Construindo e Explorando *Dataset*s Curados Sobre a Produção Leiteira do Brasil

Max Felipe S. S. Cravo [1], Pedro Vieira Cruz [2], Jorge Zavaleta [3,4], Sérgio Manuel Serra da Cruz [1,4]

[1]Programa de Pós-Graduação em Informática – Universidade Federal do Rio de Janeiro (UFRJ)
[2]Programa de Pós-Graduação em Agronomia - Ciências do Solo – Universidade Federal Rural do Rio de Janeiro (UFRRJ)
[3]Departamento de Ciências Ambientais – Universidade Federal Rural do Rio de Janeiro (UFRRJ)

[4]Alda Research Institute (ALDA)

maxcravo25@gmail.com, pedrovieira.br@gmail.com,
zavaleta@pet-si.ufrrj.br, sergioserra@ic.ufrj.br

Abstract. *This article addresses the problem of data dispersion and fragmentation in the Brazilian milk production chain, a factor that limits technological advancement and productivity in the sector. The objective of the study is to build curated and unified datasets that relate climatic, economic, quality, and productivity variables across different regions of Brazil. The methodology is based on a Python Data Engineering pipeline structured into three fundamental stages: acquisition, cleaning, and provenance annotation; the differential of this work is the adoption of the W3C PROV standard to record the retrospective data provenance in textual and graph formats, ensuring traceability and reliability. As a result, the curated datasets enabled the creation of interactive visualizations on geographic maps.*

Resumo. *Este artigo aborda o problema da dispersão e fragmentação dos dados da cadeia produtiva do leite no Brasil, fator que limita o avanço tecnológico e a produtividade do setor. O objetivo do estudo é construir datasets curados que relacionem variáveis climáticas, econômicas, de qualidade e de produtividade em diferentes regiões do país. A metodologia baseia-se em um pipeline de Engenharia de Dados em Python estruturado em três etapas fundamentais: aquisição, limpeza e enriquecimento com anotação de proveniência, o diferencial do trabalho é a adoção do padrão W3C PROV para registrar a proveniência retrospectiva dos dados em formatos textual e de grafos garantindo confiabilidade e rastreabilidade dos dados. Como resultado, os datasets curados permitiram a criação de análises e visualizações interativas em mapas geográficos.*

1. Introdução

O Brasil posiciona-se no cenário global como um grande produtor de alimentos. Nossa cadeia produtiva do leite é a sexta maior do mundo e, segundo o Centro de Estudos Avançados em Economia Aplicada (CEPEA), estima-se que em 2020 ela gerou 77,1 bilhões de reais com produção aproximada de 35,37 bilhões de litros, representando 4% do PIB do agronegócio [Grigol, 2025]. Tais números evidenciam a importância da cadeia

como elemento estratégico para a economia nacional com alto impacto social e econômico e ambiental.

No entanto, diferentemente dos Estados Unidos e da Europa, os dados dessa cadeia ainda estão desconectados e dispersos em repositórios isolados em diversas organizações, cooperativas ou instituições. Essa condição reduz o diferencial competitivo da inovação da cadeia do leite no que diz respeito ao aumento da produtividade, qualidade do produto e mitigação do alto impacto ambiental gerado pela atividade. Estudos anteriores indicam a necessidade de desenvolver novas estratégias que visem aliviar esses problemas [Cruz et al., 2019].

Outro problema está relacionado com as pressões de mercado e migração de consumidores. Atualmente, os problemas de sustentabilidade na cadeia do leite afetam diretamente a aceitação do produto. Consumidores, especialmente os mais jovens, estão demandando mais transparência na cadeia produtiva e comprovação de práticas de responsabilidade ambiental. Logo, o aumento da confiabilidade dos indicadores e a rastreabilidade de qualidade do leite é essencial. No entanto, muitos indicadores encontram-se dispersos, fragmentados, muitas vezes desatualizados e sem definição clara de origens ou autoria [Hott et al., 2019]. A oferta de *datasets* curados da cadeia do leite permitirá análises mais refinadas, oferecendo uma visão integrada de como os fatores regionais e ambientais influenciam a produtividade e a qualidade final do leite.

