

## Previsão da taxa de mortalidade infantil no Ceará: uma abordagem com aprendizado de máquina<sup>a</sup>

*Prediction of the Child Mortality Rate in the State of Ceará: a Machine Learning Approach*

Manoel Alexandre de Lucena<sup>b</sup> 

Universidade Federal do Rio Grande do Norte, Departamento de Demografia e Ciências Atuariais, Natal (RN), Brasil

Marcos Antonio Carvalho Alves de Sousa Júnior<sup>c</sup> 

Universidade Regional do Cariri, Departamento de Economia, Crato (CE), Brasil

Eliane Pinheiro de Sousa<sup>d</sup> 

Universidade Regional do Cariri, Departamento de Economia, Crato (CE), Brasil

---

**Resumo:** A Taxa de Mortalidade Infantil (TMI) é um indicador crucial do bem-estar e do desenvolvimento regional. O Ceará vem reduzindo a TMI ao longo do tempo. Nesse contexto, este trabalho objetiva prever a TMI cearense de 2001 a 2023 utilizando os modelos Autoregressive Integrated Moving Average (ARIMA), Support Vector Regression (SVR), Decision Tree Regression (DTR), Random Forest Regression (RFR) e eXtreme Gradient Boosting (XGB). O número de óbitos a cada mil nascidos vivos para tal período foi extraído do IPECEDATA. Consi derou-se 75% da série para treinamento dos modelos e o restante para teste. Dentre esses, o XGB apresenta menor erro de previsão. Usando este algoritmo, constata-se, na projeção da TMI para 6 anos à frente, que esta taxa permanece nos mesmos patamares dos seis anos anteriores, com média de 11,5 óbitos a cada

---

Editor responsável: Felipe Orsolin Teixeira

<sup>a</sup> Submissão: 30/09/2024 | Aprovação: 28/02/2025 | DOI: 10.5380/re.v46i88.97156

<sup>b</sup> manoelalex123@gmail.com

<sup>c</sup> marcos.antonio@urca.br

<sup>d</sup> pinheiroeliane@hotmail.com

Contribuição individual de cada autor/a: (b) Concepção e análise dos dados e/ou documentos. (c) Pesquisa de dados e participação ativa na discussão dos resultados. (d) Revisão e aprovação da versão final.



Esta publicação está licenciada sob os termos de Creative Commons Atribuição 4.0 Internacional

mil nascidos vivos. Portanto, recomendam-se políticas públicas no Ceará, visando reduzir a TMI a médio e longo prazos.

**Palavras-chave:** Taxa de mortalidade infantil. Previsão de séries temporais. Modelos de aprendizado de máquina. Ceará.

**Abstract:** Child Mortality Rate is a crucial indicator of the well-being and the regional development. The Brazilian state of Ceará has been reducing the Child Mortality Rate over the years. In this context, this study aims to predict the Child Mortality Rate of Ceará from 2001 to 2023 by using the following models: Autoregressive Integrated Moving Average (ARIMA), Support Vector Regression (SVR), Decision Tree Regression (DTR), Random Forest Regression (RFR) and eXtreme Gradient Boosting (XGB). The number of deaths per 1,000 live births for this period was extracted from IPECEDATA [Institute of Economic Research and Strategy of Ceará]. It was considered 75% of the series for training of the models and the remainder for testing. Among these, XGB exhibits lesser prediction error. By using this algorithm, one may observe, in the projection of the Child Mortality Rate for the following 6 years, that this rate remains in the same level of the preceding six years, with an average of 11.5 deaths per 1,000 live births. Therefore, one recommends public policies in Ceará, aiming to reduce the Child Mortality Rate in the medium and long terms.

**Keywords:** Child Mortality Rate. Time Series Forecasting. Machine Learning Model. Ceará.

**JEL:** C53.

---

## 1. Introdução

A mortalidade infantil continua sendo um dos maiores desafios da saúde pública mundial (Frota; Hasegawa; Jacinto, 2024). De acordo com um relatório da Organização das Nações Unidas (ONU), milhões de crianças morrem anualmente devido a causas que são, em grande parte, evitáveis ou tratáveis, como pneumonia, diarreia, malária, parto prematuro e complicações no momento do nascimento. Tais óbitos representam óbitos que poderiam ser mitigados através da implementação de políticas públicas mais eficazes, e de um maior acesso a serviços de saúde básicos, como vacinações, apoio à amamentação precoce e contínua, e diagnóstico e tratamento de doenças infantis (UNICEF, 2023). Nesse sentido, a mortalidade infantil não apenas representa uma perda irreparável para as famílias afetadas, mas também reflete os níveis de saúde, de desenvolvimento socioeconômico e as condições de vida de uma localidade (Ministério da Saúde, 2021). Esses aspectos são reforçados por Mendonça (2023) e por Brito e Nhamuave (2022), que apontam a taxa de mortalidade infantil como um indicador essencial do desempenho econômico de um país, uma vez que evidencia não apenas as disparidades econômicas, mas também a natureza das desigualdades sociais.

Por ser um indicador fundamental para avaliar a qualidade de vida e as condições de saúde de um país, a redução da taxa de mortalidade infantil tem sido foco de esforços internacionais, a exemplo dos Objetivos de Desenvolvimento Sustentável (ODS) das Nações Unidas, que visam acabar com as mortes evitáveis de recém-nascidos e crianças menores de 5 anos até 2030 (ONU, 2024). Embora essa e outras iniciativas para reduzir a mortalidade infantil tenham gerado certos avanços, ela ainda é uma realidade presente em países em desenvolvimento e é considerado um grave problema de saúde mundial (Sanders *et al.*, 2017).

A Taxa de Mortalidade Infantil (TMI) é definida como o número de óbitos de menores de um ano de idade, a cada mil nascidos vivos, em determinado local, no período avaliado (Ministério da Saúde, 2021). Assim, o indicador calcula o risco de morte das crianças nascidas vivas antes de completarem um ano de idade. A mortalidade infantil abrange três categorias: mortalidade neonatal precoce, que ocorre entre 0 e 6 dias de vida; mortalidade neonatal tardia, entre 7 e 27 dias; e mortalidade pós-neonatal, que abrange o período de 28 a 364 dias após o nascimento (Rabello *et al.*, 2018).

Segundo o Instituto Brasileiro de Geografia e Estatística (IBGE), a Taxa de Mortalidade Infantil (TMI) do Brasil, em 2022, é de 12,59 óbitos por mil nascidos vivos. Historicamente, esse valor tem se reduzido, porém de forma desigual. Silva (2016) destaca que, apesar da redução do indicador em âmbito nacional, essa diminuição não ocorreu de forma uniforme entre as diferentes regiões do país. Essa desigualdade é, em grande medida, reflexo das características socioeconômicas das famílias, bem como do nível de desenvolvimento econômico local, que resulta diretamente em uma maior ou menor perpetuação da mortalidade infantil. Paixão e Ferreira (2012) atestam a redução acelerada da TMI entre os anos 1980 e 2000 e associam-na à atuação de políticas públicas no âmbito do saneamento básico, educação e acesso à saúde, a exemplo dos Programas de Atenção Integral à Saúde da Mulher (PAISM) e o Programa Nacional de Imunização (PNI). Para Sousa *et al.* (2016), a redução da TMI no período de 2001 a 2011 ocorreu de forma associada ao Programa Saúde da Família (PSF), bem como a outros fatores, como: melhoria na renda *per capita*, expansão dos serviços de saneamento básico e redução do índice de Gini (desigualdade de renda).

