



UNIVERSIDADE FEDERAL DO CEARÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA-CAEN
MESTRADO PROFISSIONAL EM ECONOMIA DO SETOR PÚBLICO - MESP

THIAGO MESQUITA VIEIRA

UMA ANÁLISE PREDITIVA DA INSEGURANÇA ALIMENTAR BRASILEIRA
UTILIZANDO UM MODELO DE MACHINE LEARNING

FORTALEZA

2026

THIAGO MESQUITA VIEIRA

**UMA ANÁLISE PREDITIVA DA INSEGURANÇA ALIMENTAR BRASILEIRA
UTILIZANDO UM MODELO DE MACHINE LEARNING**

Projeto de pesquisa apresentado ao Programa de Pós-Graduação em Economia do Centro de Aperfeiçoamento de Economistas do Nordeste (CAEN) da Universidade Federal do Ceará, como requisito parcial para a obtenção do grau de Mestre em Economia do Setor Público.

Orientador Prof. Dr. João Mário Santos de França

Coorientadora Profª. Dra. Guaracyane Lima Campêlo

FORTALEZA

2026

THIAGO MESQUITA VIEIRA

UMA ANÁLISE PREDITIVA DA INSEGURANÇA ALIMENTAR BRASILEIRA
UTILIZANDO UM MODELO DE MACHINE LEARNING

Dissertação submetida à Coordenação do Programa de Economia Profissional – PEP, da Universidade Federal do Ceará - UFC, como requisito parcial para a obtenção do grau de Mestre em Economia. Área de Concentração: Economia do Setor Público.

Aprovada em: 27/03/2026.

BANCA EXAMINADORA

Prof. Dr. João Mário Santos de França (Orientador)
Universidade Federal do Ceará (UFC)

Profa. Dra. Guaracyane Lima Campelo (Coorientadora)
Universidade Federal do Ceará (UFC)

Profa. Dra. Natália Cecília de França
Universidade Federal do Ceará (UFC)

RESUMO

A presente pesquisa analisou os perfis de risco associados à insegurança alimentar para domicílios brasileiros, usando os microdados da PNADC 2024 e a modelagem de um modelo de aprendizado de máquina *XGBoost*. O modelo econométrico *Logit* foi empregado para fins comparativos, permitindo examinar diferenças de desempenho entre abordagens inferenciais e preditivas. A análise considera fatores associados à renda domiciliar, composição familiar, características do chefe do domicílio, condições habitacionais, localização geográfica e recebimento de benefícios sociais. Os resultados indicam que o modelo *XGBoost* apresenta desempenho preditivo superior ao *Logit*, com acurácia maior e melhor equilíbrio entre sensibilidade e especificidade. Evidencia-se a renda domiciliar *per capita* como principal preditor da insegurança alimentar, seguida dos recebimento de benefícios sociais, das características regionais, da cor e da escolaridade do chefe de domicílio e dos adultos, da inserção ocupacional e da composição familiar. O estudo contribui para o aprofundamento do debate sobre insegurança alimentar no Brasil ao incorporar técnicas de modelagem preditiva ainda pouco exploradas na literatura nacional, oferecendo evidências empíricas que podem subsidiar o aprimoramento de políticas públicas voltadas à mitigação de vulnerabilidades alimentares.

Palavras-chave: Insegurança alimentar; *XGBoost*; Modelos de aprendizado de máquina.

ABSTRACT

This research analyzed the risk profiles associated with food insecurity for Brazilian households, using microdata from PNADC 2024 and the modeling of a machine learning model XGBoost. The Logit econometric model was used for comparative purposes, allowing to examine differences in performance between inferential and predictive approaches. The analysis considers factors associated with household income, family composition, characteristics of the head of household, housing conditions, geographical location and receipt of social benefits. The results indicate that the XGBoost model presents better predictive performance than Logit, with greater accuracy and better balance between sensitivity and specificity. Household income per capita is highlighted as the main predictor of food insecurity, followed by receipt of social benefits, regional characteristics, color and education of the head of household and adults, occupational insertion and family composition. The study contributes to the deepening of the debate on food insecurity in Brazil by incorporating predictive modeling techniques still little explored in the national literature, offering empirical evidence that can subsidize the improvement of public policies aimed at mitigating food vulnerabilities.

Keywords: Food insecurity; XGBoost; Machine learning.

LISTA DE ABREVIATURAS E SIGLAS

2SLS	<i>Two-Stage Least Squares</i>
BPC-LOAS	Benefício Assistencial de Prestação Continuada
CSG	<i>Child Support Grant</i>
EBIA	Escala Brasileira de Insegurança Alimentar
FAO	<i>Food and Agriculture Organization</i>
GMM	<i>Generalized Method of Moments</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
IPEM	Índice de Pobreza Energética Multidimensional
ODS	Objetivos de Desenvolvimento Sustentável
ONU	Organização das Nações Unidas
PNAD	Pesquisa Nacional por Amostra de Domicílios
PNADC	Pesquisa Nacional por Amostra de Domicílios Contínua
POF	Pesquisa de Orçamentos Familiares
PPC	Paridade do Poder de Compra
<i>SHAP</i>	<i>Shapley Additive exPlanations</i>
<i>SMOTE</i>	<i>Synthetic Minority Over-sampling Technique</i>
<i>XGBoost</i>	<i>Extreme Gradient Boosting</i>

LISTA DE SÍMBOLOS

%	Porcentagem
%	Porcentagem
β	Vetor de coeficientes estimados
ε	Termo de erro aleatório
Pr()	Probabilidade de ocorrência de evento
R^2	Proporção da variância explicada
kWh	Unidade de consumo elétrico
F1	Métrica de desempenho classificatório
%	Porcentagem
\mathbb{R}	Números Reais
Σ	Somatório

SUMÁRIO

1. INTRODUÇÃO	5
2. REVISÃO DE LITERATURA	8
2.1 Relação entre insegurança alimentar e outras variáveis socioeconômicas e estruturais	8
2.2 Relação entre programas sociais e insegurança alimentar.....	12
2.3 Estudos sobre insegurança alimentar utilizando modelos de <i>machine learning</i>	14
3. METODOLOGIA.....	19
3.1 Base de dados.....	19
3.2 Pobreza energética multidimensional	21
3.3 Modelo <i>XGBoost</i>	22
3.4 Análise do modelo de previsão	25
4. RESULTADOS.....	27
4.1 Estatísticas descritivas	27
4.2 Modelagem e desempenho preditivo	29
4.3 Valores <i>SHAP</i>	31
5. CONSIDERAÇÕES FINAIS.....	36
REFERÊNCIAS	37
APÊNDICE A – Estimativas das Razões de Chance do Modelo <i>Logit</i>	42

1. INTRODUÇÃO

A insegurança alimentar caracteriza-se por ser uma condição em que uma pessoa não acessa, de forma regular, alimentos, além de seguros e nutritivos, em quantidade calórica suficiente (FAO, 2025). Sendo assim, por ser um fenômeno que reflete a insuficiência tanto quantitativa como qualitativa alimentar, ocasiona efeitos adversos sobre condições de vida, saúde e bem-estar das famílias. Como elemento agravante da situação para o atual momento, surge uma combinação de estagnação econômica e inflação, que acaba por restringir ainda mais o acesso aos alimentos, principalmente, para aqueles em situação de maior vulnerabilidade social (ONU, 2025).

Dados da FAO apresentam números alarmantes sobre a fome a nível mundial. Conforme o estudo, em 2023, aproximadamente, 733 milhões de pessoas passaram fome. Um fato que se apresenta como preocupante é o crescimento vertiginoso entre 2019 e 2023, tendo em vista que, conforme o estudo, houve um incremento de 152 milhões de pessoas no interstício de 4 anos. Além disso, merece destaque o achado que relata que, aproximadamente, um em cada cinco africanos passou fome em 2023. Já a Ásia, apesar de ter um dos menores números de forma relativa, 6,2%, apresenta a maior quantidade, em números absolutos, de pessoas que passaram fome, 384,5 milhões (FAO *et al*, 2024).

Nesse sentido, a literatura internacional e nacional tem avançado na compreensão da insegurança alimentar como um fenômeno multidimensional, cuja explicação extrapola a insuficiência de renda monetária e envolve um conjunto mais amplo de condicionantes socioeconômicos, demográficos e institucionais. Embora parte dos estudos enfatize recortes territoriais específicos, outros adotam abordagens mais agregadas, sem delimitação regional explícita, convergindo, contudo, ao apontar que fatores associados ao nível de riqueza, à saúde, à educação, às características do domicílio e ao acesso a serviços essenciais influenciam de forma relevante o acesso regular a alimentos em quantidade e qualidade adequadas.

No plano metodológico, parte da literatura chama atenção para os limites de abordagens baseadas exclusivamente em renda ou consumo, independentemente do recorte geográfico adotado. Hjelm, Mathiassen e Wadhwa (2016), ao compararem medidas de consumo per capita e índices de riqueza derivados de ativos e características domiciliares em diferentes contextos nacionais, demonstram que essas métricas capturam dimensões distintas do bem-estar e não são plenamente intercambiáveis na análise da segurança alimentar. Os autores convergem

ao afirmar que a escolha do indicador influencia a identificação da insegurança alimentar e o direcionamento de políticas públicas, ainda que os resultados não se restrinjam a uma única região ou território específico.

Em diálogo com essa discussão, estudos mais recentes têm enfatizado o papel de determinantes que pressionam o orçamento doméstico e afetam indiretamente o consumo alimentar, sobretudo em contextos de restrição orçamentária, ainda que com diferentes escalas territoriais. Silva *et al.* (2025), ao analisarem domicílios brasileiros com base em dados nacionais, indicam que despesas associadas a necessidades básicas competem com o gasto alimentar, ampliando a probabilidade de insegurança alimentar entre grupos vulneráveis. Esse achado converge com a literatura ao evidenciar que rigidezes orçamentárias podem gerar *trade-offs* intradomiciliares, mesmo quando não há distinção explícita entre regiões.

Por fim, a renda e a inserção produtiva permanecem como elementos centrais no debate, sobretudo em estudos que adotam recortes territoriais definidos. Evidências para o contexto brasileiro indicam que a diversificação das fontes de renda atua como mecanismo de proteção contra a insegurança alimentar, especialmente em áreas rurais. Cruz, Silva e Araújo (2025), ao analisarem domicílios do Nordeste brasileiro, mostram que a inserção em atividades não agrícolas reduz a probabilidade de insegurança alimentar e nutricional, resultado que dialoga com achados internacionais ao destacar o papel da estabilidade e da composição da renda na mitigação de vulnerabilidades alimentares.

A implicação é direta para o debate de políticas: além da transferência de renda, a geração de oportunidades e a diversificação de fontes de rendimento podem atuar como mecanismos complementares de mitigação de vulnerabilidades alimentares, sobretudo quando articuladas a instrumentos de proteção social.

Analisando especificamente o Brasil, segundo dados do Instituto Brasileiro de Geografia e Estatística (IBGE), especificamente da POF 2017-2018, de um montante aproximado de 68,9 milhões de domicílios, em torno de 25,2 milhões deles estavam na linha de insegurança alimentar, sendo essa dividida em leve (16,5 milhões de domicílios), moderada (5,6 milhões) e grave (3,1 milhões de domicílios). No que concerne ao uso de energia elétrica, teve maior frequência (60,9%) nos domicílios em suficiência alimentar e menos (33,5%) nos domicílios com insuficiência alimentar grave, o que indica a relação entre insuficiência alimentar e outras variáveis, a exemplo de pobreza energética. Por último, cabe ressaltar que as regiões com menores índices de domicílios em segurança alimentar, no referido período, ainda segundo o IBGE, foram as regiões Norte (43,0%) e Nordeste (49,7%) (IBGE, 2021).

Diante desse contexto, este estudo tem como objetivo prever a insegurança alimentar, tratada como variável dependente, a partir de um conjunto amplo de características socioeconômicas, demográficas, territoriais e de acesso à proteção social dos domicílios brasileiros, utilizando dados da Pesquisa Nacional por Amostra de Domicílios Contínua de 2024 (PNADC), comparando a eficácia do modelo de *machine learning* e *Xtreme Gradient Boosting* (XGBoost) com o modelo *Logit*.

A contribuição do estudo reside na oferta de subsídios ao aprimoramento de políticas públicas de combate à insegurança alimentar, ao avaliar empiricamente a capacidade de características socioeconômicas e do acesso a programas de transferência de renda em explicar e prever sua ocorrência nos domicílios brasileiros. Ao empregar, de forma comparativa, métodos econométricos tradicionais e técnicas de aprendizado de máquina, o trabalho contribui para preencher uma lacuna da literatura nacional, ainda marcada pela escassez de estudos que utilizam abordagens preditivas baseadas em *machine learning* para a análise da insegurança alimentar no Brasil, em consonância com os Objetivos de Desenvolvimento Sustentável relacionados à erradicação da pobreza (ODS 1) e à fome zero (ODS 2).

No que concerne à estrutura do estudo, na segunda seção, apresenta-se a revisão de literatura, que discute os principais conceitos e evidências empíricas relacionados à insegurança alimentar, bem como sua associação com programas de transferência de renda e com determinantes estruturais de ordem socioeconômica, demográfica e territorial. Na terceira seção, descrevem-se metodologia e dados, detalhando os procedimentos de construção de indicadores e as estratégias econométricas. Na quarta seção apresentam-se os resultados, com destaque para as relações entre insegurança alimentar, proteção social e variáveis explicativas selecionadas. Por fim, na quinta seção, expõem-se as conclusões, contribuições e implicações para políticas públicas, bem como sugestões para pesquisas futuras.

