

# Análise Temporal e Predição de Nutricionistas no Ceará: Abordagens em Ciência de Dados para Suporte à Gestão em Saúde

Nator Junior Carvalho da Costa  
Escola de Saúde publica do Ceará  
Faculdade SENAC Ceará  
Fortaleza, Ceará, Brasil  
natorjuniorcc@gmail.com

Igor Dayan Barbosa Lucas  
Escola de Saúde publica do Ceará  
Fortaleza, Ceará, Brasil  
igor.dayan@outlook.com

Caio Gabriel Paiva Vale  
Faculdade SENAC Ceará  
Fortaleza, Ceará, Brasil  
caiogabriel.pv@hotmail.com

Domingos Bruno Sousa Santos  
Universidade Estadual do Ceará  
Fortaleza, Ceará, Brasil  
domingosbruno4764@gmail.com

Ryan Cavalcante de Lima  
Faculdade SENAC Ceará  
Fortaleza, Ceará, Brasil  
ryancavalcante.rc@gmail.com

Larissa Torres Ferreira  
Faculdade SENAC Ceará  
Fortaleza, Ceará, Brasil  
larissa\_tf@hotmail.com

## ABSTRACT

A ciência da nutrição ganhou destaque no início do século XX, sendo considerada um fenômeno recente no Brasil. A formação de profissionais de nutrição iniciou-se na década de 1940, com a criação dos primeiros cursos de graduação profissionalizante e o surgimento dos médicos nutricionistas. Em outubro de 2024, o Ceará contabilizava 1.623 nutricionistas registrados na base de dados do CNES, totalizando 2.028 vínculos, este trabalho apresenta uma análise temporal com o objetivo de quantificar o estado atual da distribuição dos profissionais de nutrição no estado do Ceará e suas perspectivas futuras, considerando especialidades e áreas de atuação. Os dados foram organizados em tabelas em que foi possível avaliar as flutuações dos vínculos empregatícios dos profissionais de saúde no estado em diferentes períodos. Adicionalmente, foram testados diferentes modelos de aprendizado de máquina para prever a evolução dos vínculos de nutricionistas ao longo do tempo. O modelo *Exponential Smoothing* destacou-se como a abordagem mais precisa, apresentando os menores valores de erro, como MAE de 30,87 e MAPE de 2,1%. Esses resultados demonstram a eficácia desse modelo para capturar padrões temporais e sazonais na força de trabalho. Os achados reforçam a relevância de ferramentas preditivas no suporte à tomada de decisão estratégica, permitindo aos gestores de saúde otimizar a alocação de recursos, reduzir desigualdades regionais e planejar políticas públicas fundamentadas em evidências.

## KEYWORDS

Força de trabalho em Saúde; Ciência de Dados; Gestão em saúde

## 1 INTRODUÇÃO

A Sociedade Brasileira de Nutrição foi fundada em 1940, e com ela, a criação dos Arquivos Brasileiros de Nutrição (1944) e a fundação da Associação Brasileira de Nutricionistas (ABN) (1949) [1]. Nesse ponto a Nutrição inicia seu processo de institucionalização no Brasil enquanto ciência, área de política pública e campo profissional. [1]

Com o Parecer nº 265, de 19 de outubro de 1962, o curso de nutrição como formação em nível superior teve seu devido reconhecimento através do Conselho Federal de Educação (CFE). [2]

A presença do nutricionista nos serviços de saúde contribuem de maneira significativa, atuando diretamente junto a indivíduos,

famílias e comunidade; na participação de ações educacionais continuada de profissionais de saúde; na articulação de estratégias de ação, em prol da promoção da alimentação saudável, do Direito Humano à Alimentação Adequada e da Segurança Alimentar e Nutricional.[3]

Com o aumento contínuo na geração e disponibilização de dados, torna-se essencial o desenvolvimento de ferramentas que ofereçam instrumentos eficazes para a gestão e avaliação em saúde, possibilitando uma tomada de decisão mais embasada e estratégica [4]. A presença do nutricionista contribui para a qualificação das ações de atenção nutricional auxiliando as pessoas em todas as fases do curso da vida a desenvolverem práticas alimentares adequadas, com a devida assistência a indivíduos e grupos populacionais [5]. Esta contribuição do profissional nutricionista está prevista na Resolução do Conselho Federal de Nutricionistas (CFN) 600/2018 e na Política Nacional de Alimentação e Nutrição (PNAN)15 [5, 6].

Nesse sentido, é possível sugerir que o nutricionista representa uma categoria profissional com contribuições técnicas científicas significativas para a organização da atenção nutricional, desde a atenção primária até a atenção especializada hospitalar, que incluem a promoção da saúde e da alimentação adequada e saudável, a vigilância alimentar e nutricional e ações de prevenção e tratamento de doenças em todas as fases da vida [7].

A fim de orientar-nos de maneira mais eficaz em relação às questões em aberto e às propostas de solução neste trabalho, elencamos a seguir três perguntas de pesquisa:

- **RQ1:** Qual é o estado atual da distribuição geográfica e demográfica dos profissionais de nutrição no estado do Ceará, considerando especialidades e áreas de atuação?
- **RQ2:** Quais padrões ou tendências de vínculo profissional podem ser observados nas séries temporais dos profissionais de nutrição ?
- **RQ3:** De que forma essa ferramenta pode auxiliar gestores na tomada de decisões estratégicas, otimizando o planejamento e a alocação de recursos?

O documento está organizado em cinco seções. A Seção 2 apresenta os trabalhos relacionados, oferecendo uma visão geral das pesquisas existentes na área. Na Seção 3, são detalhadas a metodologia aplicada, a descrição da base de dados utilizada e as métricas

adotadas para validação dos modelos. A Seção 4 apresenta os resultados obtidos após a aplicação da proposta, discutindo os principais achados e análises realizadas. Por fim, a Seção 5 consiste na conclusão, destacando as contribuições do estudo e sugerindo direções futuras.

## 2 TRABALHOS RELACIONADOS

Na literatura, existem estudos que abordam ferramentas e análises sobre a presença e a evolução dos profissionais de nutrição. A seguir, elencamos algumas contribuições relevantes que enriquecem a discussão e servem de base para o desenvolvimento deste trabalho.

