

InSANDb: Bases de Dados Integrada, por Município Brasileiro, para Estudos da Insegurança Alimentar e Nutricional no Brasil

Arthur Henrique Chaves Oliveira¹, Sara Alves Martins¹, Daniel Rocha Franca¹
Lucas Rafael Pessoa do Nascimento¹, Heitor Ramos de Oliveira¹
Marco Paulo Gomes¹, Luis Enrique Zárate¹

¹ ¹Ciência de Dados e I.A., PUC Minas, Belo Horizonte, MG, Brasil

¹Instituto de Ciências Exatas e Informática, PUC Minas, Belo Horizonte, MG, Brasil

Abstract. *Food and nutrition insecurity (FNI) is a structural, multifaceted problem influenced by economic, sociocultural, and political issues. Understanding the conditions contributing to this insecurity can increase the effectiveness of public policies at the municipal level in Brazil. This work aims to integrate nine public data sources that help characterize FNI. As a main part of the adopted methodology, we propose the conceptual modeling this domain by identifying the factors and their multiple dimensions and, based on this, building a unified database. As a result, the database can be useful for the development of future research in data science.*

Resumo. *A insegurança alimentar e nutricional (InSAN) é um problema estrutural, multifacetado e influenciado por problemas econômicos, socioculturais e políticos. Conhecer as condições que contribuem com essa insegurança pode aumentar a eficácia de políticas públicas em nível municipal no Brasil. O objetivo deste trabalho é integrar nove fontes de dados públicas que ajudem a caracterizar a InSAN. Como parte principal da metodologia adotada, é proposta a modelagem conceitual desse domínio identificando os fatores e suas múltiplas dimensões, e com base nisso, construir uma base de dados unificada. Como resultado, a base de dados pode ser útil para o desenvolvimento de futuras pesquisas em ciência de dados.*

1. Introdução

Segundo a Organização das Nações Unidas para a Alimentação e a Agricultura (FAO), a segurança alimentar e nutricional (SAN) é um direito fundamental que assegura a todas as pessoas o acesso físico, econômico e social a alimentos em quantidade e qualidade adequadas para atender às suas necessidades biológicas, suficiente para levar uma vida ativa e saudável. A SAN está diretamente relacionada à promoção da saúde, ao desenvolvimento socioeconômico e à redução das desigualdades sociais [FAO 1996].

A existência da segurança alimentar e nutricional pressupõe também o seu contraponto, a insegurança alimentar e nutricional (InSAN), que se define pela dificuldade ou impossibilidade de acesso regular e permanente a alimentos de qualidade. A InSAN pode se manifestar em diferentes formas, desde a preocupação com a disponibilidade futura de alimentos até a fome propriamente dita. Essa condição está associada a fatores socioeconômicos, políticas públicas ineficientes e vulnerabilidades estruturais que afetam principalmente grupos de baixa renda [Segall-Corrêa et al. 2007], [Hoffmann 2013].

Contudo, a partir de um estudo aprofundado sobre a InSAN, percebe-se que se trata de um tema multifacetado, podendo ser analisado através da ótica de diversas áreas do conhecimento, o que o torna um assunto de grande complexidade. Entre essas perspectivas, pode-se citar a visão da nutrição e da saúde pública, que analisam os impactos da InSAN na saúde da população. A visão das ciências sociais e econômicas, que investigam a influência das desigualdades sociais e econômicas, a visão da agronomia e da sustentabilidade, que abordam questões como produção, distribuição e desperdício de alimentos, a visão da psicologia e da educação, que buscam entender os impactos da falta de uma nutrição adequada no desenvolvimento cognitivo e no aprendizado, entre outras diversas dimensões.

Apesar da complexidade e da relevância do tema, há poucas pesquisas que utilizam abordagens da ciência de dados para compreender e caracterizar a InSAN. A integração e a análise de dados sobre segurança alimentar ainda são desafiadoras devido à segmentação das informações em diferentes bancos de dados, sendo muitos deles públicos. A falta de um repositório unificado dificulta a análise e a formulação de diagnósticos precisos, e a elaboração de políticas públicas mais eficazes para o enfrentamento da InSAN.

Iniciativas em outros setores mostram como é importante e possível integrar dados públicos. O Laboratório de Dados Educacionais (LDE), por exemplo, organiza e oferece dados sobre educação de vários anos, ajudando quem pesquisa e quem decide políticas. O jeito que o LDE unifica dados de diferentes épocas e lugares pode ser um bom exemplo para criar plataformas parecidas em outras áreas. Afinal, as ideias para combinar dados de tipos diferentes e com detalhes variados são bem parecidas, não importa a área. Nosso trabalho, que cria o InSANDb, segue essa linha de integrar dados. Isso mostra que ter informações unidas é muito importante para a pesquisa e para que as políticas públicas funcionem melhor. Aprender com projetos de outras áreas, como o LDE, ajuda a melhorar como integramos dados e a criar soluções que funcionam bem juntas, gerando mais conhecimento e ajudando a sociedade [Yamanaka et al. 2024].