2. REVISÃO DE LITERATURA

2.1 Relação entre insegurança alimentar e outras variáveis socioeconômicas e estruturais

A insegurança alimentar tem sido amplamente reconhecida na literatura como um fenômeno multidimensional, cuja ocorrência não se explica apenas pela insuficiência de renda monetária. Estudos recentes indicam que fatores associados à forma de mensuração da pobreza, à inserção produtiva, à estrutura domiciliar, às condições territoriais, à governança e ao acesso a serviços e mercados desempenham papel relevante na determinação do acesso regular a alimentos em quantidade e qualidade adequadas (Hjelm, Mathiassen e Wadhwa, 2016; Cruz, Silva e Araújo, 2025; Okolie *et al.*, 2025; Anser *et al.*, 2021). Nesse sentido, a análise da relação entre insegurança alimentar e outras variáveis socioeconômicas e estruturais torna-se fundamental para compreender os diferentes mecanismos que produzem vulnerabilidade alimentar.

Um primeiro eixo de investigação diz respeito às métricas utilizadas para captar a condição socioeconômica dos domicílios. Hjelm, Mathiassen e Wadhwa (2016), com base em inquéritos domiciliares realizados entre 2006 e 2012 em cinco países (Nepal, Malawi, Tanzânia, Uganda e Madagascar), analisaram a segurança alimentar a partir da comparação entre medidas de fluxo, como o consumo per capita, e medidas de estoque, como índices de riqueza construídos a partir da posse de ativos e características habitacionais. Os autores demonstram que, embora correlacionadas, essas medidas capturam dimensões distintas da vulnerabilidade: enquanto o consumo se mostra mais sensível à privação calórica imediata, os índices de riqueza refletem restrições estruturais mais persistentes. Como resultado, diferentes métricas levam à identificação de grupos distintos como mais expostos à insegurança alimentar, com implicações relevantes para a formulação e avaliação de políticas públicas.

No contexto brasileiro, a relação entre insegurança alimentar e inserção produtiva foi explorada por Cruz, Silva e Araújo (2025), que analisaram o impacto das atividades não agrícolas sobre a segurança alimentar e nutricional de famílias rurais do Nordeste, utilizando microdados da PNAD 2013 e a Escala Brasileira de Insegurança Alimentar (EBIA). Para lidar com o viés de seleção associado à participação em atividades não agrícolas, o estudo aplicou o método *de Propensity Score Matching*, com estimação do escore por modelo *Logit* e pareamentos por *Kernel* e *Nearest Neighbor*. Os resultados indicam que a inserção em atividades não agrícolas está associada a uma condição mais favorável de segurança alimentar,

sugerindo que a diversificação das fontes de renda atua como mecanismo relevante de mitigação da vulnerabilidade alimentar em áreas rurais marcadas por limitações produtivas e climáticas.

Ao ampliar o escopo para determinantes institucionais e tecnológicos, o estudo elaborado por Anser *et al.* (2021), no período de 2005 a 2018, combinando indicadores de segurança alimentar com medidas de adoção de tecnologias da informação e comunicação (TIC) e de governança, estimadas por meio do método system-GMM. Os resultados apontam efeitos positivos tanto da adoção de TIC quanto da qualidade institucional, em especial da efetividade governamental e do controle da corrupção, sobre os níveis de segurança alimentar. Ademais, os autores destacam que a interação entre TIC e governança apresenta efeito estatisticamente significativo, sugerindo que avanços tecnológicos tendem a produzir melhores resultados quando combinados a ambientes institucionais mais robustos.

Em contextos em que a dimensão territorial do acesso a alimentos assume centralidade, a literatura recente tem recorrido a modelos preditivos para identificar áreas com maior risco de insegurança alimentar. Okolie *et al.* (2025) utilizaram o *USDA Food Access Research Atlas* para classificar 72.531 *census tracts* nos Estados Unidos quanto à condição de baixo acesso a alimentos, com foco em populações de baixa renda. O estudo comparou modelos de Regressão Logística, *Random Forest* e *XGBoost*, observando desempenho superior das abordagens baseadas em árvores em relação ao modelo linear. A análise de importância de variáveis destaca fatores como taxa de pobreza, renda mediana familiar, participação no SNAP¹ e acesso a veículos, reforçando que desigualdades socioeconômicas e barreiras de mobilidade permanecem centrais na explicação das assimetrias territoriais de acesso alimentar.

Ainda no contexto norte-americano, Chilukuri (2025) analisou dados do *Food Access Research Atlas* de 2019 com o objetivo de avaliar o desempenho comparativo de diferentes famílias de algoritmos na classificação de áreas de baixo acesso a alimentos. O autor comparou cinco modelos, incluindo Regressão Logística, *Random Forest*, *XGBoost* e uma arquitetura tabular baseada em atenção, constatando desempenho superior dos modelos baseados em árvores, sem que abordagens mais complexas superassem consistentemente os métodos tradicionais. No plano substantivo, o estudo aponta como variáveis relevantes a composição racial dos territórios, a condição urbano-rural e o número de domicílios sem acesso

¹ O Programa de Assistência Nutricional Suplementar (*SNAP*, na sigla em inglês) ajuda pessoas de baixa renda a comprar alimentos nutritivos”. Disponível em <https://www.ssa.gov/pubs/EN-05-10101.pdf>. Acesso em 08/02/2026.

a veículos, indicando que características demográficas e estruturais do espaço urbano influenciam a distribuição do risco alimentar.

Ao deslocar a análise para realidades rurais em países em desenvolvimento, Zhao *et al.* (2025) investigaram a insegurança alimentar entre pequenos agricultores na China, com base em dados coletados em 2022 em três províncias (Jilin, Shandong e Hubei), totalizando 1.080 produtores. A insegurança alimentar foi mensurada por meio da ingestão calórica diária per capita, construída a partir de registros alimentares de três dias, e as variáveis foram organizadas em blocos que incluem características socioeconômicas do domicílio, condições de mercado alimentar rural e fatores comunitários. A estratégia metodológica combinou a seleção por LASSO e predição por *Random Forest*, destacando como fatores relevantes a renda domiciliar, o tamanho do mercado alimentar local e a disponibilidade de mercados de agricultores, o que reforça o papel da infraestrutura de abastecimento e do contexto econômico local na determinação da segurança alimentar.

Em uma linha de pesquisa próxima, o estudo apresentado no âmbito do FiSAED 2024² (BIO *Web of Conferences*, 2025) analisou dados de 25.873 domicílios, após procedimentos de limpeza, operacionalizando a insegurança alimentar de forma binária. A modelagem comparou *Random Forest* e *Generalized Random Forest*, com divisão da amostra em conjuntos de treinamento e teste e avaliação por métricas como acurácia, sensibilidade, especificidade e *balanced accuracy*. Os resultados indicam melhor equilíbrio entre sensibilidade e especificidade no modelo GRF, especialmente em contextos de dados desbalanceados. Entre os preditores relevantes, destacam-se características da moradia, escolaridade do chefe do domicílio e indicadores de inclusão financeira, sugerindo que privações estruturais e restrições de acesso a serviços e ativos financeiros estão associadas a maior risco de insegurança alimentar.

Nessa esteira, a literatura tem incorporado o conceito de pobreza energética como uma dimensão relevante na análise da insegurança alimentar. A pobreza energética pode estar associada à ausência de consumo energético suficiente para atender às necessidades essenciais das famílias, como preparo de alimentos, aquecimento, iluminação, refrigeração, acesso à comunicação e lazer. Tal condição impõe restrições materiais que afetam diretamente a

² A Conferência The Frontier in Sustainable Agromaritime and Environmental Development (FiSAED) 2024 é inspirada pelo surgimento do Agromarítimo 4.0, que integra tecnologias avançadas às práticas agrícolas e marítimas com o objetivo de ampliar a sustentabilidade e a eficiência. Disponível em https://www.bio-conferences.org/articles/bioconf/pdf/2025/22/bioconf_FiSAED2024_About.pdf. Acesso 08/02/2026.

organização do consumo doméstico e a capacidade das famílias de garantir uma alimentação adequada.

Dados da Organização Mundial da Saúde indicam que aproximadamente um terço da população mundial ainda utiliza fogões ineficientes, alimentados por querosene ou biomassa — como madeira, estrume animal e resíduos agrícolas — além do carvão, o que ocasiona poluição ambiental prejudicial à saúde das famílias (WHO, 2024). Estudos apontam, ainda, que milhões de domicílios continuam utilizando lenha ou carvão como principal fonte de energia doméstica, o que limita o preparo seguro de alimentos e compromete condições básicas de bem-estar. Ademais, a OMS destaca que a poluição do ar no âmbito doméstico foi responsável por cerca de 3,2 milhões de mortes em 2020, evidenciando a interseção entre restrições energéticas, saúde e qualidade de vida.

Nesse compasso, a pobreza energética e a insegurança alimentar configuram problemas interligados, conforme apontado por diferentes estudos, ao refletirem múltiplas dimensões da exclusão socioeconômica, sobretudo em países em desenvolvimento (François 2022; Kadhim et al., 2024). Parte da literatura recente reconhece que a deficiência energética domiciliar impacta diretamente a segurança alimentar das famílias, não apenas ao pressionar o orçamento doméstico, mas também ao limitar condições materiais fundamentais para o armazenamento, a preparação e o consumo adequado de alimentos, além de restringir o potencial de desenvolvimento humano e econômico (Koirala; Rahut, 2024).

Em continuidade a essa discussão, Silva *et al.* (2025) analisaram dados da POF 2017-2018, abrangendo 48.925 domicílios brasileiros, com o objetivo de identificar a relação causal entre as despesas com eletricidade e a insegurança alimentar. O estudo utilizou modelos *Logit* binário e ordenado generalizado, além do método de variáveis instrumentais (IV-2SLS), tendo o preço médio da eletricidade (R\$/kWh) como instrumento exógeno.

As variáveis analisadas incluíram escolaridade, gênero, idade, cor/raça, número de cômodos, renda e região geográfica. Os resultados indicaram que famílias em situação de vulnerabilidade energética apresentaram, aproximadamente, 43% a 48% mais probabilidade de insegurança alimentar moderada ou grave, revelando que o aumento das despesas com energia reduz o consumo e a qualidade dos alimentos, consolidando um *trade-off* entre energia e alimentação no contexto brasileiro.

A nível internacional, Görbig et al. (2025) complementa essas evidências ao investigar o risco de pobreza energética na Europa, com foco no dilema “aquecer ou comer”. O estudo, baseado em dados de 2020 provenientes de fontes como Eurostat, *OECD* e *FAO*, abrangeu 25 países europeus por meio da *Household Budget Survey*. Utilizando abordagem

quantitativa e multidimensional, com análises estatísticas, regressões e Análise de Componentes Múltiplos (*MCA*, sigla em inglês), os autores identificaram correlação negativa entre renda e despesa energética. Os resultados evidenciam maior vulnerabilidade entre idosos, famílias com crianças e grupos de baixa renda, especialmente em países das regiões dos Bálcãs e do Báltico, onde a necessidade de escolha entre comer ou se aquecer se apresenta com maior intensidade.

Em conjunto, os estudos analisados indicam que a insegurança alimentar se articula a um conjunto amplo de variáveis socioeconômicas, territoriais e institucionais, que extrapolam a renda monetária isoladamente. Aspectos como inserção produtiva, estrutura domiciliar, acesso a mercados, governança, tecnologia e localização espacial interagem de forma complexa na determinação do risco alimentar, o que reforça a necessidade de abordagens analíticas integradas e de políticas públicas capazes de capturar essa multiplicidade de dimensões.

2.2 Relação entre programas sociais e insegurança alimentar

Uma das respostas mais recorrentes, nas últimas décadas, ao enfrentamento da pobreza e de seus desdobramentos, entre eles a insegurança alimentar, tem sido a adoção de políticas de proteção social, em especial transferências de renda e programas afins voltados a famílias em situação de vulnerabilidade. Nesse enquadramento, a literatura sugere que a capacidade de acesso a alimentos (em quantidade e qualidade adequadas) não decorre apenas da disponibilidade física, mas também de restrições de renda, de exposição a choques e de custos incompressíveis do orçamento doméstico, o que recoloca os programas sociais como instrumentos de estabilização do consumo e de redução de riscos de privação alimentar.

Antes de discutir, propriamente, impactos de políticas, é útil notar que a forma de mensurar pobreza pode alterar o diagnóstico sobre quem está exposto à insegurança alimentar e, por consequência, afetar a focalização de programas. Hjelm, Mathiassen e Wadhwa (2016) exploram esse ponto ao comparar medidas de bem-estar baseadas em consumo per capita (fluxo) e índices de riqueza construídos por ativos e características habitacionais (estoque), utilizando bases de dados do Nepal, Malawi, Tanzânia, Uganda e Madagascar.

