

Modelagem Multinível de Indicadores de Insegurança Alimentar Baseados em Dados Socioeconômicos do Ceará

Daniel de Amaral da Silva¹, Davi Queiroz Albuquerque¹,
Antonio Rafael Braga², Anna Karine Gurgel³, Danielo G. Gomes¹

¹Grupo de Redes de Computadores, Engenharia de Software e Sistemas (GREat)
Departamento de Engenharia de Teleinformática
Centro de Tecnologia, Universidade Federal do Ceará (UFC), Fortaleza - CE

²Redes de Computadores – Campus Quixadá,
Universidade Federal do Ceará (UFC), Quixadá-CE

³Iris LabGov - Laboratório de Inovação e Dados do Governo do Ceará
Fortaleza – CE

[daviqueiroz2002, danielamaral]@alu.ufc.br
karine.gurgel@casacivil.ce.gov.br
[rafaelbraga, danielo]@ufc.br

Abstract. *This paper addresses the issue of food insecurity in Brazil, focusing on the state of Ceará. Using socioeconomic data collected at the national level, we applied multilevel models to analyze how different factors impact food insecurity in various geographic regions, including micro and mesoregions. We found that the proportion of individuals living in households with poor conditions, such as dirt floors and using fuel for lighting, is significantly linked to higher chances of nutritional deficiencies in children. Here we also emphasize the importance of considering socioeconomic and demographic variables when developing policies to combat food insecurity.*

Resumo. *Este artigo aborda a questão da insegurança alimentar no Brasil, com foco no estado do Ceará. Utilizando dados socioeconômicos coletados em nível nacional, aplicamos modelos multiníveis para analisar como diferentes fatores impactam a insegurança alimentar em diversas regiões geográficas, incluindo micro e mesorregiões. Nossa pesquisa identificou que a proporção de indivíduos em domicílios com condições precárias, como piso de terra e uso de combustíveis para iluminação, está associada a um aumento significativo nas chances de deficiências nutricionais em crianças. Este artigo também destaca a importância de considerar variáveis socioeconômicas e demográficas ao elaborar políticas para combater a insegurança alimentar.*

1. Introdução

A insegurança alimentar continua a ser um desafio global significativo, afetando milhões de pessoas em todo o mundo. No Brasil, dados recentes apontam que 21,1 milhões de brasileiros enfrentaram insegurança alimentar grave em 2022, caracterizada pelo estado de fome [FAO et al. 2023]. A insegurança alimentar é definida como o acesso insuficiente a alimentos seguros e nutritivos, necessários para uma vida ativa e saudável, bem

como para um crescimento e desenvolvimento normais. Este problema pode ser causado tanto pela indisponibilidade de alimentos quanto pela falta de recursos para sua obtenção [FAO 2013].

As consequências da insegurança alimentar são abrangentes e impactantes. Além dos efeitos imediatos na saúde física, como desnutrição e deficiências nutricionais, a falta de acesso adequado aos alimentos pode levar a prejuízos cognitivos, atraso no desenvolvimento físico e uma predisposição a diversas doenças [Sousa and Diniz 2024]. Esses impactos não se restringem apenas ao indivíduo, mas se estendem à sociedade como um todo, afetando a produtividade econômica, o desempenho escolar e elevando os custos com saúde pública.

Neste contexto, a identificação precisa de áreas e populações em risco de insegurança alimentar torna-se crucial para o desenvolvimento e implementação de políticas públicas eficazes. Tradicionalmente, a avaliação da insegurança alimentar tem se baseado em pesquisas diretas e na coleta de indicadores antropométricos, como *déficit* de altura para idade, *déficit* de peso para altura e *déficit* de peso para idade. No entanto, essas abordagens muitas vezes são custosas, demoradas e podem não fornecer uma cobertura abrangente da população.

Estudos anteriores exploraram a relação entre fatores socioeconômicos e insegurança alimentar. Por exemplo, Subianto et al. (2023) utilizaram o algoritmo CatBoost para classificar domicílios em categorias de segurança alimentar na Indonésia, identificando variáveis importantes para prever a insegurança alimentar. NJ et al. (2020) revelaram que fatores como renda do chefe de família, nível educacional, posse de terras e participação em programas de assistência social estão significativamente associados à insegurança alimentar no Quênia. No Brasil, Oliveira et al. (2023) destacaram a importância de fatores como renda familiar, escolaridade materna, acesso a serviços de saúde e condições de saneamento básico na determinação do estado nutricional das crianças no Nordeste do país.