Ceccon *et al.* (2014) investigaram a associação estatística entre a cobertura da Estratégia Saúde da Família (ESF) e a mortalidade infantil, apontando que os óbitos em crianças menores de um ano, em sua maioria, poderiam ser evitados com as próprias ações realizadas na ESF. Apesar da ausência de relação significativa entre a cobertura da ESF e a redução da TMI, os autores apontaram a discrepância entre as taxas de mortalidade registradas nas Regiões Norte e Sul do Brasil; discrepância esta que pode ser influenciada pelas características geográficas dos Estados, pela extensão territorial do país e por desigualdades sociais históricas. Essa relação de desigualdade também foi citada por Paixão e Ferreira (2012), ao apontarem que os estados das regiões Norte e Nordeste historicamente registram as maiores TMI, por serem localidades onde os problemas sociais e econômicos são mais intensos, comparativamente, às demais regiões. Segundo o Boletim Epidemiológico do Ministério da Saúde, no período entre 1990 e 2019, a região Nordeste apresentou declínio na TMI de 75,8 para 15,2 óbitos por mil nascidos vivos (NV); no Norte, a redução foi de 45,9 para 16,6; no Sudeste, o declínio foi de 32,6 para 11,9; no Centro-Oeste, de 34,3 para 13,0; e no Sul, de 28,3 para 10,2. As maiores reduções ocorreram em estados nordestinos, com destaque para Alagoas (86%), Pernambuco (83%) e Ceará (83%).

No que tange ao Ceará, em 2023, segundo o Instituto de Pesquisa e Estratégia Econômica do Ceará (IPECE), a Taxa de Mortalidade Infantil era de 11,72 óbitos a cada mil nascidos vivos, ao passo que, em 2001, era 24,60 óbitos a cada mil nascidos vivos. Historicamente, o estado vem conquistando uma redução desse indicador, fato que tem sido objeto de estudos como em Mendonça (2023), Brito e Nhamuave (2022), Sanders *et al.* (2017), Feitosa *et al.* (2015), Lima *et al.* (2014), Sousa *et al.* (2012), Silva *et al.* (2010) e Bezerra Filho *et al.* (2007). Entretanto, mesmo com a redução da mortalidade infantil, há uma constante necessidade de melhorar as políticas públicas de assistência à saúde.

Uma vez destacada a importância da Taxa de Mortalidade Infantil, faz-se necessária analisar a trajetória histórica desse indicador a fim de fazer previsões, que servem de monitoramento por parte da sociedade civil e dos gestores públicos; já que, nos últimos anos, a previsão da TMI tem se tornado uma ferramenta importante para os formuladores de políticas públicas e profissionais de saúde. A capacidade de identificar tendências e áreas de alto risco possibilita intervenções preventivas e a alocação mais eficaz de recursos, que ocorrem por meio de políticas públicas. Nessa linha, estudos recentes foram realizados para prever a mortalidade infantil por meio da metodologia ARIMA (*Autoregressive Integrated Moving Average*), como, por exemplo, Melchior, Zanini, Oliveira (2024); Silva *et al.* (2021) e Chaib (2019), ou utilizando modelos de aprendizado de máquina, como Frota, Hasegawa e Jacinto (2024), Freitas (2023), Soares *et al.* (2021) e Aguiar (2019).

Conforme Costa *et al.* (2001), a mortalidade infantil é envolvida por fatores complexos e dinâmicos, apresentando constantes mudanças, o que torna seu monitoramento essencial para avaliar a situação de saúde local e orientar medidas de controle pertinentes. Nesse contexto, a previsão da TMI torna-se uma ferramenta essencial para gestores e formuladores de políticas públicas, permitindo a formulação de políticas públicas preventivas e educativas destinadas ao risco de um nascido vivo morrer durante o seu primeiro ano de vida no estado do Ceará. Qual seria, então, a metodologia mais precisa para realizar tal previsão?

Diante disso, considerando que o modelo de aprendizado de máquina seja um algoritmo que faz previsões baseadas no treinamento com dados históricos, este estudo busca identificar qual modelo de aprendizado de máquina apresenta maior eficácia na previsão da TMI do Ceará, analisando diferentes algoritmos de

aprendizado de máquina. A escolha do modelo mais adequado pode fornecer subsídios valiosos para a tomada de decisão, contribuindo para o desenvolvimento de estratégias preventivas e interventivas voltadas à redução desse indicador no estado.

Além desta introdução, o artigo apresenta em seguida uma revisão da literatura de estudos anteriores, seguido pela descrição da metodologia utilizada no presente estudo, bem como por uma análise dos resultados encontrados, e por fim, as considerações finais.

## **2. Revisão bibliográfica**

Na revisão bibliográfica apresentada a seguir, com efeito, apresentam-se duas subseções. A primeira trata de evidências empíricas acerca da TMI no Ceará, ao passo que a segunda recorre a pesquisas acerca da previsão desta taxa utilizando modelos ARIMA e de aprendizado de máquina.

### **2.1 Taxa de mortalidade infantil no Ceará**

Mendonça (2023) analisou a taxa e a tendência temporal da mortalidade infantil e dos componentes etários no Ceará e nas cinco Regiões de Saúde (Fortaleza, Norte, Cariri, Sertão Central e Litoral Leste), no período de 2010 a 2021. Trata-se de um estudo ecológico em série temporal, no qual foi utilizado o método da regressão linear. Os resultados apontaram uma tendência decrescente da taxa de mortalidade infantil (TMI) no Ceará, com maior frequência no período neonatal precoce. Brito e Nhamuave (2022) verificaram a tendência temporal da TMI na 7ª Região de Saúde do Ceará, pertencente à macrorregião do Litoral Leste e Jaguaribe, que compõe os municípios de Aracati, Icapuí, Fortim e Itaiçaba, no período de 2009 a 2019. Para tal, aplicaram o método de regressão linear simples. Os resultados confirmaram a tendência de queda da TMI no estado do Ceará, porém essa tendência não foi estatisticamente significativa na 7ª Região de Saúde do Ceará. Embora a TMI da região considerada tenha sido inferior à do Ceará nesse período investigado, não segue a tendência de redução do Estado, apontando para a necessidade de reavaliar as ações destinadas à prevenção da mortalidade infantil.

Sanders *et al.* (2017) analisaram os fatores relacionados à TMI no município de Fortaleza, capital cearense, no período de 2005 a 2010. Trata-se de um estudo

do tipo caso (óbitos) – controle (não óbitos), em que foi considerada a razão de três controles para cada caso, a fim de fornecer maior confiabilidade estatística aos dados. Utilizou-se a regressão logística, onde a variável dependente se referiu à ocorrência de óbito em menores de um ano, em forma dicotômica (sim; não), enquanto as explicativas corresponderam as variáveis sociodemográficas e socioeconômicas maternas; variáveis relacionadas à assistência na gestação e no parto; e variáveis relacionadas à criança ao nascer. Ademais, realizou-se também a regressão logística múltipla considerando apenas as variáveis significativas. Os resultados constataram que a baixa escolaridade materna afeta, fortemente, o nível socioeconômico, o acesso a serviços de saúde e a vulnerabilidade da mulher grávida, e do recém-nascido. Verificou-se que crianças cujas mães que tiveram um número de consultas de pré-natal inferior a quatro apresentaram risco duas vezes superior de mortalidade no primeiro ano de vida. O parto cesáreo apresentou-se como fator de proteção para o óbito infantil e o baixo peso ao nascer teve associação com o óbito infantil.

Feitosa *et al.* (2015) analisaram os fatores associados aos óbitos infantis nos municípios da Região Metropolitana do Cariri, no Sul do Ceará, no período de 2009 a 2013. Trata-se de um estudo transversal e descritivo, em que foram consideradas variáveis do recém-nascido, tais como sexo, raça/cor do bebê, peso ao nascer; variáveis de parte maternas, como escolaridade e tipo de gestação; e assistenciais (tipo de parto). Os resultados apontaram que, no quinquênio avaliado, os fatores da TMI estavam relacionados, em sua maioria, com mortes ocorridas em meninos, de cor parda, com peso ao nascer abaixo do normal e cujas mães eram jovens, com boa escolaridade e gestação única, de parto vaginal.

