patos não segue o mesmo padrão epidemiológico do restante do Brasil, conforme os dados dos boletins epidemiológicos. O período que apresentou maior índice de epidemia foram os meses de março a julho. Entretanto, no Brasil o período de outubro a maio é epidêmico para as doenças transmitidas pelo *Aedes aegypti*, devido ao calor e as chuvas, que são condições ideais para a proliferação do mosquito vetor da dengue.

Na cidade de Patos a doença dengue não apresenta um comportamento contínuo no que diz respeito à padronização da frequência dos casos notificados. No Gráfico 2 (a) os anos de 2006, 2007, 2009, 2010, 2014 apresentam os níveis de casos baixos ou quase nulos, demonstrando que a ocorrência da dengue entre os anos não apresenta um comportamento capaz de ser previsto através de uma análise inicial dos dados. As características observadas anteriormente podem estar relacionadas a medidas preventivas adotadas pela população como também campanhas governamentais, de modo que não se observa grandes surtos em anos consecutivos

---

<sup>1</sup> Os códigos e *dataset* utilizados podem ser encontrados em: [github.com/joosenetoo/tccdengue](https://github.com/joosenetoo/tccdengue)

**Gráfico 2** - Variação anual da chuva e casos de dengue

**Fonte:** Elaborado pelo autor, 2019.

O Gráfico 2 (b) apresenta a variação dos dados em relação às chuvas durante os meses do ano, em todo o período do estudo, demonstrando que existe um aumento das chuvas no início do ano e diminuição a partir da segunda metade do ano.

Como pode ser visto nos gráficos anteriores o período de dengue não apresenta um surto frequente entre os anos, destacando-se que a cada período de surto da doença são seguidos por dois anos com índice baixo de notificação de casos da dengue. Entretanto, o período chuvoso não segue esse padrão, tendo como destaque o ano de 2009 que houve uma quantidade considerável no aumento das chuvas, porém não houveram notificações significativas de casos de dengue no município nesse período.

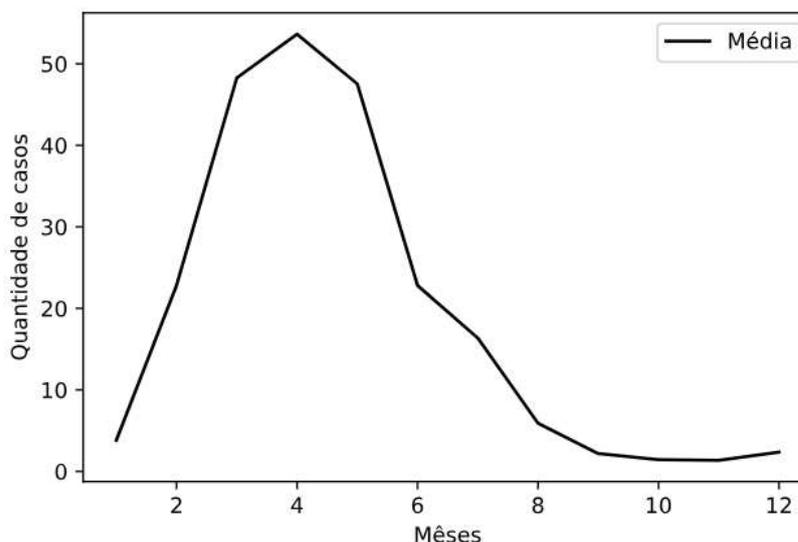
Souza, Vianna e Moraes (2007) relatam que no Estado da Paraíba, localizado no Nordeste do Brasil, o clima é marcado pela existência exclusivamente de duas estações climáticas: a chuvosa e a seca, de modo que a pluviometria é o principal fator de alteração climática da região ao longo do ano. Os autores enfatizam que, no Estado da Paraíba, o período chuvoso geralmente é curto, ocorrendo no verão nas regiões mais áridas e no inverno nas áreas mais úmidas.

Na estação do outono as chuvas são pouco frequentes, mas, no entanto, este período é o de maior incidência do dengue no estado. Esse fator não é observado em outros estados do Brasil onde a sazonalidade da dengue coincide com o período chuvoso, segundo o Ministério da Saúde (FARHAT, 2003).

Relacionado ao comportamento sazonal, os dados demonstraram uma frequência no primeiro semestre do ano, apresentando uma intensidade maior de

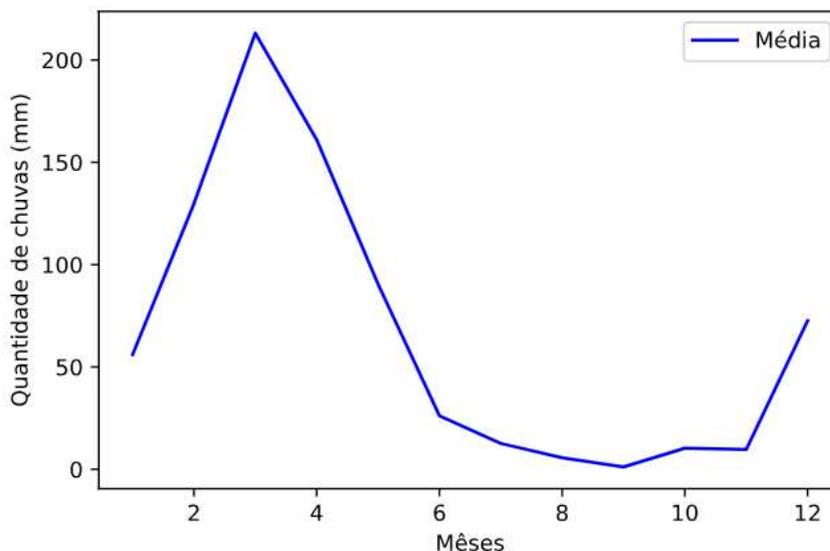
casos começando em fevereiro, demonstrando redução a partir de maio. O Gráfico 3 expõe a média dos casos de dengue durante o período de 2005 a 2015 exibindo o comportamento durante os doze meses do ano.

