

# Estudo Comparativo de Técnicas de Previsão para Casos de Dengue


*Comparative study of forecasting techniques for Dengue cases*

Geovanne Oliveira Alves <sup>1</sup>  [orcid.org/0000-0002-2084-5516](https://orcid.org/0000-0002-2084-5516)

Thomás Tabosa de Oliveira <sup>1</sup>  [orcid.org/0000-0001-8224-5922](https://orcid.org/0000-0001-8224-5922)

Gleyson Rhuan Nascimento Campos <sup>1</sup>  [orcid.org/0000-0003-1245-3106](https://orcid.org/0000-0003-1245-3106)

Lubnnia Moraes Florêncio de Souza <sup>1</sup>  [orcid.org/0000-0002-2188-6272](https://orcid.org/0000-0002-2188-6272)

Sebastião Rogério da Silva Neto <sup>1</sup>  [orcid.org/0000-0001-8109-697X](https://orcid.org/0000-0001-8109-697X)

<sup>1</sup> Escola Politécnica de Pernambuco, Universidade de Pernambuco, Recife, Brasil.

E-mail do autor principal: Geovanne Oliveira Alves [goa@ecomp.poli.br](mailto:goa@ecomp.poli.br)

## Resumo

---

A dengue é uma infecção viral que se espalha rapidamente e é endêmica em mais de 100 países tropicais e subtropicais da África, América e regiões da Ásia-Pacífico. Geralmente, a epidemiologia da dengue é influenciada por uma complexa interação de fatores que incluem rápida urbanização e aumento da densidade populacional, capacidade dos sistemas de saúde, eficácia dos sistemas de controle de vetores, limpeza urbana etc. No Brasil há uma incidência de casos crescente nos últimos anos, em específico na capital Pernambucana Recife há um ambiente propício para o aumento expressivo dos casos de dengue como são apresentados pelo Boletim Epidemiológico da Secretaria de Vigilância em Saúde do Ministério da Saúde. O presente trabalho apresenta um estudo comparativo de técnicas de Mineração de Dados seguindo a abordagem CRISP-DM, para um modelo de previsão de casos de Dengue no Recife-PE.

**Palavras-Chave:** Previsão de dengue; Aprendizagem de máquina; Regressão Linear.

## Abstract

---

*Dengue is a viral infection that spreads rapidly and is endemic in more than 100 tropical and subtropical countries in Africa, America and Asia-Pacific regions. Generally, the epidemiology of dengue is influenced by a complex interplay of factors including rapid urbanization and increased population density, capacity of health systems, effectiveness of vector control systems, urban cleanliness, etc. In Brazil there is an increasing incidence of cases in recent years, specifically in the capital Pernambucana Recife there is a favorable environment for the significant increase in dengue cases as presented by the Epidemiological Bulletin of the Health Surveillance Secretariat of the Ministry of Health. The present work presents a comparative study of Data Mining techniques following the CRISP-DM approach, for a model for predicting Dengue cases in Recife-PE.*

**Key-words:** Dengue Prevision, Machine Learning; Linear Regression.

### 1 Introdução

A dengue é uma infecção viral que se espalha rapidamente e é endêmica em mais de 100 países tropicais e subtropicais da África, América e regiões da Ásia-Pacífico. É causada por qualquer um dos quatro sorotipos do vírus da dengue, e a infecção de um sorotipo do vírus da dengue não fornece imunidade cruzada contra os outros três sorotipos. Os vírus da dengue são transmitidos por mosquitos *Aedes* através da alimentação de sangue em hospedeiros humanos.

Os pacientes de dengue sofrem repentinamente febre, erupções cutâneas, dores musculares, dor nas articulações e leucopenia, geralmente se recuperando em 14 dias. No entanto, alguns pacientes desenvolvem dengue grave, que é uma complicação potencialmente letal caracterizada por manifestações hemorrágicas, vazamento plasmático grave e comprometimento grave dos órgãos [1][2].

Geralmente, a epidemiologia da dengue é influenciada por uma complexa interação de fatores que incluem rápida urbanização e aumento da densidade populacional, capacidade dos sistemas de saúde, eficácia dos sistemas de controle de vetores, limpeza urbana etc.

Até que surja a vacina ou medicamento para dengue esteja, as operações de controle de vetores que eliminam mosquitos adultos e/ou suas larvas através da redução da fonte de reprodução continuam sendo o único método eficaz para conter a transmissão de dengue.