Lima *et al.* (2014) buscaram descrever o perfil epidemiológico dos óbitos infantis no Ceará entre 1996 e 2011 por meio de um estudo descritivo de evolução temporal. Os resultados mostraram um declínio lento nos valores do coeficiente de mortalidade, com maior predominância de óbitos em recém-nascidos com 28 a 36 semanas de gestação, sendo que as causas de óbito mais prevalentes ocorreram no período perinatal devido às malformações congênitas.

Sousa *et al.* (2012) debruçaram-se sobre o Programa Trevo de Quatro Folhas, um conjunto de ações de atenção materno-infantil implementado em 2001, no município cearense de Sobral. Esse programa é uma política pública permanente do município, que busca reduzir a mortalidade materna e infantil.

Além de salvar vidas, o programa se propõe a melhorar a qualidade dos vínculos entre profissionais de saúde, usuários e parceiros da rede para a promoção da saúde e cidadania.

Silva *et al.* (2010) buscaram avaliar o impacto do PSF sobre a TMI dos municípios cearenses. Para tal, consideraram como variável dependente a TMI e como variáveis explicativas o PSF, representadas por um vetor de *dummies* para captar a existência ou não de pessoas atendidas pelo programa no município; número de médicos residentes por mil habitantes; taxa de analfabetismo de pessoas com 15 anos ou mais; proporção de pessoas que vivem em domicílios com banheiro e água encanada; número médio de filhos nascidos vivos tido por uma mulher no final do seu período reprodutivo, residente em dado município; e renda ou Produto Interno Bruto per capita. Os resultados revelaram que as ações do PSF foram efetivas na redução da TMI dos municípios cearenses.

Bezerra Filho *et al.* (2007), por sua vez, visaram identificar os padrões de distribuição espacial dos componentes neonatal e pós-neonatal da TMI dos municípios cearenses e áreas de risco elevado, assim como discutir os principais fatores socioeconômicos, demográficos e de atenção à saúde que contribuíram para uma dependência espacial destes componentes. Para tal, utilizaram médias trienais para o período de 2000 a 2002. Trata-se de um estudo ecológico transversal, com o emprego de regressão linear múltipla, cuja análise espacial desses componentes foi obtida por meio do índice de Moran. Os resultados apontaram que a assistência à gravidez, ao parto e ao recém-nascido, acrescidas de uma melhor distribuição de renda, são decisivas para a sobrevivência no primeiro mês de vida, enquanto má alimentação, imunização, saneamento, educação e condição econômica são possíveis determinantes da mortalidade pós-neonatal.

## **2.2 Previsão da taxa de mortalidade infantil aplicando modelos ARIMA e de aprendizado de máquina**

Melchior, Zanini, Oliveira (2024) compararam a evolução das taxas de mortalidade infantil no estado do Rio Grande do Sul para o período de 2000 a 2017 por meio da modelagem de séries temporais utilizando as metodologias ARIMA e ARMA. Os resultados indicaram que o modelo ARMA (4,6) registrou a melhor qualidade de ajuste considerando os critérios de informação Akaike Information Criterion (AIC) e Bayesian Information Criterion (BIC), porém as melhores

medidas de acuracidade foram obtidas no modelo SARIMA. Ao comparar os dois modelos, constataram que a diferença dos valores previstos e originais foram muito próximos para ambos.

Silva *et al.* (2023) estimaram as taxas de mortalidade nas primeiras 24 horas de vida e por causas evitáveis no estado de Pernambuco no período de 2000 a 2021 por meio da metodologia ARIMA. Os resultados apontaram que o modelo escolhido para a mortalidade nas primeiras 24 horas foi o ARIMA(1,1,1)(2,0,0) e para a mortalidade por causas evitáveis nas primeiras 24 horas foi o ARIMA(1,1,2)(1,0,0). Portanto, os modelos ARIMA apresentaram estimativas satisfatórias para as taxas de mortalidade e por causas evitáveis nas primeiras 24 horas de vida.

Silva *et al.* (2021) discutiram os aspectos conceituais e metodológicos da série temporal por meio da modelagem ARIMA e a sua aplicabilidade na mortalidade infantil no Brasil, considerando o período de 2000 a 2018.

Chaib (2019) apresentou a trajetória da TMI no estado de São Paulo, no período de 1996 a 2016, e realizou a previsão do índice para os cinco primeiros meses de 2017. Para tal, utilizou o modelo ARIMA, uma vez que este explica uma variável a partir de valores passados. Os resultados revelaram comportamento sazonal da mortalidade infantil, já que registrou quedas em períodos específicos, logo identificou um modelo SARIMA. A previsão realizada evidenciou tendência de queda da taxa.

A literatura tem destacado estudos que tratam de modelagens de aprendizado de máquina, sobretudo, utilizando classificadores, ou seja, em que se analisa quais fatores afetam a probabilidade de sobrevivência ou mortalidade infantil no Brasil. Alguns destes estudos são examinados a seguir.

Frota, Hasegawa e Jacinto (2024) utilizaram análise de sobrevivência com modelos de aprendizado de máquina, como *Survival Support Vector Machines*, *Random Survival Forest* e *eXtreme Gradient Boosting* (XGB) para prever quais bebês têm o maior risco de não sobreviver o primeiro ano de vida no Brasil e os fatores de risco associados à mortalidade. Os resultados indicaram que variáveis como semanas de gestação, baixo peso e cesariana interagem de forma não linear no impacto da mortalidade. Constataram que os modelos de aprendizado de máquina são eficientes e podem ser incorporados a uma estrutura política, podendo mitigar a mortalidade infantil.

Freitas (2023) desenvolveu uma análise geral de fatores que influenciam na identificação da mortalidade infantil. Para tal, empregou modelos de aprendizado de máquina, tais como o *XGB*, *Decision Tree* e o *Random Forest*, para prever a mortalidade infantil no Ceará, considerando dados de 2020. Verificou que o classificador *XGB* obteve resultados mais equilibrados para as métricas de precisão de sensibilidade em comparação aos outros classificadores.

Soares *et al.* (2021) investigaram os dados de mortalidade infantil nos estados de Amapá e Santa Catarina, respectivamente, com a maior e a menor TMI dentre os estados brasileiros, com relação ao ano de 2016, a partir dos algoritmos *C4.5*, *RIPPER* (*Repeated Incremental Pruning to Produce Error Reduction*), *Random Forest*, *SVM* (*Support Vector Machine*) e *RNA* (Redes Neurais Artificiais). Constataram que os algoritmos *Random Forest*, *SVM* e *RNA* obtiveram os melhores resultados, com valores muito semelhantes e que peso, idade e escolaridade da mãe, gestação e quantidade de filhos vivos foram atributos comuns no *ranking* dos dois estados analisados.

Aguiar (2019) desenvolveu modelos preditivos de mortalidade infantil usando algoritmos de inteligência artificial e mineração de dados no estado do Ceará para o período de 2013 a 2017. Utilizando algoritmos de Redes Neurais em 20 variáveis das características de 8.159 óbitos infantis e de 10.235 nascidos vivos que sobreviveram ao primeiro ano de vida, o estudo obteve acurácia média de 97,41% de acertos com margem de erro de +/- 0,41% na predição do futuro.

### **3. Metodologia**

Esta seção é dedicada à metodologia do presente trabalho. Neste sentido, a primeira subseção apresenta a descrição dos modelos empregados, ao passo que a segunda subseção se dedica ao detalhamento do percurso da aplicação de tais modelos.