Nosso estudo se baseia nessas pesquisas anteriores e as expande, propondo uma abordagem inovadora que utiliza dados socioeconômicos amplamente disponíveis para estimar indicadores de insegurança alimentar. Especificamente, focamos em três indicadores principais: *déficit* de altura para idade (DAI), *déficit* de peso para altura (DPA) e *déficit* de peso para idade (DPI) em crianças menores de 5 anos, os quais são marcadores importantes de desnutrição crônica e insegurança alimentar.

A principal contribuição deste artigo está no uso de um modelo linear misto que permite capturar a estrutura hierárquica dos dados e que leva em conta as variações entre Estados, mesorregiões e microrregiões. Isso é particularmente importante no contexto brasileiro, onde as disparidades regionais em termos de desenvolvimento socioeconômico e acesso a serviços públicos são significativas.

2. Material e Métodos

Nossa abordagem metodológica combina dados de duas fontes principais: o Cadastro Único para Programas Sociais (CadÚnico), qual fornece informações socioeconômicas detalhadas, e o Sistema de Vigilância Alimentar e Nutricional (SISVAN), que registra dados sobre o estado nutricional da população brasileira. Para lidar com o grande volume

de variáveis potenciais, implementamos um processo de seleção em dois estágios: primeiro, utilizamos o teste de correlação de Pearson para selecionar preditores com forte correlação com a variável alvo; em seguida, aplicamos a técnica *Stepwise*. Esta abordagem nos permite identificar as variáveis mais relevantes para o modelo final, oferecendo uma ferramenta potencialmente valiosa para a identificação rápida de áreas de risco de insegurança alimentar.

2.1. Modelos Multiníveis

Modelos multiníveis, também conhecidos como modelos hierárquicos ou de efeitos mistos, são uma extensão dos modelos lineares generalizados (GLMs) que permitem a inclusão de efeitos aleatórios além dos efeitos fixos tradicionais. Esses modelos são particularmente úteis em dados que possuem uma estrutura hierárquica ou agrupada, como em estudos longitudinais ou em pesquisas que envolvem diferentes níveis de agregação espacial ou temporal [Agresti 2014].

Os Modelos Lineares Generalizados Mistos (GLMM) expandem os GLMs ao incorporar efeitos aleatórios no preditor linear. Em um GLMM, a resposta Y_{ij} para a observação j no grupo i é modelada condicionalmente a um efeito aleatório u_i . A forma geral do preditor linear em um GLMM é dada por:

$$g(\mu_{ij}) = \beta_0 + \beta_1 X_{ij1} + \dots + \beta_p X_{ijp} + u_i, \quad (1)$$

em que $g(\cdot)$ é a função de ligação, μ_{ij} é a média condicional de Y_{ij} , $\beta_0, \beta_1, \dots, \beta_p$ são os parâmetros de efeitos fixos, X_{ijk} são os valores das variáveis explicativas, e u_i representa o efeito aleatório específico do grupo i . No contexto do presente estudo, u_i pode corresponder aos efeitos aleatórios associados às diferentes regiões geográficas, como estados, microrregiões ou mesorregiões. Esses efeitos aleatórios são assumidos como normalmente distribuídos com média zero e variância σ_u^2 [Agresti 2014].

Os modelos multiníveis apresentam várias vantagens em relação aos modelos tradicionais. Eles permitem a modelagem explícita da correlação intragrupo, melhorando a precisão das estimativas dos efeitos fixos. Além disso, fornecem uma decomposição clara da variabilidade em diferentes níveis hierárquicos e podem lidar com dados desbalanceados, comuns em estudos longitudinais e observacionais [Agresti 2014].

2.2. Obtenção dos Dados

Nossa pesquisa utilizou dados do Sistema de Vigilância Alimentar e Nutricional (SISVAN¹) de 2023, analisando indicadores de desenvolvimento para crianças menores de 5 anos em todos os municípios brasileiros. Focamos em três dos quatro principais indicadores disponibilizados pela plataforma para essa faixa etária: *Déficit* de Altura para Idade (DAI), *Déficit* de Peso para Altura (DPA) e *Déficit* de Peso para Idade (DPI), contabilizando o número de crianças abaixo do ponto de corte da OMS para cada nível geográfico.