O trabalho de [7] estudou a tendência temporal da inserção de nutricionistas no Sistema Único de Saúde (SUS) entre os anos de 2009 e 2018. Nesse período, constatou-se que em 2009 existiam apenas 14.502 nutricionistas no país, enquanto em 2018 esse número subiu para 30.226 profissionais, demonstrando um crescimento significativo na força de trabalho. Embora os estados do Sudeste apresentem a maior concentração de nutricionistas, as regiões Nordeste e Centro-Oeste registraram os maiores crescimentos proporcionais, com aumentos de 3,5 vezes e 2,3 vezes, respectivamente. Em contrapartida, a região Norte apresentou o menor crescimento, de apenas 1,8 vezes no mesmo período, refletindo desigualdades regionais na distribuição desses profissionais.

No trabalho de [8] analisou a variação no número de cursos de graduação em nutrição no Brasil entre os anos de 2000 e 2010. O estudo evidenciou um crescimento expressivo na oferta de cursos, com destaque para a região Nordeste, que apresentou um incremento considerável. Nesse intervalo, o Nordeste obteve 757 novos cursos, o que representou um aumento significativo em comparação com outras regiões, reforçando sua relevância para a formação de novos profissionais. Já a região Centro-Oeste obteve incremento de 267%, a Sudeste de 223% e a Sul de 146%. A região norte ficou com os menores índices de nutricionistas em comparação as outras regiões, porém apresentou um incremento com variação de 700%, refletindo esforços para atender à crescente demanda por nutricionistas em áreas antes menos atendidas.

Além desses estudos, [9] realizou uma análise histórica sobre a evolução da profissão de nutricionista no Brasil, destacando a criação dos primeiros cursos de graduação nos anos 1940 e os marcos legais que regulamentaram a profissão, como o parecer nº 265 de 1962, que reconheceu a formação em nutrição como curso de nível superior. Essa análise destaca como os avanços na formação e regulamentação contribuíram para o crescimento do campo.

[10] apresentam uma revisão sistemática sobre sistemas de alerta precoce para epidemias e pandemias baseados em Inteligência Artificial (IA). Estes sistemas podem ser aliados da gestão em saúde. O estudo analisa 33 trabalhos recentes, destacando o uso de técnicas de aprendizado de máquina e modelos preditivos para a detecção de surtos de doenças como COVID-19, Dengue e Malária. Os autores identificam desafios como a qualidade dos dados, viés nos modelos e falta de padronização. Concluem que a integração de dados ambientais, sociais e clínicos, aliada à transparéncia e mitigação de viés, é fundamental para aprimorar a eficácia desses sistemas.

Esses trabalhos destacam a relevância de estudos voltados para a distribuição e formação de nutricionistas, bem como o papel de métodos analíticos avançados na formulação de políticas públicas.

Eles reforçam a necessidade de estratégias integradas para equilibrar a distribuição dos profissionais no Brasil e melhorar a cobertura e a qualidade dos serviços prestados.

## 3 METODOLOGIA

O primeiro passo da metodologia envolve a coleta de dados relacionados aos profissionais de nutrição que atuam no estado do Ceará. Essas informações são extraídas do Cadastro Nacional de Estabelecimentos de Saúde (CNES), uma plataforma que disponibiliza dados detalhados sobre os profissionais vinculados a instituições de saúde [11–14].

O Cadastro Brasileiro de Ocupações (CBO) é um documento oficial que lista todas as ocupações reconhecidas no mercado de trabalho brasileiro [15, 16]. Sua função principal é identificar e classificar as ocupações dos trabalhadores, servindo como referência para registros administrativos, com destaque para os Registros Administrativos do Ministério da Previdência e Assistência Social (MPAS) [17–19].

### 3.1 Análise estática

Nesta seção, será descrito o processo para a realização da análise estática. Essa abordagem permite examinar características fixas, ou seja, aquelas que podem ser delimitadas e analisadas de forma a identificar e distinguir aspectos gerais de uma determinada competência.

A identificação dos profissionais é realizada utilizando o Cadastro Nacional de Saúde (CNS), que funciona como um identificador único para os usuários em todas as competências avaliadas. Para trazer maior clareza ao processo e às regras envolvidas, listamos abaixo algumas diretrizes importantes para compreender os resultados e as análises que serão apresentadas a seguir.

No contexto da análise referente ao estado do Ceará, conforme descrito na Equação 1, os profissionais que atuam em mais de uma cidade são contabilizados apenas uma vez no total estadual. Essa abordagem evita duplicações e assegura que os índices estaduais refletem o número exato de profissionais ativos, independentemente da quantidade de vínculos em diferentes municípios. Seja o conjunto  $S_j = \{\{CNS_1\}, \{CNS_2\}, \dots, \{CNS_{184}\}\}$  um município do estado contendo todos os vínculos ativos, cada conjunto  $S_i$  contém os profissionais de saúde identificados no município  $j$ .

$$N_{\text{total}} = \left| \bigcup_{j=1}^{184} S_j \right| \quad (1)$$

$N_{\text{total}}$  representa o número total de profissionais de saúde contabilizados em todos os municípios incluídos na análise.  $\bigcup_{i=1}^{184} S_j$  simboliza a união dos conjuntos de profissionais de saúde de cada um dos 184 municípios do estado do Ceará.  $\left| \bigcup_{j=1}^{184} S_j \right|$  denota a cardinalidade da união desses conjuntos, ou seja, o número total de elementos distintos na união. Esse valor representa o número total de profissionais de saúde únicos em todo o estado.

Para a análise municipal, cada cidade em que o profissional possui vínculo é considerada separadamente, resultando em uma contagem individual para cada localidade. Ao distinguir os vínculos por município, é possível avaliar de forma detalhada a distribuição

geográfica dos profissionais, permitindo identificar diferenças regionais e atender às necessidades específicas de cada cidade.

$$N_j^{\text{profissional}} = \left| \bigcup_{i=1}^m \text{CNS}_i \right| \quad (2)$$

Na equação 2,  $j$  representa o município em questão, enquanto  $\text{CNS}_i$  refere-se a um vínculo profissional, denotado por  $i$ . O termo  $\bigcup_{i=1}^m$  denota a união dos conjuntos unitários dos vínculos de profissionais de nutrição ( $\text{CNS}_i$ ) vinculados a cada município  $j$ , em que  $m$  é o número total de vínculos profissionais do município.

Por fim,  $\left| \bigcup_{i=1}^m \text{CNS}_i \right|$  indica a cardinalidade da união dos conjuntos, ou seja, o número total de elementos únicos nos conjuntos, garantindo que cada profissional seja contado uma única vez no município, mesmo que possua múltiplos vínculos.

Essa estrutura permite uma análise clara e detalhada da distribuição dos profissionais em cada cidade, promovendo a individualização das contagens e possibilitando a identificação de variações regionais com maior precisão.