#### **3.1 Modelos analíticos**

##### **3.1.1 Autoregressive Integrated Moving Average (ARIMA)**

O processo *Autoregressive Integrated Moving Average* (ARIMA), também conhecido como metodologia de Box-Jenkins, é a modelagem tradicional de séries temporais, composta por um processo autorregressivo (AR), um processo de média

móvel (MA) e um processo de integração (I). Particularmente, considerando uma série temporal,  $y_t$ , no período  $t$ , afirmar que  $y_t$  é um processo autorregressivo, implica dizer que seu valor em dado período depende dos seus valores nos períodos anteriores, acrescido de um termo de erro. Assim, para  $p$  defasagens, escreve-se o processo AR( $p$ ). Quando a série apresenta média e variância constante, tem-se um processo AR( $p$ ) estacionário.

Os valores da série temporal também podem ser influenciados pelos erros das previsões anteriores. Este processo é denominado de Média Móvel (MA) (Agustin, 2019). Assim, escreve-se MA( $q$ ) para um processo de média móvel de ordem  $q$ . A união entre os processos AR( $p$ ) e MA( $q$ ) é denominada de ARMA( $p,q$ ). Contudo, para a estimação do ARMA( $p, q$ ) normalmente exige-se que a série seja estacionária. Um procedimento comum para obter a estacionariedade da série é a diferenciação.

Assim, o modelo ARIMA considera também um parâmetro  $d$ , que indica quantas vezes a série precisa ser diferenciada. Como destacado por Gujarati (2011), uma grande quantidade das séries temporais econômicas não é estacionária, logo, precisa ser diferenciada para assim se tornarem. Assim, o modelo ARIMA ( $p,d,q$ ) é descrito, respectivamente, pela ordem da parte autorregressiva ( $p$ ), pelo número de vezes que uma dada série deve ser diferenciada antes de tornar-se estacionária ( $d$ ) e pela ordem da parte de média móvel ( $q$ ).

### 3.1.2 *Support Vector Regression (SVR)*

O modelo *Support Vector Regression* (SVR), ou Regressão por Vetor de Suporte, é uma técnica de aprendizado de máquina supervisionada utilizada para tarefas de regressão. Ela é uma extensão do *Support Vector Machine* (SVM), diferenciando-se, porém, por ser utilizada para problemas de regressão, e não de classificação, como a SVM (Drucker, 1997).

O SVR, como o nome sugere, funciona a partir de vetores de suportes, que são os pontos mais próximos da linha de regressão. A finalidade desse modelo é gerar vetores de suporte com valores mínimos de erro e que separem os dados da forma mais adequada possível. Tal processo ocorre através de um processo de otimização que encontra o hiperplano ótimo, aquele com maior margem de separação e que comete poucos erros marginais (Lorena; Carvalho, 2003).

Particularmente, a otimização da SVR depende da escolha adequada dos hiperparâmetros, como, por exemplo, o erro de tolerância ( $\epsilon$ ), uma função de transformação do tipo *kernel*, e a constante de regularização ( $C$ ) (Lucena, 2024).

Para aplicação da SVR em problemas de séries temporais, é necessário adaptar as variáveis de entrada e saída do modelo. Neste caso, a *target*, ou seja, a série que se deseja prever, é  $y_t$ , ao passo que as variáveis independentes correspondem as defasagens da própria série temporal, que pode ser definida por  $x_t = (y_{t-1}, \dots, y_{t-p})$ , isto é, tem-se  $p$  defasagens (Cunha Júnior, 2022).

### 3.1.3 *Decision Tree Regression (DTR)*

O modelo *Decision Tree Regressor* (DTR), ou Regressão de Árvore de Decisão, é uma técnica de aprendizado de máquina supervisionada muito utilizada para a modelagem preditiva. É assim chamada porque seus algoritmos utilizam a estrutura (fluxograma) análoga a de uma árvore para modelar as relações entre as variáveis e os possíveis resultados advindos entre elas. Nesse sentido, cada nó representa uma condição baseada numa característica dos dados, e cada ramo representa o resultado dessa relação. À medida que os dados são trabalhados, portanto, novos ramos e nós vão se constituindo. Apesar da conceituação aparentemente simples, o DTR é um modelo poderoso e capaz de encontrar relações complexas não lineares nos dados (Géron, 2021).

A ramificação da árvore acontece à medida que os nós de decisão se estabelecem, até serem atingidos os *leaf nodes*. Segundo Gomboski (2019), esse processo termina quando uma das seguintes situações é atingida: (1) todos os resultados dos nós pertencem à mesma classe; (2) nenhuma outra característica pode ser diferenciada entre os resultados atingidos pelos nós; (3) a árvore atingiu o limite pré-estabelecido. A previsão no modelo DTR é feita com base nos valores anteriores - que retornam até o chamado nó raiz - e são avaliados antes do próximo nó ser feito.

### 3.1.4 *Random Forest Regression (RFR)*

O *Random Forest Regressor* (RFR) ou Regressão de Floresta Aleatória é uma técnica de aprendizado de máquina supervisionada que combina várias árvores de decisão em um único modelo. O modelo também se baseia no método

*Ensemble* (normalmente do tipo *bagging*), ou seja, é um modelo que funciona a partir de um aglomerado de outros modelos a fim de realizar previsões com maior acurácia (Géron, 2021). Cada árvore da floresta é formada por um subconjunto diferente de dados e é treinada de forma aleatória. A previsão final é baseada na média das previsões de todas as árvores da floresta. A construção de múltiplas árvores de decisão a partir do método de *Ensemble Learning* pode possibilitar a obtenção de resultados robustos, que muitas vezes não seriam obtidos por meio de modelos individuais.

Para o treinamento do RFR, alguns hiperparâmetros são utilizados para o ajuste do modelo, podendo-se destacar: *n\_estimators* (define o número de árvores de decisão que será utilizado no modelo); *max\_depth* (define a profundidade máxima das árvores); *max\_features* (define o número máximo de características que o modelo utilizará para realizar uma divisão; e *min\_samples\_split* (determina o número mínimo de amostras necessárias para dividir um nó) (Lucena, 2024).

### 3.1.5 *eXtreme Gradient Boosting (XGB)*

O algoritmo *eXtreme Gradient Boosting (XGB)* é uma técnica de aprendizado de máquina pertencente ao grupo de algoritmos *boosting*. Esse algoritmo faz parte dos métodos *ensemble* (modelos combinados) e opera a partir de um treinamento sequencial de previsores, que se corrigem sequencialmente (Géron, 2021). O termo *eXtreme* deriva de várias melhorias e otimizações que tornam o XGB mais eficiente e poderoso em comparação com implementações anteriores do algoritmo *boosting*.

A técnica de XGB pode ser utilizada tanto para problemas de classificação quanto de regressão. O XGB constrói vários previsores fracos que corrigem o seu antecessor, até atingir um ponto de parada; cada novo previsor é ajustado aos erros residuais do preditor que o precede. A construção de previsores mais fortes a partir da agregação de outros mais fracos pode ser feita, por exemplo, por meio de árvores de decisão.