Complementamos nossa análise com informações socioeconômicas da plataforma CECAD², que fornece 56 variáveis para cada município. Considerando valores absolutos e percentuais, obtivemos 592 variáveis predictoras. Como a plataforma não permite filtrar

¹<https://sisaps.saude.gov.br/sisvan/relatoriopublico/index>

²https://cecad.cidadania.gov.br/tab_cad.php

os valores das variáveis em relação ao tempo, coletamos apenas os dados atualizados referentes ao mês de maio de 2024.

Ademais, obtivemos através da API de localidades do IBGE³ os dados referentes aos níveis geográficos de cada município, isto é, a Mesorregião e a Microrregião em que aquele município está inserido.

Após excluir amostras com informações incompletas, nossa análise final incluiu dados de 5.528 municípios brasileiros, combinando indicadores de desenvolvimento infantil, dados socioeconômicos e informações geográficas.

2.3. Pré-processamento e Seleção de Características

Para reduzir a complexidade na escolha do modelo, realizamos a seleção de variáveis preditoras em duas etapas. Primeiramente, identificamos as variáveis com forte correlação com a variável alvo utilizando o teste de correlação de Pearson. Avaliamos a significância estatística de cada preditor com base nos valores-p. Variáveis com valor-p $\leq 0,15$ foram consideradas estatisticamente significativas e selecionadas para a próxima etapa.

Na segunda etapa, aplicamos a técnica *Stepwise* para refinar a seleção das variáveis preditoras. Este método envolve um processo em dois passos: na seleção direta, adicionamos preditores a um modelo inicial vazio, enquanto na seleção reversa, removemos termos do modelo resultante da seleção direta. Utilizamos o Critério de Informação Bayesiano modificado (mBIC2), conforme descrito por Frommlet et al. (2011), para decidir a adição ou remoção de variáveis, garantindo a inclusão das variáveis mais relevantes para a predição dos indicadores de insegurança alimentar.

2.4. Modelagem

A partir das variáveis selecionadas, construímos um modelo de efeitos mistos para capturar a estrutura hierárquica dos dados. Consideramos os efeitos fixos para as variáveis socioeconômicas selecionadas e os efeitos aleatórios para os níveis geográficos (estado, mesorregião e microrregião):

- Efeitos fixos: Incluímos todas as variáveis socioeconômicas selecionadas no processo de seleção *stepwise*.
- Efeitos aleatórios: Consideramos efeitos aleatórios para os níveis geográficos, especificando uma estrutura hierárquica para estado, mesorregião e microrregião.

Para a estimação dos parâmetros do modelo, utilizamos um modelo de regressão multinível, com uma família de distribuição beta-binomial para capturar a superdispersão nos dados. Implementamos o modelo utilizando a biblioteca `glmmTMB` do software R.

2.5. Métricas

Para entender a influência das variáveis preditoras nos modelos, utilizamos razões de chances (*Odds Ratios*), que indicam a influência da proporção do preditor na chance de *déficit* do indicador analisado: valores maiores que 1 aumentam a chance, menores que 1 diminuem, e igual a 1 não altera a chance (Agresti, 2014). Contudo, devido à possível assimetria na distribuição das razões de chances, preferimos o logaritmo natural das razões de chances (*Log Odds Ratios*), especialmente para intervalos de confiança.

³<https://servicodados.ibge.gov.br/api/docs/localidades>

Para analisar os modelos, construímos *forest plots* mostrando os preditores e seus valores de logaritmo das razões de chances, além de analisar a significância estatística com valores-p, considerando preditores com valor-p $\leq 0,001$ como altamente relevantes, assinalados com três asteriscos nos gráficos. Assim, a significância estatística das variáveis foi assinalada nos gráficos de acordo com as seguintes condições: $p < 0,001$ (***) ; $p < 0,01$ (**) ; $p < 0,05$ (*) .

3. Resultados

A Tabela 1 apresenta os valores de R^2 marginal e R^2 condicional para os modelos de DAI e DPA. Observa-se que o R^2 marginal, que representa a proporção de variância explicada pelos efeitos fixos (variáveis socioeconômicas), é consideravelmente inferior ao R^2 condicional. Isso sugere que, embora as variáveis socioeconômicas expliquem uma parte relevante da variação nos déficits nutricionais, há outros fatores em jogo que não são capturados apenas pelos efeitos fixos.