A proposta deste trabalho é singular na medida em que busca integrar, a partir do entendimento conceitual do fenômeno InSAN, diferentes fontes de dados públicas em nível municipal, possibilitando a construção de uma base consolidada (InSANDb) que reúna informações relevantes para a caracterização da insegurança alimentar e nutricional. Essa integração, realizada a partir de nove bases distintas, permite não só uma análise mais abrangente do fenômeno em diferentes municípios do Brasil, mas também a identificação de padrões regionais e suas possíveis correlações com diversos fatores, contribuindo para um entendimento mais amplo e aprofundado da problemática. O banco de dados resultante será disponibilizado à comunidade científica, permitindo o uso de diferentes técnicas, como as de Ciência de Dados, e apoiando iniciativas futuras de combate à InSAN.

Como metodologia adotada, o estudo inicia a partir do entendimento do domínio de problema (a InSAN) e da esquematização conceitual desse domínio por meio do método CAPTO [Gonçalves et al. 2024]. O método busca identificar dimensões (diversas perspectivas do problema) e aspectos (fatores que podem influenciar o problema em níveis mais específicos). O esquema conceitual é um guia para a identificação de fontes de dados relevantes para o estudo. Após este processo, são descritos os procedimentos de

integração empregados para construir a base de dados alvo.

Este trabalho está organizado em três seções. A primeira seção trata desta breve introdução onde é apresentado o problema. Na segunda seção, apresentamos a metodologia adotada, com a descrição das fontes de dados consideradas, o entendimento do problema e o processo de integração das fontes de dados. Na terceira seção, a base de dados integrada é descrita. Finalmente, na seção de conclusões, perspectivas e futuros trabalhos para o banco de dados integrado são apontados.

2. Metodologia

2.1. Materiais: Descrição das Bases de Dados Originais

Para a análise da insegurança alimentar e nutricional no Brasil em nível municipal, foram coletadas diversas bases de dados de fontes oficiais. Entre elas, utilizamos informações provenientes da base de dados da Divisão Territorial Brasileira (DTB), da Produção Agrícola Municipal (PAM), do Censo Agropecuário, do Cadastro Nacional de Estabelecimentos de Saúde (CNES), do Cadastro Central de Empresas (CEMPRE), do Sistema de Vigilância Alimentar e Nutricional (SISVAN), do Produto Interno Bruto Municipal (PIB Munic). Além disso, buscaram-se dados oriundos de duas fontes de dados sendo elas o Ministério do Desenvolvimento Social (MDS) e do Censo Demográfico. Estes dados são essenciais para compreender fatores socioeconômicos, agrícolas, de saúde e de desenvolvimento municipal. É importante ressaltar que estas bases de dados vinculadas foram identificadas após a etapa de modelagem conceitual do domínio da InSAN.

Tabela 1. Bases de dados por ano e fonte

Bases	Fonte	2017	2018	2019	2020	2021	2022	2023	2024
PAM	IBGE	X	X	X	X	X	X	X	
Censo Agropecuário	IBGE	X							
CNES	DATASUS	X	X	X	X	X	X	X	X
CEMPRE	IBGE				X	X	X		
Novo CAGED	GOV				X	X	X	X	X
MDS	GOV	X	X	X	X	X	X	X	X
PIB Munic	IBGE	X	X	X	X	X			
Censo Demográfico	IBGE						X		
DTB	IBGE	X	X	X	X	X	X	X	X

A Tabela 1 mostra as bases de dados para os anos de 2017 a 2024 e suas respectivas fontes. As bases analisadas são provenientes de instituições como IBGE, DATASUS e GOV, abrangendo dados econômicos, demográficos e de saúde.

A Tabela 1 também permite visualizar a disponibilidade dos registros ao longo do tempo, evidenciando períodos de maior ou menor disponibilidade. Enquanto o Censo Demográfico do IBGE se revelou como a fonte de dados que apresenta a maior descontinuidade, o Cadastro Nacional de Estabelecimentos de Saúde (CNES) do DataSUS foi a base mais contínua. Estes dados são essenciais para a realização de análises históricas e estudos comparativos. É importante ressaltar que o banco de dados resultante da integração não recebeu uma etapa de pós-processamento, como o tratamento de dados ausentes, o que se espera que ocorra quando aplicado a ela um processo completo de preparação de dados para fins específicos da análise de dados resultante da integração.