Diante dessas limitações, elaborou-se uma estratégia computacional de baixo custo capaz de integrar dados dispersos e criar *datasets* curados a partir de séries de dados da cadeia do leite oriundas das diferentes regiões brasileiras. Essa investigação contribui para compreender a dinâmica da produção do leite, aspectos de qualidade e o desempenho das regiões do Brasil ao longo do tempo.

O objetivo deste trabalho é criar e disponibilizar *datasets* de dados curados e enriquecidos com metadados de proveniência, agregando indicadores econômicos, climáticos e de produtividade da cadeia do leite, integrando-os com parâmetros de qualidade para que analistas aprofundem análises sobre o processo produtivo do leite. Adicionalmente, oferecem-se formas de rastreabilidade e visualização de dados que permitem uma interpretação inicial do panorama nacional para diferentes tipos de especialistas.

Este artigo está organizado da seguinte forma. A seção 2 apresenta os principais trabalhos relacionados ao tema. A seção 3 apresenta os processos envolvidos na criação do *pipeline* de alta reprodutibilidade computacional para a geração dos *datasets* de dados curados e enriquecidos. A Seção 4 apresenta exemplos de visualização de dados e análises básicas. Por fim, a seção 5 apresenta as considerações finais, limitações e trabalhos futuros.

2. Trabalhos Relacionados

Atualmente existem poucos trabalhos relacionados a preparação de *datasets* curados sobre a cadeia do leite no Brasil. Em geral os trabalhos se baseiam em análises de planilhas ou arquivos CSV isolados. Dentre os principais trabalhos destacamos: [Spers et al. 2013] analisam tendências de produção e consumo de leite no Brasil utilizando o método *Delphi* e elaboraram projeções quantitativas, porém sem aplicar análises estatísticas profundas sobre os dados brutos.

[Telles et al. 2020] investigam a dinâmica e a estrutura dos sistemas de produção leiteira no Sul do Brasil no período de 2000-2015, aplicando Análise de Componentes Principais e *Clustering* nos dados. [Barros et al. 2022] realizam uma avaliação do ciclo de vida dos animais para analisar como os fatores ambientais se relacionam com a cadeia do leite em Minas Gerais e no Paraná, comparando-o com padrões internacionais. [Cordeiro et al. 2022] estudam a influência de instituições assistência técnica e extensão Rural no desempenho de fazendas leiteiras na fronteira oeste do Rio Grande do Sul. [Nääs et al. 2008] focam na predição do ciclo de estro em vacas holandesas utilizando lógica *fuzzy* e mineração de dados coletados por sensores.

Os artigos estrangeiros exploram diferentes abordagens. [Cesarini et al. 2024]. propõem modelos que correlacionam a produção leiteira na França com variáveis climáticas e econômicas como dados entrada, os autores concluem que esses modelos podem superar algoritmos tradicionais de *machine learning.* Seguindo a abordagem de aplicação de modelos, [Rajini e Sravani 2025] aplicaram 14 modelos supervisionados de aprendizado de máquina para avaliar como sete fatores afetam a qualidade do leite (pH, temperatura, sabor, odor, percentual de gordura, cor e turbidez). Os autores indicam que o odor e a turbidez são os parâmetros mais críticos para a classificação da qualidade.