#### 1.1 Descrição do Problema

O Brasil é situado em uma área predominantemente tropical, com extensas florestas na Região Amazônica, além de florestas no Leste, sudeste e litoral sul. Como também, uma grande região de pântano (Pantanal) no centro-oeste, uma região de savana (Cerrado), na área do planalto central, e uma região seca (Caatinga) no interior nordestino. A maior parte do País tem um clima tropical, sendo um local adequado para existência do vetor da dengue e sua proliferação [3].

A capital pernambucana Recife, possui a quarta maior cidade urbana do Brasil em população, com relação a sua geologia e relevo da cidade, está situada sobre uma planície aluvial (fluviomarinha), constituída

por ilhas, penínsulas, alagados e manguezais envolvidos por cinco rios.

Diante desse contexto, Recife apresenta um ambiente propício para o aumento expressivo dos casos de dengue como são apresentados pelo Boletim Epidemiológico da Secretaria de Vigilância em Saúde do Ministério da Saúde [4][5]. Dado a combinação da localização e os aspectos do clima tropical, torna a dinâmica de transmissão em Recife um fator alarmante.

Sendo assim, o trabalho pretende responder a seguinte questão de pesquisa: como um modelo de preditivo pode auxiliar na identificação do comportamento de casos de dengue na cidade do Recife-PE.

#### 1.2 Objetivo

Este estudo tem como objetivo desenvolver um modelo de previsão da dengue para aprimorar a vigilância e o controle da doença no estado de Pernambuco, especificamente na cidade de Recife.

O modelo de previsão de dengue se baseia na correlação de séries temporais levando em consideração (número de casos confirmados, mês e ano). Visando compreender o período temporal com maior probabilidade de elevados casos de dengue, possibilitando assim uma possível ferramenta de apoio aos órgãos públicos responsáveis tomarem as medidas preventivas necessárias.

#### 1.3 Justificativa

Apesar de ser uma doença já conhecida, a dengue ainda é muito presente em nossas vidas, e se agravou ainda mais dado aos seus transmissores, mosquitos (*Aedes aegypti* e outros), adaptam-se facilmente a climas tropicais, apesar das técnicas para prevenção de surtos dessas doenças já serem conhecidas pela população e pelos departamentos de controle, e seu tratamento disponível no SUS, esta doença ainda pode levar pessoas a óbito. Segundo o Ministério da Saúde *et al*, no boletim epidemiológico de 2019 foram registrados 1.439.471 casos no Brasil, destes casos, 591 vieram a óbito [6]. Dados a esses fatos, foi observado em entrevistas com o *stakeholder* que não foi encontrado um mecanismo computacional preditivo para casos de dengue na Prefeitura do Recife, para tanto, a implementação deste modelo proporciona vantagens aos órgãos administrativos públicos, para elaboração de medidas para

minimização dos impactos da proliferação do mosquito.

## 1.4 Escopo Negativo

Este trabalho não tem o objetivo de implementar painéis com apresentação de métricas e indicadores de maneira automática que são conhecidos como: dashboard. Como também, o trabalho não pretende apresentar ou identificar os casos de dengue através dos sintomas relacionados. Não foram coletadas informações de bases de coleta de lixo da cidade, pavimentação, terrenos baldios a fim de projetar possíveis casos de dengue ou surtos.

## 2 Fundamentação Teórica

### 2.1 Dengue

De acordo com a *World Health Organization*, a dengue é uma doença viral transmitida por mosquitos mais amplamente distribuída e de rápida disseminação no mundo, sua incidência aumentou 30 vezes nos últimos 50 anos, estima-se que entre 50 e 100 milhões de infecções ocorram anualmente em mais de 100 países endêmicos, colocando quase metade da população mundial em risco [1].

Os mosquitos vetores da doença agem principalmente nas regiões tropicais e subtropicais. Estes também estão bem adaptados aos ambientes urbanos, o que permite que os vírus se espalhem facilmente pelas cidades. Além disso, as condições climáticas locais desempenham um papel crítico no desenvolvimento de populações de vetores nos principais centros urbanos [7].

A dengue é causada por um vírus da família *Flaviviridae* e existem quatro sorotipos distintos, mas intimamente relacionados, do vírus causador da dengue (DENV-1, DENV-2, DENV-3 e DENV-4).