Os autores mostram que, embora as medidas sejam correlacionadas, não são intercambiáveis: a prevalência de insegurança alimentar declina do grupo mais pobre ao mais rico em todos os países e indicadores, mas a deficiência energética varia mais nitidamente entre quintis de consumo do que entre quintis de riqueza; em contrapartida, desigualdades no peso do gasto alimentar aparecem mais pronunciadas quando estratificadas por riqueza. Resumindo,

o resultado sugere que programas orientados por cadastros “mais leves” (ativos e moradia) podem capturar dimensões relevantes de vulnerabilidade, mas tendem a refletir aspectos distintos daqueles captados por medidas de consumo, exigindo cautela na definição de elegibilidade e na avaliação de resultados em segurança alimentar.

No caso brasileiro, a discussão ganha enlevo quando se observa como transferência as monetárias afetam, direta ou indiretamente, os componentes do bem-estar que condicionam a alimentação. Cavalcanti *et al.* (2013) avaliaram o Programa Bolsa Família (PBF) no Nordeste, com dados da PNAD de 2004 e 2006, aplicando *Propensity Score Matching* para reduzir viés de seleção e, adicionalmente, utilizando o índice FGT para verificar efeitos sobre pobreza e desigualdade entre os pobres. O foco recaiu sobre dois objetivos centrais do programa: alívio imediato da pobreza/desigualdade e estímulo à frequência escolar.

Os resultados indicaram efeito positivo sobre a renda das famílias beneficiárias e apontaram impacto relevante sobre a frequência escolar quando comparadas às famílias elegíveis, com indicação de que o aumento do valor do benefício exerce influência mais expressiva do que a mera ampliação da cobertura. Ainda que o desfecho do estudo não seja a insegurança alimentar, a evidência é coerente com a hipótese de que programas de transferência elevam capacidade de compra e reduzem fragilidades de curto prazo.

Entretanto, a insegurança alimentar não depende apenas da renda monetária, mas também das condições estruturais de inserção produtiva e da composição das fontes de rendimento no domicílio. Nesse sentido, Cruz, Silva e Araújo (2025) examinaram a situação de famílias rurais do Nordeste brasileiro com base na PNAD 2013, utilizando a Escala Brasileira de Insegurança Alimentar como indicador de insegurança alimentar e aplicando *Propensity Score Matching* (com pareamentos Kernel e *Nearest Neighbor*) para mensurar o efeito da escolha ocupacional por atividades não agrícolas sobre o desfecho.

A hipótese testada foi a de que a diversificação ocupacional, especialmente em um contexto de vulnerabilidade climática no meio rural, poderia reduzir a propensão à insegurança alimentar ao elevar a renda do trabalho. Os autores relatam que as atividades não agrícolas contribuem para uma situação mais favorável de segurança alimentar e nutricional entre domicílios rurais nordestinos, com resultados estatisticamente significativos e boa qualidade de pareamento, sugerindo que políticas públicas que estimulem oportunidades não agrícolas (ou que estabilizem renda em períodos adversos) podem atuar como complemento relevante aos programas tradicionais de assistência.

Ao lado de renda e ocupação, outro mecanismo importante na discussão sobre segurança alimentar diz respeito à pressão de gastos obrigatórios sobre o orçamento doméstico.

Nessa linha, o estudo realizado por Silva *et al.* (2025) investiga o caso brasileiro com microdados da POF 2017–2018 (amostra final de 48.925 domicílios, representando cerca de 66,8 milhões de domicílios), explorando a hipótese de que o aumento dos gastos com eletricidade pode deslocar recursos do orçamento e ampliar a insegurança alimentar.

Os resultados apontam relação positiva e estatisticamente significativa entre preços de eletricidade e insegurança alimentar. A evidência permanece consistente com associação reforçada em grupos mais vulneráveis (por exemplo, domicílios chefiados por mulheres e por pessoas não brancas), o que sugere que políticas de alívio tarifário, quando integradas a estratégias de proteção social, podem repercutir sobre a segurança alimentar ao reduzir a restrição orçamentária das famílias.

Esses achados ajudam a sustentar uma leitura mais ampla: programas sociais podem atuar tanto pelo canal direto da renda quanto por canais indiretos que reduzem gastos de difícil redução e ampliam margem de consumo alimentar. Embora o desenho institucional varie entre países, a literatura sobre transferências monetárias também tem buscado identificar efeitos causais com estratégias de identificação mais rígidas.

Adofo e Koomson (2025), por exemplo, analisam a experiência sul-africana do *Child Support Grant (CSG)*, “um dos maiores programas de subsídios sociais na África”, valendo-se de mudanças nos critérios etários de elegibilidade (2010 e 2012) como fonte de variação exógena, em um arranjo próximo a uma descontinuidade “fuzzy” estimada via IV. Com dados de três ondas do *National Income Dynamics Panel Study*, os autores estimam que receber o benefício reduz a probabilidade de pobreza energética em 7,5 pontos percentuais, com efeitos mais pronunciados em domicílios chefiados por mulheres e por população negra, operando, entre os mecanismos, via acumulação de ativos. Ainda que o desfecho não seja a insegurança alimentar, o estudo é relevante porque explicita um mecanismo de fortalecimento de bem-estar (aumento de capacidade de investimento e de poupança) que tende a reduzir a exposição a choques e, em cenários de restrição orçamentária, pode influenciar a estabilidade do consumo de alimentos.

2.3 Estudos sobre insegurança alimentar utilizando modelos de *machine learning*

No campo de aplicações voltadas ao diagnóstico e à classificação da insegurança alimentar, Okolie *et al.* (2025), analisando a realidade norte americana, empregaram uma base pública de grande abrangência (*USDA Food Access Research Atlas*) para classificar grupos censitários, segundo critérios de restrição de acesso a alimentos, em especial para populações

de baixa renda. A comparação envolveu Regressão Logística, *Random Forest* e *XGBoost*, com desempenho superior dos modelos baseados em árvores Ping quando comparados ao padrão linear. No conjunto de variáveis interpretadas, os autores destacam determinantes como taxa de pobreza, renda mediana familiar, participação no SNAP e acesso a veículo, reforçando a utilidade de modelos interpretáveis para orientar intervenções em desigualdades de acesso.

Ainda no contexto dos Estados Unidos, Chilukuri (2025) adota um enfoque metodológico ao utilizar o *Food Access Research Atlas (FARA)* de 2019 para avaliar o desempenho comparativo de diferentes famílias de algoritmos na classificação de *census tracts*, segundo o critério de “baixo acesso a comida”, definido pelo *USDA*³. O autor testa cinco modelos, incluindo Regressão Logística, *Random Forest*, *XGBoost* e um modelo tabular baseado em atenção (TabNet), com o objetivo de verificar se abordagens de maior complexidade seriam capazes de superar métodos tradicionais em dados censitários estruturados.

Os resultados indicam desempenho superior dos algoritmos baseados em árvores, especialmente *XGBoost* e *Random Forest*, enquanto o TabNet não apresentou ganhos preditivos relevantes, sugerindo que a sofisticação do modelo não se traduz, necessariamente, em melhor capacidade de classificação nesse tipo de aplicação. Como complemento interpretativo, o estudo aponta que variáveis como a proporção de população branca, o número de domicílios sem acesso a veículo e a condição urbano/rural figuram entre os atributos mais informativos, funcionando como evidência de coerência substantiva dos modelos com melhor desempenho.

Por sua vez, analisando o contexto chinês, Zhao *et al.* (2025) propõem um desenho aplicado a pequenos agricultores em áreas rurais da China, com dados coletados em 2022 em três províncias (Jilin, Shandong e Hubei), totalizando 1.080 produtores. O estudo mensura insegurança alimentar por uma medida objetiva de consumo - ingestão calórica diária per capita e seu desvio do recomendado - construída com base em registro alimentar de 3 dias, bem como organiza os preditores em quatro blocos: características do decisor alimentar, fatores socioeconômicos do domicílio, condições de mercado alimentar rural e fatores comunitários.

O estudo combina seleção de variáveis por LASSO e predição por *Random Forest*, com divisão amostral em treinamento (70%) e validação (30%), além de comparação com OLS.

³ USDA (sigla em inglês) significa Departamento de Agricultura dos Estados Unidos, sendo responsável por “liderar em questões relacionadas à alimentação, agricultura, recursos naturais e assuntos correlatos”. Disponível em <https://www.usa.gov/agencies/u-s-department-of-agriculture>. Acesso em 08/02/2026.

Nos resultados, observa-se que o *Random Forest* alcança desempenho superior ao LASSO e ao modelo de referência, com acurácia de 0,657 e recall médio de 0,893 para identificar domicílios com pobreza calórica. Em contrapartida, os autores ponderam que medidas como precisão e F1 tornam mais visível a dificuldade de separar adequadamente os casos positivos, sugerindo um cenário em que a sensibilidade do modelo supera sua capacidade de identificar falsos positivos. Ao final, os autores enfatizam, entre variáveis centrais, renda domiciliar, tamanho do mercado alimentar rural e disponibilidade de mercados de agricultores como elementos-chave do desempenho preditivo.

Em uma linha próxima, considerando a classificação de domicílios com e sem insegurança alimentar, a partir de variáveis socioeconômicas e de infraestrutura, o estudo conduzido por Fransiska *et al.* (2025) e apresentado no contexto do FiSAED 2024 trabalhou com 25.873 domicílios na Indonésia, após procedimento de limpeza de dados, operacionalizando a variável resposta de forma binária (0 = sem insegurança; 1 = com insegurança). A modelagem comparou *Random Forest (RF)* e *Generalized Random Forest (GRF)*, com divisão em 70% treinamento e 30% teste e avaliação por acurácia, sensibilidade, especificidade e *balanced accuracy*, repetida em 100 iterações.

Nos principais resultados, embora o *RF* tenha apresentado tendência a maior acurácia e sensibilidade, o *GRF* se destacou por especificidade superior e, sobretudo, por melhor *balanced accuracy*, sendo apontado como mais adequado em cenário desbalanceado, justamente por equilibrar melhor a detecção de casos positivos e negativos. Entre os preditores importantes listados, aparecem variáveis de condição da moradia (tamanho da casa; tipo de piso; saneamento), além de características do chefe (idade e escolaridade) e indicadores de inclusão financeira (conta poupança), o que reforça o caráter estrutural da insegurança alimentar no nível domiciliar.

Adicionalmente, mudando o foco de domicílios e extratos sociais para abordagens voltadas ao monitoramento e antecipação de crises, Busker, Hurk e Homberg (2023) apresentam um modelo de *XGBoost* destinado a prever crises de insegurança alimentar no Chifre da África, região historicamente marcada por secas e episódios recorrentes de crise. O estudo menciona que o modelo foi treinado com mais de 20 bases de dados e utiliza como dados de referência estimativas correntes, buscando prever crises com antecedência de até 12 meses; como resultado, destaca-se maior capacidade de capturar dinâmicas com até 3 meses de antecedência, especialmente em zonas pastoris e agropastoris. Além disso, há uma comparação com as

previsões de 8 meses do *FEWS NET*⁴ (período de teste 2019–2022), na qual se sustenta desempenho semelhante em certas zonas de subsistência, embora inferior em áreas predominantemente agrícolas, apontando, assim, tanto o potencial quanto os limites práticos da automação preditiva.

Na mesma seara temática, Westerveld *et al.* (2019) desenvolveram um estudo aplicado à Etiópia, com o objetivo explícito de criar um modelo de *machine learning*, também utilizando XGBoost, para prever transições em insegurança alimentar, utilizando variáveis escaláveis, a exemplo de imagens de satélite, e comparando o desempenho do modelo a *benchmarks* operacionais. Os autores reportam que, após procedimentos de reamostragem, engenharia de atributos e ajuste de hiper parâmetros, o modelo alcançou F1 macro $\approx 0,526$, aproximando-se do benchmark referido (F1 macro $\approx 0,637$), além de apresentar melhor capacidade de identificar movimentos positivos nos níveis de segurança alimentar em comparação ao modelo de referência. Como implicação metodológica, o estudo ressalta que variáveis relacionadas a condições climáticas e a padrões de uso e cobertura do solo tendem a assumir papel relevante no desempenho preditivo, bem como a necessidade de incorporar dependências espaciais entre zonas de subsistência para o aprimoramento das estimativas.

Em outra vertente, mas ainda usando modelos para mensurar e antecipar condições conjuntas, Saragih e Wibowo (2025) discutem a previsão do Índice de Segurança Alimentar na Indonésia com base em dados do *Food Security and Vulnerability Atlas (FSVA)*⁵ no período de 2018 a 2024, em nível de distritos e cidades, totalizando 3.598 registros. O estudo compara Regressão Linear Múltipla, LASSO, *Random Forest*, XGBoost, *Support Vector Regression* e modelos de conjuntos, avaliando desempenho por R², RMSE e MAE, indicando o *XGBoost* como o método mais acurado. Além disso, os autores realizaram validação cruzada estratificada por ano e apontaram discussões sobre relevância de indicadores (inclusive variáveis

⁴ A Rede de Sistemas de Alerta Precoce de Fome (*FEWS NET*, sigla em inglês) “tem como principal objetivo fornecer alertas precoces de crises alimentares aos tomadores de decisão, bem como por organizações como a Federação Internacional da Cruz Vermelha e do Crescente Vermelho e o Escritório das Nações Unidas para a Coordenação de Assuntos Humanitários, que têm um mandato mais amplo, mas que também inclui a segurança alimentar”. Disponível em <https://www.pnas.org/doi/10.1073/pnas.2119399119>. Acesso em 08/02/2026.