Ao comparar os trabalhos da literatura nacional e internacional, observa-se que os textos nacionais focam majoritariamente nos estados ou em poucos municípios produtores e, têm ênfase na qualidade do leite e limitada preocupação com fatores ambientais. Com relação aos modelos computacionais, no Brasil e no exterior, verifica-se que os modelos estão concentrados na identificação de padrões de qualidade ou na predição de produtividade considerando variáveis ambientais. Também se verificou que apesar de existirem modelos preditivos consistentes, os trabalhos ressaltam que adotam pequenos *datasets* e mesmo assim enfrentaram problemas relacionados com a qualidade de dados e indicam a necessidade de integração de diferentes tipos de *datasets* para desenvolver estudos mais abrangentes.

3. Engenharia de Dados

Esta subseção descreve os métodos e materiais utilizados nesta pesquisa. Inicialmente, descrevemos as origens dos dados brutos utilizados nos processos de engenharia de dados (ED) concebidos para a construir os *datasets* da cadeia do leite.

Atualmente, três órgãos disponibilizam dados primários sobre indicadores econômicos relacionados à cadeia produtiva: CEPEA, Centro de Inteligência do Leite (CILeite) e o Instituto Brasileiro de Geografia e Estatística (IBGE). Além deles, o Ministério da Agricultura e Pecuária (MAPA), juntamente com a Rede Brasileira de Qualidade do Leite (RBQL), fornece, por meio do Observatório da Qualidade do Leite informações e parâmetros qualitativos do produto em âmbitos estadual e regional.

Os processos de ED permitem a consolidação de diferentes tipos de arquivos; adotou-se metodologia indicada por [Dong e Srivastava 2013] que, explicita a necessidade de rastrear, indexar, e extrair e integrar os dados em *datasets* curados através da técnica de *Record Linkage*. Para enriquecer o *Record Linkage,* foi utilizada a especificação PROV da W3C [Groth e Moreau 2013] que provê uma abordagem padronizada para registrar a rastreabilidade dos dados e processos. A especificação explicita a correlação entre dados, agentes e processos na Web. Ela prove interpretação através de anotações dos dados com metadados ou geração de grafos de proveniência. Em

última análise, a proveniência enriquece os dados e amplia a confiabilidade, auditabilidade e a reprodutibilidade de experimentos computacionais.

A construção de datasets curados e enriquecidos com metadados de proveniência permite aos especialistas da cadeia leiteira avaliar disparidades regionais e ambientais com maior precisão por meio de estatística e IA. Essa integração facilita a identificação de correlações complexas, a criação de modelos preditivos e o desenvolvimento de visualizações que antes demandavam maior esforço. Para ilustrar esse processo, a Figura 1 apresenta o pipeline conceitual do projeto, estruturado em três etapas principais: ingestão, limpeza/transformação e consolidação dos datasets fundamentados no padrão PROV.

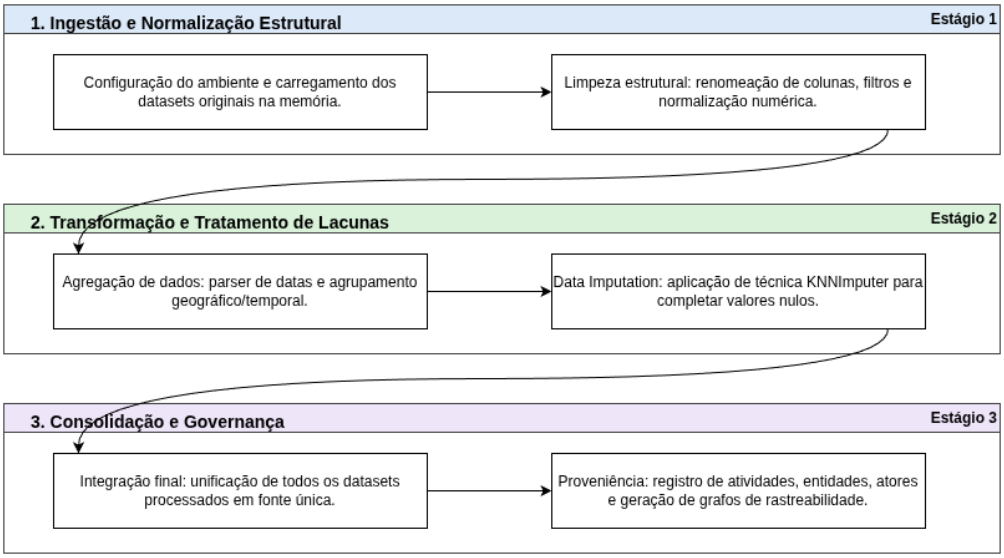


Figura 1. Representação conceitual do *pipeline* de construção de *datasets* de dados curados da cadeia do leite.