Os primeiros casos de dengue no Brasil datam do final do século XIX e, apesar da eliminação do *Aedes aegypti* em 1955, o mosquito foi reintroduzido no país na década de 70. Um surto ocorreu em 1981 em Boa Vista, no estado de Roraima, após vários surtos na América Central envolvendo os sorotipos DENV-1 e DENV-4 [8][9].

### 2.2 Mineração de Dados

Com o intuito de auxiliar na tomada de decisão, a mineração de dados é o conjunto de esforços realizados para descobrir padrões em bases de dados [10]. A extração das informações é dividida em Preditivas e Descritivas. As técnicas Preditivas são

aquelas que utilizam os dados para prever valores futuros, é onde são aplicadas classificações, regressões. Já as técnicas descritivas têm como objetivo encontrar formas que descrevam os dados, é onde são aplicadas técnicas de agrupamento, sumarização, modelagem de dependências e detecção de desvios [11].

#### 2.2.1 Regressão Linear Simples

É uma técnica de análise de dados entre duas variáveis, que analisa a amplitude de variação de dados de uma variável em decorrência de outra [10]. A equação é representada abaixo:  $y=a+bx$  onde  $y$  é a variável que se deseja prever, também chamada de dependente,  $x$  é a variável independente. A variável  $a$  é o intercepto do eixo  $y$  e, por fim, a variável  $b$  é a inclinação da reta gerada.

#### 2.2.2 Support Vector Machine

A *support vector machine* (SVM) é um classificador baseado na teoria de aprendizado estatístico de Vapnik [12]. Para efetuar classificações o SVM constrói hiperplanos em um espaço multidimensional objetivando separar casos de diferentes classes. Cada hiperplano é considerado como uma separação ótima que separa os vetores das classes sem erro e com distância máxima para com os vetores mais próximos [13].

#### 2.2.3 Multi-Layer Perceptron

O *multi-Layer Perceptron* (MLP) são um dos modelos de *machine learning* mais comuns, basicamente é um modelo de *perceptron* com uma ou mais camadas ocultas, cada camada possuindo uma determinada quantidade de neurônios, que são conectados pelos pesos [14].

#### 2.2.4 Random Forest

É uma técnica muito popular entre as técnicas de mineração, pois é uma técnica que pode ser utilizado tanto para predição como para classificação e são fáceis para o treinamento. Basicamente é uma técnica que unifica várias árvores de decisão referente aos dados de entrada da base de dados. Para cada variável de entrada da base, uma árvore de decisão sobre os seus dados é criada, que pode ser por um intervalo dos valores ou o valor específico. Desta forma, é desenvolvido um "caminho" por onde o dado irá caminhar até chegar na variável de saída.

## 2.3 Métricas

Existem várias métricas que podem ser utilizadas para avaliar o desempenho de algoritmos de aprendizagem de máquina, classificação, bem como algoritmos de regressão. Abaixo são apresentados algumas destas métricas.

**RMSE:** é a métrica mais amplamente usada para tarefas de regressão e é a raiz quadrada da diferença quadrática média entre o valor de destino e o valor previsto pelo modelo. É mais preferido em alguns casos porque os erros são primeiro ao quadrado antes da média, o que representa uma grande penalidade para erros grandes. Isso implica que o RMSE é útil quando erros grandes são indesejados.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Predito_i - Atual_i)^2}{n}}$$

**EMPA:** Média Absoluta Percentual: A Acurácia do erro em porcentagem. Este erro é calculado como a média do erro percentual. Quanto menor o valor de EMPA, melhor para o modelo.

$$EMPA = \frac{\sum_{i=1}^n |(Real - Previsão)/Real|}{n} * 100$$

**R<sup>2</sup>:** Coeficiente de determinação ou R<sup>2</sup> é outra métrica usada para avaliar o desempenho de um modelo de regressão. A métrica nos ajuda a comparar nosso modelo atual com uma linha de base constante e nos diz o quanto nosso modelo é melhor. A linha de base constante é escolhida tomando a média dos dados e desenhando uma linha na média. R<sup>2</sup> é uma pontuação sem escala que implica que não importa se os valores são muito grandes ou muito pequenos, o R<sup>2</sup> será sempre menor ou igual a 1.

$$R^2 = 1 - \frac{MSE(modelo)}{MSE(base)}$$

## 2.4 Trabalhos Relacionados

As condições climáticas desempenham um papel crítico no desenvolvimento de populações de vetores nos principais centros urbanos. Diversos trabalhos na literatura abordam a relação entre o clima e o surgimento do mosquito *Aedes aegypti*.