⁵ O Atlas de Segurança e Vulnerabilidade Alimentar (*FSVA*, sigla em inglês) é um mapa temático que apresenta a visualização geográfica das áreas vulneráveis à insegurança alimentar. O *FSVA* é elaborado com o objetivo de fornecer informações precisas, abrangentes e sistematizadas sobre a segurança alimentar, a fim de apoiar ações de prevenção e enfrentamento da insegurança alimentar e nutricional, bem como de orientar e subsidiar os tomadores de decisão na formulação de programas, políticas públicas e intervenções, tanto em nível central quanto regional/local. Disponível em <https://data.acehprov.go.id/dataset/peta-fsva-provinsi-tahun-2022>. Acesso em 08/02/2026.

relacionadas a acesso à eletricidade), ao mesmo tempo em que apresentaram projeções para a média nacional do índice em 2025 e 2026.

No contexto brasileiro, Silva *et al.*, 2025 trabalharam com um levantamento conduzido pela Secretaria de Proteção Social do Ceará (SPS), no âmbito do Cartão Mais Infância Ceará (CMIC)⁶, que, até agosto de 2022, havia alcançado 184 municípios e reunido informações de quase 50 mil famílias em condição de elevada vulnerabilidade. A mensuração do desfecho foi realizada pela Escala Brasileira de Insegurança Alimentar, com modelagem por métodos de *machine learning* voltados a identificar fatores associados a diferentes níveis de insegurança, tomando como parâmetros explicativos aspectos de acesso a alimentos, emprego e renda, habitação e serviços públicos. Os autores reportam um desempenho do melhor modelo em torno de 0,75 de acurácia e $F1 \approx 0,80$, com capacidade de diferenciar insegurança severa de não severa, além de ressaltarem o peso explicativo de fatores como renda e condições habitacionais.

⁶ O CMIC é um programa do Governo do Estado do Ceará que “desenvolve ações para combater a desigualdade e promover a justiça social e o desenvolvimento integral e integrado da criança”. Disponível em <https://www.sps.ce.gov.br/institucional/secretarias-executivas/infancia-familia-e-combate-a-fome/mais-infancia-ceara-destaque/>. Acesso em 08/02/2026.

3. METODOLOGIA

3.1 Base de dados

A base de dados utilizada foi a Pesquisa Nacional por Amostra de Domicílios Contínua – PNADC para o 4º trimestre de 2024, realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE), pois nesse trimestre se encontra o suplemento referente a Segurança Alimentar. Os domicílios serão classificados em situação de segurança alimentar, situação em que os moradores possuem acesso aos alimentos em termos de quantidade suficiente como em qualidade adequada. Os programas assistenciais utilizados serão: Benefício Assistencial de Prestação Continuada (BPC-LOAS), Programa Bolsa Família e outros programas sociais do governo (como seguro-desemprego, entre outros).

Utilizou-se o *software* Stata, versão 17, para o tratamento do dados, sendo que a extração foi feita com o pacote *Data Zoom* (2023). Foram excluídas da amostra as pessoas cuja condição no domicílio era pensionista, empregado doméstico e parente de empregado doméstico. A base de dados é composta por crianças de 5 a 13 anos de idade e contém 131.673 observações.

Com os microdados da PNADC 2024, computou-se a variável de interesse que representa a situação de insegurança alimentar do domicílio e conjunto de variáveis explicativas contempla características socioeconômicas, demográficas e regionais dos domicílios, atributos da pessoa de referência do domicílio, número de crianças e idosos, pobreza energética, renda, benefícios sociais, sexo, dentre outras. A inclusão de indicadores regionais e de pobreza energética permite capturar as diversidades regionais e vulnerabilidades estruturais. As variáveis selecionadas alinham-se com a literatura, que demonstra desempenho superior de algoritmos de aprendizado de máquina baseados em árvores de decisão como *XGBoost* e *Random Forest* quando alimentados com conjuntos preditores socioeconômicos e estruturais (Zhao *et al.*, 2025; Okolie *et al.*, 2025; Chilukuri, 2025).

Para o cálculo da variável resposta (y_i), definiu-se uma variável binária construída pela seguinte função indicadora:

$$y_i = \begin{cases} 1, & \text{se o domicílio } i \text{ está em insegurança alimentar} \\ 0, & \text{se o domicílio } i \text{ está em segurança alimentar} \end{cases} \quad (1)$$

A variável dependente, neste estudo, é uma *dummy* de insegurança alimentar construída a partir da Escala Brasileira de Insegurança Alimentar (EBIA). Inicialmente, os domicílios da amostra foram classificados segundo a presença de moradores menores de 18

anos, tendo em vista que a EBIA adota números de itens e pontos de corte distintos conforme a composição etária do domicílio.

As respostas às questões da EBIA foram recodificadas de forma binária e somadas para a obtenção de um escore domiciliar, variando de 0 a 14 para domicílios com menores de 18 anos e de 0 a 8 para domicílios sem menores de 18 anos. Com base nesse escore, os domicílios foram classificados em segurança alimentar ou insegurança alimentar leve, moderada ou grave, seguindo os pontos de corte oficiais da EBIA conforme o Quadro 1.

Quadro 1 - Pontos de corte da EBIA segundo presença de menores de 18 anos no domicílio

Classificação	Com menores de 18 anos	Sem menores de 18 anos
Segurança alimentar	0	0
Insegurança alimentar leve	1 a 5	1 a 3
Insegurança alimentar moderada	6 a 9	4 a 5
Insegurança alimentar grave	10 a 14	6 a 8

Fonte: elaboração própria

O Quadro 2 mostra a definição das variáveis explicativas utilizadas nos modelos implementados.

Quadro 2: Definição das variáveis explicativas

Variável	Definição
aluguel	<i>Dummy</i> que assume valor 1 se o domicílio é alugado e 0, caso contrário
pobre_mult33	<i>Dummy</i> que assume valor 1 se o domicílio está em situação de pobreza energética multidimensional e 0, caso contrário ⁷
norte	<i>Dummy</i> que assume valor 1 se o domicílio está situado na região Norte e 0, caso contrário
nordeste	<i>Dummy</i> que assume valor 1 se o domicílio está situado na região Nordeste e 0, caso contrário
sudeste	<i>Dummy</i> que assume valor 1 se o domicílio está situado na região Sudeste e 0, caso contrário
centro_oeste	<i>Dummy</i> que assume valor 1 se o domicílio está situado na região Centro Oeste e 0, caso contrário
urbana	<i>Dummy</i> que assume valor 1 se o domicílio está situado na zona urbana e 0, caso contrário
lrdpc	Logaritmo da renda domiciliar per capita (rdpc) construída como $\text{lrdpc} = \log(\text{rdpc}+1)$
dom_benef	<i>Dummy</i> que assume valor 1 se alguém no domicílio recebe benefício social e 0, caso contrário

⁷ Explicitado na próxima seção

Variável	Definição
dom_aposent	<i>Dummy</i> que assume valor 1 se alguém no domicílio recebe aposentadoria e 0, caso contrário
chefe_fem	<i>Dummy</i> que assume valor 1 se o chefe do domicílio é do sexo feminino e 0, caso contrário
chefe_branc	<i>Dummy</i> que assume valor 1 se o chefe do domicílio é branco e 0, caso contrário
chefe_idad	Idade do chefe do domicílio medida em anos
chefe_estudo	Anos de estudo do chefe do domicílio
chefe_ocupad	<i>Dummy</i> que assume valor 1 se o chefe do domicílio está ocupado no mercado de trabalho e 0, caso contrário
chefe_previd	<i>Dummy</i> que assume valor 1 se o chefe do domicílio contribui para previdência e 0, caso contrário
chefe_agric	<i>Dummy</i> que assume valor 1 se o chefe do domicílio trabalha no setor agrícola e 0, caso contrário
adult_estud_dom	Anos de estudo médios dos demais adultos do domicílio
fam_estend	<i>Dummy</i> que assume valor 1 se a espécie da unidade doméstica é estendida e 0, caso contrário
fam_comp	<i>Dummy</i> que assume valor 1 se a espécie da unidade doméstica é composta e 0, caso contrário
n_adult_sem_ocup	Número de adultos sem ocupação no domicílio
dom_criança	Número de crianças no domicílio
dom_idoso	Número de idosos no domicílio

Fonte: elaboração própria a partir dos dados da PNADC 2024

3.2 Pobreza energética multidimensional

Esta seção descreve a metodologia empregada para identificar os domicílios em situação de pobreza energética multidimensional, baseada no arcabouço desenvolvido por Nussbaumer, Bazilian e Modi (2012), sendo esta uma extensão do índice de pobreza multidimensional proposto por Alkire e Foster (2011).

Considere uma população composta por n domicílios e d indicadores que determinam serviços de energia. Represente por y_{ij} a dotação do domicílio i ($i = 1, \dots, n$) no indicador j ($j = 1, \dots, d$). A matriz de dotações é dada por $Y = [y_{ij}]$, com dimensão $n \times d$. O vetor linha $Y_i = [y_{i1}, \dots, y_{id}]$, de dimensão $1 \times d$, mostra as dotações do domicílio i nos diferentes indicadores.

Suponha que cada indicador tenha um peso associado w_j , tal que:

$$\sum_{j=1}^d w_j = 1, \quad 0 \leq w_j \leq 1 \quad (2)$$

A pobreza energética multidimensional pode ser identificada através de duas etapas de corte. Inicialmente, deve ser considerada uma linha de corte, Z_j , para cada um dos d

indicadores. A partir disso, formula-se a matriz de privação, $G = [g_{ij}]$, cuja dimensão é $n \times d$. A condição de privação do domicílio i no indicador j pode ser escrita como:

$$g_{ij} = \begin{cases} w_j, & \text{se } y_{ij} \leq Z_j \text{ (domicílio privado no indicador } j) \\ 0, & \text{se } y_{ij} > Z_j \text{ (domicílio não privado no indicador } j) \end{cases} \quad (3)$$

Em seguida, computa-se a soma de privações ponderada (C_i) do domicílio i :

$$C_i = \sum_{j=1}^d g_{ij} \quad (4)$$

Na segunda etapa, emprega-se uma segunda linha de corte para identificar os domicílios que vivem sob pobreza energética multidimensional. Um domicílio i é pobre energético multidimensional se sua soma de privações ponderada, C_i , supera a linha de corte, k . No presente trabalho, utilizou-se $k = 33\%$. Assim, a variável indicadora que identifica se o domicílio está em situação de pobreza energética multidimensional é expressa pela seguinte função:

$$pobre_{mult33i} = \begin{cases} 1, & \text{se } C_i \geq k \text{ (domicílio pobre energético multidimensional)} \\ 0, & \text{se } C_i < k \text{ (domicílio não pobre energético multidimensional)} \end{cases} \quad (5)$$

O Quadro 3 apresenta os indicadores selecionados, condições de privação e pesos para a identificação dos domicílios em condição de pobreza energética multidimensional.

Quadro 3 - Indicadores, condições de privação e pesos para a pobreza energética multidimensional

Indicador	Condição de privação	Peso
Energia	Domicílio não possui acesso à energia elétrica	1/6
Combustível	Pessoas no domicílio utilizam lenha ou carvão para preparar as refeições	1/6
Refrigeração	Ausência de geladeira no domicílio	1/6
Comunicação	Não há telefone celular, nem fixo no domicílio	1/6
Entretenimento	Ausência de TV no domicílio	1/6
Informação	Não há computador no domicílio	1/6

Fonte: elaboração própria elaboração própria a partir dos dados da PNADC 2024.

3.3 Modelo XGBoost

O algoritmo *Extreme Gradient Boosting* (*XGBoost*), proposto por Chen e Guestrin (2016), é utilizado para estimar a probabilidade de insegurança alimentar domiciliar, sendo um modelo de aprendizado de máquina supervisionado fundamentado em árvores de decisão construídas de modo sequencial. Ele combina vários modelos fracos em sequência para construir um modelo forte com o intuito de minimizar o erro residual de iterações anteriores.

O modelo foi escolhido em virtude da sua elevada velocidade de processamento e desempenho superior em comparação a outros algoritmos de *machine learning* em vários

campos de aplicação, incluindo estudos de previsão de secas e segurança alimentar (Zhang *et al.*, 2023; Foini, 2023).

Além disso, foi considerada sua capacidade de capturar relações complexas e não lineares entre as variáveis explicativas, aspecto relevante na análise de fenômenos multidimensionais como a insegurança alimentar. O método baseia-se em técnicas de *gradient boosting* aplicadas a árvores de decisão, permitindo a construção de modelos preditivos por meio da combinação sequencial de árvores, nas quais cada nova árvore é ajustada para reduzir os erros das anteriores. Conforme Chen e Guestrin (2016), o XGBoost tem apresentado desempenho superior em diversos problemas de classificação, sendo amplamente utilizado em aplicações empíricas e competições de machine learning, o que evidencia sua robustez e capacidade de generalização.