3.1. Ingestão de Dados

Os dados primários foram coletados separadamente de diversas fontes relativas ao período de 2000 a 2023. As principais variáveis são: média de produção de leite (em milhões de litros); variação do índice IPCA; a temperatura média anual das regiões produtoras; taxas de precipitação média anual; média do preço do leite pago ao produtor (por litro); e a média do preço do leite comercializado entre produtores.

Agregar as variações de temperatura e umidade é importante pois variações bruscas geram estresse térmico no rebanho e afeta diretamente a produção leiteira [Barros et al. 2022]. Para as variáveis de temperatura e precipitação, utilizou-se dados do Instituto Nacional de Meteorologia (INMET), através do portal BDMET[1] e englobando as estações meteorológicas disponíveis em cada estado avaliado. Cabe ressaltar a disparidade observada na cobertura das estações meteorológicas, sendo que a região Sudeste possui

[1]https://www.gov.br/agricultura/pt-br/assuntos/defesa-agropecuaria/laboratorios-credenciados/laboratorios-credenciados/produtos-de-origem-animal/rede-brasileira-de-qualidade-do-leite-rbql

um número consideravelmente superior. O processo envolveu a carga de centenas de arquivos em formato CSV.

A variação do IPCA foi extraída do IBGE. No entanto, dada a dificuldade em localizar a variação de forma regionalizada, optou-se por utilizar o índice nacional, com o objetivo de avaliar como a produção de cada região reage à inflação nacional, tais dados foram obtidos através de arquivos CSV. Os valores de média de produção de leite e a média do preço do leite pago ao produtor nas regiões Sul e Sudeste foram provenientes do CILeite, os dados já estavam segmentados em regiões, permitindo a análise comparativa entre as produções e a influência de outras variáveis. A aquisição de dados ocorreu através de técnicas de *web scraping* e *downloads*. Adicionalmente, foram utilizados os dados de "Preços do leite ao produtor no Brasil (deflacionado)", que ajustam os valores para remover o efeito inflacionário.

Quanto aos dados de produção e qualidade do leite identificou-se que apenas duas instituições públicas, a Embrapa e a RBQL, disponibilizam informações atualizadas periodicamente. As variáveis de qualidade do leite são relacionadas à análise microbiana, a saber: Contagem de Células Somáticas (CSS) e Contagem Padrão em Placas (CPP), já as variáveis relacionadas a análise física de sólidos do leite são, Extrato Seco Total (EST) e o Extrato Seco Desengordurado (ESD). Os dados relacionados a essas variáveis são obtidos através de uma coleta manual e reunidos em arquivos CSV agrupados por região/estado e ano.

3.2. Limpeza e Preenchimento de Falhas

A limpeza de dados processou centenas de arquivos CSV através de um *pipeline Python*. A limpeza ajustou campos com dados numéricos e textuais e detectou de erros e falhas. Identificou-se que os anos de 2000 e 2006 havia ausências de dados relativos à média do preço do leite pago ao produtor e a média do preço do leite comercializado entre produtores.