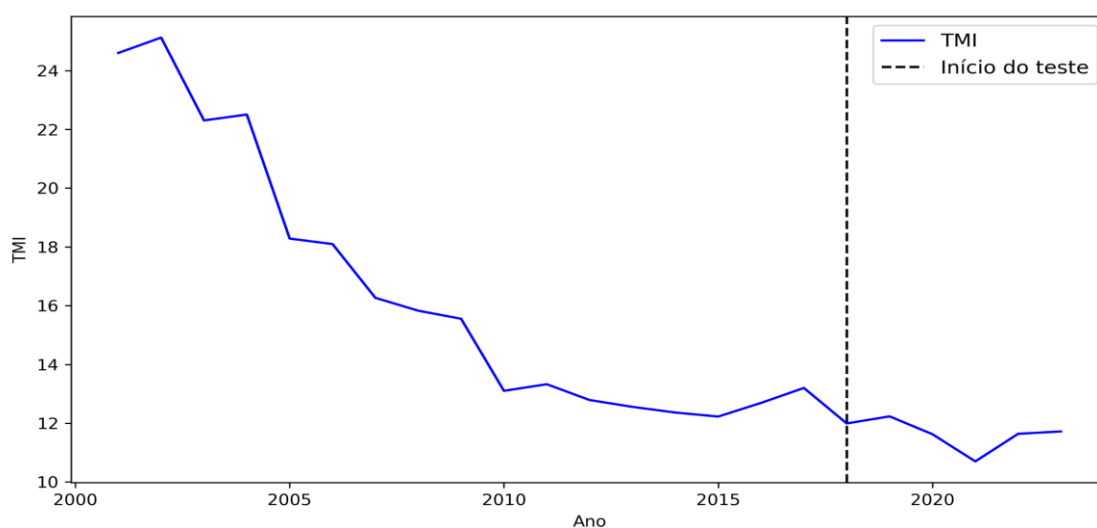
O processo de ajuste do XGB ocorre por meio de otimização de hiperparâmetros. Particularmente, alguns hiperparâmetros mais comuns e considerados são: *learning\_rate* (define o quanto cada árvore de decisão impacta na previsão final); *n\_estimators* (define o número de árvores do modelo);

*max\_depth* (controle a profundidade de cada árvore); *subsample* (define a porcentagem de dados utilizada para gerar cada árvore) (Lucena, 2024).

### 3.2 Percurso metodológico

Para modelar e prever a Taxa de Mortalidade Infantil (TMI) do Ceará no período de 2001 a 2023, extraiu-se a TMI, dada em número de óbitos a cada mil nascidos vivos, do IPECEDATA. Em seguida, o conjunto de dados foi dividido em dois subconjuntos: treinamento e teste. O conjunto de treinamento correspondeu a 75% da série e o restante foi alocado para testar os modelos estimados. A Figura 1 apresenta o gráfico da série da TMI, com a separação entre treinamento e teste.

**Figura 1 – Série histórica da TMI do Ceará considerando a divisão entre treinamento e teste**



Fonte: Lucena, Sousa Júnior e Sousa (2024).

Na etapa seguinte, de conjunto de treinamento, aplicou-se a padronização de escala, em que, de cada observação, realiza-se a subtração de seu valor pela média e divide-se pelo desvio padrão. Ademais, neste mesmo procedimento, aplicou-se a função *StandardScaler* da biblioteca *sklearn* na linguagem *Python*. Em seguida, utilizou-se o conjunto de treinamento para o ajuste dos modelos.

Inicialmente, o ARIMA foi ajustado aplicando a função *auto\_arima* da biblioteca *pmdarima.arima* (Smith *et al.*, 2017). Para tornar os modelos de

aprendizado de máquina autoregressivos, utilizou-se os lags do modelo ARIMA. Desta forma, foram utilizadas as defasagens com entradas dos modelos e a série temporal da TMI com variável dependente. Ademais, no ajuste dos modelos de aprendizado de máquina, utilizou-se otimização de hiperparâmetros. Particularmente, empregou-se um otimizador do tipo *grid search*, o qual busca testar todas as combinações de hiperparâmetros em um espaço de busca definido para determinar a combinação que leva ao melhor desempenho do modelo (Ogunsanya; Isichei; Desai, 2023).

Além disso, a otimização com *grid search* é combinada com a técnica de *Cross Validation* (CV), denominada de *Grid Search Cross Validation* (GSCV). A técnica de GSCV divide o conjunto de treinamento em diversos subconjuntos, em que são treinados e avaliados modelos em tais subamostras, sendo selecionado o melhor modelo (Adnan *et al.*, 2022). Assim, a função *GridSearchCV* da biblioteca *sklearn* (Pedregosa *et al.*, 2011) foi utilizada neste estudo. Esta técnica também foi aplicada no estudo de Aguiar (2019) para a previsão da mortalidade infantil no Ceará. Neste estudo, para todos os modelos, utilizou-se CV igual a 3. A Tabela 1, por seu turno, apresenta a distribuição dos hiperparâmetros utilizados nos modelos de aprendizado de máquina.

**Tabela 1 – Distribuição dos hiperparâmetros utilizados nos modelos de aprendizado de máquina**

Hiperparâmetro	SVR	DTR	RFR	XGB
<i>C</i>	[0, 1, 1,5]			
<i>epsilon</i>	[0, 01, 0,05, 0, 10]			
<i>kernel</i>	[ <i>rbf</i> , <i>sigmoid</i> , <i>poly</i> ]			
<i>gamma</i>	[0, 05, 0,10, 0, 25]			
<i>max_depth</i>		[2, 4, 7, 9]	[2, 4, 9, 15]	[11, 18, 27]
<i>min_samples_split</i>		[2, 9, 11]	[2, 7, 9]	
<i>max_features</i>		[2,4,6,8]	[2, 3, 4, 7]	
<i>n_estimators</i>			[3, 10, 45]	[10, 25, 50]
<i>learning_rate</i>				[0, 1, 0,25, 0,5]
<i>subsample</i>				[0, 25, 0,75]

Fonte: Lucena, Sousa Júnior e Sousa (2024).

Assim, após o ajuste dos modelos, foi realizada a previsão considerando o horizonte do conjunto de teste. Este estudo adotou a técnica de previsão recursiva, a qual é realizada um passo à frente, considerando o valor imediatamente anterior da série. Nas previsões, aplicou-se a padronização reversa para reverter o procedimento aplicado no conjunto de treinamento. Em seguida, foram calculadas métricas de desempenho para escolha do melhor modelo para a previsão da TMI do Ceará. Neste estudo, foram utilizadas o MAE (Erro Absoluto Médio, do inglês, *Mean Absolute Error*), RMSE (Raiz do Erro Quadrático Médio, do inglês, *Root Mean Squared Error*), apresentadas nas equações (1) e (2), respectivamente.

$$MAE = \frac{1}{N} \sum_{t=1}^N \hat{y}_t - y_t \quad (1)$$

$$RMSE = \sqrt{\quad} \quad (2)$$

Em que:  $N$  diz respeito ao tamanho da série temporal;  $y_t$  é o valor da série no instante  $t$ ;  $\hat{y}_t$  é valor previsto no instante  $t$ .

De forma geral, quanto menores os valores do MAE e RMSE, mais adequado é o modelo para prever a série temporal da TMI no Ceará.

#### 4. Resultados

Em conformidade com os procedimentos metodológicos, realizou-se a divisão em dois conjuntos após a extração da série temporal da TMI do IPECEDATA, sendo o primeiro com 75% da série para o treinamento dos modelos e o segundo para o teste. Aplicando-se a transformação de escala no conjunto de treinamento, realizou-se a modelagem da série utilizando os modelos ARIMA, SVR, DTR, RFR e XGB. A distribuição dos parâmetros de modelos supracitados é apresentada na Tabela 2.

