

Alimentando Informações: Um Estudo De Dados Nutricionais

Gustavo Silva Rios
Silas de Souza Ferreira
Israel Soares do N. Viana

Faculdade de Computação e Informática (FCI)
Universidade Presbiteriana Mackenzie São Paulo, SP – Brasil

2025

Resumo

Este projeto tem como objetivo analisar os dados do Sistema de Vigilância Alimentar e Nutricional (*Sisvan*) para compreender padrões do estado nutricional da população atendida na Atenção Primária à Saúde (APS). A relevância do estudo se dá pela necessidade de fornecer *insights* para profissionais de saúde e gestores públicos, auxiliando na formulação de políticas eficazes contra desnutrição e obesidade.

A base de dados utilizada contém informações detalhadas sobre medidas antropométricas, dados demográficos e classificações nutricionais de indivíduos desde o ano de 2008. Para garantir uma análise abrangente e longitudinal, os registros do ano de 2018 a 2023 serão integrados, criando uma base consolidada e permitindo o estudo de tendências ao longo do tempo.

Os principais objetivos incluem a exploração da estrutura dos dados, análise de distribuições por idade e região, desenvolvimento de modelos preditivos, detecção de outliers e criação de visualizações interativas. A partir disso, espera-se identificar grupos vulneráveis, entender e propor melhorias para o monitoramento nutricional da população brasileira.

1 Introdução

A análise de dados em saúde pública é essencial para a formulação de políticas e estratégias voltadas à promoção do bem-estar da população. No Brasil, o Sistema de Vigilância Alimentar e Nutricional (*Sisvan*), gerenciado pelo Ministério da Saúde, desempenha um papel central no monitoramento do estado nutricional e dos hábitos

alimentares da população atendida na Atenção Primária à Saúde (APS). Esse sistema consolida informações individualizadas e anonimizadas, permitindo avaliar indicadores nutricionais e subsidiar ações governamentais para a promoção da alimentação adequada e a prevenção de doenças associadas à nutrição.

Nos últimos anos, o Brasil tem enfrentado o duplo desafio da desnutrição em bolsões de vulnerabilidade social e do aumento do excesso de peso na população. Dados do (*Sisvan*) indicam que, em 2022, cerca de 14,2% das crianças menores de 5 anos apresentavam excesso de peso, ao mesmo tempo em que déficits nutricionais persistem em regiões de difícil acesso. Diante desse cenário, este projeto tem como objetivo analisar os dados nacionais do Sisvan entre 2018 e 2023, buscando identificar padrões e tendências no estado nutricional da população atendida pelo SUS.

Os principais objetivos incluem calcular a prevalência de desnutrição e excesso de peso por ano, idade, sexo e região; relacionar esses achados a variáveis demográficas e socioeconômicas, como a vinculação ao programa Auxílio Brasil; desenvolver visualizações e modelos preditivos para classificar o estado nutricional e identificar fatores de risco; e gerar recomendações para aprimorar as ações nutricionais na APS. A análise integrará dados complementares do e-SUS APS e do Auxílio Brasil para enriquecer as conclusões e apoiar a formulação de políticas públicas baseadas em evidências, reforçando o papel do (*Sisvan*) na tomada de decisão em saúde.

2 Motivações e Justificativa

A análise do estado nutricional da população é um fator fundamental para a formulação de políticas públicas de saúde e assistência social. O Sistema de Vigilância Alimentar e Nutricional (*Sisvan*) fornece dados essenciais sobre a condição antropométrica e hábitos alimentares da população atendida na Atenção Primária à Saúde (APS), permitindo identificar padrões de desnutrição, sobrepeso e obesidade.

Este projeto se justifica pela necessidade de aprimorar a análise desses dados para fornecer *insights* que auxiliem profissionais de saúde e gestores públicos na tomada de decisão. Através da análise exploratória, modelagem de dados e visualização interativa, será possível entender padrões epidemiológicos e propor soluções para melhorar o acompanhamento nutricional da população brasileira.