Na Equação 3, é apresentada a forma de contabilização dos vínculos, distinguindo-os da contagem de indivíduos. Diferentemente dos indivíduos, um mesmo profissional pode possuir múltiplos vínculos. Assim, para cada município  $j$ , somam-se todos os vínculos atribuídos aos profissionais registrados nesse município.

$$N_j^{\text{vinculo}} = \sum_i \text{CNS}_i \quad (3)$$

As formulações descritas nas Equações 1 e 2 permitem uma análise precisa da distribuição dos profissionais em nível municipal e estadual, assegurando a individualidade de cada registro e promovendo maior clareza na avaliação das diferenças regionais e nas necessidades locais.

Essas regras simplificam o processo de contagem e categorização dos profissionais, assegurando uma abordagem consistente e padronizada na análise dos dados, o que contribui para maior precisão e clareza nos resultados apresentados.

A Equação 4 calcula a medida relativa da presença de profissionais de saúde em relação à população total, normalizada por dez mil habitantes.  $N_j^{\text{relativo}}$  representa a relação entre o número de profissionais vinculados e a população total do município  $j$ .

$$N_j^{\text{relativo}} = \frac{N_j^{\text{profissional}}}{pop_j} \cdot 10000 \quad (4)$$

- $N_j^{\text{relativo}}$  é a medida relativa que representa a relação entre o número de profissionais vinculados e a população total no município  $j$ .
- $N_j^{\text{profissional}}$  denota o número de profissionais de saúde identificados pelo Cadastro Nacional de Saúde (CNS) vinculados ao município  $j$ .
- $pop_j$  representa a população total do município  $j$ .
- A multiplicação por 1000 tem como objetivo normalizar a medida, expressando-a em termos de mil habitantes, o que facilita a interpretação e a comparação entre municípios.

### 3.2 Modelos

O *Exponential Smoothing* é uma técnica amplamente utilizada para previsão em séries temporais [20–22]. Ele atribui pesos exponencialmente decrescentes aos valores históricos, permitindo capturar padrões de nível, tendência e sazonalidade de forma eficiente. Conforme é descrito na Equação 1. Os modelos AutoARIMA, SARIMA e Prophet são ferramentas populares para previsão de séries temporais. O AutoARIMA automatiza a identificação dos parâmetros ideais de um modelo ARIMA (*AutoRegressive Integrated Moving Average*), otimizando os componentes de autoregressão, integração e média móvel para ajustar os dados de forma eficiente [23].

$$S_t = \alpha \cdot Y_t + (1 - \alpha) \cdot S_{t-1} \quad (5)$$

Em que:

- $S_t$ : Valor suavizado ou previsão no instante  $t$ .
- $Y_t$ : Valor observado no instante  $t$ .
- $\alpha$ : Fator de suavização ( $0 < \alpha < 1$ ).

### 3.3 Métricas

As métricas utilizadas para avaliar os modelos incluem o erro absoluto médio (MAE), erro quadrático médio (MSE), raiz do erro quadrático médio (RMSE) e erro absoluto percentual médio (MAPE), definidos pelas seguintes equações:

O MAE, descrito na Equação 6 mede a média das diferenças absolutas entre os valores observados ( $y_i$ ) e os valores previstos ( $\hat{y}_i$ ) pelo modelo. É uma métrica que avalia o quanto próximo o modelo está dos dados reais, sem considerar o sinal do erro (positivo ou negativo) [24, 25].

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (6)$$

Em que  $n$  representa o número total de observações,  $y_i$  é o valor observado, e  $\hat{y}_i$  corresponde ao valor predito pelo modelo. O *Mean Squared Error* (MSE) é uma métrica amplamente utilizada para avaliar o desempenho de modelos de regressão [26–29]. Ele mede a média dos quadrados dos erros, conforme demonstrado na Equação 7 onde cada erro é definido como a diferença entre o valor real ( $y_i$ ) e o valor predito pelo modelo ( $\hat{y}_i$ ). Essa métrica penaliza erros maiores de forma mais severa devido à elevação ao quadrado, tendo portanto, maior sensibilidade aos outliers [30].

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (7)$$

O *Root Mean Squared Error* (RMSE), demonstrado na Equação 8 é uma métrica amplamente utilizada para avaliar o desempenho de modelos preditivos [31], especialmente em problemas de regressão. Ele é definido como a raiz quadrada da média dos quadrados dos erros, em que cada erro corresponde à diferença entre o valor real ( $y_i$ ) e o valor predito pelo modelo ( $\hat{y}_i$ ).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (8)$$

O *Mean Absolute Percentage Error* (MAPE), descrito na Equação 9 é uma métrica amplamente utilizada para avaliar o desempenho de

modelos preditivos, especialmente em contextos onde a escala dos dados pode variar [32–34]. Ele é definido como a média percentual do erro absoluto, calculado como a diferença entre o valor real ( $y_i$ ) e o valor predito ( $\hat{y}_i$ ), dividido pelo valor real, e multiplicado por 100 para expressar o erro em termos percentuais.

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (9)$$

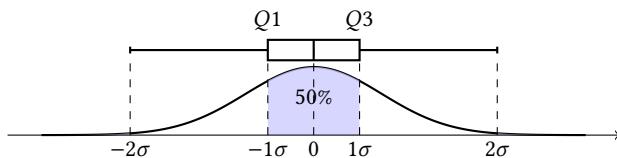
O erro residual é definido como a diferença entre o valor real ( $y_i$ ) e o valor predito pelo modelo ( $\hat{y}_i$ ) para uma determinada observação, conforme descrito na Equação 10.

$$ErroResidual = y_i - \hat{y}_i \quad (10)$$

Os resíduos fornecem informações importantes sobre a qualidade do ajuste do modelo aos dados [35]. Idealmente, os resíduos devem ser distribuídos de maneira aleatória em torno de zero, sem apresentar padrões ou tendências sistemáticas.