**Tabela 2 – Distribuição dos hiperparâmetros dos modelos ajustados no conjunto de treinamento para a previsão da TMI do Ceará (2001-2023)**

Modelo	Hiperparâmetros
ARIMA	ARIMA(2, 1, 0), $ar_1 = -0,0205$ , $ar_2 = 0,6930$
SVR	$C = 5$ , $\varepsilon = 0,01$ , $\gamma = 0,05$ , $kernel = rbf$
DTR	$max_{depth} = 7$ , $max_{features} = 2$ , $min = 2$
RFR	$max_{depth} = 9$ , $max_{features} = 2$ , $min = 2$ , $n_{estimators} = 10$
XGB	$learning_{rate} = 0,25$ , $max_{depth} = 11$ , $subsample = 0,75$ , $n_{estimators} = 50$

Fonte: Lucena, Sousa Júnior e Sousa (2024).

Particularmente, no caso do ARIMA, encontrou-se um modelo com ordem autoregressiva igual a 2 e nenhum componente de médias móveis. Desta forma, para os algoritmos de aprendizado de máquina, utiliza-se como entrada a série histórica da TMI com duas defasagens, ao passo que a saída (*target*) é a série no instante  $t$ . Recorrendo à Tabela 2, observa-se que a distribuição dos hiperparâmetros dos modelos de aprendizado de máquina de árvores é próxima (DTR e RFR). A maior profundidade dos algoritmos baseados em árvores

(*max\_depth*), bem como o maior número de modelos (*n\_estimators*), foi utilizado pelo XGB.

Na Tabela 3, são apresentadas as métricas de erros dos modelos de previsão da TMI do Ceará no período de 2001 a 2023. Conforme se observa, em todas as medidas consideradas o XGB apresenta menor erro (MAE e RMSE iguais a 0,8772 e 0,9843, respectivamente). Desta forma, o melhor modelo para a previsão da TMI do Ceará é o XGB. Ressalta-se que o XGB é um *ensemble* do tipo *boosting*, ou seja, cada modelo é treinado sequencialmente buscando modelar os erros dos modelos anteriores. Particularmente, a literatura destaca a performance de modelagens combinadas para a previsão em diferentes tipos de problemas (Firmino; Mattos Neto; Ferreira, 2014; Géron, 2022). Especificamente, para a mortalidade infantil, a pesquisa de Rodríguez *et al.* (2022) destaca o XGB na previsão na Região Metropolitana do Vale do Paraíba e Litoral Norte, no estado de São Paulo. Ademais, o estudo de Freitas (2023) destaca a acurácia do XGB, neste caso, porém, aplicada à classificação da mortalidade infantil.

Após o XGB, pode-se destacar o desempenho da SVR. Neste sentido, no estudo de Santos e Carvalho (2017), a SVR superou o modelo de regressão linear na previsão da TMI do Paraná, no Brasil. Na mesma direção, Junhui *et al.* (2009) constataram melhor desempenho da SVR, comparada a *Radial Basis Function Artificial Neural Network* (RBFNN), na previsão da mortalidade infantil de menores de 5 anos em Sichuan, na China.

**Tabela 3 – Métricas de erros no conjunto de teste dos modelos de previsão da TMI do Ceará (2001-2023)**

Modelo	MAE (por mil nascidos vivos)
ARIMA	2,4784
SVR	0,9198
DTR	0,9856
RFR	1,1757
XGB	0,8772

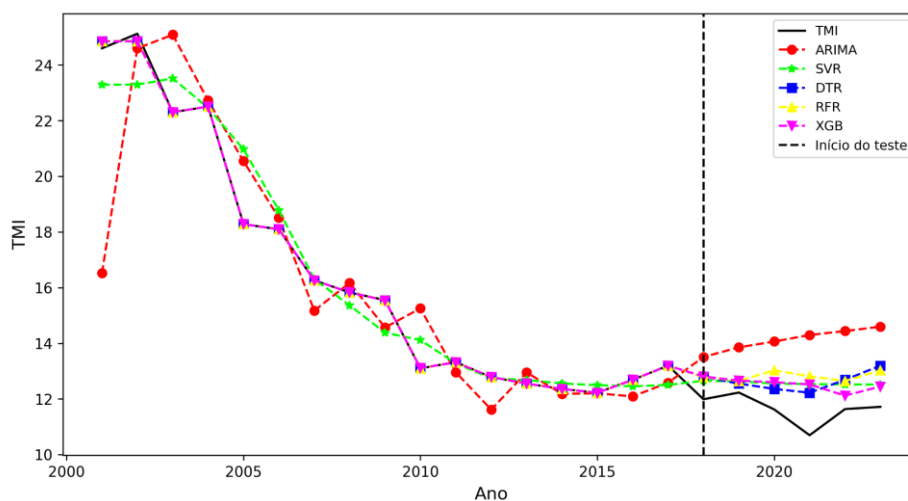
Fonte: Lucena, Sousa Júnior e Sousa (2024).

Em seguida, destacam-se o DTR e RFR, nesta ordem. Iqbal *et al.* (2023), aplicaram modelos de aprendizado de máquina como classificadores na modelagem da mortalidade infantil para o Paquistão, e encontraram DTR como sendo o melhor modelo, em termos de exatidão e precisão nas previsões. Em último lugar, destaca-se o modelo ARIMA. Vale ressaltar que o modelo ARIMA é tradicionalmente a abordagem de séries temporais mais empregada para as tarefas de modelagem e previsão de séries históricas de diversos fenômenos. Particularmente, diversas pesquisas têm realizado previsões da TMI aplicando modelos da família ARIMA, como por exemplo, Chaib (2019) em São Paulo, no Brasil; Raziq *et al.* (2021) no Paquistão; Ogundunmade, Daniel e Awwal (2023) na Nigéria.

Na Figura 2, apresenta-se as previsões realizadas para os modelos nos conjuntos de treinamento e teste. Inicialmente, é pertinente pontuar as divergências observadas nas previsões para ambos os conjuntos supracitados. Notadamente, as previsões realizadas para o treinamento tendem a acompanhar a série temporal, ao passo que, nos testes, não se verifica o mesmo padrão. Com efeito, portanto, o comportamento da série histórica difere-se ao longo do tempo, afetando os modelos treinados. Ademais, no presente estudo, utilizam-se séries pequenas, o que pode afetar o desempenho dos modelos, conforme aponta Lucena (2024).

Todavia, como se observa, no conjunto destinado à avaliação dos preditores, os modelos tiveram resultados próximos, destacando o XGB que se aproximou da série modelada no final do conjunto de testes. Este resultado, portanto, ratifica as métricas de erros da Tabela 2. Assim, a projeção da TMI foi realizada com a XGB.

**Figura 2 – Previsão da TMI do Ceará nos conjuntos de treinamento e de teste**



Fonte: Lucena, Sousa Júnior e Sousa (2024)

Na Tabela 3, são apresentadas as projeções para a TMI no período de 2024 a 2029, bem como a média e o desvio padrão desta projeção. Comparando o primeiro ano da projeção, 2024, com o último ano, (2029), observa-se um crescimento desta taxa em 2,47%. A média para este período é de 11,52 óbitos a cada mil nascidos vivos, valor este próximo àqueles observados na série no período de 2018 a 2023.

Neste sentido, constata-se um quadro diferente da literatura revisada. Em outros termos, mantidas as condições observadas na série estudada, a projeção revela um rompimento da tendência de declínio da TMI no Ceará. Na contramão, os estudos desenvolvidos por Brito e Nhamuave (2022) e Mendonça (2023) relataram essa tendência de declínio. Essas inferências podem sinalizar uma perspectiva de desalinhamento de políticas destinadas à mitigação deste tipo de óbito. Conforme destacam Brito e Nhamuave (2022), estes óbitos podem ser evitados através do fortalecimento das linhas de cuidado materno-infantil.