3 Objetivos

O objetivo projeto é utilizar a análise dos dados do *Sisvan* para identificar padrões de desnutrição, sobrepeso e obesidade, que são fatores críticos para a saúde pública. Outros objetivos específicos incluem:

- Explorar e compreender a estrutura dos dados do *Sisvan*;
- Analisar distribuições e tendências de indicadores nutricionais por região, idade e gênero;
- Criar visualizações interativas para facilitar a análise dos dados;
- Desenvolver modelos preditivos para identificar fatores de risco associados ao estado nutricional;

- Gerar *insights* que possam subsidiar políticas públicas de saúde alimentar e nutricional

4 Descrição da Base de dados

A base de dados utilizada neste projeto é proveniente do *Sisvan*, mantido pelo Ministério da Saúde do Brasil. Os dados são anonimizados e contêm informações detalhadas sobre o estado nutricional de indivíduos atendidos na APS desde 2008. A base de dados possui 212.343 registros que variam entre os anos de 2008 a 2023. Os principais atributos disponíveis na base incluem:

- Identificação geográfica: Estado (*SGUF*) e município (*NOMUNICIPIO*);
- Dados demográficos: Idade (*NUIDADEANO*), fase da vida (*DSFASEVIDA*), sexo (*SGSEXO*), raça/cor (*DSRACACOR*) e comunidade/povo tradicional (*DSPOVOCOMUNIDADE*);
- Nível educacional: Escolaridade do indivíduo (*DSESCOLARIDADE*);
- Data e periodicidade do acompanhamento: Data do acompanhamento (*DTACOMPANHAMENTO*) e competência do registro (*NUCOMPETENCIA*);
- Medidas antropométricas: Peso (*NUPESO*), altura (*NUALTURA*), índice de massa corporal (*DSIMC*);
- Classificações nutricionais: Estado nutricional para diferentes faixas etárias e critérios de análise (*PESO X IDADE*, *PESO X ALTURA*, *CRIANÇA. ALTURA X IDADE*, *CRI. IMC X IDADE*, *ADULTO. ALTURA X IDADE*, *ADULTO. IMC X IDADE*, *COESTADONUTRIADULTO*, *COESTADONUTRIIDOSO*);
- Fonte do acompanhamento: Sistema de origem do dado (*SISTEMAORIGEMA-COMP*).

A base de dados possui registros individuais anonimizados, permitindo a análise de frequências relativas e prevalências dos estados nutricionais na população atendida. Os registros podem se sobrepor em casos de múltiplos acompanhamentos para um mesmo indivíduo, sendo recomendada a priorização dos dados provenientes do *Sisvan* e Auxílio Brasil/Bolsa Família.

A análise desta base permitirá uma visão abrangente sobre a evolução do estado nutricional da população brasileira ao longo dos anos, auxiliando na formulação de estratégias de saúde pública.

5 Referencial Teórico

A Vigilância Alimentar e Nutricional (VAN) é definida pelo Ministério da Saúde como um processo contínuo de coleta, análise e interpretação de dados sobre consumo alimentar e estado nutricional, visando apoiar ações de saúde pública. Inserida no âmbito do Sistema Único de Saúde (SUS), a VAN integra a Atenção Primária à Saúde como uma ferramenta estratégica para promoção da alimentação saudável e prevenção de agravos nutricionais. Seus instrumentos geram um conjunto de indicadores epidemiológicos que orientam gestores na formulação de políticas e programas de nutrição. Entre os principais

indicadores antropométricos acompanhados pela VAN destacam-se peso, estatura e idade, combinados em índices como peso-para-idade, estatura-para-idade e índice de massa corporal (IMC) para idade.

Esses indicadores permitem avaliar o estado nutricional individual e coletivo segundo referências padronizadas. Por exemplo, em crianças, utiliza-se as curvas de crescimento da Organização Mundial da Saúde (OMS) para classificar desnutrição (déficit pondero-estatural), eutrofia (adequação) ou excesso de peso (sobrepeso/obesidade) de acordo com os desvios-padrão do IMC-para-idade. Em adultos, o IMC ($\text{peso}/\text{altura}^2$) é empregado para diagnosticar sobrepeso ($IMC \geq 25 \text{ kg/m}^2$) e obesidade ($IMC \geq 30 \text{ kg/m}^2$), enquanto desnutrição em adultos pode ser inferida por IMC baixo ($IMC \leq 18,5 \text{ kg/m}^2$) ou perda de peso involuntária.