**Gráfico 3** - Média mensal de casos de dengue entre 2005 a 2015



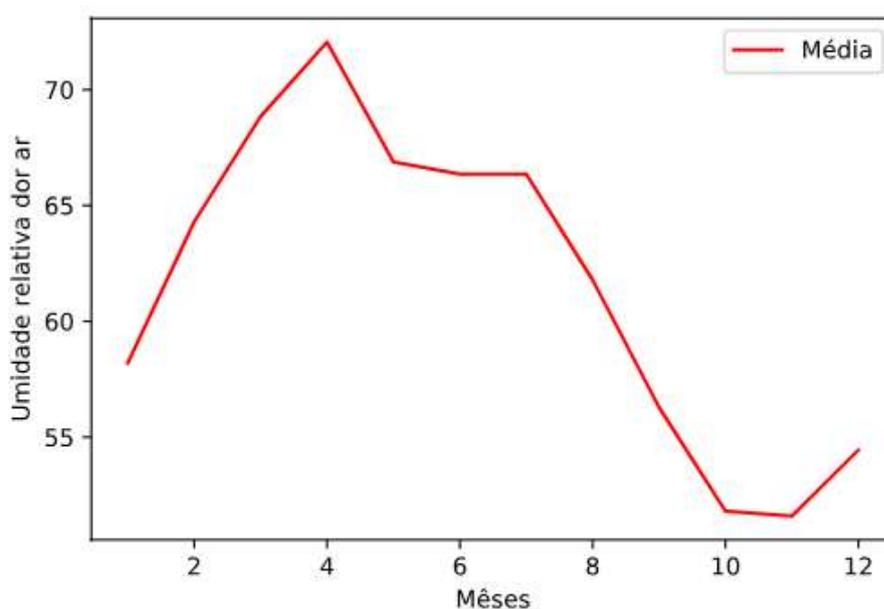
**Fonte:** Elaborado pelo autor, 2019.

Os dados apresentados no Gráfico 3 permitiram relacionar o período chuvoso na cidade de Patos com a ocorrência da dengue, tendo em vista que o período chuvoso em Patos demonstrou um comportamento regular em relação a todos os anos estudados. Ficou evidenciado que do mesmo modo que os casos de dengue, a chuva em Patos se apresenta de forma mais constante no primeiro semestre do ano. O Gráfico 4 demonstra o comportamento da chuva durante o período de estudo, iniciando-se no começo do verão, se estendendo até meados do outono, repetindo o ciclo contínuo a cada ano.

**Gráfico 4** - Média mensal das chuvas entre 2005 a 2015

**Fonte:** Elaborado pelo autor, 2019.

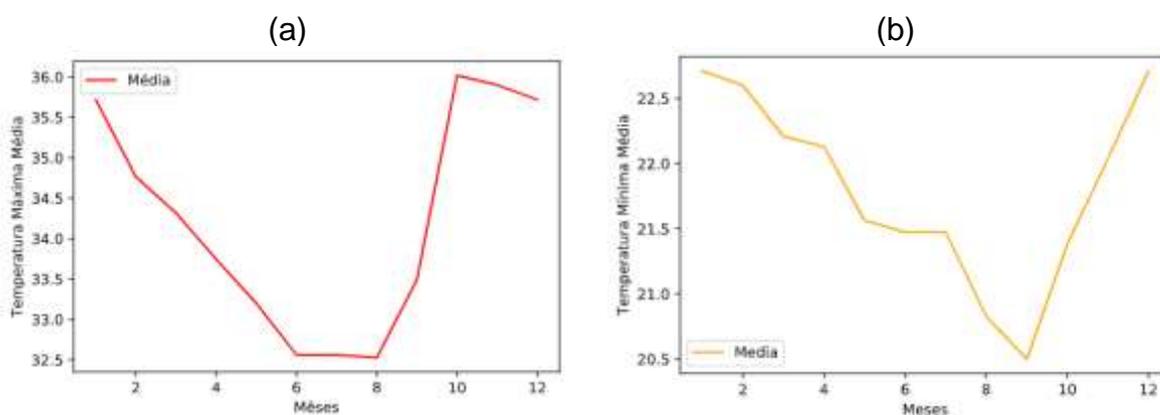
A umidade relativa do ar foi uma das variáveis climáticas citadas no referencial teórico como tendo relação com a proliferação do vetor da dengue, estando ligada aos padrões epidemiológicos da dengue. No Gráfico 5 é demonstrado a variação média durante o período de estudo, apresentando de maneira também semelhante ao padrão de distribuição do comportamento da dengue. Como pode ser visto, existe um aumento no início do primeiro semestre do ano e um declínio no segundo semestre.

**Gráfico 5** - Média mensal da umidade entre 2005 a 2015

**Fonte:** Autor

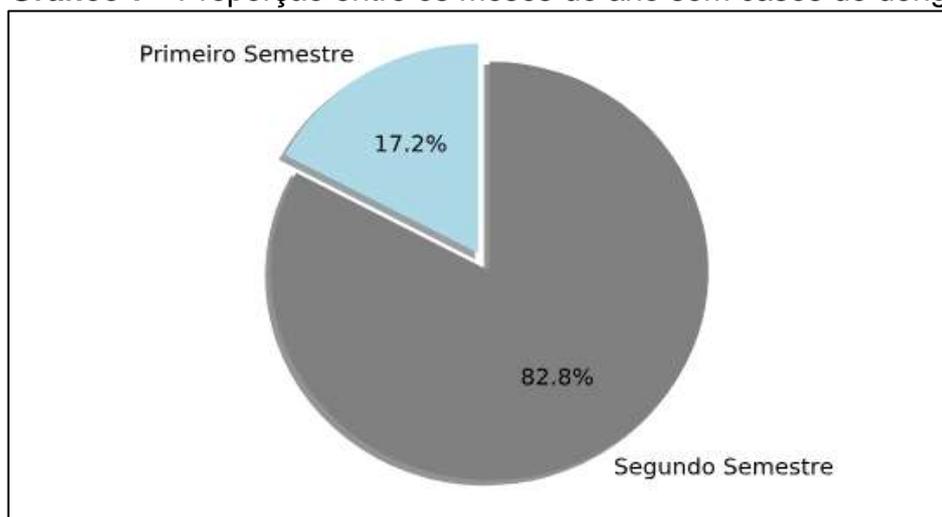
Outro fator importante citado como facilitador para a proliferação da dengue foi a temperatura. No Gráfico 6 (a) vê-se a variação média da temperatura máxima média no decorrer de todos os anos analisados e a variação média da temperatura mínima no Gráfico 6 (b). É possível notar que, comparando o Gráfico 5 anterior percebe-se a relação inversa entre as variáveis, isto é, a medida em que decorre os meses, a temperatura mínima e máxima decaem no primeiro semestre, e voltam a subir a partir do segundo semestre do ano, esta relação é inversa a da umidade, que tem seu aumento na primeira parte do ano e decréscimo a partir do segundo semestre. Logo, conforme a temperatura baixa, a umidade relativa do ar aumenta.