Além disso, o modelo incorpora uma função objetivo regularizada, que penaliza a complexidade das árvores e contribui para a mitigação de problemas de sobreajuste, favorecendo maior estabilidade das estimativas. O algoritmo também apresenta elevada eficiência computacional, com suporte a paralelização, tratamento de dados esparsos e otimizações de memória, características importantes para o processamento de bases de grande dimensão e com variáveis heterogêneas, como é o caso dos microdados da PNADC. Adicionalmente, estratégias como *shrinkage* e amostragem de variáveis contribuem para aprimorar o desempenho preditivo do modelo (Chen; Guestrin, 2016).

Por fim, embora modelos baseados em árvores apresentem maior complexidade estrutural em comparação a modelos lineares, a utilização de métodos de interpretabilidade, como o SHAP, permite decompor as previsões em contribuições individuais das variáveis explicativas, assegurando consistência teórica na atribuição de importância (Lundberg; Lee, 2017). Dessa forma, a escolha do XGBoost neste estudo permite conciliar elevado desempenho preditivo com capacidade de interpretação dos resultados, alinhando-se aos objetivos da análise empírica proposta.

Detalhando mais o modelo, considere que para um conjunto de dados com n observações e p características, $D = \{(x_i, y_i)\}$, com $(|D| = n, x_i \in \mathbb{R}^p, y_i \in \mathbb{R})$, a previsão final, após K iterações, pode ser representada como a soma das contribuições de cada árvore $f_k(x_i)$:

$$\hat{y}_i = \varphi(x_i) = \sum_{t=1}^T f_t(x_i), f_t \in \mathcal{F} \quad (6)$$

Onde x_i representa o vetor de características e y_i a variável resposta.

Em que $\mathcal{F} = \{f(x) = w_{q(x)}\}$ ($q: \mathcal{R}^p \rightarrow T$, $w \in \mathcal{R}^T$) representa o espaço das árvores de regressões. O parâmetro q denota a estrutura de cada árvore que mapeia uma observação para o índice de folha correspondente e T é número de folhas da árvore. Cada f_k corresponde a uma estrutura de árvore independente q e a um vetor de pesos w .

O *XGBoost* minimiza uma função objetivo regularizada (L):

$$L(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{t=1}^T \Omega(f_t) \quad (7)$$

Essa função denominada Modelo de Agrupamento de Árvores (*Tree Ensemble Model*) é composta por dois componentes:

- Função de perda $l(y_i, \hat{y}_i)$: Uma função de perda diferenciável e convexa que quantifica a discrepância entre as observações reais \hat{y}_i e as previsões y_i ;
- Termo de regularização $\Omega(f_t)$: Define a complexidade do modelo para prevenir o sobreajuste (*overfitting*), dado por:

$$\Omega(f_t) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (8)$$

Em que γ e λ são parâmetros de regularização, T é número de folhas da árvore e w são os pesos das folhas. Dada a natureza aditiva do modelo, a otimização é realizada de forma iterativa. Na etapa t , o modelo busca a função f_t que melhor minimiza o objetivo;

$$L^t = \sum_{i=1}^n l(y_i, \hat{y}_i^{t-1} + f_t(x_i)) + \Omega(f_t) \quad (9)$$

O *XGBoost* se diferencia dos demais algoritmos de *boosting* por utilizar a expansão de Taylor de segunda ordem para aproximar a função de perda, permitindo uma otimização mais eficiente. (Chen; Guestrin, 2016; Friedman, 2001)

O desempenho preditivo do modelo *XGBoost* é condicionado por um grupo de hiper parâmetros, os quais governam a estrutura das arvores de decisão e a dinâmica estocástica durante o processo de estimação. (Chen; Guestrin, 2016; Chen *et al*, 2026). Em um contexto com a variável dependente desbalanceada (como é o caso de domicílios em insegurança alimentar com baixa proporção de resultados positivos), aplicou-se o hiper parâmetro *scale_pos_weight*⁸. Este parâmetro opera o reponderação das observações positivas na função de perda, permitindo elevar a sensibilidade do modelo na identificação da insegurança

⁸ $scale_pos_weight = \frac{\text{total de domicílios em segurança alimentar}}{\text{total de domicílios em insegurança alimentar}}$

alimentar, preservando a distribuição original dos dados e evitando vieses. Não foram aplicadas técnicas de *oversampling* como *SMOTE* (*Synthetic Minority Over-sampling Technique*), que gera novos exemplos sintéticos da classe minoritária para equilibrar a base, para evitar o aumento do tamanho da amostra. O modelo foi treinado sem o uso de pesos de expansão da PNAD Contínua por meio do *software* R-4.5.2 para Windows.

No arcabouço de modelos de aprendizado de máquina, a explicabilidade das previsões assume relevância similar a acurácia do algoritmo. Todavia, diversos algoritmos de *machine learning*, como o *XGBoost*, atuam como “caixas-pretas” devido à sua complexidade estrutural e interpretabilidade dos resultados. Para lidar com esse desafio, Lundberg e Lee (2017) propõem uma abordagem para interpretação de previsões denominado *SHAP* (*Shapley Additive exPlanations*). Os valores *SHAP* são calculados para cada observação e indicam a magnitude e a direção do efeito de cada característica sobre a previsão. A análise da distribuição desses valores permite detectar a variabilidade dos efeitos e a presença de relações não lineares complexas no processo preditivo.

O *SHAP* médio é uma medida derivada da agregação das contribuições marginais individuais de todas as observações do conjunto de dados. Formalmente, para cada atributo explicativo, computa-se a média aritmética dos valores absolutos dos seus respectivos valores *SHAP* consoante sugerido por Lundberg e Lee (2017). Essa métrica permite mensurar a magnitude média do impacto de cada variável sobre a previsão final, neutralizando a direção da contribuição (positiva ou negativa). Dessa forma, essa medida fornece um arcabouço comparativo para o ordenamento da relevância relativa dos preditores, destacando aqueles que mais influenciam na variabilidade do modelo.

Cabe salientar que as variáveis mais importantes em termos do *SHAP* médio não são necessariamente aquelas que apresentam os maiores impactos ao longo da distribuição. Algumas variáveis podem causar um grande impacto em um pequeno grupo de pessoas, e outras podem ter um efeito menor em cada caso individual, mas por afetarem quase todo mundo da mesma forma, acaba sendo mais importante na média *SHAP* (*Shapley Additive exPlanations*) geral, apresentando um efeito generalizado e constante.

3.4 Análise do modelo de previsão

A etapa seguinte à construção do modelo consiste em examinar o seu comportamento preditivo, passo fundamental para medir a capacidade de generalização do algoritmo diante dos dados não observados durante o treinamento. Quando um modelo

apresenta desempenho superior nos dados do treinamento, no entanto, demonstra performance insatisfatória no processamento do teste, tem-se um cenário de sobreajuste (*overfitting*), comprometendo seu poder analítico e preditivo. (Chen; Guestrin, 2016; Hastie, 2009)

Para viabilizar essa análise, constrói-se a matriz de confusão (Quadro 4), bidimensional que compara o resultado da previsão com o que foi realmente observado, apontando os acertos (diagonal principal da matriz) e os erros (diagonal secundária da matriz) do modelo.

Quadro 4: Matriz de confusão

	Classe Prevista	
Classe observada	Segurança alimentar ($\hat{y} = 0$)	Insegurança alimentar ($\hat{y} = 1$)
Segurança alimentar ($y = 0$)	Verdadeiro negativo (VN)	Falso positivo (FP)
Insegurança alimentar ($y = 1$)	Falso negativo (FN)	Verdadeiro positivo (VP)

Fonte: Elaboração própria elaboração própria a partir dos dados da PNADC 2024.

Com base nessas comparações entre os valores previstos e observados dessa matriz, é possível calcular algumas medidas de desempenho do modelo:

- *Acurácia*: é uma performance geral do modelo e mostra quantas previsões o modelo acertou em relação ao total.

$$acurácia = \frac{VP + VN}{Total}$$

- *Precisão*: proporção de previsões corretas de uma categoria em relação a todas as previsões feitas dessa categoria.

$$precisão = \frac{VP}{VP + FP}$$

- *Recall ou sensibilidade*: proporção de previsões corretas da categoria alvo (VP) em relação a soma dos verdadeiros positivos com os falsos negativos.

$$sensibilidade = \frac{VP}{VP + FN}$$

- *F1-Score*: média harmônica entre precisão e sensibilidade, representando em um número único a qualidade geral do modelo.

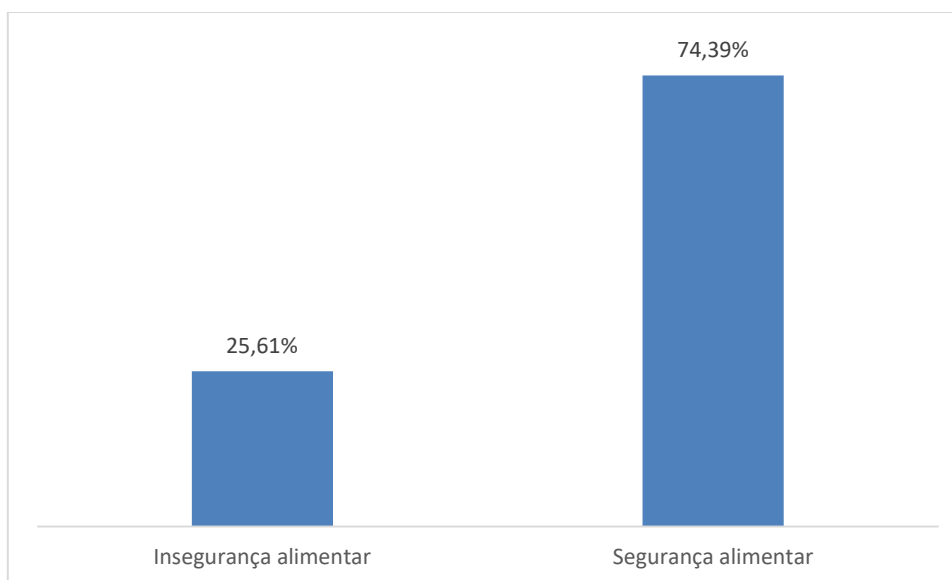
$$F1 = \frac{2(precisão)(sensibilidade)}{(precisão + sensibilidade)}$$

4. RESULTADOS

4.1 Estatísticas descritivas

A Figura 1 mostra a incidência da insegurança alimentar na base de dados, indicando o desbalanceamento entre as duas classes. O percentual de domicílios em situação de insegurança alimentar é de 25,61%, enquanto 74,39% referem-se aos domicílios fora dessa condição.

Figura 1 - Incidência da insegurança alimentar



Fonte: elaboração própria com base nos dados da PNAD Contínua 2024.

A Tabela 1 compara as médias e desvios-padrão das variáveis entre domicílios em situação de segurança e de insegurança alimentar, mostrando diferenças entre os dois grupos. Em termos de distribuição regional, há uma grande concentração de domicílios em insegurança alimentar na região Nordeste (45,5%).

Os domicílios em insegurança alimentar apresentam piores condições socioeconômicas. A incidência de pobreza energética multidimensional é substancialmente mais elevada nesse grupo (35,7% contra 22,1%), assim como a dependência de benefícios sociais (47,3% versus 20,4%). Além disso, a renda domiciliar *per capita* (em logaritmo) é consideravelmente menor entre os domicílios em insegurança alimentar, reforçando o papel central da restrição de renda. Este resultado corrobora com os estudos de Okolie *et al.* (2025) e Zhao *et al.* (2025).

No que se refere às características do domicílio, há uma maior presença de crianças nos domicílios em insegurança alimentar, enquanto a presença de idosos é mais frequente entre os domicílios em segurança alimentar. Ademais, domicílios com insegurança alimentar apresentam maior número de adultos sem ocupação, o que sinaliza maior vulnerabilidade no mercado de trabalho.

As características do chefe do domicílio também se diferenciam entre os grupos. Domicílios em insegurança alimentar são mais frequentemente chefiados por mulheres, por pessoas não brancas, com menor escolaridade média e menor taxa de ocupação. A participação do chefe em regimes previdenciários é significativamente menor entre os domicílios em insegurança alimentar, o que indica menor proteção social associada ao trabalho formal ou à aposentadoria. Tais achados convergem com o trabalho de Fransiska *et al.* (2025).

Tabela 1: Estatísticas descritivas das variáveis independentes

Variável	Segurança alimentar		Insegurança alimentar		Total	
	Média	Desvio padrão	Média	Desvio padrão	Média	Desvio padrão
aluguel	0,1685	0,3743	0,1839	0,3874	0,1724	0,3778
pobre_mult33	0,2209	0,4149	0,3571	0,4792	0,2558	0,4363
norte	0,1056	0,3073	0,1623	0,3687	0,1201	0,3251
nordeste	0,2817	0,4499	0,4551	0,4980	0,3261	0,4688
sudeste	0,2977	0,4572	0,2103	0,4075	0,2753	0,4467
centro_oeste	0,1047	0,3062	0,0827	0,2754	0,0991	0,2987
urbana	0,7494	0,4334	0,7006	0,4580	0,7369	0,4403
lrdpc	7,3139	1,3077	6,5616	1,3800	7,1213	1,3666
dom_benef	0,2041	0,4031	0,4731	0,4993	0,2730	0,4455
dom_aposent	0,3642	0,4812	0,2823	0,4501	0,3432	0,4748
chefe_fem	0,5005	0,5000	0,5991	0,4901	0,5258	0,4993
chefe_branc	0,4341	0,4956	0,2545	0,4356	0,3881	0,4873
chefe_idad	51,4902	16,5967	49,3128	15,4885	50,9327	16,3477
chefe_estudo	9,6276	4,9153	7,8525	4,6927	9,1731	4,9206
chefe_ocupad	0,5892	0,4920	0,4878	0,4999	0,5632	0,4960
chefe_previd	0,3837	0,4863	0,2076	0,4056	0,3386	0,4732
chefe_agric	0,0886	0,2841	0,0860	0,2803	0,0879	0,2831
adult_estud_dom	10,2363	4,3546	8,6565	4,2033	9,8436	4,3711
fam_estend	0,1380	0,3450	0,1820	0,3858	0,1493	0,3564
fam_comp	0,0099	0,0991	0,0100	0,0993	0,0099	0,0991
n_adult_sem_ocup	0,0501	0,2406	0,1274	0,3873	0,0699	0,2874
dom_criança	0,4244	0,7488	0,6503	0,9706	0,4822	0,8173
dom_idoso	0,5545	0,7562	0,4367	0,6769	0,5244	0,7385

Fonte: elaboração própria com base nos dados da PNAD Contínua 2024.