### 3.4 Visão de Violin e Boxplot

A visualização de violino (*violin plot*) e o *boxplot* são ferramentas complementares para a análise de distribuições de dados [36]. O gráfico de violino combina a funcionalidade do *boxplot* com uma estimativa de densidade. Ele apresenta a dispersão e simetria, destacando picos e caudas que podem não ser evidentes em um *boxplot* tradicional. A integração dessas duas abordagens fornece uma análise mais rica [37–39]. O *boxplot* resume os dados com cinco medidas principais:



Relação entre Boxplot e Curva de Densidade

Em que o  $Q_1$  representa o primeiro quartil,  $Q_2$  a mediana e  $Q_3$  o terceiro quartil. O intervalo interquartil (IQR) é dado por:

$$IQR = Q_3 - Q_1 \quad (11)$$

Na equação 12, é demonstrado o IQR, que é um intervalo usado para identificar *outliers*, que são valores discrepantes em relação à distribuição dos dados [36]. Os *outliers* são definidos como aqueles que estão fora do intervalo interquartil ampliado por 1,5 vezes o seu tamanho.

$$f(x) = \begin{cases} \text{Outlier,} & \text{se } x < Q_1 - 1.5 \cdot IQR \\ \text{Outlier,} & \text{se } x > Q_3 + 1.5 \cdot IQR \\ \text{Não Outlier,} & \text{caso contrário.} \end{cases} \quad (12)$$

Já o gráfico de violino expande essa análise, incluindo a estimativa de densidade  $f(x)$ , com base no núcleo gaussiano:

$$f(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad (13)$$

Em que  $K$  é o núcleo gaussiano,  $h$  é o parâmetro de suavização e  $n$  é o número de observações. O gráfico de violino, portanto, combina

o resumo do *boxplot* com o KDE, resultando em uma visualização para compreender as medidas estatísticas e a forma completa da distribuição.

$$\text{Violin Plot} = \text{Boxplot} + \text{Kernel Density Estimation (KDE)} \quad (14)$$

## 4 RESULTADOS E DISCUSSÃO

Nesta seção, são apresentados os resultados obtidos e uma discussão das questões de pesquisa inicialmente formuladas. A primeira questão de pesquisa (RQ1) busca compreender: **Qual é o estado atual da distribuição geográfica e demográfica dos profissionais de nutrição no estado do Ceará, considerando especialidades e áreas de atuação?** Essa análise é fundamental para identificar padrões de concentração e lacunas na distribuição dos profissionais, bem como para traçar o perfil demográfico e as principais áreas de atuação no estado. O que pode contribuir para a compreensão das desigualdades regionais e dos desafios na oferta de serviços.

Em outubro de 2024, o Ceará contabilizava 1.623 nutricionistas registrados na base do CNES, totalizando 2.028 vínculos. Essa diferença ocorre porque um mesmo profissional pode atuar em mais de uma unidade ou cidade, resultando em múltiplos vínculos sendo registrados.

Outra questão importante é que os dados do CNES refletem apenas os vínculos ativos, sejam eles públicos ou privados, e não representam o número total de nutricionistas formados no estado.

Nas Figuras 1 e 2, é possível observar os municípios com maior número de vínculos de nutricionistas (Top 10). Destacam-se Fortaleza, com 824 vínculos (40,6%); Sobral, com 117 (5,8%); Juazeiro do Norte, com 54 (2,7%); e Caucaia, com 50 (2,5%).

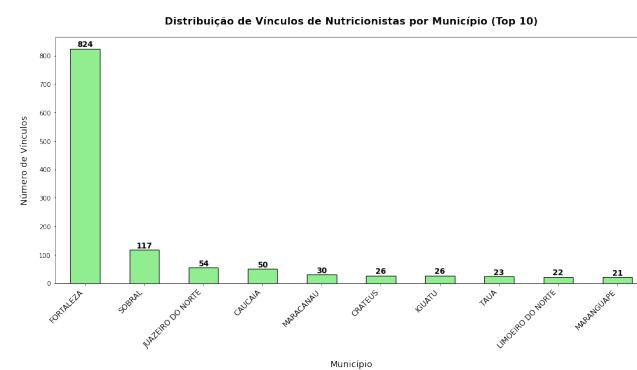


Figure 1: Nutricionistas por Município (Top 10)

Embora a análise de números absolutos forneça uma visão geral importante, ela carece de proporcionalidade à população local, o que limita a interpretação da real cobertura dos profissionais de nutrição em cada município. Por essa razão, projetamos o mesmo Top 10 ajustado para números relativos à população, proporcionando uma perspectiva mais equitativa da distribuição dos vínculos por habitantes. Na Figura 3, os números estão normalizados por 10 mil habitantes. Nesse contexto, destaca-se o município de Varjota, com a maior proporção, registrando 11 nutricionistas para cada 10 mil habitantes, evidenciando uma cobertura significativa em relação à sua população.

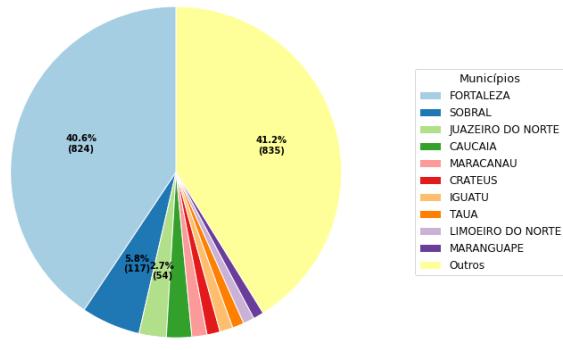


Figure 2: Nutricionistas por Município (Top 10+Outros)

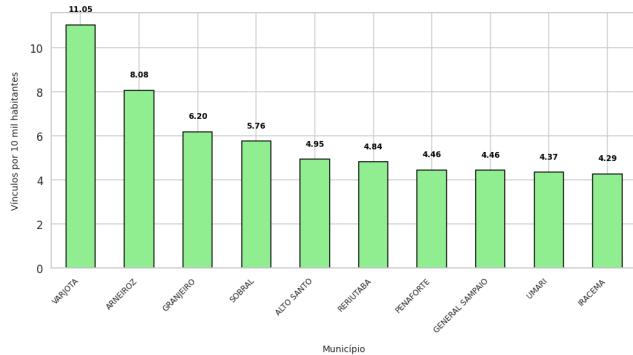


Figure 3: Nutricionistas por 10 mil habitantes

Outra questão importante abordada neste trabalho refere-se à identificação de padrões e tendências nas séries temporais da quantidade de vínculos profissionais de nutricionistas no Ceará. Para essa análise, utilizamos modelos de aprendizado de máquina, com o objetivo de capturar tendências e sazonalidades presentes nos dados ao longo do tempo. Essa abordagem visa responder à **RQ2: Quais padrões ou tendências de vínculo profissional podem ser observados nas séries temporais dos profissionais de nutrição?** Os resultados permitem compreender dinâmicas de crescimento, flutuações sazonais e possíveis fatores que influenciam a variação no número de vínculos ao longo dos anos.