**Gráfico 6** - Média mensal da temperatura entre 2005 a 2015



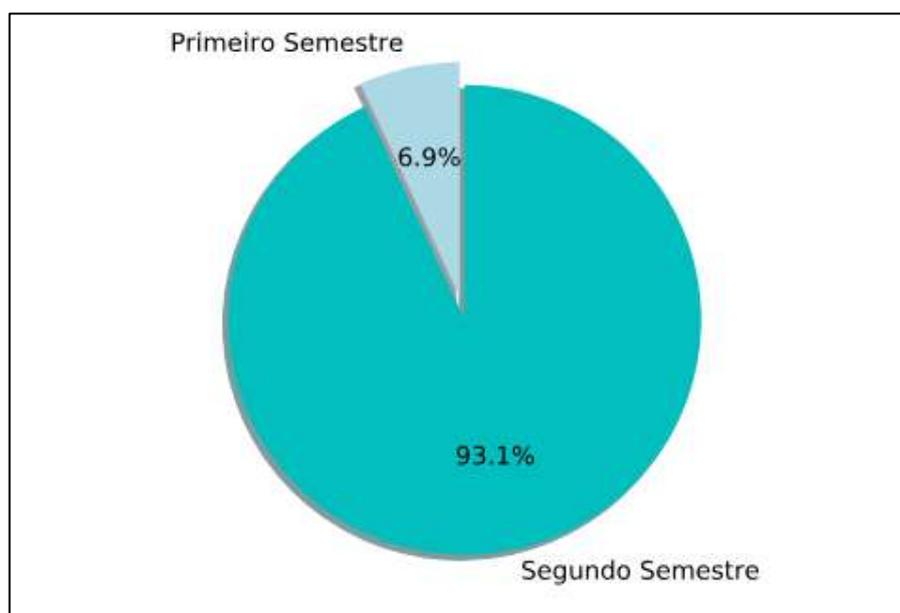
**Fonte:** Elaborado pelo autor, 2019.

Observa-se, nesta situação, que a dengue possui um comportamento sazonal durante o primeiro semestre do ano, em que os casos aumentam na condição da diminuição da temperatura e aumento da umidade. A partir do conjunto de dados foi possível identificar esse período, e no Gráfico 7 é apresentada a proporção dos meses em que não houveram notificações de casos de dengue no período de 2005 a 2015. Foram identificados 29 meses no qual possuíam zero casos registrados, do total 82,8% destes meses estão presentes no segundo semestre do ano demonstrando que o período de ocorrência da dengue se apresenta com mais frequência no primeiro semestre do ano.

**Gráfico 7** - Proporção entre os meses do ano sem casos de dengue

**Fonte:** Elaborado pelo autor, 2019.

Seguindo o mesmo padrão comportamental da dengue identificado no período entre 2005 e 2015, a chuva apresenta de maneira similar o comportamento de maior frequência no primeiro semestre do ano. Foi verificado que dos meses analisados, 29 destes não possuíram ocorrência de chuva, sendo 93,1% destes meses se encontram no segundo semestre do ano. O Gráfico 8 apresenta a proporção desta análise.

**Gráfico 8** - Proporção entre os meses do ano sem chuvas

**Fonte:** Elaborado pelo autor, 2019.

A partir da similaridade dos últimos Gráficos 7 e 8, no Quadro 2 é exposto os pontos de pico entre duas variáveis. Foram selecionados os meses em que ocorreram o maior pico do surto da dengue (em vermelho) e o maior pico de ocorrência de chuvas (em azul) durante o período de 2005 e 2015.

**Quadro 2** - Picos de casos de dengue e chuvas entre 2005 e 2015

	JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO	SET	OUT	NOV	DEZ
2005			384		128							
2006			244		27							
2007		264					27					
2008			491	211								
2009				641	3							
2010			123			37						
2011	221		286									
2012		111		91								
2013				67								87
2014			514			9						
2015		187			178							

Fonte: Elaborado pelo autor, 2019.

Como pode-se perceber, o período chuvoso em Patos transcorre entre os meses de dezembro a abril com maior índice de chuva (mm) destacado de fevereiro a março e o de dengue entre os meses de março e julho. A partir da representação do quadro 2 pode-se chegar à premissa de que seguindo um padrão comportamental, os maiores surtos de dengue se apresentam de forma regular após o período chuvoso na cidade de Patos.

#### 4.2 Análise de correlação de dados

A aplicação do método de Pearson para correlação de variáveis foi aplicada nos dados presentes no *dataset* e apresentados no Gráfico 9. No referido gráfico é possível identificar o grau de relação entre as variáveis estudadas. O gráfico é apresentado em cores que indicam o quanto uma variável pode estar relacionada

linearmente com outra. A cor pura amarela indica uma forte relação positiva (+1), enquanto a cor azul escura demonstra que as variáveis possuem uma relação forte negativa (-1). As cores intermediárias variam conforme a relação cresce ou diminui, tendo uma cor clara quando o coeficiente de correlação se aproxima de 0, dito como uma relação desprezível.

**Gráfico 9** - Correlação entre as variáveis do estudo



**Fonte:** Elaborado pelo autor, 2019.

Os índices de correlação podem ser interpretados da seguinte forma:

- +0.9 a +1.0 | -0.9 a -0.9 indica-se que existe uma correlação muito forte;
- +0.7 a +0.9 | -0.7 a -0.9 indica-se uma correlação forte;
- +0.5 a +0.7 | -0.5 a -0.7 indica-se uma correlação moderada;
- +0.3 a +0.5 | -0.3 a -0.5 indica-se uma correlação fraca;
- +0.0 a +0.3 | -0.0 a -0.3 indica-se uma correlação desprezível.