2.2. Métodos

2.2.1. Entendimento do Problema e Esquema Conceitual

Com o objetivo de construir uma base de dados integrada e multifatorial para a insegurança alimentar em nível municipal, como primeira etapa da metodologia adotada, buscou-se um melhor entendimento acerca do domínio do problema. Para isso, foi utilizado o método CAPTO proposto em [Gonçalves et al. 2024] para construção de um esquema conceitual. O método objetiva a captura de conhecimento tácito, proveniente do especialista de domínio, e do conhecimento explícito, obtido de fontes distintas como literatura, relatórios técnicos e dicionários de dados. O método busca a criação de um esquema conceitual unificado para servir como guia na seleção conceitual de dimensões, aspectos e atributos que representem um domínio de problema específico.

O esquema conceitual (Figura 1) é estruturado em dimensões (perspectivas mais abrangentes dentro do domínio do problema), aspectos (uma divisão menor da dimensão) e atributos, (possíveis variáveis vinculadas a cada aspecto do domínio), os quais são vinculados aos campos das bases de dados disponíveis para o estudo. No artigo [Gonçalves and Gálvez 2024] é citada e demonstrada a importância da aplicação do método CAPTO para a construção de modelos representativos.

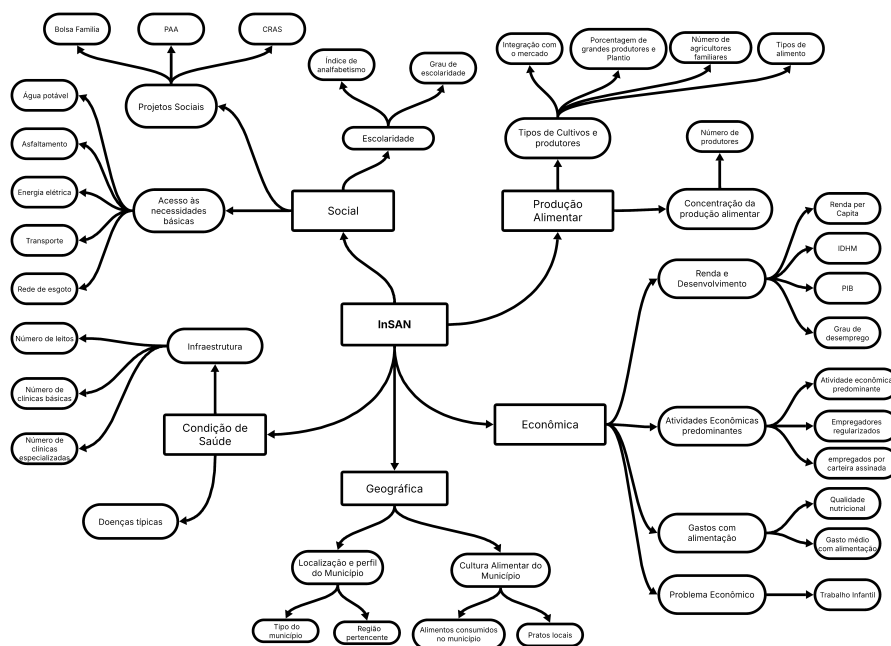


Figura 1. Esquema conceitual da InSAN: dimensões, aspectos e atributos

A partir do esquema conceitual, foram identificadas as seguintes dimensões associadas à InSAN: a) Social; b) Produção Alimentar; c) Econômica; d) Geográfica; e e) Condição de Saúde. A Tabela 2 mostra as dimensões, os aspectos, e os atributos vinculados às tabelas das bases de dados consideradas. Cada aspecto é justificado a partir de referências bibliográficas.