4.2 Modelagem e desempenho preditivo

A base de dados foi estratificada de forma aleatória em amostra de treinamento (80% das observações) e amostra de teste (20% das observações), garantindo a avaliação da execução preditiva em dados não utilizados durante a estimação dos parâmetros.

O aprimoramento do algoritmo consiste no ajuste de hiperparâmetros, viabilizado por meio de uma validação cruzada com cinco subconjuntos. Além disso, utilizou-se como indicador de comparação a métrica área sob a curva ROC (AUC). Por fim, selecionou-se o conjunto de hiperparâmetros com maior robustez preditiva para configurar o modelo final.

Após o treinamento, o desempenho preditivo do modelo selecionado foi validado no conjunto de teste a partir das probabilidades estimadas. A curva ROC foi construída e o ponto de corte ótimo foi definido com base no critério de Youden ($J = \text{Sensibilidade} + \text{Especificidade} - 1$), que maximiza simultaneamente a sensibilidade e a especificidade. A partir desse critério, as probabilidades foram transformadas em categorias binárias, permitindo a elaboração da matriz de confusão (Quadro 5).

Os hiperparâmetros do modelo final foram:

- Subamostra de observações utilizada em cada árvore: 0,8;
- Número máximo de árvores no processo de *boosting*: 2.000;
- Profundidade máxima das árvores: 6;
- Taxa de aprendizado (*learning rate*): 0,05;
- Proporção de variáveis selecionadas em cada árvore: 0,8;
- Peso mínimo para divisão de nós internos: 1; e
- Parâmetro de regularização gama: 0.

Quadro 5: Matriz de confusão do modelo *XGBoost*

		Classe Prevista	
		Segurança alimentar ($\hat{y} = 0$)	Insegurança alimentar ($\hat{y} = 1$)
Classe Observada	Segurança alimentar ($y = 0$)	13.195	6.397
	Insegurança alimentar ($y = 1$)	1.935	4.808
Acurácia: 0,6836		Valor preditivo positivo: 0,4291	
Intervalo de confiança de 95%: (0,678; 0,6892)		Valor preditivo negativo: 0,8721	

Taxa sem informação (<i>No Information Rate</i>): 0,744 Valor-p [Acurácia > NIR]: 1 Índice Kappa: 0,3176 Valor-p do teste de McNemar: <2e-16 Sensibilidade: 0,7130 Especificidade: 0,6735	Prevalência: 0,2560 Taxa de detecção: 0,1826 Prevalência de detecção: 0,4255 Acurácia balanceada: 0,6933
--	---

Fonte: elaboração própria com base nos dados da PNAD Contínua 2024.

Como procedimento de validação e comparação, estimou-se adicionalmente um modelo de regressão logística sob as mesmas condições de treinamento e conjunto de preditores do algoritmo *XGBoost*. Os resultados expressos no Quadro 6 indicam um desempenho competitivo, porém inferior ao modelo *XGBoost*.

Quadro 6: Matriz de confusão do modelo *Logit*

		Previsto	
		Segurança alimentar ($\hat{y} = 0$)	Insegurança alimentar ($\hat{y} = 1$)
Observado	Segurança alimentar ($y = 0$)	9.547	5.240
	Insegurança alimentar ($y = 1$)	1.306	3.550
Acurácia: 0,6668 Intervalo de confiança de 95%: (0,6601; 0,6733) Taxa sem informação (<i>No Information Rate</i>): 0,7528 Valor-p [Acurácia > NIR]: 1 Índice Kappa: 0,2961 Valor-p do teste de McNemar: <2e-16 Sensibilidade: 0,7311 Especificidade: 0,6456		Valor preditivo positivo: 0,4039 Valor preditivo negativo: 0,8797 Prevalência: 0,2472 Taxa de detecção: 0,1807 Prevalência de detecção: 0,4475 Acurácia balanceada: 0,6883	

Fonte: elaboração própria com base nos dados da PNAD Contínua 2024.

Os resultados apresentados no Quadro 5 e no Quadro 6 revelam que o modelo *XGBoost* apresenta acurácia (68,4%) e acurácia balanceada (69,3%) marginalmente superior às do modelo *Logit* (66,7% e 68,8% respectivamente), sugerindo melhor equilíbrio entre a identificação correta de domicílios em segurança e em insegurança alimentar. Este achado está alinhado com os trabalhos de Chilukuri (2025) e Okolie *et al.* (2025).

O incremento na acurácia balanceada e no índice de Kappa (0,318 no *XGBoost* contra 0,296 no *Logit*) aponta um melhor ajuste ao panorama de desbalanceamento da base. Tal indício converge com as conclusões de Fransiska *et al.* (2025).

Em termos de capacidade de identificação do desfecho de interesse, o *Logit* apresenta sensibilidade ligeiramente maior (73,1% contra 71,3%), sinalizando maior capacidade de detectar corretamente os domicílios em insegurança alimentar. Por outro lado, o *XGBoost* apresenta maior especificidade (67,4% contra 64,6%), diminuindo os casos de falsos positivos. Ainda assim, ambos os modelos proporcionam valores preditivos positivos baixos (42,9% no *XGBoost* e 40,4% no *Logit*), mostrando a dificuldade de prever corretamente a insegurança alimentar em um panorama de prevalência relativamente menor (cerca de 25%).

4.3 Valores SHAP

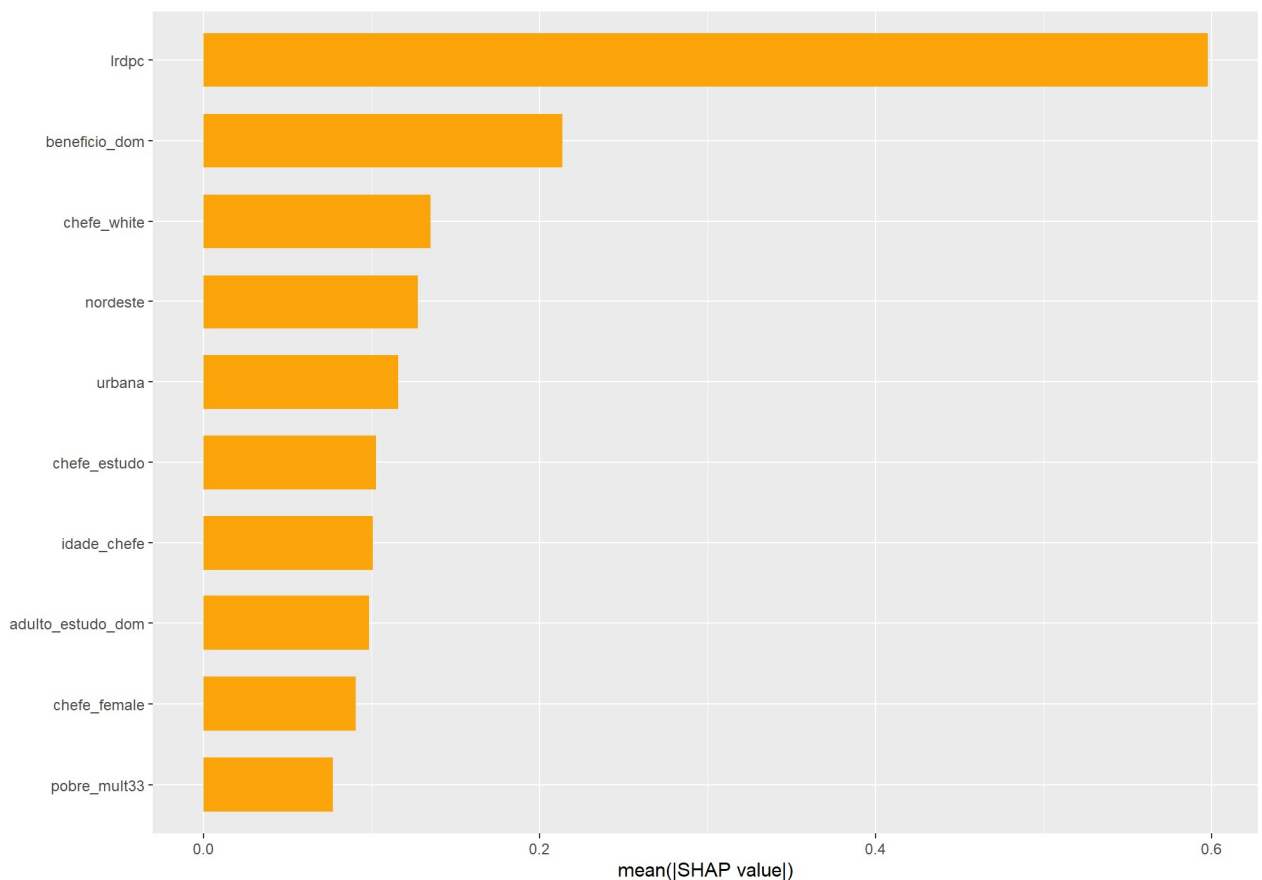
Os valores *SHAP* (*SHapley Additive exPlanations*) foram empregados neste trabalho com a finalidade de complementar a interpretação do modelo *XGBoost* e calculados com base no conjunto de dados de treinamento. Essa abordagem permite decompor cada predição do modelo em contribuições individuais das variáveis explicativas, indicando como e em que magnitude cada característica aumenta ou reduz a probabilidade estimada de ocorrência da insegurança alimentar. Ao incorporar os valores *SHAP*, a análise avança além do desempenho preditivo, oferecendo transparência ao funcionamento do modelo e possibilitando a comparação direta entre padrões identificados pelo aprendizado de máquina e os resultados obtidos por métodos econométricos tradicionais.

A importância global das 10 variáveis mais importantes no modelo de previsão da insegurança alimentar, medida pelo valor médio absoluto dos valores *SHAP*, pode ser visualizada pela Figura 2. A renda domiciliar *per capita* é o principal determinante das previsões do modelo, concentrando a maior parte da explicação da insegurança alimentar corroborando com estudos de Zhao et al. (2025) e Okolie et al. (2025).

Em seguida, destaca-se o recebimento de benefícios no domicílio, o que é consistente com o fato de que esses programas estão fortemente associados a contextos de vulnerabilidade social, ainda que possam exercer um papel mitigador do risco, convergindo com os resultados de Okolie et al. (2025). Variáveis sociorraciais e territoriais, como a condição de o chefe ser branco e a localização no Nordeste, também apresentam relevância, assim como o contexto urbano, indicando que desigualdades regionais e espaciais influenciam a probabilidade de insegurança alimentar. Tal fato está alinhado com os trabalhos de Chilukuri (2025) e Silva et al. (2025).

Também aparecem características relacionadas ao capital humano e à composição do domicílio, como escolaridade e idade do chefe, anos médios de estudo dos adultos e chefia feminina. Tais evidências estão em concordância com Zhao *et al.* (2025) e Barbosa e Nelson (2016). Por fim, a pobreza energética multidimensional, embora com menor peso relativo, permanece informativa, indicando que privações não monetárias seguem relevantes mesmo após o controle pela renda.

Figura 2 - Importância global das variáveis na previsão da Insegurança Alimentar (SHAP médio)



Fonte: elaboração própria com base nos dados da PNAD Contínua 2024.

A Figura 3 ilustra a distribuição dos valores *SHAP* em função variáveis explicativas selecionadas. No eixo das abcissas, constam os valores assumidos por cada variável. Enquanto, no eixo das ordenadas, aparece o valor *SHAP*, sendo que valores positivos indicam uma contribuição para o aumento no risco da insegurança alimentar e valores negativos revelam um efeito protetivo da variável em questão.

O primeiro diagrama da Figura 3 revela uma correlação negativa entre os anos médios de estudo dos adultos do domicílio e risco de insegurança alimentar. Esse resultado é coerente com o *Logit*, em que a razão de chance é 0,972, indicando que mais escolaridade dos

adultos no domicílio reduz significativamente o risco de insegurança alimentar. Residir em área urbana está associado a valores *SHAP* positivos, ou seja, maior probabilidade de insegurança alimentar em comparação à área rural. A regressão logística está nessa mesma linha, com razão de chance de 1,4088, sugerindo maior risco de insegurança alimentar em áreas urbanas, possivelmente refletindo custo de vida mais elevado e maior vulnerabilidade urbana.