A Figura 4 apresenta a evolução temporal da contagem de CNS por competência (ano e mês) no estado do Ceará. A linha azul ilustra a contagem mensal de CNS, enquanto a linha pontilhada verde exibe a média móvel de 3 meses, que suaviza as variações de curto prazo, evidenciando as tendências gerais. Nota-se um crescimento constante na contagem de CNS ao longo do período analisado, atingindo um ponto máximo de 1.623. Após 2020, um ano marcado pela pandemia e seguido por um período de alta incidência de casos

no Ceará [40–43], observa-se um crescimento mais acentuado no número de profissionais cadastrados no CNES.

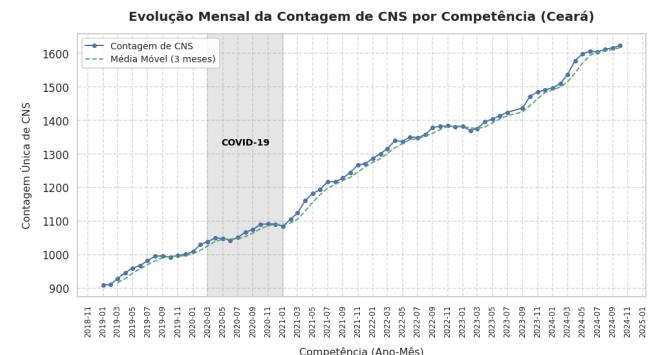


Figure 4: Contagem de CNS por Competência (Ceará).

A Figura 5 apresenta um *violin plot* que ilustra a distribuição anual da quantidade única de CNS no período de 2019 a 2024. Observa-se uma tendência crescente na mediana e no intervalo interquartil ao longo dos anos, sugerindo não apenas um aumento consistente na contagem geral, mas também uma maior variabilidade nos registros, com valores mínimos e máximos mais distantes. Essa dinâmica reflete a inclusão de novos profissionais cadastrados e uma maior heterogeneidade no ritmo de crescimento. O ano de 2021 se destaca por apresentar a maior concentração e amplitude dos valores, o que pode indicar tanto um aumento significativo na cobertura quanto mudanças nos processos de coleta e registro de dados. Em 2020, sugestivamente devido a pandemia, tem uma variação menor no número destes profissionais durante o ano.

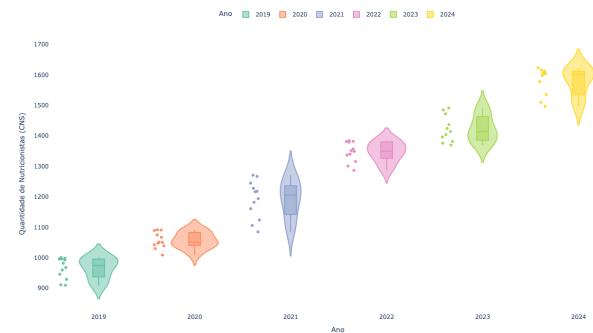


Figure 5: Quantidade de profissionais nutricionistas por ano (Ceará).

A análise da distribuição geográfica e demográfica dos nutricionistas no Ceará revelou uma concentração significativa dos vínculos profissionais em municípios maiores, como Fortaleza, que sozinho representa 40,6% dos vínculos totais. É um concentração, que reflete a centralização dos serviços em áreas urbanas e mais populosas. Porém, também expõe disparidades regionais na cobertura de profissionais. Ajustando os dados pela população, o cenário se torna mais equilibrado, com municípios menores como Várzea destacando-se por uma alta proporção de nutricionistas por

habitante (11 para cada 10 mil). Esse ajuste é essencial para uma análise mais justa da cobertura, pois revela que algumas localidades menores podem ter uma melhor densidade relativa, apesar de apresentarem números absolutos mais baixos. A Tabela 1 apresenta as métricas de desempenho dos modelos utilizados para previsão de vínculos profissionais de nutrição no estado do Ceará. É fornecido uma visão abrangente da precisão de cada abordagem preditiva.

No contexto temporal, os resultados das séries históricas apontam para uma tendência crescente no número de vínculos de nutricionistas no estado, refletindo um aumento na oferta e na demanda por serviços nutricionais.

Os resultados apresentados na Tabela 1 demonstram uma variação significativa no desempenho dos modelos testados, evidenciada pelas métricas de erro calculadas. O modelo ExponentialSmoothing destacou-se como a melhor abordagem, com os menores valores de MAE (30,87), MSE (1409,92), RMSE (37,53) e MAPE (0,021), indicando uma boa precisão e menor desvio médio absoluto em relação aos valores reais.

Table 1: Métricas de Desempenho dos Modelos

| # | Modelo               | MAE    | MSE      | RMSE   | MAPE  |
|---|----------------------|--------|----------|--------|-------|
| 1 | ExponentialSmoothing | 30.87  | 1409.92  | 37.53  | 0.021 |
| 3 | AutoARIMA            | 36.78  | 1849.69  | 43.01  | 0.02  |
| 4 | TBATS                | 42.04  | 2307.96  | 48.04  | 0.02  |
| 5 | Theta                | 48.30  | 3720.60  | 60.98  | 0.03  |
| 6 | PolynomialTrend      | 145.81 | 24090.76 | 155.24 | 0.09  |
| 7 | SeasonalNaive        | 170.41 | 36877.25 | 191.99 | 0.11  |

Conforme a Tabela 1, O modelo que teve o melhor desempenho foi o ExponentialSmoothing, demonstrado na Figura 6 que revela insights sobre o desempenho preditivo do modelo testados para a previsão da contagem de profissionais de saúde (CNS) por competência no estado do Ceará. A figura demonstra a evolução temporal dos dados reais (divididos em treinamento e teste) e as previsões realizadas pelo modelo **ExponentialSmoothing**, que apresenta boa aderência às tendências observadas nos dados reais de teste.

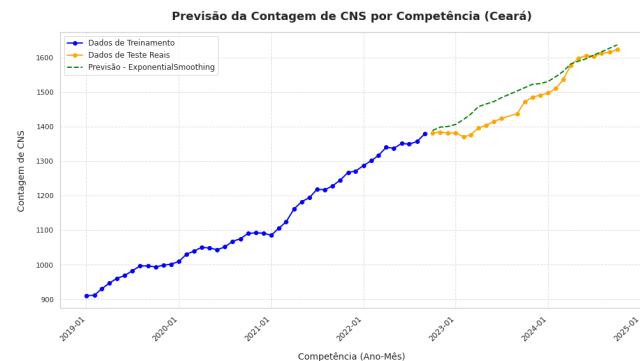


Figure 6: Evolução da contagem de profissionais de saúde (CNS) por competência no estado do Ceará. A figura apresenta os dados de treinamento (em azul), os dados reais de teste (em laranja) e as previsões realizadas pelo modelo ExponentialSmoothing (linha verde tracejada).