Tabela 1. R^2 marginal e R^2 condicional dos modelos multiníveis para DAI e DPA.

Índice	R^2 marginal	R^2 condicional
DAI	0,70	≈ 1
DPA	0,41	≈ 1

Por outro lado, o R^2 condicional, que incorpora tanto os efeitos fixos quanto os efeitos aleatórios (UF, microrregião e mesorregião), é muito próximo de 1, indicando que a maior parte da variabilidade é explicada quando as variáveis de efeitos aleatórios são consideradas. Essa alta explicação da variância reflete as disparidades regionais significativas na prevalência de déficits nutricionais, destacando o papel crucial da localização geográfica na determinação desses indicadores.

As Figuras 1 e 2 apresentam os *forest plots* dos modelos para DAI e DPA, respectivamente. Observamos duas variáveis presentes em ambos os modelos analisados:

- “Cor ou Raça - Branca”, apresentando razões de chances $\exp(-0,28) \approx 0,76$ e $\exp(-0,38) \approx 0,68$ para DAI e DPA, respectivamente. Logo, o aumento da proporção de pessoas de cor branca está associado a reduções de, aproximadamente, 24% ($1 - 0,76$) e de 32% ($1 - 0,68$) nas chances de crianças do município apresentarem *déficit* de altura/idade e peso/altura, respectivamente;
- “Material predominante no piso do domicílio - Terra”, apresentando razões de chances $\exp(0,28) \approx 1,32$ e $\exp(0,62) \approx 1,86$ para DAI e DPA, respectivamente. Dessa forma, o aumento de pessoas no município que residem em domicílios com piso de terra está associado a um aumento nas chances de *déficit* de altura/idade e peso/altura de, aproximadamente, 32% e 86%, respectivamente.

Além das variáveis comuns a ambos os modelos, destacamos também a influência do aumento da proporção de pessoas em domicílios com iluminação movida a combustíveis (óleo, querosene ou gás), estando esse preditor associado a um aumento das chances de *déficit* de altura/idade em quase 5 vezes (razão de chances igual a $\exp(1,55) \approx 4,71$). Outra variável que se destacou na nossa análise para DPA foi a

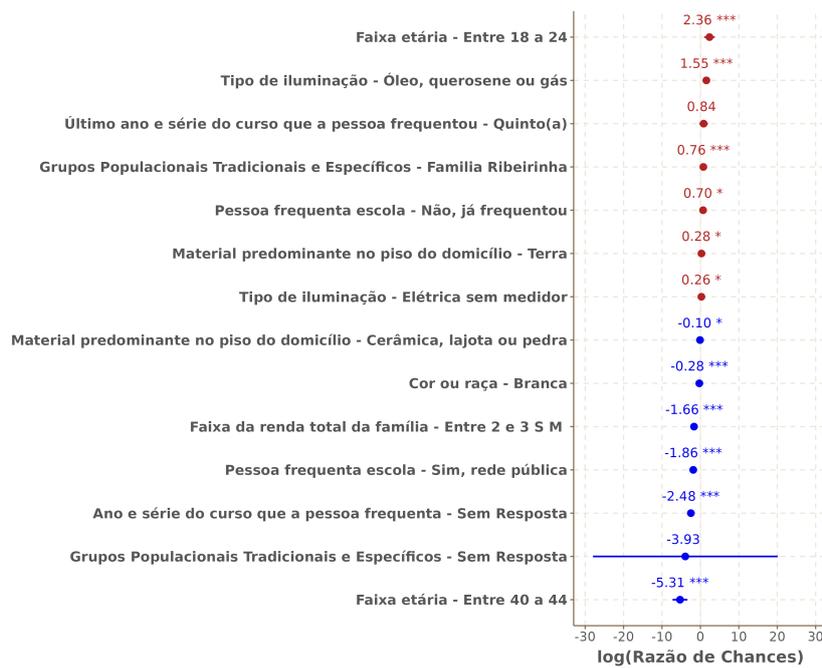


Figura 1. Variáveis preditoras no modelo de *déficit* de altura/idade (DAI).