Tabela 2. Mapa Conceitual Tabular

Descrição do Mapa Conceitual – Domínio de problema: InSAN	
Dimensão: Social	
Referências para os Aspectos e Atributos Associados	Bases de Dados
Acesso às necessidades básicas: [Mattos et al. 2021] e [de Araújo et al. 2020]. Atributos: Transporte, Saneamento, Esgoto, Água potável, Energia elétrica, Asfaltamento	Tabela 9922 - Total; Urbana; Rural; Fonte: Censo Demográfico(2022) Tabela 9543 - Total; Fonte: Censo Demográfico(2022) Tabela 6326 - Total; Casa; Casa de vila ou em condomínio; Apartamento; Habitação indígena sem paredes ou maloca; Fonte: Censo Demográfico (2022) Tabela 6892 - Total; Coletado; Fonte: Censo Demográfico (2022) Tabela 6804 - Total; Rede geral de distribuição; Canalizada até dentro da casa; apartamento ou habitação; Canalizada, mas apenas no terreno; Sem água canalizada; Fonte: Censo Demográfico(2022) Tabela 6805 - Total; Rede geral, rede pluvial ou fossa ligada à rede; Fonte: Censo Demográfico(2022) Tabela 9542 - Alfabetizadas; Não alfabetizadas; Fonte: Censo Demográfico(2022)
Projetos Sociais: [Mattos et al. 2021] e [Bezerra et al. 2017] Atributos: Bolsa Família, PAA (Programa de Aquisição de Alimentos), Quantidade de pessoas registradas no CRAS, BPC	Quantidade De Famílias Beneficiárias, Valor Repassado e Valor Médio dos benefícios pagos do PBF - qtd.familias.beneficiarias.bolsa.familia.s; valor.repassado.bolsa.familia.s; pbf.vlr.medio.benef.f; Fonte: MDS(2017-2024) Beneficiários BPC - bpc.ben; bpc.pcd.ben; bpc.val; bpc.pcd.val; bpc.idoso.val; Fonte: MDS(2017-2024)
Escolaridade: [Mattos et al. 2021]. Atributos: Índice de analfabetismo, Grau de escolaridade	Tabela 9543 - Total; Fonte: Censo Demográfico(2022) Tabela 9542 - Total; Alfabetizadas; Não Alfabetizadas; Fonte: Censo Demográfico(2022)
Dimensão: Produção Alimentar	
Tipos de Cultivos e produtores: [Bernardes et al. 2021] e [Patel 2009]. Atributos: Tipos de alimento, Número de agricultores familiares, Diversidade de plantio, Porcentagem de grandes produtores, Integração com o mercado	Tabela 5457 - Abacate; Abacaxi*; Açaí; Alfafa fenada; Algodão arbóreo (em caroço); Algodão herbáceo (em caroço); Alho; Amendoim (em casca); Arroz (em casca); Aveia (em grão); Azeitona; Banana (cacho); Batata-doce; Batata-inglesa; Cacau (em amêndoa); Café (em grão) Total; Café (em grão) Arábica; Café (em grão) Canephora; Caju; Cana para forragem; Caju; Castanha de caju; Cebola; Centeio (em grão); Cevada (em grão); Chá-da-índia (folha verde); Coco-da-baía*; Dendê (cacho de coco); Erva-mate (folha verde); Ervilha (em grão); Fava (em grão); Feijão (em grão); Figo; Fumo (em folha); Girassol (em grão); Goiaba; Guaraná (semente); Juta (fibra); Laranja; Limão; Linho (semente); Maçã; Malva (fibra); Mamão; Mamona (baga); Mandioca; Manga; Maracujá; Marmelo; Melancia; Melão; Milho (em grão); Noz (fruto seco); Palmito; Pera; Pêssego; Pimenta-do-reino; Rami (fibra); Sisal ou agave (fibra); Soja (em grão); Sorgo (em grão); Tangerina; Tomate; Trigo (em grão); Triticale (em grão); Tungue (fruto seco); Urucum (semente); Uva; Fonte: PAM(2017-2023) Tabela 6906 - Total; Agricultura familiar - não; Agricultura familiar - sim; Fonte: Censo Agropecuário (2017)
Concentração da produção alimentar: [Campi et al. 2021]. Atributos: Número de produtores	Tabela 6906 - Total; Fonte: Censo Agropecuário (2017)
Dimensão: Condição de Saúde "Típico"	
Infraestrutura: [de Araújo et al. 2020]. Atributos: Número de leitos, Número de clínicas básicas, Número de clínicas especializadas	Recursos Físicos Ambulatório Consultórios - Clínica Básica; Clínica Especializada; Fonte: CNES(2017-2024) Recursos Físicos Ambulatório Leitos de Repouso e Observação - Leitos Repouso/Observação Ped; Leitos Repouso/Observação Masc; Leitos Repouso/Observação Femin; Leitos Repouso/Observação Indif; Fonte: CNES(2017-2024) Recursos Físicos Hospitalar Leitos de internação - Quantidade existente; Fonte: CNES(2017-2024) Recursos Físicos Urgência Leitos de Repouso/Observação - Leitos Repouso/Observação Ped; Leitos Repouso/Observação Masc; Leitos Repouso/Observação Femin; Leitos Repouso/Observação Indif; Fonte: CNES(2017-2024)
Doenças típicas: [de Oliveira Gallo and Jaime 2024].	Não foram encontradas bases para esse aspecto
Dimensão: Geográfica	
Localização e perfil do Município: [Mattos et al. 2021]. Atributos: Tipo do município (rural ou urbano), Região pertencente(Norte, Sul, Sudeste, etc)	RELATORIO.DTB.BRASIL.MUNICIPIO - UF; Nome.UF; Código Município Completo; Nome.Município; Fonte: DTB(2023) Tabela 9922 - Total; Urbana; Rural; Fonte: Censo Demográfico(2022)
Cultura Alimentar do Município: [Burity et al. 2010] e [Sambuichi et al. 2022]. Atributos: Alimentos consumidos no município, Pratos locais	Não foram encontradas bases para esse aspecto
Dimensão: Econômica	
Renda e Desenvolvimento: [Allee et al. 2021] e [Dasgupta and Robinson 2022]. Atributos: Renda per Capita, IDHM, PIB, Grau de desemprego	Tabela 13 - Unidades da Federação; Municípios; Número de unidades locais; Pessoal ocupado Total; Pessoal ocupado Assalariado; Pessoal assalariado médio; Salários e outras remunerações (1 000 R\$); Salário médio mensal (salários mínimos) (1) (2); Salário médio mensal (R\$) (2); Número de empresas atuantes; Fonte: CEMPRE (2020-2022) Tabela 5938 - Código do Município; Produto Interno Bruto per capita, a preços correntes(R\$ 1,00); Fonte: PIB Munic (2017-2021)
Atividades Econômicas predominantes: [Mattos et al. 2021]. Atributos: Atividade econômica predominante, Porcentagem de empregados por carteira assinada, Empregadores regularizados	Tabela 5938 - Código do Município; Atividade com maior valor adicionado bruto; Fonte: PIB Munic (2017-2021) Tabela 8.1 - Código do Município; UF; Município; Estoque; Fonte: Novo-CAGED (2020-2024)
Gastos com alimentação: [de Araújo et al. 2020]. Atributos: Gasto médio com alimentação, Qualidade nutricional	Não foram encontradas bases para esse aspecto
Problema Econômico: [Bernal et al. 2014]. Atributos: Trabalho Infantil	Não foram encontradas bases para esse aspecto