Tais definições seguem protocolos técnicos do Ministério da Saúde e diretrizes internacionais (OMS), assegurando comparabilidade dos indicadores. A vigilância antropométrica é complementada pelo monitoramento de marcadores de consumo alimentar (por exemplo, frequência de ingestão de frutas, ultraprocessados, aleitamento materno), compondo um panorama abrangente da situação nutricional. Diversos trabalhos correlatos na literatura exploram os dados do Sisvan ou bases similares para entender os padrões nutricionais da população. O próprio Sisvan foi instituído na década de 1990 como ferramenta do SUS para coletar e consolidar dados nutricionais na APS, e sua cobertura vem se ampliando mediante estratégias como a vinculação das condicionalidades nutricionais ao Programa Bolsa Família, atual Auxílio Brasil.

Essa integração dos programas sociais com a vigilância nutricional aumentou o acesso de grupos vulneráveis (crianças e gestantes de baixa renda) aos serviços de saúde, contribuindo para maior equidade no acompanhamento nutricional. Potenciais e limitações do Sisvan têm sido discutidos: por um lado, abrange segmentos populacionais vulneráveis normalmente pouco representados em inquéritos nacionais (por exemplo, populações indígenas), fornecendo dados contínuos e desagregados localmente; por outro, apresenta lacunas de cobertura e qualidade dos dados, já que nem todas as pessoas são atendidas ou inseridas no sistema, especialmente em áreas remotas, e há sub-registro em alguns municípios. Mrejen, Cruz e Rosa (2023) analisaram o Sisvan como ferramenta de monitoramento infantil e constataram que, embora a cobertura nacional tenha se expandido entre 2008 e 2019, ela sofreu queda em 2020 e permanece heterogênea regionalmente, sendo maior no Norte/Nordeste e entre crianças menores.

Além disso, esses autores compararam os resultados do Sisvan com pesquisas de base populacional e identificaram divergências, sinalizando a necessidade de aprimorar os instrumentos de vigilância nutricional para melhor refletir a realidade do país. Outros estudos recentes utilizaram dados do Sisvan para avaliar tendências e perfis nutricionais em diferentes contextos: *Apelini et al. (2021)* estudaram a evolução do sobrepeso e obesidade em crianças de 5 a 9 anos no Espírito Santo, evidenciando o aumento dessas prevalências no período de 2009 a 2018; *Silva et al. (2022)* examinaram a expansão da cobertura do Sisvan e do estado nutricional em adultos brasileiros de 2008 a 2019; *Barbosa et al. (2023)* investigaram a população idosa, relacionando a cobertura do Sisvan e os indicadores de baixo peso/obesidade a desigualdades sociais.

Tais trabalhos ressaltam as vantagens do uso de grandes bases administrativas de saúde – abrangência e periodicidade dos dados – ao mesmo tempo em que discutem limitações metodológicas, como vieses de seleção (apenas usuários do SUS são incluídos) e inconsistências nos registros antropométricos individuais. No tocante a outras abordagens

de análise de dados em saúde pública, tem-se observado uma incorporação crescente de métodos estatísticos avançados e de inteligência artificial para aprimorar a vigilância nutricional. Modelos de séries temporais e análises espaciais, por exemplo, já foram aplicados para detectar padrões geográficos de desnutrição infantil e prever tendências de obesidade em determinadas regiões. Mais recentemente, técnicas de *machine learning* vêm sendo exploradas para identificar fatores de risco nutricional e estratificar populações prioritárias. Iniciativas nacionais, como o programa *Grand Challenges Explorations* em Ciência de Dados para Saúde Materno-Infantil, incentivaram o uso de aprendizado de máquina e *big data* para descobrir determinantes chave da saúde nutricional e auxiliar na avaliação de políticas públicas. Da mesma forma, o desenvolvimento de *dashboards* interativos e sistemas de apoio à decisão tem ampliado a transparência e agilidade na vigilância em saúde.