Adotou-se o método *k-nearest neighbors* (KNN) para o preenchimento de dados numéricos faltantes devido a sua eficácia comprovada em séries temporais [Ahn et al. 2021]. A técnica foi validada comparativamente aos métodos tradicionais de média e mediana utilizando dados do trabalho. O método KNN (k=5) apresentou o melhor desempenho na imputação de dados, registrando as menores taxas de erro do teste (RMSE=0,033 e MAE=0,024). O modelo KNN com k=10 obteve resultados muito próximos, enquanto as abordagens tradicionais por Média e Mediana mostraram-se as menos eficientes, com erros significativamente maiores (RMSE=0,20 e MAE=0,17).

3.3. Consolidação e Enriquecimento de Dados com Proveniência

Durante os processos de ingestão, limpeza e preenchimento de falhas, cada item de dado é relacionado a parâmetros em dois *datasets;* esse procedimento foi realizado através da biblioteca Pandas onde os parâmetros foram distribuídos em dois *datasets* ("quality_milk.csv" para a qualidade e o "production_climate_economic_milk.csv" para a produção do leite). Os metadados de proveniência são do tipo retrospectiva, utilizam a biblioteca PROV do *Python* que registram todo os processos e transformações até a construção do *dataset* final.

Cada parâmetro processado possui suas respectivas anotações de proveniência disponibilizadas em formato textual. A Figura 2 indica um fragmento de um grafo de

proveniência que descreve a sequência de execução dos *scripts* por um pesquisador e que utilizados na produção do *dataset* de produção leiteira e dados climáticos e econômicos.

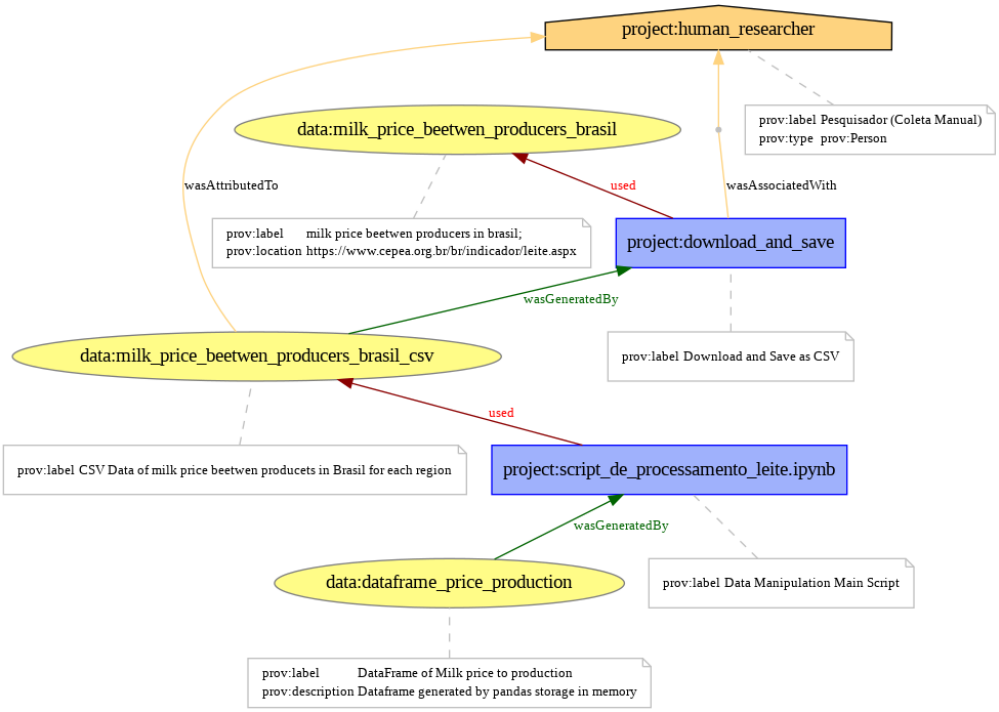


Figura 2. Fragmento de grafo de proveniência que descreve as etapas da criação desenvolvido do *dataset* de fatores de produção leiteira.