2.3. Transformação dos dados

Antes do processo de integração das bases de dados originais foram submetidas a um processo de transformação. Este processo foi desenvolvido na linguagem python, identificando os dados a serem extraídos de cada uma delas, formatando-os sem alterar o significado e a informação, excluindo os atributos desnecessários e transformando dados em métricas básicas de atributos específicos, sendo eles as médias, a soma dos valores, o valor máximo e o valor mínimo. Para representar o processo de transformação, o infográfico da Figura 2 é mostrado. A seguir as etapas do fluxograma são sucintamente descritas.

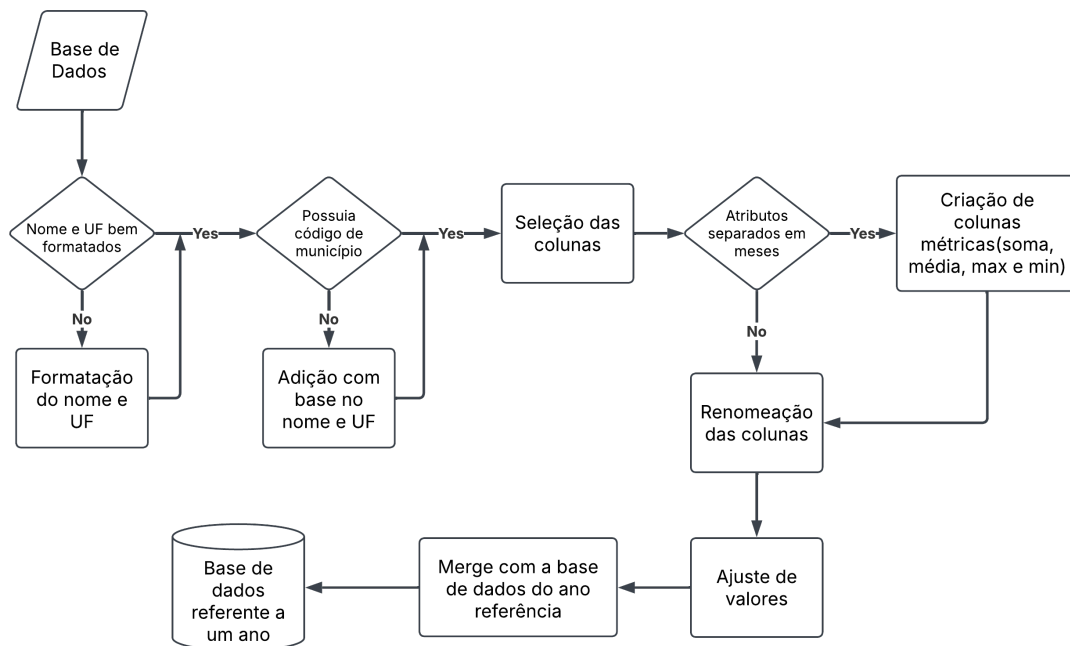


Figura 2. Fluxograma Transformação dos dados

- Para lidar com os diferentes tipos de formatações das bases de dados escolhidas, foi necessário realizar um processo de transformação para específicos campos (nomes de município, siglas da unidade federativa, formatações numéricas, etc.). Isso com o objetivo de padronizar os atributos correspondentes e possibilitar, então, a integração por meio do método "merge". Foram definidas funções para cada processo de transformação realizado nas bases de dados.
- O primeiro passo do processo de transformação correspondeu ao carregamento das bases de dados referentes a um ano específico e a seleção dos atributos vinculados ao esquema conceitual. O passo seguinte foi avaliar se a formatação de nomes do município e estado encontravam-se separados em colunas diferentes. Caso estes valores estivessem juntos, eles eram separados e formatados corretamente. Após estas etapas, foi verificada a existência de um atributo contendo o código do município e, caso esse atributo não existisse, era utilizado o nome e o estado para identificar o código em uma tabela de referência.
- Algumas bases de dados não encontravam-se agregadas em nível anual, mas sim em períodos mensais de cada ano. Para consolidar os dados, foram calculadas e inseridas as colunas que representam o total, a média e os valores máximo e mínimo para o ano específico. Utilizando estas colunas, os valores foram agregados por ano sem alterar os seus significados.
- Campos (atributos) que não poderiam ser facilmente compreendidos por interpretação humana, tiveram seu nome alterado para facilitar o seu entendimento. Adicionalmente, os valores de alguns campos foram formatados para cumprir o mesmo objetivo. Essa formatação envolvia alteração do formato numérico americano para brasileiro e a troca dos tipos das colunas que estavam incorretos.
- O último passo foi juntar as diferentes bases de dados em uma única base integrada e unificada para cada ano separadamente, usando a função "merge" da biblioteca