A Figura 7 apresenta os resíduos do modelo ExponentialSmoothing ao longo do tempo, evidenciando as diferenças entre os valores reais e as previsões realizadas pelo modelo para cada competência (ano-mês). O comportamento pode indicar que o modelo não conseguiu capturar completamente certos componentes da série temporal, como sazonais ou mudanças estruturais nos dados. Esses padrões nos resíduos sugerem que uma análise complementar ou ajustes no modelo poderiam melhorar o desempenho preditivo, especialmente em períodos com desvios mais acentuados. Isso é devido a uma queda acentuada a partir desse período, que pode foguear o padrão seguido nos últimos meses.

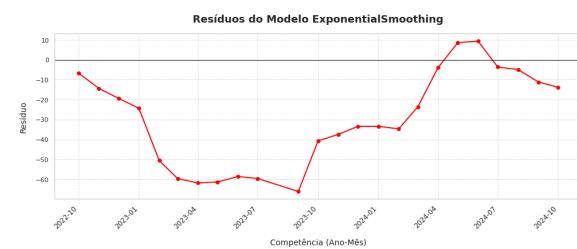


Figure 7: Análise dos resíduos do modelo ExponentialSmoothing. O gráfico apresenta os resíduos ao longo do tempo, calculados como a diferença entre os valores reais e as previsões do modelo para cada competência (ano-mês).

O gráfico de dispersão apresentado na Figura 8 mostra a relação entre os valores reais e as previsões do modelo ExponentialSmoothing. No eixo x estão os valores reais, enquanto no eixo y estão os valores previstos, com a linha pontilhada  $y = x$  representando a situação ideal em que as previsões coincidem exatamente com os valores reais. A proximidade da maioria dos pontos em relação à linha indica que o modelo conseguiu realizar previsões consistentes e próximas dos valores observados. No entanto, a presença de pequenos desvios em torno da reta evidencia a existência de erros residuais, indicando que algumas variações nos dados podem não ter sido completamente capturadas pelo modelo. Esses desvios, embora pequenos, reforçam a importância de explorar possíveis ajustes ou refinamentos no modelo.

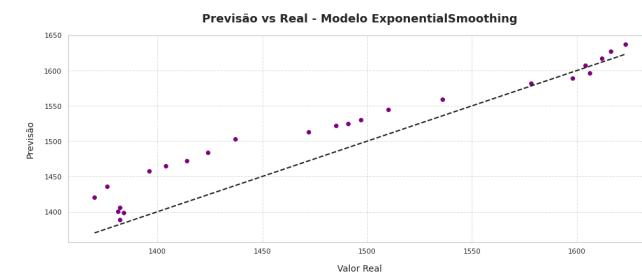


Figure 8: Relação entre valores reais e previstos pelo modelo ExponentialSmoothing. O gráfico de dispersão apresenta os valores reais no eixo x e os valores previstos no eixo y, com a linha pontilhada representando a reta  $y = x$ , que indica a perfeição das previsões.

Para abordar a pergunta de pesquisa **RQ3: De que forma essa ferramenta pode auxiliar gestores na tomada de decisões estratégicas, otimizando o planejamento e a alocação de recursos?**, apresentamos uma discussão técnica sobre o potencial das abordagens de ciência de dados e ferramentas preditivas como suporte na tomada de decisão no âmbito da saúde pública.

Os resultados obtidos ao longo deste trabalho demonstram um aumento contínuo na quantidade de profissionais de nutrição no estado do Ceará. Contudo, a análise revela uma concentração significativa desses profissionais em grandes regiões metropolitanas, como Fortaleza, enquanto municípios de menor porte apresentam escassez ou ausência desses profissionais, evidenciando disparidades geográficas na distribuição dos recursos humanos em saúde.

Nesse contexto, a utilização de modelos de aprendizado de máquina e outras ferramentas preditivas emerge como uma abordagem promissora. Tais modelos possuem a capacidade de analisar grandes volumes de dados históricos, identificar padrões temporais e geográficos, capturar sazonais e prever demandas futuras por profissionais de nutrição. Isso possibilita aos gestores antecipar cenários críticos e propor intervenções mais assertivas, direcionando recursos de forma eficiente e equilibrada entre regiões com maior carência.

Além disso, essas ferramentas permitem simular diferentes cenários de alocação de profissionais, embora essa abordagem não tenha sido amplamente discutida nesse trabalho há um potencial significativo pois esses modelos podem considerar variáveis demográficas, econômicas e de infraestrutura local. Com essa capacidade analítica, gestores municipais, estaduais e federais podem tomar decisões baseadas em evidências, priorizando investimentos nas áreas mais necessitadas.

Por fim, a integração dessas soluções em estratégias de planejamento de políticas públicas possibilita não apenas a correção das desigualdades regionais, mas também a otimização da força de trabalho, garantindo cobertura adequada dos serviços de nutrição e contribuindo para a melhoria da qualidade de vida da população. Essa abordagem reforça a importância de adotar a ciência de dados como uma ferramenta estratégica para enfrentar os desafios na gestão de recursos humanos em saúde.

## 5 CONCLUSÃO

O presente estudo analisou a distribuição demográfica e geográfica dos nutricionistas no estado do Ceará, em que podemos observar uma concentração significativa dos vínculos profissionais em municípios maiores, como Fortaleza, que sozinho representa 40,6% dos vínculos totais, sendo visível as desigualdades regionais na oferta de serviços de nutrição. Do ponto de vista populacional, o cenário é mais equilibrado, onde observamos que localidades menores apresentam uma densidade relativa de nutricionistas superior, como Varjota, que apresentou uma alta quantidade de nutricionistas por habitante (11 para cada 10 mil).

Com a utilização de modelos de aprendizagem de máquina foi possível identificar padrões nos vínculos profissionais ao longo dos anos. O modelo ExponentialSmoothing se mostrou como a melhor abordagem preditiva, com uma maior precisão, e menores valores de erros, indicando uma boa precisão e menor desvio em relação aos valores reais. Podemos ver desigualdades significativas

na distribuição de nutricionistas no Ceará. Esses Resultados são fundamentais para a formulação de políticas públicas que busquem uma oferta mais justa de serviços nutricionais em todo o estado. Diante disso, o uso de modelos de aprendizado de máquina se mostra uma ferramenta promissora para projetar a demanda de novos profissionais e identificar padrões sazonais. Tal abordagem pode contribuir para um planejamento mais eficiente dos serviços de nutrição.