3.4. Materiais

Os materiais utilizados no desenvolvimento dos *pipelines* envolvem diversos recursos do ecossistema Python 3.11. As principais bibliotecas são: Pandas 2.2.2; NumPy 2.0.2; Scikit-Learn 1.6.1; PROV 2.1.1 e Folium 0.12. Os *scripts* foram executados no ambiente Google Colab[2] por intermédio de um *hardware* de baixo custo com processador de 2 núcleos Xeon(R) de 2.20ghz e 12.7 GB de memória RAM. A descrição completa dos códigos do *pipeline*, bibliotecas e descrição dos atributos do *dataset* e proveniência utilizados nesta pesquisa estão disponíveis em <https://zenodo.org/records/17714633>.

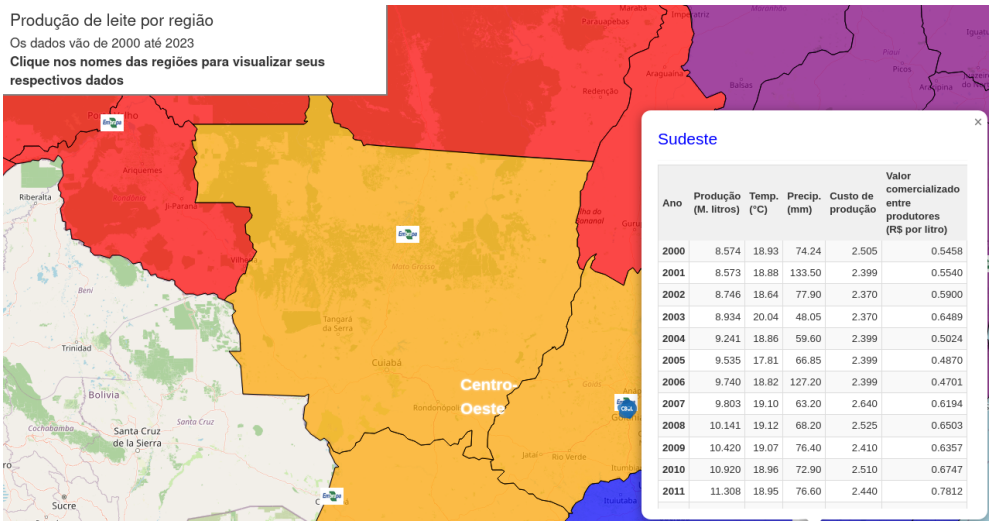
3.5. *Datasets* de Dados Curados

Os *datasets* de fatores econômicos-climáticos e de qualidade, gerados a partir dos processos supracitados, tem as características apresentadas nas Tabelas 1 e 2. Os dados do estudo estão organizados em dois conjuntos principais com características distintas. O dataset de fatores econômico-climáticos engloba uma série histórica contínua de 24 anos (2000 a 2023), contendo 24 observações agrupadas em 20 colunas, com abrangência para todas as cinco regiões do Brasil. Já o dataset de qualidade cobre um recorte temporal de 11 anos (2013 a 2023), sendo formado por 133 observações distribuídas em 10 colunas; sua cobertura geográfica possui maior granularidade, descendo ao nível estadual, com exceção dos estados das regiões Norte e Nordeste.

[2] https://colab.google/

4. Visualizações dos *Datasets*

O *dataset* resultante reúne dados de produção, fatores econômicos e climáticos associados aos processos de produção e comercialização de leite nas diferentes regiões do Brasil. A visualização do *dataset* foi elaborado utilizando-se a biblioteca *Folium*. Os dados são agrupados por territórios (região/estado/município), ano e médias climáticas de temperatura e precipitação, onde cada território é representado no mapa, e, ao ser selecionado, são apresentados com um *snapshot* dos dados presentes no *dataset*.