Na tabela 2 são expostos de todos os resultados dos coeficientes da aplicação do método de correlação de Pearson entre as variáveis analisadas.

**Tabela 2** - Coeficientes de correlação entre as variáveis

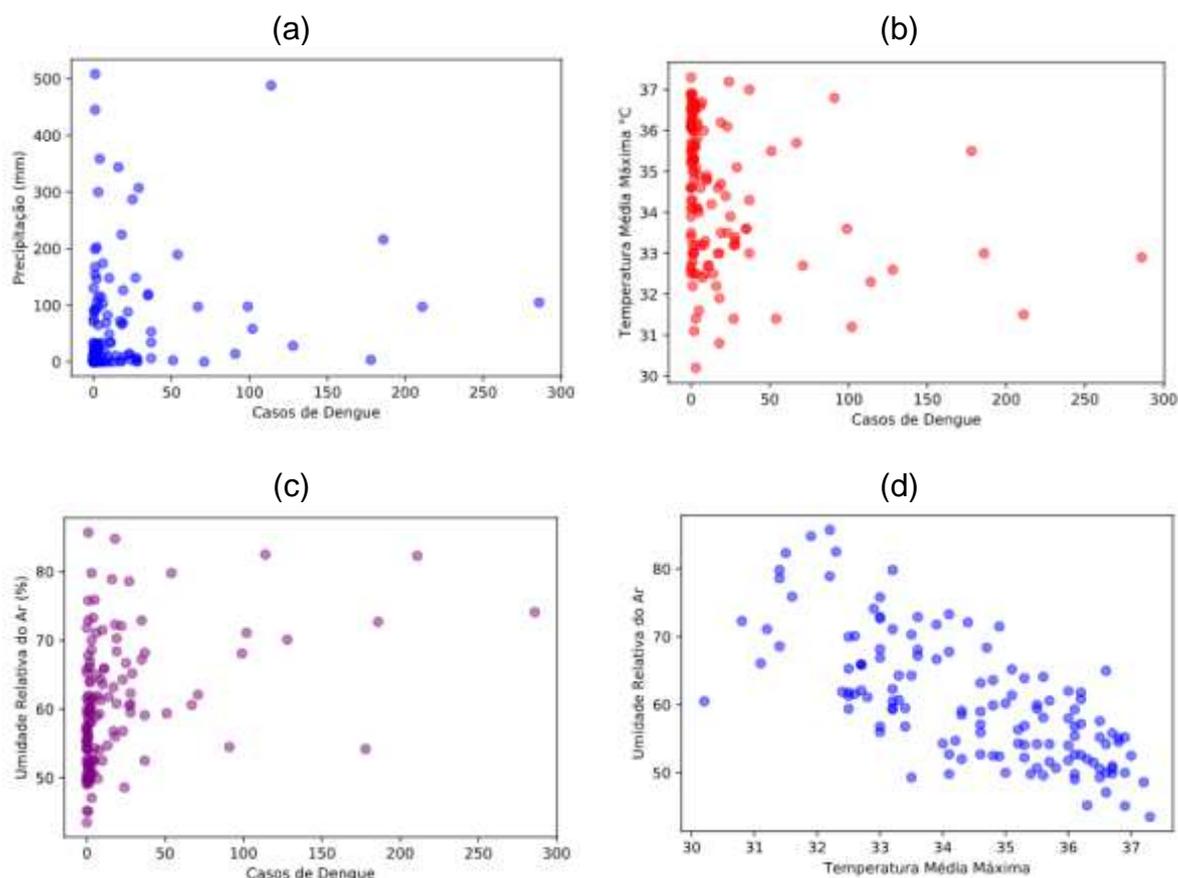
	MÊS	EVAPO	INSOL	PRECIP	TMAXM	TMINM	UMID	CASOS	CHUVAC
MÊS	1,000	-0,546	0,462	-0,440	0,258	-0,270	-0,535	-0,264	-0,388
EVAPO	-0,546	1,000	-0,386	0,661	-0,370	0,067	0,656	0,277	0,578
INSOL	0,462	-0,386	1,000	-0,310	0,638	0,003	-0,606	-0,130	-0,299
PRECIP	-0,440	0,661	-0,310	1,000	-0,301	0,079	0,658	0,161	0,931
TMAXM	0,258	-0,370	0,638	-0,301	1,000	0,371	-0,711	-0,257	-0,295
TMINM	-0,270	0,067	0,003	0,079	0,371	1,000	-0,195	0,035	0,020
UMID	-0,535	0,656	-0,606	0,658	-0,711	-0,195	1,000	0,360	0,613
CASOS	-0,264	0,277	-0,130	0,161	-0,257	0,035	0,360	1,000	0,148
CHUVAC	-0,388	0,578	-0,299	0,931	-0,295	0,020	0,613	0,148	1,000

Fonte: Elaborado pelo autor, 2019.

Com isso, pode-se analisar a correlação entre as variáveis, tendo como maior destaque as seguintes relações: mês, evapotranspiração e umidade relativa do ar, evapotranspiração, precipitação e chuvas acumuladas, insolação total, temperatura máxima média e umidade, precipitação e umidade, temperatura máxima média e umidade, umidade e chuva acumuladas. Os demais coeficientes apresentaram uma relação fraca ou desprezível.

A análise de correlação não indica a causalidade da relação entre as variáveis, mas demonstra estatisticamente a relação, servindo como base para análises posteriores. Com isso, a partir dos dados obtidos, não demonstrou uma relação linear entre os casos de dengue e as variáveis climáticas, assim a correlação linear mostrou-se não existir.

O coeficiente de correlação demonstrou-se abaixo do ideal, tendo o resultado exposto nos Gráficos 10 de dispersão. Como pode ser visto nos gráficos abaixo, a relação entre as variáveis climatológicas não se apresenta de maneira linear ou crescente. O Gráfico 10 (a) evidencia que, entre 0 e 50 ocorrências de casos de dengue notificados são mais frequentes quando a precipitação se apresenta de até 200mm, assim como os casos de dengue que são mais regulares quanto a temperatura máxima está entre 33°C e 37°C conforme exposto no Gráfico 10 (b).