Em [15] os autores apresentaram um método para a previsão da distribuição espacial de novos

casos de dengue nos estados de Alagoas e Paraíba (nordeste do Brasil). Em Martinez *et al.* os autores avaliaram o desempenho do modelo sazonal ARIMA (SARIMA) na descrição e previsão do número mensal de casos notificados de dengue em Ribeirão Preto (São Paulo) [16].

A capacidade de identificar locais urbanos com alto risco de infecção por doenças é um aspecto central das políticas públicas que visam o controle dessas doenças. De acordo com Andersson *et al.*, foi investigado o uso de imagens no nível da rua, como as do *Google Street View*, juntamente com as Redes Neurais Convolucionais para prever as taxas de dengue e dengue hemorrágica em locais urbanos [17].

Em Appice *et al.* podemos observar que foi formulado uma estratégia holística de aprendizado de máquina para analisar a dinâmica temporal dos dados de temperatura e dengue e foi utilizado esse conhecimento para produzir previsões precisas da dengue, com base na temperatura em uma escala anual [18]. Já em Halim *et al.* e Azevedo *et al.*, foram utilizados modelos de regressão para prever surtos de dengue [19][20].

Por fim, observamos que também foram utilizados modelos de redes neurais para previsão de casos de dengue. De acordo com Chovatiya *et al.*, foram usadas as redes neurais recorrentes para previsão de casos de dengue [21]. Os dados utilizados para a previsão incluirão as condições climáticas, a poluição e as estatísticas dos pacientes diagnosticados com dengue nos anos anteriores. O modelo aprenderá usando esses dados e preverá a possibilidade de um surto se condições climáticas semelhantes ocorrerem no futuro.

## 3 Materiais e Métodos

### 3.1 Descrição da Base de Dados

Na condução desse trabalho foi utilizada a base de dados de casos de dengue. Com o intuito de atender o objetivo da pesquisa, foi necessário realizar alguns agrupamentos e filtros na base de dados. Segue abaixo o detalhamento da base de dados e os procedimentos aplicados.

Os dados referentes à dengue foram retirados dos conjuntos de dados disponíveis ao público do Sistema de Dados Abertos da Prefeitura do Recife [22]. Esta base de dados é populada com as notificações de dengue da cidade do Recife e contém dados não

sensíveis dos pacientes, como raça, idade e sexo do mesmo, além de dados clínicos com possíveis sintomas sobre a doença, como febre, vômito e náuseas. Esse estudo contempla as bases no período de 2013 a 2018, devido a disponibilidade dos dados de Dengue da cidade do Recife.

### 3.2 Pré-processamento dos Dados

Foram selecionadas as colunas da data da notificação e tipo de classificação final na base de dengue, pois o objetivo é selecionar apenas a quantidade de casos confirmados da doença. Em seguida, a base de dengue foi agrupada por mês, totalizando a quantidade de casos de dengue que ocorreram em cada mês do ano de 2013 até 2018. Resultando em uma base com 67 registros e 3 atributos, dos quais são os anos, os meses e os números de casos de dengue. Segue abaixo o dicionário da base final (Tabela 1).

**Quadro 1:** Dicionário da base de dados.

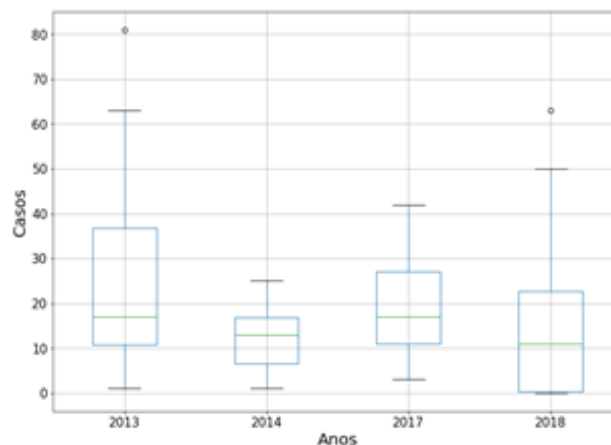
Código	Descrição	Tipo	Observação
ano	Ano referente aos dados	Numérico	Ex.: 2018
mês	Mês referente aos dados	Numérico	Janeiro equivalente a 1 e assim sucessivamente.
num_casos	Número de casos de dengue	Numérico	Ex.: 50