Órgãos governamentais e pesquisadores acadêmicos têm criado painéis *on-line* que permitem visualizar indicadores do Sisvan em diferentes recortes (por município, faixa etária, ano), facilitando a identificação de situações críticas e o monitoramento de metas. O *Sisvan Web*, por exemplo, disponibiliza relatórios públicos automáticos, consolidando as informações coletadas em gráficos e tabelas acessíveis aos gestores locais. Ferramentas assim potencializam a utilização dos dados de vigilância, ao transformar bancos volumosos em *insights* acionáveis. Em suma, o referencial teórico indica que a conjugação de conceitos epidemiológicos clássicos com novas técnicas de ciência de dados pode aprimorar significativamente a detecção de padrões nutricionais e, conseqüentemente, embasar intervenções mais eficazes em saúde pública.

Diversos estudos recentes têm aplicado técnicas de inteligência artificial para fortalecer a análise dos dados nutricionais em saúde pública. Por exemplo, Gama et al. (2021) utilizaram algoritmos de *machine learning*, como *Random Forest* e *XGBoost*, para prever o risco de sobrepeso em crianças com base em variáveis sociodemográficas e de consumo alimentar extraídas do *Sisvan*. Os modelos demonstraram alta acurácia na identificação de grupos de risco, destacando o potencial preditivo dos dados administrativos. Em outro estudo, Ferreira et al. (2022) aplicaram redes neurais artificiais (*Artificial Neural Networks* - ANNs) para classificar o estado nutricional de escolares utilizando dados antropométricos e de hábitos alimentares, alcançando bons resultados em termos de sensibilidade e especificidade. Além disso, Oliveira e Souza (2020) desenvolveram um modelo de *clustering* (*K-means*) para segmentar perfis nutricionais de gestantes cadastradas no *Sisvan*, identificando padrões associados a maior vulnerabilidade social.

Esses trabalhos demonstram como técnicas de IA podem ser integradas à vigilância nutricional para automatizar diagnósticos, antecipar tendências e melhorar a alocação de recursos. A aplicação de modelos supervisionados e não supervisionados permite não só identificar casos críticos, mas também revelar padrões ocultos nos dados, ampliando a capacidade de resposta do *SUS*. A tendência atual aponta para uma vigilância mais inteligente, que alia volume de dados à capacidade analítica avançada, um cenário em que a ciência de dados e a saúde pública convergem para decisões mais eficazes e baseadas em evidências.

6 Pipeline da Solução

Para atingir os objetivos propostos, delineou-se um pipeline de análise de dados composto por etapas sequenciais, desde a obtenção das bases brutas até a geração de *insights* para formulação de políticas. As etapas previstas no projeto são as seguintes:

- **Aquisição de dados:** Inicialmente, será realizada a coleta e consolidação dos dados relevantes. Isso inclui a extração de registros antropométricos e demográficos do Sisvan Web (referentes aos anos de 2018 a 2023) por meio das bases disponibilizadas pelo Ministério da Saúde, bem como a obtenção de dados complementares do *e-SUS APS* (sistema de prontuário eletrônico da atenção primária) e do cadastro do Programa Auxílio Brasil. A integração desses conjuntos de dados permitirá cruzar informações nutricionais com variáveis de assistência e contexto socioeconômico, enriquecendo as análises. Essa etapa garante a matéria-prima necessária para o estudo, reunindo em um só repositório todos os dados brutos que serão explorados;
- **Pré-processamento:** Com os dados em mãos, procede-se ao tratamento e organização para análise. Serão aplicadas rotinas de limpeza, corrigindo inconsistências e erros de digitação, eliminando duplicidades e filtrando registros irrelevantes (por exemplo, medidas fisiologicamente impossíveis). Também haverá tratamento de valores faltantes (nulidade), empregando técnicas adequadas como imputação ou exclusão, conforme o caso e a proporção de faltantes. Nessa fase ocorre ainda a conversão de tipos de dados, padronização de unidades de medida e cálculo de novos campos derivados – por exemplo, cálculo do IMC e classificação do estado nutricional de cada indivíduo segundo faixas etárias específicas. O pré-processamento é crucial para assegurar a qualidade e a consistência dos dados, evitando distorções nas etapas seguintes e alinhando os dados brutos aos requisitos das análises estatísticas;
- **Análise Exploratória de Dados (AED):** Com a base já tratada, será conduzida uma análise exploratória abrangente. Nessa etapa, calculam-se estatísticas descritivas (médias, medianas, frequências) dos principais indicadores antropométricos e sociodemográficos. Serão estimadas as distribuições de variáveis como idade, sexo e região dos indivíduos monitorados, bem como das variáveis de desfecho nutricional (percentual de baixo peso, eutrofia, sobrepeso e obesidade). Gráficos iniciais de dispersão, histogramas e *boxplots* auxiliarão na visualização da distribuição dos dados e na identificação de *outliers* ou padrões inusitados. Também serão realizadas comparações entre grupos: por exemplo, avaliar diferenças na prevalência de obesidade entre regiões (Norte vs. Sul) ou tendências de redução da desnutrição ano a ano. A AED permitirá gerar hipóteses e direcionar o foco das etapas subsequentes, ao evidenciar quais fatores parecem mais relevantes ou quais disparidades merecem investigação aprofundada. Além disso, essa análise fornecerá um primeiro panorama do desempenho do Sisvan em termos de cobertura – calculando, por exemplo, a porcentagem da população-alvo efetivamente acompanhada pelo sistema em cada localidade – e da qualidade dos registros (distribuição de valores faltantes, etc.);
- **Visualização de dados:** Para melhor comunicar os achados e possibilitar interações dinâmicas, o projeto incluirá a construção de gráficos e painéis interativos. Utilizando ferramentas de *business intelligence* (p. ex. *Power BI*, *Tableau*) ou bibliotecas de visualização em *Python/R* (como *Plotly*, *Bokeh* ou *ggplot2*), serão desenvolvidos *dashboards* que exibam os indicadores-chave. Espera-se criar visualizações como mapas temáticos (mostrando a distribuição geográfica da desnutrição e obesidade), séries temporais (linha do tempo 2018–2023 das prevalências nutricionais), gráficos de barras comparando categorias (por exemplo, estado nutricional por faixa etária) e diagramas de dispersão correlacionando variáveis (como cobertura do Sisvan vs. índice de desenvolvimento municipal). Tais visualizações facilitam a interpretação dos dados por parte de gestores e demais interessados, evidenciando padrões de forma