**Gráfico 10** - Dispersão entre as variáveis selecionadas

**Fonte:** Elaborado pelo autor, 2019.

A relação entre a umidade relativa do ar e os casos de dengue demonstrada no Gráfico 10 (c) explica que, de 0 a 50 ocorrências de casos de dengue são mais recorrentes quando o nível de umidade está entre 50% e 70%. Já no Gráfico 10 (d) é possível identificar uma correlação inversa (negativa), no qual a medida em que a temperatura (x) aumenta a umidade (y) diminui, essa relação foi identificada através do coeficiente de Pearson.

Em um estudo realizado no *Institute for Animal Health*, Mellor e Leake (2000) afirmam que a distribuição geográfica e taxa de desenvolvimento de insetos vetores está fortemente relacionada à temperatura, chuva e umidade. Um aumento na temperatura acelera a taxa metabólica do inseto, aumenta a produção de ovos e torna a alimentação de sangue mais frequente. Outro aspecto levantado por Mellor e Leake (2000) aponta que por definição, as doenças transmitidas por vetores em seu estado inicial são sensíveis a mudanças nas variáveis climáticas, citando temperatura, umidade, chuva e vento. Além desses efeitos diretos, as variáveis climáticas também

podem ter efeitos indiretos importantes sobre a abundância e distribuição dos vetores e a ocorrência de doenças.

A Organização Mundial da Saúde (*World Health Organization*, WHO) (2004) ao referir-se ao assunto, diz que a influência da chuva também é significativa, embora menos fácil de prever. A precipitação tem um efeito indireto na longevidade do vetor através de seu efeito na umidade. Condições relativamente úmidas podem criar habitats favoráveis para insetos, aumentando assim a distribuição geográfica e a abundância sazonal de vetores de doenças. Em outros casos, o excesso de chuvas pode também ter efeitos catastróficos nas populações de vetores locais se as inundações eliminarem locais de reprodução.

Ainda nesta mesma linha de considerações, a Organização Mundial da Saúde (2004) argumenta que, mesmo onde as ligações entre doença e clima são relativamente fortes, outros fatores não climáticos também podem ter um impacto significativo no momento e gravidade dos surtos de doenças. Um desses fatores é a vulnerabilidade da população, destacando problemas sociais e econômicos.

Desta forma, a utilização exclusivamente de variáveis climáticas não se demonstrou eficaz em encontrar relações diretas com a doença da dengue, pois os fatores climáticos estão relacionados diretamente ao vetor da doença.

### **4.3 Aplicação e resultados da Regressão Linear**

Após a análise de dados a fim de visualizar melhor as relações e padrões entre algumas variáveis foi necessário investigar os coeficientes de correlação das variáveis selecionadas para verificar qual se adequa melhor ao problema e que possa ser capaz de produzir um resultado satisfatório.

Com o resultado das análises de correlação não foi possível identificar uma relação linear direta entre os casos de dengue e as variáveis meteorológicas. Inicialmente era proposto a utilização, a partir do resultado do coeficiente de correlação, selecionar pares de variáveis para identificar a possibilidade da predição de casos de dengue utilizando aprendizado de máquina.

Dito isso, foi feito a aplicação do algoritmo de aprendizado de máquina, utilizando regressão linear simples com os pares de variáveis, sendo elas: chuvas e casos de dengue, umidade relativa do ar e casos de dengue, temperatura máxima

média e casos de dengue, com finalidade de verificar a possibilidade de prever casos de dengue a partir destas variáveis meteorológicas.

As justificativas relativas à qualidade dos dados, interferindo no sucesso do modelo de Regressão Linear, podem-se destacar: (1) a coleta e tabulação e armazenamento dos dados dos casos notificados da doença dengue; (2) ausência de informação de outros fatores inerentes à incidência da dengue; (3) a diferença de grandezas e intervalos dos dados disponíveis.

Após a análise inicial dos dados, verificou-se que a partir dos dados utilizados, as correlações iniciais entre as variáveis do estudo, apesar de não indiciar causalidade, os coeficientes de correlação se demonstraram de maneira oposta ao esperado, pois esperava-se encontrar inicialmente alguma relação direta entre os dados.

Após a análise, foi aplicado o algoritmo de Regressão Linear nos dados utilizados a fim de verificar o coeficiente de determinação  $R^2$  e a relação identificada na análise prévia. Utilizou-se como variável preditora (variável que tem o poder de influenciar o resultado) os dados de chuvas, temperaturas e umidade relativa do ar e como variável alvo (ou 0, variável a ser prevista) os casos de dengue.

O coeficiente  $R^2$  pode ser visto como uma tentativa de explicar as variações da regressão linear nas variáveis preditoras que resultam alterações nas variáveis alvos. O resultado do coeficiente  $R^2$  dá o percentual da variação que é explicada pela regressão linear. O coeficiente de determinação  $R^2$  varia entre 0 e 1, em percentagem, quanto maior o coeficiente melhor o modelo é explicado e melhor se ajusta à amostra de dados utilizados.

Portanto, a partir dos coeficientes  $R^2$  foi possível verificar o quanto o modelo utilizado explica os dados analisados e se o modelo é eficaz. Inicialmente testou-se a correlação, como foi demonstrado na seção anterior, entre os pares de variáveis, os dados dos casos de dengue e as demais variáveis meteorológicas. Os resultados dos coeficientes de correlação não foram satisfatórios, contudo, foi verificado cada variável preditora separadamente com os casos de dengue com o objetivo de observar as relações entre pares. Os resultados da análise do coeficiente  $R^2$  entre os pares de variáveis é exibido na tabela 3 abaixo.

**Tabela 3** – Coeficientes de determinação entre as variáveis

COEFICIENTE		CHUVAS	TMAXM	UMIDADE	CASOS
R <sup>2</sup> _X	<b>CASOS</b>	0,03	0,07	0,13	1,00
R <sup>2</sup> _Y		0,00	0,00	0,01	1,00
R <sup>2</sup> _A		0,07	0,11	0,14	1,00
R <sup>2</sup> _B		0,09	0,06	0,06	1,00

Fonte: Elaborado pelo autor, 2019.