Como trabalhos futuros, é possível explorar de forma mais aprofundada o potencial dos modelos preditivos em níveis municipais e/ou regionais. Isso permitiria ampliar as análises e avaliar a capacidade dos modelos em diferentes cenários, incluindo municípios com poucos ou nenhum profissional de nutrição, onde a escassez de dados representa um desafio adicional.

Além disso, há um potencial significativo em expandir o estudo para realidades similares em outros municípios do Brasil, especialmente aqueles com características demográficas, econômicas e sociais próximas, possibilitando uma comparação mais abrangente. Outra vertente promissora envolve a análise dos resultados em estados circunvizinhos, o que permitiria identificar padrões regionais, variações geográficas e possíveis fluxos de profissionais entre diferentes unidades federativas. Essa abordagem integrada contribuiria para a formulação de políticas públicas mais eficazes e alinhadas às necessidades específicas de cada localidade, promovendo maior equidade na distribuição e no acesso aos serviços de nutrição.

## REFERÊNCIAS

- [1] Associação Brasileira de Nutrição. *Histórico do nutricionista no Brasil, 1939 a 1989: coletânea de depoimentos e documentos*. Atheneu editora São Paulo, 1991.
- [2] Francisco de Assis Guedes de Vasconcelos. O nutricionista no brasil: uma análise histórica. *Revista de Nutrição*, 15:127-138, 2002.
- [3] E Recine, M Leão, and M de F Carvalho. O papel do nutricionista na atenção primária à saúde [organização conselho federal de nutricionistas]. *Brasília, DF*, 3, 2015.
- [4] Bernadete dos Santos Pereira and Elaine Tomasi. Instrumento de apoio à gestão regional de saúde para monitoramento de indicadores de saúde. *Epidemiologia e Serviços de Saúde*, 25:411-418, 2016.
- [5] Luanda de Souza Conrado. Associação entre políticas públicas de alimentação e saúde no ambiente escolar e marcadores de consumo de estudantes brasileiros: resultados da pesquisa nacional de saúde do escolar (pense), 2019. 2024.
- [6] Eugênio Vilaça Mendes. As redes de atenção à saúde. *Ciência & saúde coletiva*, 15:2297-2305, 2010.
- [7] Letícia Batista de Azevedo, Tatielle Rocha de Jesus, and Erika Cardoso dos Reis. Tendência temporal da inserção de nutricionistas no sistema único de saúde segundo regiões do brasil no período de 2009 a 2018: 10.15343/0104-7809.202145024033. *O Mundo da Saúde*, 45(s/n):024-033, 2021.
- [8] Janaina das Neves, Clorine Borda Zanolrensi, Semiramis Martins Álvares Domene, Beatriz Batista, Carmen Lúcia de Araújo Calado, and Francisco de Assis Guedes de Vasconcelos. Eighty years of undergraduate education in nutrition in brazil: an analysis of the 2009-2018 period. *Revista de Nutrição*, 32: e180158, 2019.
- [9] Francisco de Assis Guedes de VASCONCELOS. The nutritionist in brazil: a historical analysis. *Revista de Nutrição*, 15:127-138, 2002.
- [10] Christo El Morr, Deniz Ozdemir, Yasmeen Asdaah, Antoine Saab, Yahya El-Lahib, and Elie Salem Sokhn. Ai-based epidemic and pandemic early warning systems: A systematic scoping review. *Health Informatics Journal*, 30(3):14604582241275844, 2024.
- [11] Thiago Augusto Hernandes Rocha, Núbia Cristina da Silva, Allan Cláudius Queiroz Barbosa, Pedro Vasconcelos Amaral, Elaine Thumé, João Victor Rocha, Viviane Alvares, and Luiz Augusto Faccini. Cadastro nacional de estabelecimentos de saúde: evidências sobre a confiabilidade dos dados. *Ciência & Saúde Coletiva*, 23:229-240, 2018.
- [12] Larissa Riani Costa Tavares, José Luiz Riani Costa, Jorge Oishi, and Patricia Driusso. Inserção da fisioterapia na atenção primária à saúde: análise do cadastro nacional de estabelecimentos de saúde em 2010. *Fisioterapia e Pesquisa*, 25:9-19, 2018.
- [13] Larissa Riani Costa Tavares, José Luiz Riani Costa, Jorge Oishi, and Patricia Driusso. Distribuição territorial de fisioterapeutas no brasil: análise do cadastro