### 3.3 Análise Descritiva dos Dados

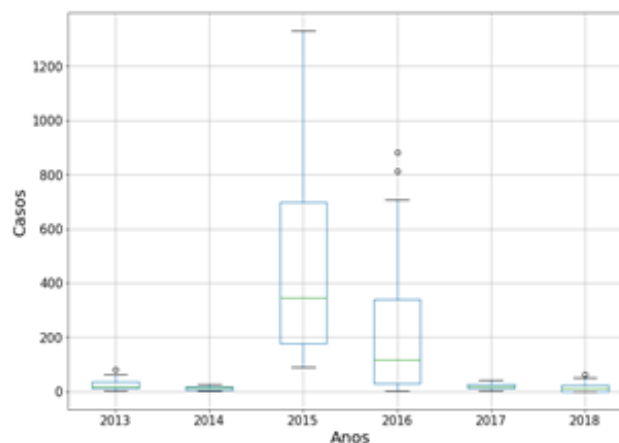
A análise descritiva representa a área de investigação nos dados que busca tanto descrever fatos relevantes, não-triviais e desconhecidos dos usuários, como analisar a base de dados, principalmente pelo seu aspecto de qualidade, para validar todo o processo da mineração e seus resultados, ou seja, o conhecimento encontrado.

A Figura 1 está representando o *boxplot* sobre a variável de casos de dengue sobre os anos, devido aos anos de 2015 e 2016 possuírem valores atípicos, resolvemos gerar uma segunda figura (Figura 2), apenas com anos de 2013, 2014, 2017 e 2018, para uma melhor visualização. Analisando a figura, podemos observar que, removendo os anos dos surtos de dengue (2015 e 2016), o ano que possui mais casos é o ano de 2013, seguido do ano de 2018, e mesmo como sendo os anos com maiores dados,

esses anos também possuem valores atípicos (outliers), especificamente, um em cada um.

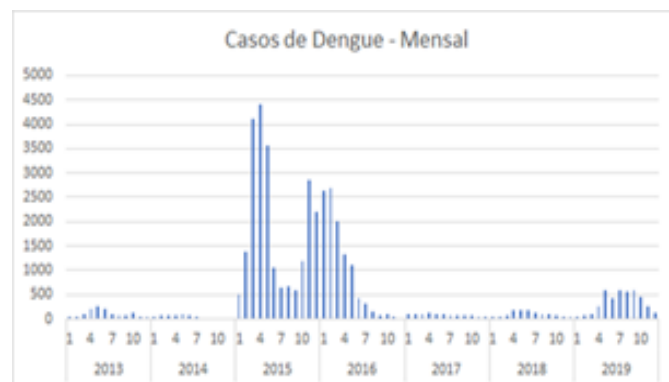


**Figura 1:** Boxplot casos por ano.



**Figura 2:** Boxplot casos por ano (sem surto).

Através da figura mensal de casos de Dengue (Figura 3) podemos observar os valores atípicos referente ao surto de casos em 2015 e 2016.



**Figura 3:** Casos de dengue por mês.



### 3.4 CRISP-DM

O **CRISP-DM** (*Cross-industry standard process for data mining*) é uma metodologia que fornece uma abordagem estruturada para processos de mineração de dados, sendo amplamente utilizada devido à sua poderosa praticidade, flexibilidade e utilidade ao usar a análise para resolver problemas comerciais complexos.

Esse processo, consiste em seis etapas estas são: entendimento do negócio, entendimento dos dados, preparação dos dados, modelagem, avaliação e implantação e pode ser considerado uma implementação do KDD [20]. A seguir será descrito cada uma das etapas desta metodologia.

#### 1. Entendimento do negócio:

Avaliar e comparar meios de realizar previsão de números de casos de arboviroses (dengue) na cidade do Recife.

#### 2. Entendimento dos dados:

Bases de dados de notificações de arboviroses obtidas através do portal de transparência de Recife, do ano 2013 até 2018. Dentre as informações cedidas nas bases, encontrava-se a data da notificação, sintomas, confirmação ou diagnóstico se o paciente está com alguma arbovirose.

#### 3. Preparação dos dados:

Nesta etapa selecionamos as colunas da data da notificação e tipo de classificação final na base de dengue, pois o objetivo é selecionar apenas a quantidade de casos confirmados da doença. Em seguida, as bases de dengue foram unificadas e então agrupadas por semana, totalizando a quantidade de casos de dengue que ocorreram em cada semana do ano de 2013 até 2018.