Adicionalmente, a referida figura mostra um aumento expressivo nos valores *SHAP* quando a variável que indica localização no Nordeste assume valor 1, indicando que domicílios nessa região têm maior risco previsto de insegurança alimentar. Totalmente consistente com o *Logit*, em que o Nordeste apresenta uma das maiores razões de chance 2,06, sinalizando a continuidade das disparidades territoriais.

No que se refere ao número de adultos sem ocupação no domicílio, quanto maior essa variável, maior o valor *SHAP*, e logo mais elevada a probabilidade de insegurança alimentar, indicando uma relação linear e positiva.

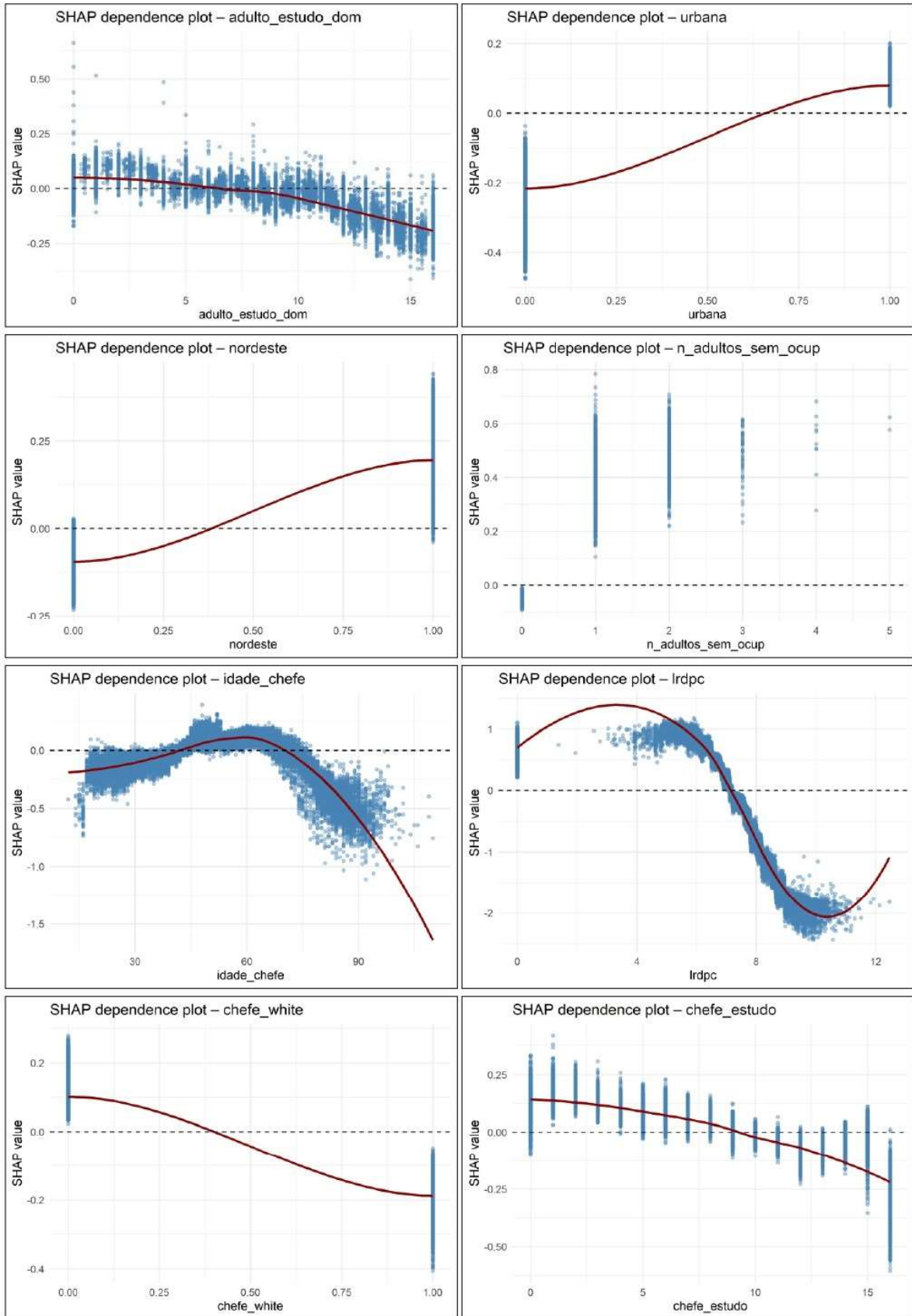
Esse resultado converge com o modelo *Logit*, que estima uma razão de chance elevada (1,64), expressando que a ausência inserção laboral é um dos fatores de vulnerabilidade alimentar. Com relação à idade do chefe do domicílio, os valores *SHAP* revelam uma relação não linear, com o risco aumentando levemente até idades intermediárias do chefe e caindo fortemente em idades mais avançadas. Este achado alinha-se com os estudos de Zhao *et al.* (2025) e Okolie *et al.* (2025). Por sua vez, a regressão logística captura apenas o efeito médio, com razão de chance próxima de 1 (1,002), enquanto o *SHAP* mostra que o impacto da idade varia ao longo do ciclo de vida.

O diagrama também indica que os valores *SHAP* associados à renda domiciliar *per capita* mostram que domicílios com níveis inferiores de renda tendem a apresentar maior probabilidade de insegurança alimentar, enquanto níveis maiores de renda reduzem esse risco. O efeito é intenso e não linear, coerente com o modelo *Logit*, que aponta uma razão de chance 0,758, confirmando a renda como fator de proteção contra a insegurança alimentar. Tal fato está em consonância com os resultados obtidos por Zhao *et al.* (2025) e Okolie *et al.* (2025) e Silva *et al.* (2025).

Os valores *SHAP* indicam que quando o chefe do domicílio é branco, há menor risco de insegurança alimentar. Em consenso com o *Logit*, cuja razão de chance é 0,767, comprovando as desigualdades raciais estruturais associadas à insegurança alimentar. Por fim, a Figura 3 ilustra que elevados níveis de escolaridade do chefe do domicílio estão relacionados a valores *SHAP* negativos, apontando redução na probabilidade de insegurança alimentar. O

impacto é gradual, duradouro e robusto, convergindo com as estimativas do *Logit*, que aponta razão de chance de 0,962. Tal indício ratifica as conclusões de Fransiska *et al.* (2025).

Figura 3 - Valores SHAP para algumas variáveis explicativas



Fonte: elaboração própria com base nos dados da PNAD Contínua 2024.

5. CONSIDERAÇÕES FINAIS

A presente pesquisa analisou os perfis de risco associados à insegurança alimentar para domicílios brasileiros, usando os microdados da PNADC 2024 e a modelagem de um modelo de aprendizado de máquina *XGBoost*. As evidências revelaram que a insegurança alimentar ainda se manifesta como um desafio estrutural, sendo influenciada por um conjunto complexo de fatores socioeconômicos e regionais.

No que tange ao desempenho preditivo, o modelo *XGBoost* apresentou marginalmente resultados superiores ao modelo de regressão Logística, demonstrando que métricas não lineares são instrumentos poderosos para tratar dados sociais complexos. O exame dos valores *SHAP* ratificou a preponderância da renda domiciliar *per capita* como principal preditor da insegurança alimentar. Adicionalmente, o recebimento de benefícios sociais foi o segundo preditor mais importante, seguido das características regionais, a cor e a escolaridade do chefe de domicílio e dos adultos, a inserção ocupacional e a composição familiar. A localização do domicílio na região Nordeste indicou um efeito substancial sobre o risco de insegurança alimentar. Esses achados corroboram com a literatura teórica e empírica que evidencia o poder dos modelos de *machine learning* em tratar padrões mais complexos do que modelos econométricos tradicionais.

Em termos de implicação das políticas públicas, os resultados obtidos recomendam ações múltiplas para o enfrentamento da insegurança alimentar, como a implementação de políticas públicas que articulem os benefícios sociais com investimento em educação, qualificação profissional, redução das disparidades regionais e enfrentamento das privações multidimensionais, como a pobreza energética. A aplicação do modelo de aprendizado de máquina apresentou potencial para aperfeiçoar a detecção de domicílios vulneráveis a insegurança alimentar e auxiliar políticas de focalização mais eficientes.

Em síntese, o estudo apresenta limitações como a transversalidade dos dados que não captura a trajetória dinâmica dos domicílios ao longo dos anos. A Escala Brasileira de Insegurança Alimentar (EBIA) é uma variável subjetiva sujeita a vieses no momento da entrevista. Pesquisas futuras podem explorar dados longitudinais e verificar transições de estados de segurança e insegurança alimentar e comparar outros algoritmos de *machine learning* que não foram testados.

REFERÊNCIAS

- ADOFO, Josephine O.; KOOMSON, Isaac. Unconditional Cash Transfers and Energy Poverty in South Africa. ScienceDirect, 2025. Disponível em: <<https://www.sciencedirect.com/science/article/abs/pii/S0140988325006097>>. Acesso em: 21 jun. 2025.
- ALKIRE, Sabina; FOSTER, James. Counting and multidimensional poverty measurement. Science Direct, 2011. Disponível em: <<https://www.sciencedirect.com/science/article/abs/pii/S0047272710001660>>. Acesso em: 05 jun. 2025.
- ANSER, Muhammad Khalid et al. Impact of ICT Adoption and Governance Interaction on Food Security in West Africa. MDPI Open Access Joirns, 2021. Disponível em: <<https://www.mdpi.com/2071-1050/13/10/5570>>. Acesso em: 08 fev. 2026.
- BARBOSA, R. M.; NELSON, D. R. The use of support vector machine to analyze food security in a region of Brazil. 2016. Disponível em: <https://www.researchgate.net/publication/303533227_The_Use_of_Support_Vector_Machine_to_Analyze_Food_Security_in_a_Region_of_Brazil>. Acesso em 05 fev. 2026.
- BRASIL. Análise de experiências estatais internacionais relativas à pobreza e justiça energética: definições, indicadores, medidas e governança. Empresa de Pesquisa Energética, Brasília, 2024. Disponível em: <https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-802/NT%20Experiências%20internacionais_finalSMA19_03_2024.pdf>. Acesso em: 05 jun. 2025.
- BUSKER, Tim ; HURK, Bart V. D.; HOMBERG, Marc V. D. Predicting Food-Security Crises in the Horn of Africa Using Machine Learning. Research Gate, 2023. Disponível em: <https://www.researchgate.net/publication/375515177_Predicting_Food-Security_Crises_in_the_Horn_of_Africa_Using_Machine_Learning>. Acesso em: 08 fev. 2026.
- CAVALCANTI, Daniella M.; COSTA, Edward M.; SILVA, Jorge Luiz M. D. Prorama Bolsa Família e o Nordeste: impactos na renda e na educação , nos anos de 2004 e 2006, 2013. Disponível em: <<https://www.scielo.br/j/rec/a/TyXwSPnBQn9pZhw86SYmQj/?format=pdf&lang=pt>>. Acesso em: 11 jun. 2025.
- CHEN, Tianqi; HE, Tong; BENESTY, Michael; KHOTILOVICH, Vadim; TANG, Yuan; CHO, Hyunsu; CHEN, Kailong; MITCHELL, Rory; CANO, Ignacio; ZHOU, Tianyi; LI, Mu; XIE, Junyuan; LIN, Min; GENG, Yifeng; LI, Yutian; YUAN, Jiaming; CORTES, David. *xgboost: Extreme Gradient Boosting*. R package version 3.3.0.0, 2026. Disponível em: <<https://cran.r-project.org/package=xgboost>>. Acesso em: 12 jan. 2026
- CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. In: ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 22., 2016, [s. l.]. Proceedings [...]. [S. l.]: ACM, 2016. p. 785-794.

CHILUKURI, Joshua. Predicting Low Food Access: An Empirical Assessment of Machine Learning Algorithms and Socio-Demographic Predictors. TechRxiv, 2025. Disponível em: <<https://www.techrxiv.org/users/961463/articles/1330464-predicting-low-food-access-an-empirical-assessment-of-machine-learning-algorithms-and-socio-demographic-predictors>>. Acesso em: 08 fev. 2026.

CHURCHILL, Sefa A.; SMYTH, Russell. Locus of control and energy poverty. Energy Economics, 2021. Disponível em: <https://www.researchgate.net/profile/Sefa-Awaworyi-Churchill/publication/352136015_Locus_of_Control_and_Energy_Poverty/links/616b56a1951b3574c6536b72/Locus-of-Control-and-Energy-Poverty.pdf?origin=publication_detail&_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB>. Acesso em: 10 jun. 2025.

CRUZ, Marcos Paulo M. D.; SILVA, Vitor Hugo M. C.; ARAÚJO, Jair A. D. Impacto das atividades não agrícolas sobre a segurança alimentar e nutricional das famílias rurais do Nordeste brasileiro. Colóquio - Revista do Desenvolvimento Regional, 2025. Disponível em: <<https://seer.faccat.br/index.php/coloquio/article/view/3156>>. Acesso em: 30 maio 2025.

EUROPEAN COMMISSION. Energy Poverty. European Commission, 2025. Disponível em: <https://energy.ec.europa.eu/topics/markets-and-consumers/energy-consumers-and-prosumers/energy-poverty_en>. Acesso em: 04 jun. 2025.