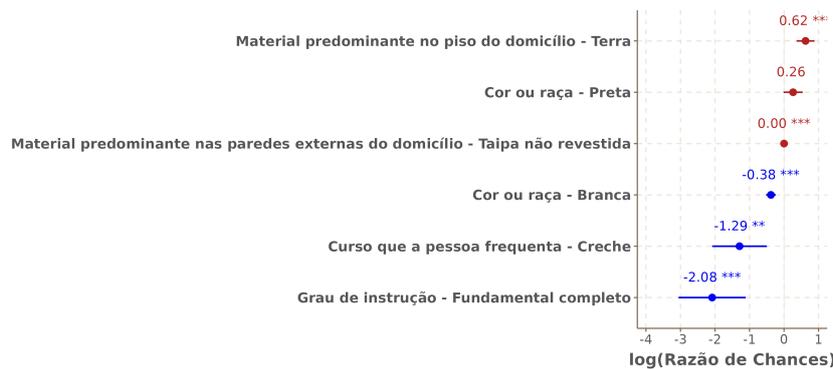


Figura 2. Variáveis preditoras no modelo de *déficit* de peso/altura (DPA).

proporção de pessoas com fundamental completo no município, cujo aumento dessa parcela esteve associado a uma redução de, aproximadamente, 88% das chances de *déficit* de peso/altura (razão de chances igual a $\exp(-2,08) \approx 0,12$).

Além dos efeitos fixos, também analisamos a influência dos efeitos aleatórios, isto é, como a subdivisão geográfica influencia nas chances de *déficit* para os índices analisados. Neste caso, visando uma discussão mais simples e objetiva considerando a nossa localidade, escolhemos o estado do Ceará para análise. Para construir os mapas da Figura 3, obtivemos os dados geoespaciais do pacote *geobr*⁴, desenvolvido pelo Instituto de Pesquisa Econômica Aplicada (IPEA).

Em relação ao DAI, na Figura 3a observamos que as mesorregiões mais ao norte apresentaram efeito de redução das chances para esse tipo de *déficit* em crianças, enquanto houve um aumento nas regiões dos sertões e ao Sul. Quando analisamos pelo nível das

⁴<https://ipeagit.github.io/geobr/>

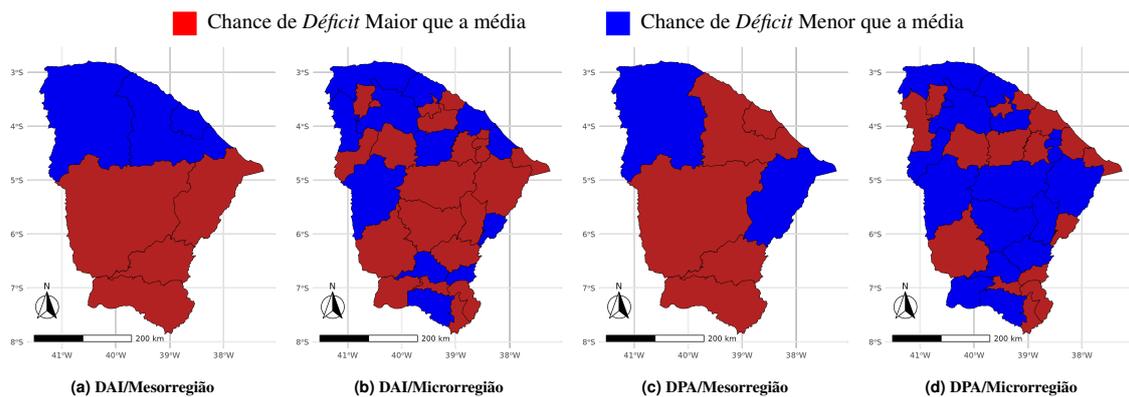


Figura 3. Efeitos aleatórios das subdivisões geográficas do Estado do Ceará na modelagem dos indicadores.

microrregiões (Figura 3b), essas tendências também permanecem, mas com uma maior heterogeneidade em relação ao nível geográfico superior.

No caso da DPA (Figura 3c), a tendência de aumento das chances do *déficit* nas mesorregiões dos sertões e ao Sul, porém esse comportamento se estendeu para algumas mesorregiões litorâneas, incluindo a Metropolitana de Fortaleza. Já analisando sob a esfera das microrregiões (Figura 3d), observamos uma grande disparidade em relação ao nível da mesorregião, pois muitas microrregiões dos sertões apresentaram efeito de diminuição das chances, indicando que mesmo dentro de uma mesorregião há microrregiões de realidades muito distintas.