#### 4. Modelagem:

Nesta etapa aplicamos as seguintes técnicas de modelagem: random forest, regressão linear, SVM e a rede neural com função de ativação tangente hiperbólica.

#### 5. Avaliação:

Dentre as técnicas de modelagem utilizadas foi utilizado as métricas de RMSE (raiz quadrada do erro-médio), R2 (coeficiente de determinação), e EMPA (média percentual absoluta do erro).

#### 6. Implantação:

Este trabalho tem como objetivo realizar uma entrega intelectual de uma análise de algoritmos para realizar a previsão de casos de dengue na cidade de Recife. Abrindo um novo campo de abordagem para uma ferramenta de apoio de decisão, visão diminuir os números de casos de dengue na cidade

### 3.5 Metodologia Experimental

Para a previsão de um determinado mês foi utilizado os três meses anteriores a ele, e os meses correlatos dos anos anteriores. Indiferente do algoritmo a ser utilizado, seguiu esse procedimento como demonstrado no diagrama abaixo (Figura 4).

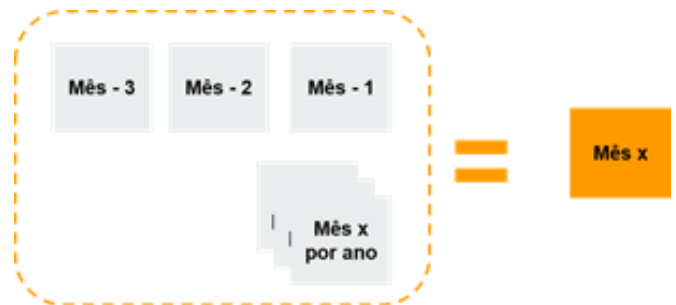


Figura 4: Procedimento para previsão.

O ambiente experimental utilizado foi um processador *intel* core i5 8º geração e placa de vídeo GeForce MX110 e 8GB de memória RAM no sistema Windows 10. Através do software Orange<sup>1</sup> foi aplicado os algoritmos SVM, Regressão Linear, Random Forest e MLP.

O algoritmo da MLP foi aplicado com 50 neurônios na camada escondida, utilizando a função de ativação tangente hiperbólica e o otimizador L-BFGS-B com no máximo 500 números de iterações. O Random Forest foi aplicado com 200 árvores e reaplicável para treino. Os algoritmos SVM e Regressão Linear foram utilizadas as configurações padrão do Orange. Logo após foi utilizado métricas para conseguir assim comparar a eficiência das previsões realizadas pelos diferentes algoritmos.

## 4 Análise de Discussão dos Resultados

### 4.1 Resultados

Após os experimentos realizados com o conjunto de dados sem o surto, ou seja, como dados mensais de 2013, 2014 e 2017. Esses dados alimentaram o modelo para a validação da previsão de casos de dengue do ano de 2018. A Figura 5 mostra a predição dos algoritmos representados pelas linhas e os números de casos reais representado pelas barras.

Nota-se uma aproximação do estimado pelos algoritmos para o real, em específico, a regressão linear que mostrou uma melhor adequação do modelo, embora, o SVM também apresenta bons resultados, observa-se maior distanciamento entre valores previsto e real nos meses de maiores casos.



Figura 5: Casos reais e casos previstos pelos classificadores.

Para uma melhor comparação segue as métricas na Tabela 1.

Tabela 1: Métricas de análise dos modelos

Métrica	Redes Neurais	Regressão Linear	SVM	Random Forest
RMSE	34,4504	28,5598	44,7753	33,3366
R2	0,6690	0,8071	0,3601	0,6868
EPMA	36,35%	28,73%	29,11%	34,43%

As métricas utilizadas para a avaliação dos modelos foram a RMSE, EPMA e R2. A métrica RMSE calcula o erro ponderando valores maiores para as diferenças mais discrepantes, logo considera-se que quanto mais próximo de zero menor o erro, nessa métrica a regressão linear se destaca.

A EPMA é uma métrica que calcula o erro em percentual, o que podemos inferir que quanto menor a taxa, melhor será o modelo. Nessa situação podemos destacar a regressão linear e o SVM, embora que, através da métrica anterior, observa-se que o SVM não prediz com precisão meses com maiores quantidades de casos.