pandas, considerando os códigos de município como atributo chave de ligação entre os dados.

2.4. Integração

Com as bases de dados devidamente tratadas, iniciou-se a etapa final, que consistiu em reuni-los em um única base de dados tratada. Para isso, as bases tratadas e integradas para cada ano foram registradas em tabelas diferentes, em um único arquivo em formato Excel xlsx. Este formato foi escolhido por sua facilidade de manipulação. Todo esse processo foi realizado utilizando a linguagem python e pode ser observado por meio do infográfico da Figura 3.

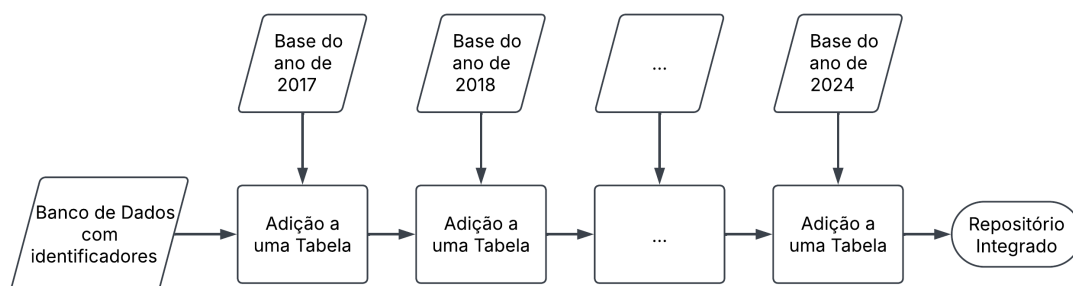


Figura 3. Integração das bases de dados anuais

As tabelas resultantes apresentam dimensionalidade diferente quanto ao número de campos. Isso se deve ao fato de diversas pesquisas não terem sido realizadas em alguns anos, fazendo com que alguns atributos não estivessem presentes. A base(Tabela_Integrada_InSAN.xlsx) e a disponibilidade de atributos por ano(features_year_InSAN.pdf) podem ser acessados a partir do link: <https://github.com/licapLaboratory/DataBaseInSAN>

3. Exploração da base de dados integrada

Após a integração, a base de dados final contém cinco categorias de dados: identificadores de município e estado, que são utilizados como chaves primárias, dados econômicos do município, dados de infraestrutura do município, dados de saúde do município e dados sociais do município. Os campos identificadores (código do município, nome e UF) atuam como chaves primárias e são a base para qualquer análise regional ou comparativa.

Quanto aos dados econômicos reunidos, a base integrada fornece informações na escala de município de sua produção econômica de frutas, grãos e cereais, tubérculos e raízes de culturas industriais, bem como da relação destas produções com o PIB local. Vale ressaltar que o IBGE, em suas publicações, como a Produção Agrícola Municipal (PAM), frequentemente separa culturas como cana-de-açúcar, fumo, algodão e outras matérias-primas industriais em categorias distintas das demais. Para este trabalho, foram integradas estas diferentes categorias, com o intuito de oferecer um panorama mais amplo, principalmente para a visualização destes dados e para futuras análises e aplicações de mineração de dados e aprendizado de máquina.

Para o processo de validação dos dados na base integrada em àqueles nas bases originais, foi utilizado o python, mais especificamente as bibliotecas pandas e matplotlib.

Foram comparadas métricas dos valores de antes e depois do processo de integração. Dentre estas métricas tivemos, para os valores numéricos, a contagem de valores, a média e o valor mínimo e máximo do atributo. Uma exemplificação dos resultados que demonstra a qualidade do banco de dados esta na Tabela ??.