O resultado do coeficiente R<sup>2</sup>\_X foi encontrado aplicando o algoritmo de regressão linear com o *dataset* completo utilizando todos os meses dos anos. O coeficiente R<sup>2</sup>\_Y foi obtido como resultado do *dataset* que possuía somente os meses do período de dengue em Patos (março a julho). Já o coeficiente R<sup>2</sup>\_A foi obtido com a inclusão da variável preditora que indica o mês do ano em relação às demais, ou seja, os casos de dengue em relação ao mês, chuva, temperatura e umidade. E a variável R<sup>2</sup>\_B utilizando os mesmos fatores da variável R<sup>2</sup>\_A mas somente com o período de surto em Patos.

A princípio pretendia-se utilizar todos os meses do ano para verificar a relação dos casos de dengue e as variáveis climatológicas, pois esperava-se identificar padrões em todos os meses durante o ano. Assim, realizou-se teste com os dados filtrados sem os meses que não faziam parte do período de surto da dengue em Patos, sendo esses os meses de março até julho. Entretanto, com a exclusão dos meses em questão, a amostra de dados perdeu significativamente tamanho, reduzindo pela metade os dados da amostra inicial, prejudicando mais o modelo de regressão linear.

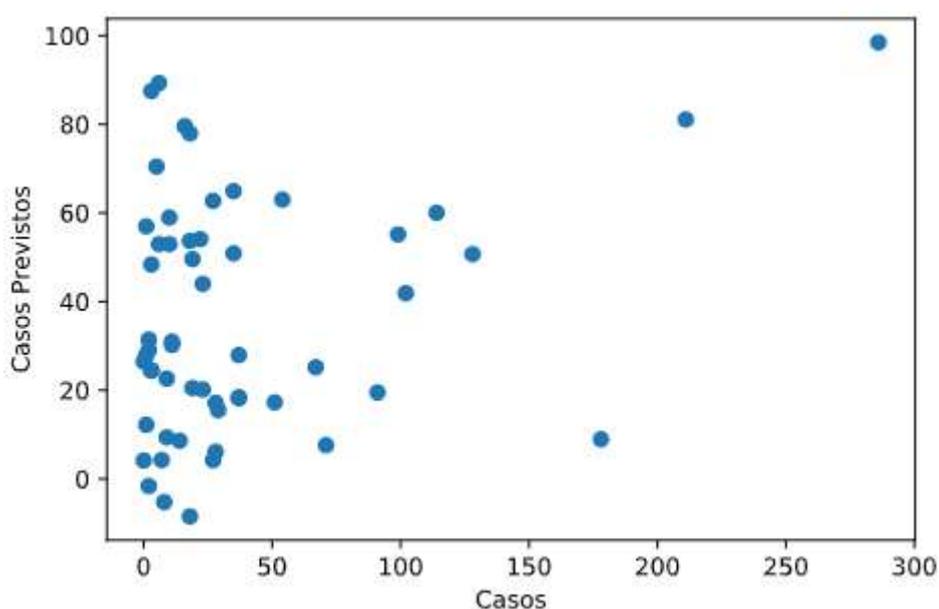
Foram realizados testes com as amostras de dados, segmentando os dados a fim de verificar o coeficiente R<sup>2</sup> do conjunto de dados utilizados. A análise foi feita entre as variáveis climáticas, os casos de dengue e o período do ano, verificando o coeficiente par a par e global.

A análise do coeficiente R<sup>2</sup> aplicado ao conjunto de dados globais forneceu um coeficiente de apenas 0.15, ou seja, o modelo de regressão só explica 15% dos casos de dengue — variável dependente — do conjunto global dos dados. Entretanto, o coeficiente gerado nos dados filtrados retornou um valor inferior, 0.12 indicando que o modelo filtrado esclarece apenas 12% dos dados, isto pode ser um indicador positivo, visto que com a exclusão de uma grande parte dos dados — os meses fora do período de surto de dengue em Patos — o coeficiente não foi alterado significativamente, sugerindo que a amostra de dados filtrada foi melhor em relação a global.

Logo, isso significa que apesar de um coeficiente  $R^2$  relativamente abaixo do esperado, o modelo se demonstrou melhor em um conjunto de dados mais específico, sugerindo a adição de dados à amostra utilizada.

Após a aplicação e avaliação do coeficiente de determinação  $R^2$  no conjunto completo dos dados, verificou-se a relação entre os casos de dengue originais e os casos previstos segundo o algoritmo de regressão linear do *scikit-learn*. Conforme o Gráfico de dispersão 11 demonstra, a relação entre as amostras originais e as previstas não apresentou nenhuma linearidade significativa, afirmando que a partir dos dados utilizados não foi possível identificar relações diretas entre as variáveis climatológicas e os casos de dengue na cidade de Patos.

**Gráfico 11** - Relação entre casos e os casos previstos



**Fonte:** Elaborado pelo autor, 2019.

Os surtos de dengue exibiram padrões distintos de distribuição ano a ano, que não eram nem relacionados à períodos nem significativamente correlacionados a qualquer variável meteorológica utilizada como base neste estudo.

Dessa forma, a partir dos resultados das aplicações dos algoritmos e das análises de dados, sugere-se que o escopo dos dados da pesquisa seja aumentado, introduzindo novos fatores inerentes a dengue como também expandir o período de de anos utilizado. A análise de ampliada em cidades circunvizinhas a Patos, é outro

fator que pode contribuir para o aperfeiçoamento do algoritmo, com o objetivo de encontrar novas relações ou comparar com as existentes no município de Patos.

## 5 CONSIDERAÇÕES FINAIS

Verificou-se através desta pesquisa, que a dengue possui padrões comportamentais complexos no município de Patos, tornando a identificação de relações entre as variáveis climatológicas e os casos de dengue difícil. Ao realizar a análise de dados, foi verificado que, a dengue apresenta um comportamento distinto no município de Patos, pois os fatores climatológicos inerentes a epidemia da dengue se demonstraram específicos da cidade, visto que as mudanças climáticas variam de maneira diferente do restante do Brasil em relação aos dados utilizados.