intuitiva. Idealmente, os gráficos serão interativos, permitindo filtragem por ano ou região, o que amplia a capacidade de exploração ad hoc das informações além do relatório estático. A etapa de visualização, portanto, contribui diretamente para o objetivo de transmitir *insights* de maneira clara e acessível, servindo de ponte entre a análise técnica e a tomada de decisão prática;

- **Modelagem preditiva:** Com base nos conhecimentos extraídos nas etapas anteriores, pretende-se desenvolver um modelo preditivo simples voltado à classificação do estado nutricional ou à previsão de ocorrências de risco. Por exemplo, pode-se treinar um modelo de classificação (como árvore de decisão ou *random forest*) que, a partir de atributos demográficos e históricos (idade, sexo, condições socioeconômicas, histórico de peso), prediga a categoria nutricional de crianças e gestantes (normal, risco de desnutrição, sobrepeso, etc.). Alternativamente, é possível ajustar um modelo estatístico (p. ex. regressão logística multinomial) para identificar os fatores mais associados a desfechos extremos (baixo peso ou obesidade) dentro da base do Sisvan. Embora o foco principal do projeto seja descritivo e exploratório, a incorporação de técnicas de *machine learning* pode oferecer *insights* adicionais, revelando padrões complexos ou interações entre variáveis que não seriam evidentes em análises univariadas. O desempenho do modelo será avaliado através de métricas apropriadas (acurácia, sensibilidade, especificidade, *AUC-ROC*, etc.), empregando validação cruzada ou separação treino-teste para verificar sua generalização. A ideia não é tanto implantar um modelo preditivo operacional, mas usar esse exercício analítico para confirmar tendências e relações nos dados (por exemplo, confirmar se determinado grupo socioeconômico tem maior probabilidade de obesidade, mesmo controlando para outras variáveis) e assim reforçar as conclusões do estudo;
- **Validação e avaliação:** Antes de reportar os resultados finais, será conduzida uma etapa de validação abrangente. Isso envolve validar os dados e os resultados obtidos sob diferentes prismas. Primeiramente, a qualidade dos dados será revisitada – checando, por exemplo, se os indicadores calculados (taxas de obesidade, etc.) estão consistentes com relatórios oficiais publicados pelo Ministério da Saúde ou em linha com estudos prévios (validação externa). Em segundo lugar, os resultados do modelo preditivo serão analisados criticamente: além das métricas de performance internas, comparar-se-á as predições com conhecimento epidemiológico estabelecido (por exemplo, se o modelo apontar que determinada região rica teria alto risco de desnutrição, isso seria um alerta de possível erro de dados ou modelo). Ademais, sempre que possível, aplicar-se-á a técnica de validação cruzada temporal para as tendências: por exemplo, usar dados até 2022 para “prever” 2023 e ver se as projeções batem com os dados observados, como forma de avaliar se as tendências captadas são robustas. A etapa de validação garante que as conclusões tenham confiabilidade e que os *insights* gerados não sejam artefatos de problemas nos dados ou de sobreajuste dos modelos. Caso incongruências sejam detectadas, ajustes no pipeline serão feitos (retorno às etapas de limpeza ou refinamento do modelo), num processo iterativo até que os resultados se mostrem coerentes;
- **Geração de *insights* e documentação:** Por fim, os achados consolidados serão interpretados à luz do referencial teórico e transformados em *insights* acionáveis. Isso significa traduzir os resultados estatísticos em implicações práticas: por exemplo, identificar quais regiões ou públicos demandam prioridade em intervenções nutricionais, quais tendências positivas devem ser fortalecidas, ou que lacunas no Sisvan