Tabela 3. Análise da qualidade dos dados

Atributo	Base de dados	Contagem	Media	Mínimo	Máximo
Produto Interno Bruto per capita, a preços correntes (R\$ 1,00)	Pib Munic(2017)	5570	22.042,72	3289,5	346739,3
	integrada	5570	22.042,72	3289,5	346739,3
Pessoal ocupado Total	CEMPRE(2020)	5570	9.460,82	115	5.636.889
	integrada	5570	9.460,82	115	5.636.889

Durante esta fase, notou-se que, nas bases de dados originais, a quantidade de municípios variava entre elas e em diferentes anos. Isso ocorreu pelo fato de que, em determinado ano, o município em questão não existia ou não estava no contingente da pesquisa. Outras bases, como as retiradas do MDS, não utilizavam o código de município como identificador, mas sim um valor diferente. A tabela de referência para estes códigos não abrangia todos os 5.570 municípios brasileiros e, por este motivo, algumas informações foram perdidas durante o processo de integração.

A Figura 4 mostra um exemplo de visualização dos dados extraídos da base de dados InSANDb referente às atividades econômicas predominantes por Unidade Federativa brasileira. O gráfico, destaca setores como agricultura, indústria e serviços, permitindo identificar características regionais que podem influenciar os níveis de segurança alimentar no país.

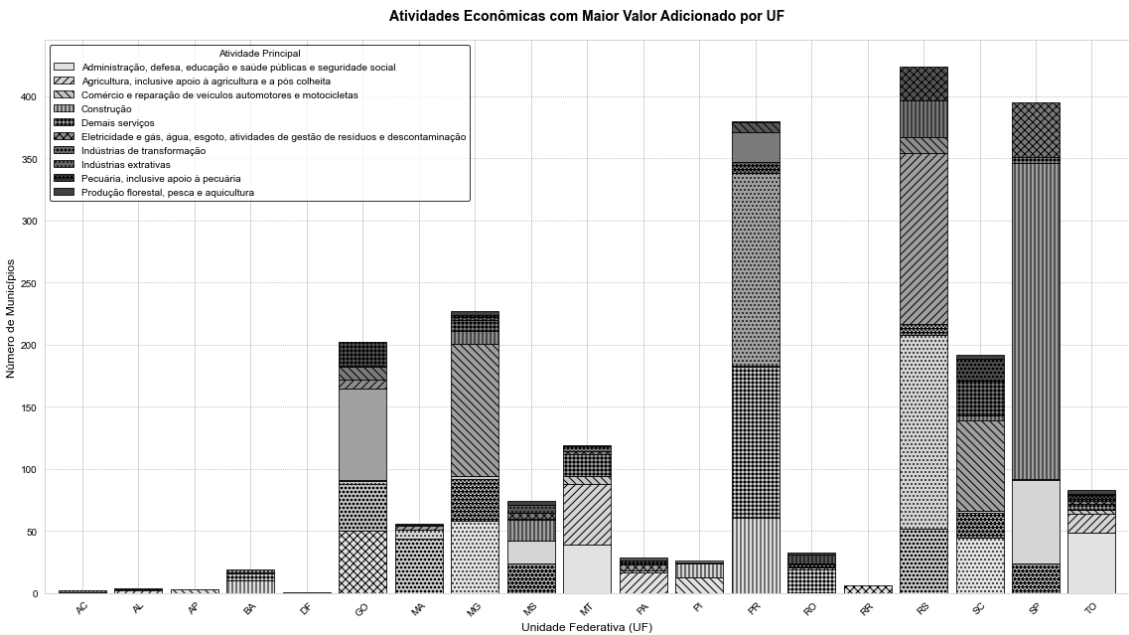


Figura 4. Atividades Econômicas com Maior Valor Adicionado por UF

Vale destacar que esta visualização é apenas um exemplo do tipo de exploração e análise de dados possíveis a partir da base de dados integrada.

Estas possíveis análises exploratórias poderiam evidenciar aspectos sociais relacionados a insegurança alimentar, podendo abordar as cinco categorias resultantes da integração das fontes de dados.

4. Conclusão

Este trabalho tratou do desafio de entender a Insegurança Alimentar e Nutricional (InSAN) no Brasil, um problema complexo e com muitas causas. A principal dificuldade encontrada na literatura era a falta de uma base de dados única em nível municipal que permitisse análises mais amplas. Os estudos anteriores costumavam ser separados ou olhar apenas para alguns aspectos, o que dificultava ter uma visão completa do problema.

A principal contribuição deste estudo foi a criação da InSANdB, uma base de dados integrada que reúne nove fontes públicas de informação, cobrindo áreas sociais, econômicas, de produção de alimentos, geográficas e de saúde para os municípios brasileiros. Diferente de trabalhos anteriores, a metodologia usada não apenas juntou dados, mas também aplicou um modelo conceitual (CAPTO) para organizar as várias dimensões da InSAN, garantindo que a base fosse consistente e representasse bem a complexidade do tema. Esse processo resultou em um recurso inédito para a comunidade científica.