**Tabela 3 – Projeção da TMI do Ceará realizada com XGB (2024-2029)**

Ano	TMI
2024	11,4508
2025	11,4413
2026	11,4244
2027	11,4208
2028	11,6651
2029	11,7340
Média	11,5227
Desvio padrão	0,1391

Fonte: Lucena, Sousa Júnior e Sousa (2024)

É necessário pontuar que a TMI já apresenta uma trajetória de incremento nos anos finais da série, saindo de 11,62 para 11,71, em 2018 e 2023, respectivamente. Na mesma direção, a projeção indica um leve incremento da TMI nos últimos dois anos. O aumento da TMI pode estar relacionado com diversos fatores, como, por exemplo, a desigualdade social que particularmente aumentou nos últimos anos, especialmente em função dos efeitos de crises econômicas e sociais. Nestes termos, a literatura corrobora a afirmação de que a desigualdade social aumenta a mortalidade infantil (Paixão; Ferreira, 2012; Ceccon *et al.*, 2014).

Adicionalmente, pode-se pontuar a relevância de tais previsões, no âmbito dos modelos de aprendizado de máquina. A este respeito, Frota, Hasegawa e Jacinto (2024) destacam a eficiência deste tipo de abordagem no que diz respeito ao auxílio de políticas para a mitigação da TMI. Com efeito, os algoritmos estudados podem ser considerados para novas projeções, bem como para acompanhamento deste indicador social.

## 5. Considerações finais

A Taxa de Mortalidade Infantil (TMI), definida como o número de óbitos em menores de um ano de idade, a cada mil nascidos vivos, é um importante indicador de desenvolvimento socioeconômico de uma região. Neste sentido, a

previsão desta taxa é fundamental para a formulação de políticas públicas com vistas a reduzi-la.

Utilizando medidas de desempenho para verificar o rol de modelos empregados, observou-se que o XGB apresentou os menores valores para o MAE e RMSE. Desta forma, o melhor algoritmo para previsão da série temporal da TMI do Ceará foi o XGB. Com o emprego deste modelo, a projeção da TMI foi realizada considerando o horizonte de 6 anos à frente. A projeção supracitada revelou que a TMI deve se manter nos patamares do período anterior, com média de 11,52 óbitos a cada mil nascidos vivos, e variabilidade baixa entre os anos. Ademais, comparando o primeiro com o último ano da projeção, verifica-se um incremento de 2,47%.

Em conclusão, verificou-se que a metodologia de aprendizado de máquina aplicada a séries temporais é útil na modelagem da TMI. Nesta direção, os modelos possibilitam a projeção desta taxa, o que fornece importantes *insights*. Em primeiro lugar, os algoritmos ajustados e selecionados constituem uma ferramenta para o acompanhamento da TMI ao longo do tempo. Com base nisso, gestores públicos e formuladores de políticas podem traçar ações com vistas a reduzir, a médio e longo prazo, a mortalidade infantil no Ceará.

Na mesma direção, considerando a constatação de um leve incremento na TMI do Ceará, bem como a relevância desta taxa em termos de indicador social, evidencia-se que é crucial a formulação de estratégias públicas para mitigação da taxa de mortalidade. Dessa forma, os algoritmos de aprendizado de máquina utilizados neste estudo são ferramentas indispensáveis. Ademais, ressalta-se as preocupações das agendas internacionais, como, por exemplo, da ONU, em termos de redução dos indicadores de mortalidade infantil.

A principal limitação deste estudo reside no fato de considerar apenas a série temporal da TMI, sem a inclusão de variáveis exógenas que influenciam tal taxa. Desta forma, para estudos futuros, sugere-se incluir outras variáveis que influenciam na TMI. Além disso, outras modelagens, como, por exemplo, redes neurais artificiais, podem ser incluídas visando complementar as previsões. Por último, a ampliação da área de estudo considerando outros estados brasileiros além dos já pesquisados pode ser útil para ampliar esta discussão em nível nacional.

## Referências

ADNAN, M. *et al.* Utilizing grid search cross-validation with adaptive boosting for augmenting performance of machine learning models. *PeerJ Computer Science*, v. 8, p. e803, 2022.

AGUIAR, W. S. Desenvolvimento de modelos preditivos de mortalidade infantil com base em inteligência artificial no estado do Ceará. Tese (Doutorado em Saúde Coletiva). Universidade Federal do Ceará, Fortaleza, 163f. 2019.

AGUSTIN, I. N. The approach of box jenkins time series analysis for predicting stock price on Iq45 stock index. *PROFIT: Journal Administrasi Bisnis*, v. 13, n. 1, p. 18-25, 2019.

BEZERRA FILHO, J. G. *et al.* Distribuição espacial da taxa de mortalidade infantil e principais determinantes no Ceará, Brasil, no período 2000-2002. *Cadernos de Saúde Pública*, v. 23, n. 5, p. 1173-1185, mai. 2007.

BRITO, A. F. S.; NHAMUAVE, E. A. Análise da mortalidade infantil de uma Região de Saúde do Ceará. *Cadernos ESP*, v. 16, n. 4, p. 10-18, 2022. DOI: 10.54620/cadesp.v16i4.825

CECCON, R. F. *et al.* Mortalidade infantil e Saúde da Família nas unidades da Federação brasileira, 1998-2008. *Cadernos Saúde Coletiva*, v. 22, n. 2, p. 177-183, abr. 2014.

CHAIB, D. C. A mortalidade infantil no estado de São Paulo: uma previsão da taxa por meio da modelagem SARIMA. *Revista de Economia da UEG*, v. 15, n. 1, p. 44-52, 2019.

COSTA, M.C.N.; AZI, P.A.; PAIM, J.S.; SILVA, L.M.V. Mortalidade infantil e condições de vida: a reprodução das desigualdades sociais em saúde na década de 90. *Cadernos de Saúde Pública* 2001, v. 17, n. 3, p. 555-567, mai-jun. 2001.

CUNHA JÚNIOR, R. O. Previsão do nível de água subterrânea usando ensemble de modelos de séries temporais. 2022. 150 f. Dissertação (Mestrado em Desenvolvimento Regional Sustentável). Universidade Federal do Cariri, Crato, 2022.

DUARTE, C. M. R. Reflexos das políticas de saúde sobre as tendências da mortalidade infantil no Brasil: revisão da literatura sobre a última década. *Cadernos de Saúde Pública*, v. 23, n. 7, p. 1511-1528, jul. 2007.

DRUCKER, H. *et al.* Support vector regression machines. *Advances in neural information processing systems*, v. 9, p. 155-161, 1997.

FEITOSA, A. C. *et al.* Fatores associados à mortalidade infantil na região metropolitana do Cariri, Ceará, Brasil. *Journal of Human Growth and Development*. Disponível em:

[\\_https://www.revistas.usp.br/jhgd/article/view/103019/103061](https://www.revistas.usp.br/jhgd/article/view/103019/103061). Acesso em: 15 jul. 2024.

FERREIRA, C. M. M. O impacto do saneamento na taxa de mortalidade infantil: uma abordagem dos investimentos da CAGECE nos municípios do Ceará – 1997 a 2001. Dissertação (Mestrado em Economia). Universidade Federal do Ceará. Fortaleza, 2004.

FIRMINO, P. R. A.; MATTOS NETO, P. S. G.; FERREIRA, T. A. E. Correcting and combining time series forecasters. *Neural Networks*, v. 50, p. 1-11, 2014.

FREITAS, D. S. Análise e predição de mortalidade infantil utilizando modelos de aprendizado de máquina. TCC (Graduação em Engenharia de Software). Universidade Federal do Ceará. Quixadá, 2023.

FROTA, L. M.; HASEGAWA, M.; JACINTO, P. Infant Mortality in Brazil: a survival analysis using Machine Learning Models. In: Encontro de Economia da Região Sul, 27, 2024. Anais... Maringá: ANPEC Sul, 2024. Disponível em: [https://www.anpec.org.br/sul/2024/submissao/files\\_I/i2-8b13f93ab3f365942738eb3c03669caa.pdf](https://www.anpec.org.br/sul/2024/submissao/files_I/i2-8b13f93ab3f365942738eb3c03669caa.pdf). Acesso em: 13 jul. 2024.