FAO. Fome e insegurança alimentar. Food and Agriculture Organization of the United Nations, 2025. Disponível em: <<https://www.fao.org/hunger/en>>. Acesso em: 12 maio 2025.

FAO; IFAD; UNICEF; WFP; WHO. The State of Food Security and Nutrition in the World. World Food Programme, 2024. Disponível em: <https://docs.wfp.org/api/documents/WFP-0000160501/download/?_ga=2.96152092.1265417301.1748906942-695795688.1748906942>. Acesso em: 02 jun. 2025.

FOINI, Pietro; TIZZONI, Michele; MARTINI, Giulia; PAOLOTTI, Daniela; OMODEI, Elisa. On the forecastability of food insecurity. Scientific Reports, v. 13, n. 1, p. 2793, 2023. Disponível em: <<https://doi.org/10.1038/s41598-023-29700-y>>. Acesso em 22 jan. 2026.

FRANÇOIS, Davi Ezequiel. Future perspectives for the energy poverty nexus in the rural areas of Ceará, Brazil. 2023. Tese (Doutorado em Engenharia) – Karlsruhe Institute of Technology (KIT), Karlsruhe, 2023. Disponível em: <<https://publikationen.bibliothek.kit.edu/1000152902>>. Acesso em: 12 dez. 2026.

FRANSISKA, Herlin et al. Evaluation of machine learning models based on household food insecurity data in Indonesia. BIO Web of Conferences, 2025. Disponível em: <https://www.bio-conferences.org/articles/bioconf/pdf/2025/22/bioconf_fisaed2025_02011.pdf>. Acesso em: 08 fev. 2026.

FRANZOLIN, CLÁUDIO JOSÉ; KLINGOR, BRUNA D. A.; BROLLO, MARIA EDUARDA A. POBREZA ENERGÉTICA DO CIDADÃO E SUA VULNERABILIDADE DIANTE DE UMA NECESSÁRIA TRANSIÇÃO ENERGÉTICA. Revista de Direito Público, 2024. Disponível em:

<<https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/download/7936/3511/28580>>. Acesso em: 11 jun. 2025.

FRIEDMAN, Jerome H. Greedy function approximation: a gradient boosting machine. *The Annals of Statistics*, v. 29, n. 5, p. 1189–1232, 2001. Disponível em <<https://doi.org/10.1214/aos/1013203451>>. Acesso em 11 jan. 2026.

GIULIO MATTIOLI, Karen L.; MARSDEN, Greg. Reprint of Transport poverty and fuel poverty in the UK: From analogy to comparison. *Transport Policy*, 2018. 114-125.
GÖRBIG, Gerard et al. Unveiling energy poverty risk: A multidimensional analysis of the heat-or-eat dilemma. *ScienceDirect*, 2025. Disponível em:
<<https://www.sciencedirect.com/science/article/pii/S2214629625001574>>. Acesso em: 19 out. 2025.

HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. *The elements of statistical learning: data mining, inference, and prediction*. 2. ed. New York: Springer, 2009.

HILLS, John. Fuel Poverty Review, março 2012. Disponível em:
<<https://sticerd.lse.ac.uk/dps/case/cr/casereport72.pdf>>. Acesso em: 18 maio 2025.

HJELM, Lisa; MATHIASSEN, Astrid; WADHWA, Amit. Measuring Poverty for Food Security Analysis: Consumption -Versis Asset - Based Approaches. *Sage Journals Home*, 2016. Disponível em: <<https://journals.sagepub.com/doi/10.1177/0379572116653509>>. Acesso em: 10 jun. 2025.

HOWELL, Anthony. Cash transfers and the Han-Minority household energy poverty gap in rural China. Arizona: Arizona State University, 2024. p. 48.

IBGE. Pesquisa de Orçamentos Familiares 2017-2018, Rio de Janeiro, 2021. Disponível em: <<https://biblioteca.ibge.gov.br/visualizacao/livros/liv101749.pdf>>. Acesso em: 04 jun. 2025.

IBGE. PNAD Contínua - Pesquisa Nacional por Amostra de Domicílios Contínua. Instituto Brasileiro de Geografia e Estatística, 2023. Disponível em:
<<https://www.ibge.gov.br/estatisticas/sociais/saude/17270-pnad-continua.html?edicao=27258&t=resultados>>. Acesso em: 14 maio 2025.

IBGE. PNAD Contínua - Pesquisa Nacional por Amostra de Domicílios Contínua. Instituto Brasileiro de Geografia e Estatística, 2024. Disponível em:
<<https://www.ibge.gov.br/estatisticas/sociais/trabalho/17270-pnad-continua.html>>. Acesso em: 08 fev. 2026.

JEAN, Wesley et al. ESTUDO SOBRE POBREZA ENERGÉTICA E SEGURANÇA ENERGÉTICA NO SEMINÁRIO BRASILEIRO: VULNERABILIDADE E RESILIÊNCIA SOCIOAMBIENTAL. *Boletim regional, urbano e ambiental - IPEA*, jul.-dez., 2024. Disponível em:
<https://repositorio.ipea.gov.br/bitstream/11058/15989/5/BRUA_32_Artigo_3_estudo_sobre_pobreza_energetica.pdf>. Acesso em: 12 jun. 2025.

LUNDBERG, Scott M.; LEE, Su-In. A unified approach to interpreting model predictions. In: *Advances in Neural Information Processing Systems*, 30., 2017, Long Beach. Proceedings

[...]. Red Hook: Curran Associates, 2017. p. 4765–4774. Disponível em: <https://papers.nips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html>. Acesso em: 09 jan. 2026.

KADHIM, Kadhim G. et al. Identifying the Effect of Energy Poverty on Income Poverty, Health. Education, Environment and Future Food Security. Cuadernos de economía, 2024. p. 10-18.

KOIRALA, Dhiroj P.; RAHUT, Dil. Multidimensional energy poverty and its effects on household consumption and savings: Evidence from Nepal. [S.l.]: Heliyon, 2024. p. 21.

MOREIRA, Rogério S. Uma medida de pobreza energética para o Brasil. Mestrado Profissional em Economia do Setor Público. ed. Fortaleza: Universidade Federal do Ceará. Faculdade de Economia, Administração, Atuária e Contabilidade, v. Dissertação (mestrado), 2023.

NUSSBAUMER, Patrick; BAZILIAN, Morgan; MODI, Vijay. Measuring energy poverty: Focusing on what matters. Renewable and Sustainable Energy Reviews, Amsterdam, julho 2012. Disponível em: <<https://www.science direct.com/science/article/abs/pii/S1364032111003972>>. Acesso em: 18 maio 2025.

OKOLIE, Awele et al. Predicting Food Insecurity Across U.S. Census Tracts: A Machine Learning Analysis Using the USDA Food Access Research Atlas. Research Gate, 2025. Disponível em: <https://www.researchgate.net/publication/398162503_Predicting_Food_Insecurity_Across_US_Census_Tracts_A_Machine_Learning_Analysis_Using_the_USDA_Food_Access_Research_Atlas>. Acesso em: 08 fev. 2026.

ONU. Objetivos de Desenvolvimento Sustentável. Nações Unidas Brasil, 2025. Disponível em: <<https://brasil.un.org/pt-br/sdgs>>. Acesso em: 05 jun. 2025.

ONU. ONU News - Perspectiva Global Reportagens Humanas. Nações Unidas, 2025.

Disponível em: <<https://news.un.org/pt/story/2025/02/1845216>>. Acesso em: 12 maio 2025.

PELZ, Setu; PACHAURI, Shonali; GROH, Sebastian. A critical review of modern approaches for multidimensional energy poverty measurement. Wires - Wiley Interdisciplinary Reviews, 22 maio 2018. Disponível em:

<<https://wires.onlinelibrary.wiley.com/doi/full/10.1002/wene.304>>. Acesso em: 18 maio 2025.

PONDIE, Thierry M. et al. Does Energy Poverty increases Starvation? Evidence from Sub-Saharan Africa, 04 mar. 2022. Disponível em:

<<https://opus.lib.uts.edu.au/handle/10453/150297>>. Acesso em: 05 jun. 2025.

PORTO VALENTE, Caroline ; MORRIS, Allan; WILKINSON, Sara. Energy poverty, housing and health: the lived experience of older low-income Australians. Building Research & Information, 31 ago. 2021. p. 1-13.

SARAGIH, Frederic M.; WIBOWO, Wahyu. Analysis of Food Security Index Predictions in Indonesia using Machine Learning Approach. E-Journal UNIPAS, 2025. Disponível em:

<<https://ejournal.unipas.ac.id/index.php/Agro/article/view/2302>>. Acesso em: 08 fev. 2026.

SHWARTZ-ZIV, Ravid; ARMON, Amitai. Tabular data: Deep learning is not all you need. *Information Fusion*, v. 81, p. 84-90, 2022.

SILVA, Lucas Adriano; BRAGA, Marcelo José; RODRIGUES, Cristiana Tristão; BEZERRA, Paula Borges da Silveira. *Electricity expenses and food insecurity in Brazilian households: a causal relationship*. 2025. Disponível em: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5315657>. Acesso em: 20 out. 2025.

SILVA, Ticiania L. Coelho da; GOMES, Laécia Gretha Amorim; FURTADO, Lara Sucupira. Machine learning methods and models to predict food insecurity levels for families in Ceará, Brazil, based on employment, housing and other social indicators. *Journal of the Brazilian Computer Society*, 2025. Disponível em: <<https://www.semanticscholar.org/paper/Machine-Learning-methods-and-models-to-predict-food-Silva-Furtado/b07507bf4133495039a8eda421276439a9da15a6>>. Acesso em: 08 fev. 2026.

WESTERVELD, Joris et al. Modelling Food Insecurity in Ethiopia: Towards a machine learning model that predicts the transitions in food security using scalable features. Utrecht University, 2019. Disponível em: <<https://studenttheses.uu.nl/handle/20.500.12932/32699>>. Acesso em: 08 fev. 2026.

ZHANG, Zhao; ZHANG, Yuyan; WEN, Y.; REN, Yaxue. Data-driven XGBoost model for maximum stress prediction of additive manufactured lattice structures. *Complex & Intelligent Systems*, v. 9, n. 5, p. 5881–5892, 2023. Disponível em: <<https://doi.org/10.1007/s40747-023-01061-z>>. Acesso em 09 jan. 2026.

ZHAO, Longqiang et al. Prediction of household food insecurity in rural China: an application of machine learning methods. Research Gate, 2025. Disponível em: <https://www.researchgate.net/publication/388516288_Prediction_of_household_food_insecurity_in_rural_China_an_application_of_machine_learning_methods>. Acesso em: 08 fev. 2026.

WHO. Household air pollution. World Health Organization, 2024. Disponível em: <<https://www.who.int/news-room/fact-sheets/detail/household-air-pollution-and-health>>. Acesso em: 13 maio 2025.

APÊNDICE A – Estimativas das Razões de Chance do Modelo *Logit*

Tabela A1: Razões de chance estimadas no modelo Logit

Coeficiente	Razão de chance	Desvio padrão	z	P-valor	Intervalo 95% confiança	
aluguel	1,2951	0,0304	11,0200	0,0000	1,2369	1,3560
pobre_mult33	1,4171	0,0303	16,3100	0,0000	1,3589	1,4777
norte	2,2813	0,0761	24,7400	0,0000	2,1369	2,4353
nordeste	2,0623	0,0602	24,7900	0,0000	1,9476	2,1838
sudeste	1,4514	0,0434	12,4700	0,0000	1,3688	1,5389
centro_oeste	1,5810	0,0578	12,5300	0,0000	1,4716	1,6984
urbana	1,4088	0,0319	15,1400	0,0000	1,3476	1,4726
lrdpc	0,7582	0,0065	-32,4300	0,0000	0,7456	0,7710
dom_benef	1,8239	0,0358	30,6400	0,0000	1,7551	1,8953
dom_aposent	1,1456	0,0301	5,1700	0,0000	1,0881	1,2061
chefe_fem	1,2665	0,0222	13,4700	0,0000	1,2237	1,3107
chefe_branc	0,7671	0,0146	-13,9700	0,0000	0,7390	0,7961
chefe_idad	1,0020	0,0008	2,4100	0,0160	1,0003	1,0036
chefe_estudo	0,9621	0,0022	-16,7400	0,0000	0,9578	0,9664
chefe_ocupad	1,2378	0,0290	9,0900	0,0000	1,1821	1,2961
chefe_previd	0,7622	0,0181	-11,4500	0,0000	0,7275	0,7984
chefe_agric	0,8504	0,0276	-4,9900	0,0000	0,7979	0,9062
adult_estud_dom	0,9720	0,0023	-12,0800	0,0000	0,9674	0,9764
fam_estend	1,1280	0,0234	5,8000	0,0000	1,0829	1,1749
fam_comp	1,0482	0,0744	0,6600	0,5070	0,9121	1,2046
n_adult_sem_ocup	1,6364	0,0385	20,9400	0,0000	1,5626	1,7135
dom_criança	1,0837	0,0111	7,8800	0,0000	1,0622	1,1056
dom_idoso	0,8355	0,0142	-10,5900	0,0000	0,8082168	0,8638
const	1,1172	0,0878	1,4100	0,1590	0,9576319	1,3033

Fonte: elaboração própria com base nos dados da PNAD Contínua 2024.