precisam ser sanadas (como melhorar a adesão ao sistema em certas localidades). Os *insights* incluirão recomendações para gestores de saúde – e.g., reforço de políticas de promoção da alimentação saudável em áreas onde o sobrepeso infantil mostrou alta crescente, ou estratégias para buscar ativamente crianças não acompanhadas onde a cobertura do Sisvan se mostrou baixa. Toda a análise será compilada em um relatório técnico-científico e em materiais de comunicação (gráficos, tabelas-síntese), documentando as metodologias aplicadas e os resultados obtidos. A documentação cuidadosa assegura transparência e reprodutibilidade, permitindo que outros profissionais confirmem ou expandam o trabalho. Essa etapa culminante fecha o ciclo do projeto, atendendo ao objetivo de produzir conhecimento aplicável – isto é, entregar informações tratadas que possam orientar a tomada de decisão e a formulação de políticas públicas na área de nutrição e saúde;

Cada etapa descrita acima desempenha um papel específico e complementar para o sucesso do projeto. As etapas iniciais (obtenção e pré-processamento) garantem a base confiável de dados sobre a qual todo o estudo se apoia. As etapas intermediárias (análise exploratória, visualização e modelagem) realizam o tratamento analítico das informações, descobrindo padrões, relações e tendências relevantes para os objetivos propostos. Finalmente, as etapas finais (validação e síntese de *insights*) asseguram que os resultados sejam robustos e úteis, fechando o *loop* entre os dados brutos e a geração de valor para a saúde pública. Assim, o pipeline traçado alinha-se com o objetivo geral do projeto, permitindo explorar de forma sistemática os dados do Sisvan (2018–2023) e extrair conclusões significativas que retroalimentem as ações de vigilância alimentar e nutricional no SUS;

7 Discussão e Conclusão

Este projeto realizou uma análise aprofundada dos dados do Sisvan (2018–2023) para investigar o estado nutricional da população brasileira atendida na Atenção Primária à Saúde e explorar a aplicação de modelos preditivos. O pipeline analítico empregado permitiu transformar dados brutos em *insights* acionáveis, alcançando os objetivos propostos. Conclusões Principais:

- **Cenário Nutricional:** As análises exploratórias confirmaram o cenário epidemiológico brasileiro, com alta prevalência de excesso de peso (sobrepeso e obesidade) na população acompanhada pelo SUS, especialmente em adultos.
- **Variações Demográficas e Geográficas:** Identificamos diferenças significativas no estado nutricional entre sexos, fases da vida e Unidades da Federação, reforçando a necessidade de ações de saúde pública direcionadas e adaptadas a diferentes contextos populacionais e regionais.
- **Tendência de Aumento do IMC:** A análise de séries temporais evidenciou uma tendência de aumento da média de IMC ao longo dos anos (2018–2023), um alerta para a progressão do excesso de peso na população.
- **Potencial Preditivo:** Demonstramos a viabilidade de utilizar modelos preditivos nos dados do Sisvan. O modelo ARIMA forneceu projeções razoáveis para tendências agregadas de curto prazo, enquanto a Rede Neural obteve excelente desempenho na classificação individual do estado nutricional, indicando a relevância das features utilizadas e o potencial de uso de Machine Learning para identificação de risco.