R2, ou também chamado de coeficiente de determinação, informa o grau de ajuste do modelo estimado para com os dados reais. Essa métrica varia de 0 a 1, onde 1 representa ótimo ajuste e 0 o pior ajuste. Destacamos novamente o modelo com regressão linear que apresentou coeficiente de 0,8071.

## 4.2 Discussão

O trabalho aborda uma análise por meio de experimentos de modelos computacionais para realizar a predição de casos de dengue na cidade do Recife, onde para meio de estudo foram abordados quatro algoritmos, sendo eles: redes neurais, regressão linear, SVM, e random forest.

Para realizar a comparação dos algoritmos de predição, selecionamos três métricas que mostram o quão o algoritmo errou e o grau de ajuste do mesmo. E com isso apresentamos os resultados dos experimentos.

Podemos notar que na métrica RMSE, temos os destaques nos algoritmos de regressão linear e SVM, onde apresentam menor e maior valor do estudo. Notamos também a similaridade no resultado de Redes neurais e Random Forest.

Na métrica R2 é possível visualizar o destaque para o maior valor no algoritmo de regressão linear com 0,8071. Apresentando assim, uma superioridade dentre os demais algoritmos.

A métrica EPMA traz a porcentagem dos algoritmos, e notamos mais uma vez o destaque para a regressão linear, mesmo tendo uma diferença de menos de 0,5% para o SVM.

Dentre os resultados obtidos damos o destaque para os resultados do algoritmo de regressão linear, o qual se manteve em destaque em todas as métricas. Mostrando assim ser uma possível ferramenta para predizer casos de dengue na cidade de Recife.

## 5. Conclusões e Trabalhos Futuros

O controle epidêmico de arboviroses, como a dengue, é um dos desafios da saúde pública no Brasil. Mesmo sendo uma doença que possui um tratamento, ela ainda aparece com recorrência ano a ano no país.

Apresentamos no trabalho meios de prever o número de casos de um determinado mês utilizando os três meses anteriores a ele, e os meses correlatos dos anos anteriores. Realizamos também um comparativo da eficácia dos algoritmos por meio das métricas RMSE, R2 e EPMA.

Por meios dos resultados das métricas identificamos e destacamos o algoritmo de regressão linear. Notamos também, que os algoritmos em geral não apresentam resultados satisfatórios em acertar o número exatos de casos reais, mas como visto na

seção 4.1, as previsões conseguem apresentar uma curva semelhante com os casos reais de dengue.

Com os dados do trabalho, podemos fortalecer que por mais que esses modelos não acertem o número de casos, ele é capaz de alertar para o surgimento de picos de casos de dengue. Auxiliando assim os órgãos responsáveis a realizar tomadas de decisão.

Como trabalhos futuros pretendemos realizar o estudo agrupando os casos de dengue por meses, e correlacionar com períodos climáticos do ano. Avaliar também os surtos dos anos de 2015, 2016, e 2019, para assim propor um modelo para previsão de possíveis surtos de dengue.

### Referências

[1] WORLD HEALTH ORGANIZATION et al. Dengue: guidelines for diagnosis, treatment, prevention and control. World Health Organization, 2009.

[2] WORLD HEALTH ORGANIZATION. Dengue and dengue haemorrhagic fever - Fact sheet 117. WHO, 2012. Disponível em: <<http://www.who.int/mediacentre/factsheets/fs117/en/>>. Acesso em: 15 jul. 2020.

[3] LOPES, Nayara; NOZAWA, Carlos; LINHARES, Rosa Elisa Carvalho. Características gerais e epidemiologia dos arbovírus emergentes no Brasil. Revista Pan-Amazônica de Saúde, v. 5, n. 3, p. 10-10, 2014.

[4] Ministério da Saúde. Monitoramento dos casos de arboviroses urbanas transmitidas pelo Aedes (dengue, chikungunya e Zika), Semanas Epidemiológicas 1 a 34. Boletim Epidemiológico 22, 2019. Disponível em: <<https://portalarquivos2.saude.gov.br/images/pdf/2019/setembro/11/BE-arbovirose-22.pdf>>. Acesso em: 15 jul. 2020.