4. Discussão

Os resultados sugerem que é possível modelar indicadores de insegurança alimentar a partir de dados socioeconômicos. O alto R^2 condicional (próximo a 1) em comparação com o R^2 marginal mais baixo indica que as disparidades regionais desempenham um papel crucial na determinação dos déficits nutricionais. Pelo lado social, destacamos que a presença de pessoas da cor branca influenciou na redução das chances de DAI e DPA, enquanto a presença de pessoas com moradias precárias, por exemplo as que possuem piso de terra, aumentou as chances de *déficit*. Além disso, a proporção de pessoas com fundamental completo também foi considerada importante para a redução das chances de déficit, revelando que políticas de ensino à educação também podem ser apresentadas como solução para reduzir a insegurança alimentar nos municípios.

No entanto, os dados do SISVAN podem apresentar viés devido à seleção das crianças com risco de desnutrição, limitando a generalização dos nossos achados. Ademais, os modelos construídos apresentaram um certo grau de sensibilidade a *outliers*, especialmente para DPI, cujo desempenho inferior levou à sua exclusão do artigo, em contraste com os melhores resultados de DAI e DPA.

5. Conclusão

Este artigo estima indicadores de insegurança alimentar a partir de dados socioeconômicos. Aplicando modelos multiníveis e técnicas avançadas de seleção de características, conseguimos identificar fatores importantes que influenciam o *Déficit* de Altura

para Idade (DAI) e o *Déficit* de Peso para Altura (DPA) em crianças menores de 5 anos. Os resultados sugerem que fatores como raça, condições de moradia e nível educacional têm um impacto significativo na insegurança alimentar. Por exemplo, uma maior proporção de pessoas brancas em um município está associada a uma menor chance de *déficit*, enquanto moradias com piso de terra batida aumentam essa chance. Além disso, a conclusão do ensino fundamental mostrou-se crucial na redução do risco de DPA.

Destacamos também a importância da localização geográfica. Observamos no estado do Ceará variações significativas nas chances de *déficit* entre diferentes mesorregiões e microrregiões. Isso sugere que políticas eficientes de combate à insegurança alimentar devem considerar as especificidades locais.

Agradecimentos

Danielo G. Gomes agradece ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela bolsa de produtividade (processo 311845/2022-3) e à FUNCAP pelo apoio financeiro na execução do projeto Cientista Chefe da Transformação Digital do Estado do Ceará (processo 06681109/2023).

Referências

- Agresti, A. (2014). *Categorical Data Analysis*. Hoboken Wiley.
- FAO (2013). *Hunger and food insecurity*. Food and Agriculture Organization of the United Nations (FAO). <https://www.fao.org/hunger/en/> [Acessado em: 16/04/2024].
- FAO, IFAD, UNICEF, WFP, and WHO (2023). *The State of Food Security and Nutrition in the World 2023: Urbanization, agrifood systems transformation and healthy diets across the rural–urban continuum*, volume 2023. Food and Agriculture Organization (FAO). <https://openknowledge.fao.org/items/445c9d27-b396-4126-96c9-50b335364d01> [Acessado em: 16/04/2024].
- Frommlet, F., Chakrabarti, A., Murawska, M., and Bogdan, M. (2011). Asymptotic bayes optimality under sparsity for generally distributed effect sizes under the alternative.
- NJ, M. K., Kisovi, L., and JA, O. (2020). Socio-economic determinants of food insecurity at household level in makueni county, kenya. *International Journal of Scientific and Research Publications*, 10:84–91.
- Oliveira, M. M. d., Lins, A. C. d. L., Bernardino, Í. d. M., and Pedraza, D. F. (2023). Fatores associados ao estado nutricional de crianças menores de 5 anos na região nordeste do brasil. *Cadernos Saúde Coletiva*, 31:e31040200.
- Sousa, I. M. L. d. and Diniz, R. B. (2024). Controle da qualidade e segurança alimentar durante a pandemia por covid-19 nos setores públicos do brasil. *Nutrivisa - Revista de Nutrição e Vigilância em Saúde*, 11(1):e12302.
- Subianto, M., ULYA, I. Y., RAMADHANI, E., SARTONO, B., and HADI, A. F. (2023). Application of shap on catboost classification for identification of variabls characterizing food insecurity occurrences in aceh province households. *Jurnal Natural*, 23(3):230–244.