GÉRON, A. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media, Inc., 2022.

GOMBOSKI, M. A utilização de algoritmos de machine learning na análise econômica. 2020. 65p. Trabalho de Conclusão de Curso (Graduação em Economia). Universidade Federal do Rio Grande do Sul, Porto Alegre, 2019.

GUJARATI, D. N.; PORTER, D. C. *Econometria Básica*. Porto Alegre: Bookman, 2011.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. Cidades e Estados: Panorama.

Disponível em: <https://cidades.ibge.gov.br/brasil/panorama>. Acesso em: 18 jul. 2024.

IQBAL, F. *et al.* Predictive analytics in smart healthcare for child mortality prediction using a machine learning approach. *Open Life Sciences*, v. 18, n. 1, p. 20220609, 2023.

JESUS DA SILVA, J. Os determinantes socioeconômicos da mortalidade infantil nos anos de 2000 e 2010: evidências empíricas para os municípios pernambucanos. *Humanas Sociais & Aplicadas*, v. 6, n. 17, 2016. DOI: 10.25242/88766172016584.

LIMA, H. M. P. *et al.* Análise da mortalidade infantil no Ceará: Evolução de 1996 a 2011. *Revista Brasileira de Pesquisa em Saúde*, v. 16, n. 3, p. 58-65, jul.-set., 2014.

LORENA, A. C.; CARVALHO, A. C. P. L. F. Introdução às máquinas de vetores suporte (support vector machines). São Carlos: ICMC-USP. 2003. Disponível em: <[https://repositorio.usp.br/directbitstream/a7ed198b-f6a3-4cec-b132-7e113bd51424/BIBLIOTECA\\_113\\_RT\\_192.pdf](https://repositorio.usp.br/directbitstream/a7ed198b-f6a3-4cec-b132-7e113bd51424/BIBLIOTECA_113_RT_192.pdf)>. Acesso em: 3 ago. 2024.

LUCENA, M. A. Modeling and forecasting small national eco-efficiency time series. 2024. 134 f. Dissertação (Mestrado em Desenvolvimento Regional Sustentável). Programa de Pós-graduação em Desenvolvimento Regional Sustentável (PRODER). Universidade Federal do Cariri, Crato, 2024.

MELCHIOR, C. Análise da evolução das taxas de mortalidade infantil de crianças com até um ano de idade residentes no estado do Rio Grande do Sul. 38 p. 2021. Monografia (Especialização em Estatística e Modelagem Quantitativa). Universidade Federal de Santa Maria, Santa Maria, 2021.

MENDONÇA, E. L. A. Mortalidade infantil no estado do Ceará: um estudo ecológico de série temporal de 2010 a 2021. Dissertação (Mestrado em Saúde Coletiva). Universidade de Fortaleza. Fortaleza, 2023.

MINISTÉRIO DA SAÚDE (MDS). Boletim Epidemiológico. Edição 2021. Disponível em: [https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/boletins/epidemiologicos/edicoes/2021/boletim\\_epidemiologico\\_svs\\_37\\_v2.pdf](https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/boletins/epidemiologicos/edicoes/2021/boletim_epidemiologico_svs_37_v2.pdf). Acesso em: 18 jul. 2024.

MINISTÉRIO DA SAÚDE (MDS). Indicadores de mortalidade. 2021. Disponível em: <http://tabnet.datasus.gov.br/cgi/idb2000/fqc01.htm>. Acesso em: 18 jul. 2024.

OGUNSANYA, M.; ISICHEI, J.; DESAI, S. Grid search hyperparameter tuning in additive manufacturing processes. *Manufacturing Letters*, v. 35, p. 1031-1042, 2023.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS. Objetivos de Desenvolvimento Sustentável. Disponível em: <https://brasil.un.org/pt-br/sdgs>. Acesso em: 30 jan. 2025.

PAIXÃO, A. N.; FERREIRA, T. Determinantes da Mortalidade Infantil no Brasil. *Informe Gepec*, v. 16, n. 2, p. 6-20, 2012.

PEDREGOSA, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825-2830, 2011.

RABELLO, D.; VINHAL, L.; MIRANDA, M.; SILVA, M.; PORTO, D.; TEIXEIRA, R.; LAMARE, T. de; MARINHO, F. Análise descritiva da mortalidade materna e na infância no Brasil, 2007 a 2016. *Consensus*, v. 28, n. 28, p. 1-6, 2018.

RODRÍGUEZ, E. *et al.* A Data-Driven Approach for Neonatal Mortality Rate Forecasting. In: IDDM. 2022. p. 86-98.

ROMA, J. C. Os objetivos de desenvolvimento do milênio e sua transição para os objetivos de desenvolvimento sustentável. *Ciência e Cultura*, v. 71, n. 1, jan./mar. 2019.

SANDERS, L. S. C. *et al.* Mortalidade infantil: análise de fatores associados em uma capital do Nordeste brasileiro. *Cadernos Saúde Coletiva*, v. 25, n. 1, p. 83-89, jan. 2017.

SILVA, A. B. S. *et al.* Modelo Autorregressivo Integrado de Médias Móveis (ARIMA): aspectos conceituais e metodológicos e sua aplicabilidade na mortalidade infantil. *Revista Brasileira de Saúde Materno Infantil*, v. 21, n. 2, p. 657-666, abr.-jun., 2021. DOI: 10.1590/1806-93042021000200016.

SILVA, A. B. S. *et al.* Análise temporal da mortalidade por causas evitáveis nas primeiras 24 horas de vida, 2000-2021. *Revista Latino-Americana de Enfermagem*, e4080, 2023. DOI: 10.1590/1518-8345.6696.4080.

SILVA, V. H. M. C. *et al.* Uma avaliação econômica do Programa de Saúde da Família sobre a taxa de mortalidade infantil no Ceará. Texto para Discussão, n. 86. Fortaleza: IPECE, 2010.

SMITH, T. G. *et al.* *pmdarima*: ARIMA Estimators for Python. 2017. Disponível em: <http://www.alkaline-ml.com/pmdarima>. Acesso em: 29 ago. 2024.

SOARES, W. L. G. *et al.* Caracterizando a mortalidade infantil utilizando técnicas de Machine Learning: um estudo de caso em dois estados brasileiros - Santa Catarina e Amapá. *Brazilian Journal of Development*, v.7, n.5, p. 45269-45290 mai. 2021.

SOUSA, J. S. *et al.* Estimacão e análise dos fatores determinantes da redução da taxa de mortalidade infantil no Brasil. *Revista Brasileira de Estudos Regionais e Urbanos*, v. 10, n. 2, p. 140-155, 2016.

Disponível em: <https://revistaaber.org.br/rberu/article/view/126/182> . Acesso em: 18 jul. 2024.

SOUSA, S. F. J. *et al.* Programa Trevo de Quatro Folhas: uma ação efetiva para a redução da mortalidade infantil em Sobral - Ceará. *SANARE*, v.11. n.1, p. 60-65, jan./jun. 2012.

Disponível em: <https://sanare.emnuvens.com.br/sanare/article/view/268/241>. Acesso em: 16 jul. 2024.

UNICEF. Mortalidade infantil atinge mínima histórica em 2022, diz relatório da ONU. 2023. Disponível em: <https://www.unicef.org/brazil/comunicados-de-imprensa/mortalidade-infantil-atinge-minima-historica-em-2022-relatorio-da-onu>. Acesso em: 30 jan. 2025.