- nacional de estabelecimentos de saúde cnes/2010. *ConScientiae saúde*, 15(1): 53–61, 2016.
- [14] Marcia Regina Pelissari. Cnes como instrumento de gestão e sua importância no planejamento das ações em saúde. *Revista de Saúde Pública do Paraná*, 2(1): 159–165, 2019.
- [15] Mariana Araujo de Albuquerque. Uma reflexão sobre a vinculação da classificação brasileira de ocupações-cbo da medicina veterinária aos sistemas de informação em saúde-sis. B.S. thesis, Universidade Federal do Rio Grande do Norte, 2022.
- [16] Nelson Hideiki Nozoe, Ana Maria Bianchi, and Ana Cristina Ablas Rondet. A nova classificação brasileira de ocupações: anotações de uma pesquisa empírica. *São Paulo em perspectiva*, 17:234–246, 2003.
- [17] Alessandra Xavier Bueno, Cibele Biehl Bossle, and Alex Branco Fraga. A classificação brasileira de ocupações do profissional de educação física no sus: da incompatibilidade à provisoriade nos serviços públicos de saúde. *Pensar a Prática*, 21(2), 2018.
- [18] Rodrigo Calado da Silva, Paulo Henrique DAngelo Seixas, and Ana Luiza d'Ávila Viana. Circularidade dos médicos nas regiões de saúde no brasil. 2017.
- [19] Alcides Silva de Miranda and Diego Azevedo Melo. Análise comparativa sobre a implantação do programa mais médicos em agregados de municípios do rio grande do sul, brasil. *Ciência & Saúde Coletiva*, 21:2837–2848, 2016.
- [20] Phillip G Enns, Joseph A Machak, W Allen Spivey, and William J Wrobleski. Forecasting applications of an adaptive multiple exponential smoothing model. *Management Science*, 28(9):1035–1044, 1982.
- [21] Doan Ngoc Bao, Ngo Duy Khanh Vy, and Duong Tuan Anh. A hybrid method for forecasting trend and seasonal time series. In *The 2013 RIVF International Conference on Computing & Communication Technologies - Research, Innovation, and Vision for Future (RIVF)*, pages 203–208, 2013. doi: 10.1109/RIVF.2013.6719894.
- [22] Ruey-chyn Tsaur. Forecasting by fuzzy double exponential smoothing model. *International journal of computer mathematics*, 80(11):1351–1361, 2003.
- [23] RJ Hyndman. *Forecasting: principles and practice*. OTexts, 2018.
- [24] T. Chai and R. R. Draxler. Root mean square error (rmse) or mean absolute error (mae)? – arguments against avoiding rmse in the literature. *Geoscientific Model Development*, 7(3):1247–1250, 2014. doi: 10.5194/gmd-7-1247-2014. URL <https://gmd.copernicus.org/articles/7/1247/2014/>.
- [25] Cort J Willmott and Kenji Matsuura. Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate research*, 30(1):79–82, 2005.
- [26] Ansari Saleh Ahmar, Fitmayanti Fitmayanti, and Ruliana Ruliana. Modeling of inflation cases in south sulawesi province using single exponential smoothing and double exponential smoothing methods. *Quality & Quantity*, pages 1–11, 2021.
- [27] A Pamungkas, R Puspasari, A Nurfarini, R Zulkarnain, and W Waryanto. Comparison of exponential smoothing methods for forecasting marine fish production in pekalongan waters, central java. In *IOP Conference Series: Earth and Environmental Science*, volume 934, page 012016, IOP Publishing, 2021.
- [28] Konstantinos P Fourkiotis and Athanasios Tsadiras. Applying machine learning and statistical forecasting methods for enhancing pharmaceutical sales predictions. *Forecasting*, 6(1):170–186, 2024.
- [29] Taufiq Dwi Cahyono, Heri Purwanto, Iwan Adhicandra, Kraugusteeliana Kraugusteeliana, and Edy Winarno. Forecasting analysis of fishermen's productivity data using single exponential smoothing. *JINAV: Journal of Information and Visualization*, 3(2):167–173, 2022.
- [30] Patrick A Thompson. An mse statistic for comparing forecast accuracy across series. *International Journal of Forecasting*, 6(2):219–227, 1990.
- [31] Sichen Yuan. Review of root-mean-square error calculation methods for large deployable mesh reflectors. *International Journal of Aerospace Engineering*, 2022(1):5352146, 2022.
- [32] Ummul Khair, Hasanul Fahmi, Sarudin Al Hakim, and Robbi Rahim. Forecasting error calculation with mean absolute deviation and mean absolute percentage error. *Journal of Physics: Conference Series*, 930(1):012002, dec 2017. doi: 10.1088/1742-6596/930/1/012002. URL <https://dx.doi.org/10.1088/1742-6596/930/1/012002>.
- [33] Arnaud De Myttenaere, Boris Golden, Bénédicte Le Grand, and Fabrice Rossi. Using the mean absolute percentage error for regression models, 2015. URL <https://arxiv.org/abs/1506.04176>.
- [34] Ida Nabillah and Indra Ranggadara. Mean absolute percentage error untuk evaluasi hasil prediksi komoditas laut. *JOINS (Journal of Information System)*, 5(2):250–255, Nov. 2020. doi: 10.33633/joins.v5i2.3900. URL <https://publikasi.dimus.ac.id/index.php/joins/article/view/3900>.
- [35] Patrícia L Espinheira, Silvia LP Ferrari, and Francisco Cribari-Neto. On beta regression residuals. *Journal of Applied Statistics*, 35(4):407–419, 2008.
- [36] Jerry L Hintze and Ray D Nelson. Violin plots: a box plot-density trace synergism. *The American Statistician*, 52(2):181–184, 1998.
- [37] René Tanious and Rumen Manolov. Violin plots as visual tools in the meta-analysis of single-case experimental designs. *Methodology*, 18(3):221–238, 2022.
- [38] Zhongmin Cui. On the cover: Violin plot. *Educational Measurement: Issues and Practice*, 39(2):7–7, 2020. doi: <https://doi.org/10.1111/emip.12339>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/emip.12339>.
- [39] Thomas Steichen. Violin: Stata module to generate violin plots. 2001.
- [40] Carlos Sanhueza-Sanzana, Italo Wesley Oliveira Aguiar, Rosa Lívia Freitas Almeida, Carl Kendall, Aminata Mendes, and Ligia Regina Franco Sansigolo Kerr. Desigualdades sociais associadas com a letalidade por covid-19 na cidade de fortaleza, ceará, 2020. *Epidemiologia e Serviços de Saúde*, 30:e2020743, 2021.
- [41] Adryane Gorayeb, Jader de Oliveira Santos, Hércules Gabriel Nascimento da Cunha, Regina Balbino da Silva, Wallason Farias de Souza, Romullo Diogo Pereira Mesquita, Lucas da Silva Libório, Francisco Douglas de Sousa Silva, Sarah Luana Maia do Nascimento, Christian Martins Mota, et al. Pesquisa participativa para o mapeamento da covid-19 na cidade de fortaleza, nordeste do brasil. *Revista Brasileira de Educação Ambiental (RevBEA)*, 15(4):568–588, 2020.
- [42] Davin Pereira da Paula, David Hélio Miranda de Medeiros, Eduardo Lacerda Barros, Renan Gonçalves Pinheiro Guerra, Jader de Oliveira Santos, João Sérgio Queiroz Lima, and Rafaela Martins Leite Monteiro. Diffusion of covid-19 in the northern metropolis in northeast brazil: territorial dynamics and risks associated with social vulnerability. *Sociedade & Natureza*, 32:663–681, 2020.
- [43] Daniele Rocha Queiros Lemos, Sarah Mendes D'Angelo, Luis Arthur Brasil Gadelha Farias, Magda Moura Almeida, Ricristhi Gonçalves Gomes, Geovana Praça Pinto, Josafa Nascimento Cavalcante Filho, Levi Ximenes Feijão, Ana Rita Paulo Cardoso, Thaisy Brasil Ricarte Lima, et al. Health system collapse 45 days after the detection of covid-19 in ceará, northeast brazil: a preliminary analysis. *Revista da Sociedade Brasileira de Medicina Tropical*, 53:e20200354, 2020.