Constatou-se que, os casos de dengue estão relacionados regularmente com as variações da precipitação, temperatura e umidade de maneira geral, tendo como fatores principais o aumento da chuva e umidade e a diminuição da temperatura durante o ano, permitindo definir o período sazonal da dengue em Patos dos meses de março a julho.

Os autores Mellor e Leake (2000) enfatizam a forte relação com o aumento da temperatura como fator de proliferação do vetor da doença dengue, entretanto as análises dos casos notificados de dengue em Patos se demonstraram contraditórias neste aspecto da relação, sugerindo que os fatores climáticos se associam de maneira desigual aos casos da doença dengue e o vetor transmissor da doença.

Ao realizar a aplicação dos métodos de aprendizado de máquina nos dados fornecidos, após os resultados dos coeficientes de correlação e aplicação no algoritmo de regressão linear, não apresentou resultados preditivos satisfatórios dentro do contexto desta pesquisa. Verificou-se que, se faz necessário a ampliação do escopo da pesquisa, incluindo novas linhas de dados à amostra, como também a possibilidade de inserção de novos fatores pertinentes à dengue. O que se sugere para trabalhos futuros.

Existe um interesse atual nas relações entre dengue e clima por um motivo principal. Há expectativa de que as previsões a médio prazo do risco de epidemias de dengue possam ser possíveis. Em termos mais gerais, espera-se que uma melhor compreensão das relações sazonais, entre o ambiente e a doença dengue possam ajudar a entender e antecipar processos de longo prazo mais complexos e indiretos, como a mudança global do clima.

Percebe-se que o desenvolvimento de modelos integrados para a avaliação do impacto na saúde das mudanças ambientais globais ainda é uma ciência

relativamente nova. Portanto, os atuais métodos epidemiológicos de pesquisa não parecem ser adaptados às análises da causa de doenças, que envolvem sistemas complexos influenciados por intervenções humanas ou processos mais simples que ocorrerão no futuro.

De modo geral, as mudanças de temperatura, umidade e precipitação podem influenciar o comportamento e a distribuição geográfica dos vetores e, assim, alterar a incidência de doenças transmitidas por vetores, mas não se pode sugerir que o clima seja fator único que afete a transmissão da dengue. Entretanto, ainda existe uma diferença fundamental entre uma abordagem explicativa baseada em dados e uma abordagem descritiva baseada em cenário. Dessa forma, a predição de casos da doença dengue utilizando ferramentas computacionais através de dados estatísticos, é necessário um modelo mais eficaz capaz de abranger fatores regionais além dos que são característicos da doença dengue, com o propósito de avaliar o impacto futuro das mudanças sociais e climáticas do ambiente e os efeitos potenciais de se antecipar o surto da doença.

Conclui-se que, os casos de dengue estão fortemente relacionados às variações meteorológicas no município de Patos. As mudanças climáticas durante o ano influenciam como os dados se comportam em relação à incidência dos casos notificados da dengue na cidade, tendo como condições favoráveis para a maior ocorrência dos casos, a diminuição da temperatura e aumento nos níveis de chuva, como também da umidade do ar.

Tendo em vista os aspectos observados, a aplicação de algoritmos de aprendizado de máquina com foco nos casos de dengue, ainda possui limitações em relação ao conjunto de dados a ser explorado, devido à grande heterogeneidade dos dados utilizados, os modelos criados para análise da regressão linear nas amostras apresentam baixo índice de sucesso. Portanto, verificou-se que, a partir dos dados utilizados, não foi possível identificar relações lineares capazes de implementar um modelo preditivo qualificado, para prever surtos de casos de dengue futuros.

Por fim, esclarece que mesmo sendo uma doença que está fortemente relacionada às mudanças climáticas, a dengue pode ser influenciada por diversos outros aspectos, como fatores ambientais, sociais e populacionais. As políticas públicas de saúde para o combate e erradicação da doença podem influenciar na maneira como os padrões epidemiológicos se comportam, tendo como foco os

programas e atividades de conscientização da população, modificando o comportamento dos dados dos casos de dengue no município de Patos.

## REFERÊNCIAS

BATISTA, Gustavo Enrique de Almeida Prado Alves. **Pré-processamento de dados em aprendizado de máquina supervisionado**. 2003. 231 f. Tese (Doutorado) - Curso de Instituto de Ciências Matemáticas e de Computação, Instituto de Ciências Matemáticas e de Computação, São Carlos, 2003.

BRASIL. Ministério da Saúde. Departamento de Vigilância Epidemiológica. **Diretrizes Nacionais para a Prevenção e Controle de Epidemias de Dengue**. Brasília: Ministério da Saúde, 2009. (A). Disponível em: <[http://bvsmms.saude.gov.br/bvs/publicacoes/diretrizes\\_nacionais\\_prevencao\\_controle\\_dengue.pdf](http://bvsmms.saude.gov.br/bvs/publicacoes/diretrizes_nacionais_prevencao_controle_dengue.pdf)>. Acesso em: 10 nov. 2017.

BRASIL. Ministério da Saúde. Departamento de Vigilância Epidemiológica. **Plano de Contingência Nacional para Epidemias de Dengue**. Brasília: Ministério da Saúde, 2015. 42 p. Disponível em: <<http://portalarquivos.saude.gov.br/images/pdf/2015/janeiro/20/plano-contingencia-dengue-19jan15-web.pdf>>. Acesso em: 09 nov. 2017.

BRASIL. MINISTÉRIO DA SAÚDE. . **VIGILÂNCIA EM SAÚDE: Dengue, Esquistossomose, Hanseníase, Malária, Tracoma e Tuberculose**. 2. ed. Brasília: Ministério da Saúde, 2008. 119 p.

CHEESEMAN, P; STUTZ, J. Bayesian classification. In: SEVENTH NATIONAL CONFERENCE OF ARTIFICIAL INTELLIGENCE, 7. **AAAI-88 Proceedings** . St. Paul: Conference Of Artificial Intelligence, 1988.

CÂMARA, Fernando Portela et al. Clima e epidemias de dengue no Estado do Rio de Janeiro. **Revista da Sociedade Brasileira de Medicina Tropical**, [s.l.], v. 42, n. 2, p.137-140, abr. 2009. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s0037-86822009000200008>.