[5] SECRETARIA ESTADUAL DE SAÚDE. Boletim Arboviroses. Secretaria Executiva de Vigilância em Saúde de Pernambuco, 2019. Disponível em: <[https://12ad4c92-89c7-4218-9e11-0ee136fa4b92.filesusr.com/ugd/3293a8\\_08db449067b6449a9d360539079ac0a6.pdf](https://12ad4c92-89c7-4218-9e11-0ee136fa4b92.filesusr.com/ugd/3293a8_08db449067b6449a9d360539079ac0a6.pdf)>. Acesso em: 15 jul. 2020.

[6] MINISTÉRIO DA SAÚDE et al. Monitoramento dos casos de arboviroses urbanas transmitidas pelo Aedes Aegypti (dengue, chikungunya e zika), Semanas Epidemiológicas 1 a 13, 2020. Secretaria de vigilância em saúde, 2020. Disponível em: <<https://www.conasems.org.br/wp-content/uploads/2020/04/Boletim-epidemiologico-SVS-14.pdf>>. Acesso em: 11 jul. 2020.

[7] STOLERMAN, Lucas M.; MAIA, Pedro D.; KUTZ, J. Nathan. Forecasting dengue fever in Brazil: An assessment of climate conditions. PloS one, v. 14, n. 8, p. e0220106, 2019.

[8] FIGUEREDO, L. T. M. Dengue in Brazil: Past. Present and Future Perspectives. dengue Bulletin, v. 27, p. 25-33, 2003.

[9] FARES, Rafaelle CG et al. Epidemiological scenario of dengue in Brazil. BioMed research international, v. 2015, 2015.

[10] DA SILVA, Leandro Augusto; PERES, Sarajane Marques; BOSCARIOLI, Clodis. Introdução à mineração de dados: com aplicações em R. Elsevier Brasil, 2017.

[11] FAYYAD, Usama M. et al. (Ed.). Advances in knowledge discovery and data mining.

[12] VAPNIK, Vladimir. The nature of statistical learning theory. Springer science & business media, 2013.

[13] SILVA NETO, Sebastião Rogério da et al. Uma abordagem computacional para identificação de indício de preconceito em textos baseada em análise de sentimentos. 2017.

[14] ZANATY, E. A. Support vector machines (SVMs) versus multilayer perception (MLP) in data classification. Egyptian Informatics Journal, v. 13, n. 3, p. 177-183, 2012..

[15] LIMA, Edivania de Araújo; FIRMINO, Janne Lúcia da Nóbrega; GOMES FILHO, Manoel F. A relação da previsão da precipitação pluviométrica e casos de dengue nos estados de Alagoas e Paraíba nordeste do Brasil. Revista brasileira de meteorologia, v. 23, n. 3, p. 264-269, 2008..



**[16]** MARTINEZ, Edson Zangiacomi; SILVA, Elisângela Aparecida Soares da. Previsão do número de casos de dengue em Ribeirão Preto, São Paulo, Brasil, por um modelo SARIMA. *Cadernos de Saúde Pública*, v. 27, n. 9, p. 1809-1818, 2011.

**[17]** ANDERSSON, Virginia Ortiz; BIRCK, Marco A. Ferreira; ARAUJO, Ricardo Matsumura. Towards predicting dengue fever rates using convolutional neural networks and street-level images. In: 2018 International Joint Conference on Neural Networks (IJCNN). IEEE, 2018. p. 1-8.

**[18]** APPICE, Annalisa et al. A multi-stage machine learning approach to predict dengue incidence: a case study in Mexico. *IEEE Access*, v. 8, p. 52713-52725, 2020.

**[19]** HALIM, Siana et al. Dengue Fever Outbreak Prediction in Surabaya using A Geographically Weighted Regression. In: 2019 4th Technology Innovation Management and Engineering Science International Conference (TIMES-iCON). IEEE, 2019. p. 1-5.

**[20]** AZEVEDO, Ana Isabel Rojão Lourenço; SANTOS, Manuel Filipe. KDD, SEMMA and CRISP-DM: a parallel overview. *IADS-DM*, 2008.

**[21]** CHOVIATIYA, Megha et al. Prediction of dengue using recurrent neural network. In: 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI). IEEE, 2019. p. 926-929.

**[22]** EMPREL. Casos de Dengue 2018. Dados Recife, 2019. Disponível em: <<http://dados.recife.pe.gov.br/dataset/casos-de-dengue-zika-e-chikungunya/resource/d754bf5a-edbe-4a60-bca9-8c4252d223c8>>. Acesso em: 12 mar. 2020.