Com a InSANdB, surgem novas possibilidades para a pesquisa em ciência de dados aplicada a políticas públicas. A base permite, pela primeira vez, investigar as relações entre diferentes fatores da InSAN em nível municipal, identificar padrões regionais e desenvolver modelos preditivos mais completos. Como próximos passos, sugere-se usar técnicas de aprendizado de máquina para identificar os principais fatores da InSAN em diferentes regiões do Brasil e criar painéis interativos de visualização que apoiem gestores públicos na tomada de decisões com base em evidências.

Agradecimentos

Os autores agradecem o apoio recebido do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Processo No 303133/2021-0, e da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), Processo No PCE-00349-25.

Referências

- Allee, A., Lynd, L. R., and Vaze, V. (2021). Cross-national analysis of food security drivers: comparing results based on the food insecurity experience scale and global food security index. *Food Security*, 13(5):1245–1261.
- Bernal, J., Frongillo, E. A., Herrera, H. A., and Rivera, J. A. (2014). Food insecurity in children but not in their mothers is associated with altered activities, school absenteeism, and stunting. *The Journal of Nutrition*, 144(10):1619–1626.
- Bernardes, M. S., Vieira, V. C. R., Francisco, P. M. S. B., Marín-León, L., Camargo, D. F. M., Segall-Corrêa, A. M., et al. (2021). (in) segurança alimentar no brasil no pré e pós pandemia da covid-19: reflexões e perspectivas:(in) segurança alimentar no pré e pós pandemia. *InterAmerican Journal of Medicine and Health*, 4.

- Bezerra, T. A., Olinda, R. A. d., and Pedraza, D. F. (2017). Insegurança alimentar no brasil segundo diferentes cenários sociodemográficos. *Ciência Saúde Coletiva*, 22(2):637–651.
- Burity, V., Franceschini, T., Valente, F., Recine, E., Leão, M., and Carvalho, M. d. F. (2010). *Direito humano à alimentação adequada no contexto da segurança alimentar e nutricional*, volume 91. Abrandh Brasília.
- Campi, M., Dueñas, M., and Fagiolo, G. (2021). Specialization in food production affects global food security and food systems sustainability. *World Development*, 141:105411.
- Dasgupta, S. and Robinson, E. J. (2022). Attributing changes in food insecurity to a changing climate. *Scientific Reports*, 12(1):4709.
- de Araújo, M., Rodrigues Nascimento, D., Souza Lopes, M., Mendes dos Passos, C., and Cristine Souza Lopes, A. (2020). Condições de vida de famílias brasileiras: estimativa da insegurança alimentar. *Revista Brasileira de Estudos de População*, 37:1–17.
- de Oliveira Gallo, C. and Jaime, P. C. (2024). O segundo objetivo de desenvolvimento sustentável: vigilância da insegurança alimentar e da má nutrição no brasil. *Linha Editorial Internacional de Apoio aos Sistemas de Saúde (LEIASS)*, pages 193–218.
- FAO (1996). Rome declaration on world food security and world food summit plan of action. *World Food Summit*, page 43 p.
- Gonçalves, L., Franca, D., and Zarate, L. (2024). Relevância do entendimento do domínio de problema na construção de modelos computacionais de aprendizado. In *Anais do XVIII Brazilian e-Science Workshop*, pages 135–142, Porto Alegre, RS, Brasil. SBC.
- Gonçalves, L. and Gálvez, L. (2024). Aplicação de técnicas de mineração de dados e aprendizado de máquina na comparação de perfis de avc entre idosos e adultos de meia-idade: Estudo da pns 2019. In *Anais Estendidos do XXXIX Simpósio Brasileiro de Bancos de Dados*, pages 308–317, Porto Alegre, RS, Brasil. SBC.
- Hoffmann, R. (2013). Determinantes da insegurança alimentar no brasil em 2004 e 2009. *Segurança Alimentar e Nutricional*, 20(2):219–235.
- Mattos, E. J. d., Bianchi, L. L., and Toigo, C. H. (2021). Avaliando a insegurança alimentar no brasil: a questão dos efeitos não simétricos. *Nova Economia*, 30(3):969–998.
- Patel, R. (2009). Grassroots voices: Food sovereignty. *Journal of Peasant Studies*, 36(3):663–706.
- Sambuichi, R. H. R., De Moura, I. F., Machado, J. G., and Perin, G. (2022). Contribuições do programa de aquisição de alimentos para a segurança alimentar e nutricional no brasil. Technical report, Texto para Discussão.
- Segall-Corrêa, A. M., Marin-León, R. S., Arantes, G., and Pérez-Escamilla, R. (2007). Adaptação da escala brasileira de insegurança alimentar para populações urbanas. *Revista de Nutrição*, 20:1–10.
- Yamanaka, M., de Paula, T. M., de Souza, F., Rocha, L. J., and Ribeiro, C. H. C. (2024). Statistical validation of column matching in the database schema evolution of the brazilian public school census. *Simpósio Brasileiro de Banco de Dados (SBBD)*, pages 1–12.