FARHAT, R. Dengue: Brasil prepare-se para a batalha. **InformeSaúde**, v. 7, n. 196, 2003.

FONSECA, J. J. S. **Metodologia da pesquisa científica**. Fortaleza: UEC, 2002.

GARSON, G. David. Structural equation modeling, from statnotes: Topics in multivariate analysis. **Retrieved May**, v. 1, p. 2010, 2009.

GERHARDT, Tatiana Engel; SILVEIRA, Denise Tolfo. **Métodos de Pesquisa**. Porto Alegre: Ufrgs, 2009. 120 p.

GIL, A. C. Métodos e técnicas de pesquisa social. 5.ed. São Paulo: Atlas, 1999.

GONZÁLEZ, M. B. e RODAS, G. G., **Predictive model of dengue focus applied to geographic information systems**, 2015 Latin American Computing Conference (CLEI), IEEE, Arequipa, 2015.

HAN, Jiawei; KAMBER, Micheline. **Data Mining: Concepts and Techniques**. 2. ed. São Francisco: Elsevier, 2006.

HALSTEAD, Scott B.. Dengue Virus–Mosquito Interactions. **Annual Review Of Entomology**, [s.l.], v. 53, n. 1, p.273-291, jan. 2008. Annual Reviews. <http://dx.doi.org/10.1146/annurev.ento.53.103106.093326>.

HALES, S et al. El Niño and the dynamics of vectorborne disease transmission. **Environmental Health Perspectives**, [s.l.], v. 107, n. 2, p.99-102, fev. 1999. Environmental Health Perspectives. <http://dx.doi.org/10.1289/ehp.9910799>.

HARRINGTON, Peter. **Machine Learning in Action**. New York: Manning Publications Co., 2012. 382 p.

Instituto Oswaldo Cruz. **O mosquito Aedes aegypti faz parte da história e vem se espalhando pelo mundo desde o período das colonizações**. Disponível em: <<http://www.ioc.fiocruz.br/dengue/textos/longatraje.html>>. Acesso em: 02 nov. 2017.

KITCHENHAM, Barbara. **Procedures for Performing Systematic Reviews**. Staffordshire: Keele University, 2004. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.122.3308&rep=rep1&type=pdf>>. Acesso em: 2 out. 2017.

MAIMON, Oded; ROKACH, Lior. **Data Mining and Knowledge Discovery Handbook**. 2. ed. [s.l.]: Springer, 2010.

MANNILA, Heikki. **Data mining: machine learning, statistics, and databases**. 1996. Disponível em: <<http://ieeexplore.ieee.org/document/505910/>>. Acesso em: 08 nov. 2017.

MARTENS, W J. **Health Impacts of Climate Change and Ozone Depletion: An Ecoepidemiologic Modeling Approach**. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1533278/>>. Acesso em: 15 abr. 2019.

MELLOR, P.S.; LEAKE, C.J.. Climatic and geographic influences on arboviral infections and vectors. **Revue Scientifique Et Technique de L'oie**, [s.l.], v. 19, n. 1, p.41-54, 1 abr. 2000. O.I.E (World Organisation for Animal Health). <http://dx.doi.org/10.20506/rst.19.1.1211>.

MILLMAN, K. Jarrod; AIVAZIS, Michael. Python for Scientists and Engineers. **Computing In Science & Engineering**, [s.l.], v. 13, n. 2, p.9-12, mar. 2011. Institute of Electrical and Electronics Engineers (IEEE). <http://dx.doi.org/10.1109/mcse.2011.36>.

MONARD, Maria Carolina, BARANAUSKAS, José Augusto. Conceitos Sobre Aprendizado de Máquina. **Sistemas Inteligentes Fundamentos e Aplicações**. 1 ed. Barueri-SP: Manole Ltda, 2003. p. 89--114. ISBN 85-204-168.

MOORE, David S.; NOTZ, William; FLIGNER, Michael A.. **The Basic Practice of Statistics**. New York: W.h. Freeman, 2003.

PEDREGOSA, Fabian et al. Scikit-learn: Machine Learning in Python. **Journal Of Machine Learning Research**. [s. L.], p. 2825-2830. 12 out. 2011. Disponível em: <<http://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>>. Acesso em: 15 nov. 2017.

PORTO, F. A. M. ; ZIVIANI, A. . Ciência de Dados. In: 3o. Seminário de Grandes Desafios da Computação no Brasil, 2014, Rio de Janeiro, RJ. Anais do 3o. Seminário de Grandes Desafios da Computação, 2014.

SANTOS, A. M dos. Profa Alcione Miranda dos Santos. **Departamento de Saúde Pública UFMA**, 2007.

SANTOS, Laura E. M; SILVA, Carlos A. et al. INTELIGÊNCIA COMPUTACIONAL APLICADA AO CONTROLE DO AEDES AEGYPTI. In: XXXVII IBERIAN LATIN-AMERICAN

CONGRESS ON COMPUTATIONAL METHODS IN ENGINEERING, 37., 2016, Brasília. **Proceedings of the XXXVII Iberian Latin-American Congress on Computational Methods in Engineering**. Brasília: Revista Interdisciplinar de Pesquisa em Engenharia, 2016. Disponível em: <<http://periodicos.unb.br/index.php/ripe/article/viewFile/23690/17014>>. Acesso em: 10 out. 2017.

SOUZA, Izabel Cristina Alcantara de; VIANNA, Rodrigo Pinheiro de Toledo; MORAES, Ronei Marcos de. 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, nov. 2007. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s0102-311x2007001100010>.

TEIXEIRA, MG; BARRETO, ML; GUERRA, Z. Epidemiologia e medidas de prevenção do Dengue. **Inf. Epidemiol. Sus**, v. 8, n. 4, p. 5-33, 1999.

VIANA, Dione Viero; IGNOTTI, Eliane. A ocorrência da dengue e variações meteorológicas no Brasil: revisão sistemática. **Revista Brasileira de Epidemiologia**, [s.l.], v. 16, n. 2, p.240-256, jun. 2013. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s1415-790x2013000200002>.

World Health Organization. **Using Climate to Predict Infectious Disease Outbreaks: a review**. Geneva: World Health Organization, 2004.