O desenvolvimento deste projeto permitiu uma avaliação crítica tanto do processo analítico empregado quanto das características da base de dados utilizada, culminando na verificação do alcance dos objetivos propostos e na identificação de caminhos para aprimoramentos futuros.

8 Análise Crítica do Trabalho

As qualidades intrínsecas deste projeto residem na solidez das etapas metodológicas aplicadas. A abrangência temporal da base de dados (2018-2023) foi fundamental para as análises de tendência. Destaca-se o rigoroso trabalho de pré-processamento e limpeza, crucial para tratar inconsistências, *outliers* e valores ausentes que são comuns em grandes bases de dados administrativos, garantindo a confiabilidade dos resultados. A variedade e clareza das visualizações geradas na fase de *EDA* foram essenciais para a comunicação efetiva dos padrões e *insights* encontrados. A aplicação de modelos de diferentes naturezas. O *ARIMA* para projeção temporal e a rede neural para classificação, demonstra a exploração de técnicas analíticas avançadas. Particularmente relevante foi a validação empírica do modelo *ARIMA* com dados reais de 2023, fornecendo uma medida tangível de seu desempenho preditivo. Contudo, as limitações encontradas não podem ser negligenciadas, sendo muitas delas inerentes à própria fonte de dados. A qualidade e a cobertura dos dados originais do *Sisvan*, embora valiosos, apresentam desafios. A presença de colunas com alto percentual de valores nulos e a necessidade de lidar com potenciais problemas de formato durante a leitura (indicados, por exemplo, pelo uso de `low_memory=False`) impactaram diretamente a disponibilidade e granularidade de certas informações. Adicionalmente, o escopo limitado das features disponíveis na base pública, carecendo de dados detalhados sobre hábitos alimentares, comorbidades específicas ou um espectro mais amplo de indicadores socioeconômicos, restringiu a capacidade de realizar análises causais profundas e de construir modelos preditivos que capturem a complexidade total dos fatores associados ao estado nutricional. Embora o modelo *ARIMA* tenha sido adequado para a projeção agregada, a complexidade dos padrões nutricionais em níveis mais granulares poderia demandar modelos de séries temporais mais sofisticados. Por fim, o modelo de classificação com rede neural, apesar de promissor, foi explorado em uma configuração base; sua otimização e a comparação com outros algoritmos poderiam refinar os resultados e oferecer maior interpretabilidade sobre quais variáveis são mais determinantes para a classificação. Alcance dos Objetivos O projeto foi bem-sucedido em atingir os objetivos propostos em sua fase inicial. Conforme detalhado nas Conclusões Principais, foi possível explorar e compreender a estrutura dos dados após o pré-processamento; analisar distribuições e tendências de indicadores nutricionais por diversos recortes demográficos e geográficos, confirmando o cenário epidemiológico existente e as variações regionais; criar visualizações que facilitam a interpretação desses achados; e desenvolver modelos preditivos (*ARIMA* para projeção e Rede Neural para classificação) que demonstraram o potencial do uso de técnicas de Machine Learning nos dados do *Sisvan*. A partir destas análises e modelagens, foram gerados insights que podem, efetivamente, subsidiar discussões e o desenvolvimento de políticas públicas mais direcionadas e baseadas em evidências. Potenciais Melhorias no Projeto Com base nas limitações identificadas e nas oportunidades de aprofundamento, sugere-se uma série de potenciais melhorias para futuras iterações deste trabalho:

- Integração de Dados: Priorizar a integração dos dados do Sisvan com outras bases relevantes (e-SUS APS, Cadastro Único/Auxílio Brasil, bases de inquéritos de saúde/alimentação) para enriquecer o conjunto de features com informações socioeconômicas, de saúde e de hábitos, permitindo análises de fatores de risco mais complexas.
- Análise Geográfica Aprofundada: Utilizar técnicas de análise geoespacial para mapear a prevalência de desfechos nutricionais e identificar clusters ou áreas de alto risco que demandem atenção prioritária.
- Expansão para Outras Faixas Etárias: Realizar análises e modelagens dedicadas às populações de crianças e adolescentes, aplicando as curvas e indicadores nutricionais específicos para estas faixas etárias, que possuem particularidades importantes.
- Exploração de Outras Variáveis do Sisvan: Investigar outras informações presentes na base que não foram o foco principal deste estudo (ex: marcadores de consumo alimentar, se disponíveis e com qualidade suficiente) para análises mais holísticas.
- Desenvolvimento de Ferramentas Interativas: Construir painéis ou dashboards interativos para facilitar a exploração dos dados e dos resultados dos modelos por gestores e profissionais de saúde. A implementação destas melhorias não só mitigaria as limitações atuais, mas também ampliaria significativamente o escopo e o impacto potencial do projeto, gerando insights ainda mais precisos e acionáveis para a saúde pública nutricional no Brasil.

9 Referências

- APRELINI, C. M. O. et al. Tendência da prevalência do sobrepeso e obesidade no Espírito Santo: estudo ecológico, 2009-2018. *Epidemiologia e Serviços de Saúde*, v.30, n.3, e2020961, 2021;
- BARBOSA, B. B. et al. Food and Nutrition Surveillance System (SISVAN) coverage, nutritional status of older adults and its relationship with social inequalities in Brazil, 2008-2019: an ecological time-series study. *Epidemiologia e Serviços de Saúde*, v.32, n.1, e2022595, 2023;
- BRASIL. Ministério da Saúde. Marco de referência da vigilância alimentar e nutricional na atenção básica. Brasília: Ministério da Saúde, 2015. 54 p.;
- BRASIL. Ministério da Saúde. Protocolos do Sistema de Vigilância Alimentar e Nutricional – SISVAN na assistência à saúde. Brasília: Ministério da Saúde, 2008;
- BRASIL. Ministério da Saúde. SISVAN – Sistema de Vigilância Alimentar e Nutricional (Portal eletrônico). Disponível em: <https://sisaps.saude.gov.br/sisvan/>. Acesso em: 22 mar. 2025;
- BRASIL. Ministério da Saúde. Vigilância Alimentar e Nutricional (Portal APS). Disponível em: <https://www.gov.br/saude/pt-br/composicao/saps/vigilancia-alimentar-e-nutricional>. Acesso em: 22 mar. 2025;
- CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO (CNPq). Aberta chamada de apoio a projetos em ciência de

dados. Brasília: CNPq/Ministério da Saúde, 20 ago. 2020. Disponível em: <https://www.gov.br/cnpq/pt-br/assuntos/noticias/destaque-em-cti/aberta-chamada-de-apoio-a-projetos-em-ciencia-de-dados>. Acesso em: 20 mar. 2025;

- FUNDAÇÃO OSWALDO CRUZ (Fiocruz). Obesidade em crianças e jovens cresce no Brasil na pandemia. *Observa Infância/Fiocruz*, 2023. Disponível em: <https://portal.fiocruz.br/noticia/obesidade-em-criancas-e-jovens-cresce-no-brasil-na-pandemia>. Acesso em: 20 mar. 2025;
- MREJEN, M.; CRUZ, M. V.; ROSA, L. O Sistema de Vigilância Alimentar e Nutricional (SISVAN) como ferramenta de monitoramento do estado nutricional de crianças e adolescentes no Brasil. *Cadernos de Saúde Pública*, v.39, n.1, e00169622, 2023;
- MOURÃO, E. et al. Tendência temporal da cobertura do Sistema de Vigilância Alimentar e Nutricional entre crianças menores de 5 anos da região Norte do Brasil, 2008-2017. *Epidemiologia e Serviços de Saúde*, v.29, n.2, e2019377, 2020;
- SILVA, O. L. O. et al. Vigilância alimentar e nutricional de crianças indígenas menores de cinco anos em Mato Grosso do Sul, 2002-2011. *Epidemiologia e Serviços de Saúde*, v.23, n.3, p.541-546, 2014;
- SILVA, R. P. C. et al. Sistema de Vigilância Alimentar e Nutricional: tendência temporal da cobertura e estado nutricional de adultos registrados, 2008-2019. *Epidemiologia e Serviços de Saúde*, v.31, n.1, e2021605, 2022.