Ano	Produção (M. litros)	Temp. (°C)	Precip. (mm)	Custo de produção	Valor comercializado entre produtores (R$ por litro)
2000	8.574	18.93	74.24	2.505	0.5458
2001	8.573	18.88	133.50	2.399	0.5540
2002	8.746	18.64	77.90	2.370	0.5900
2003	8.934	20.04	48.05	2.370	0.6489
2004	9.241	18.86	59.60	2.399	0.5024
2005	9.535	17.81	66.85	2.399	0.4870
2006	9.740	18.82	127.20	2.399	0.4701
2007	9.803	19.10	63.20	2.640	0.6194
2008	10.141	19.12	68.20	2.525	0.6503
2009	10.420	19.07	76.40	2.410	0.6357
2010	10.920	18.96	72.90	2.510	0.6747
2011	11.308	18.95	76.60	2.440	0.7812

Figura 3. Apresentação visual do *dataset* correlacionando fatores climáticos e de produção.

O estudo utiliza mapas e representações visuais para detalhar a cadeia do leite no Brasil. A Figura 4, por exemplo, cruza dados de produção, economia e clima sobre os mapas do IBGE, destacando a presença de centros de pesquisa importantes (como Embrapa e SISPEL). Além disso, o projeto criou uma visualização semelhante focada na qualidade do leite, que pode ser acessada em formato HTML no repositório Zenodo.

4.1. Avaliações Preliminares do *Dataset*

Embora o objetivo deste trabalho não seja a aplicação de métodos de descoberta do conhecimento sobre os *datasets*, algumas análises preliminares foram realizadas para avaliar a integridade dos *datasets curados*. A Figura 5 ilustra, duas análises distintas dos dois *datasets* construidos.

Na Figura 5A (esquerda) ilustra a aplicação do coeficiente correlação de Pearson (R) no recorte temporal de 2013 a 2023 para a região Sudeste. A correlação avaliou se fatores ambientais (temperatura, precipitação) e econômicos (inflação e preços ao produtor) tem poder para influenciar na produção leiteira. Identificou-se que fatores como preço ao produtor e inflação no período tem maior peso que a própria variação da precipitação.

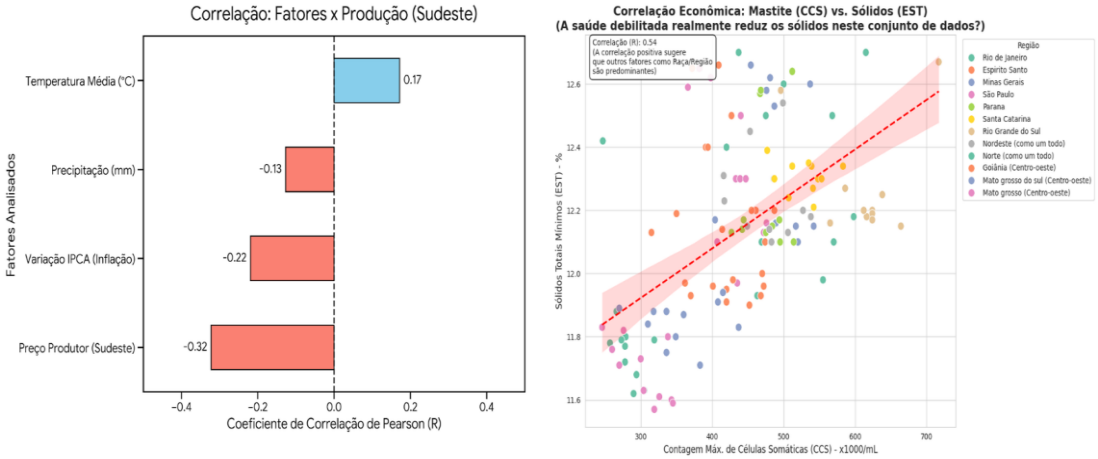


Figura 4. O gráfico a esquerda (5A) indica a correlação das variáveis de produção, ambientais e preço na região Sudeste no período de avaliado. O gráfico a direita (5B) indica a correlação entre saúde do animal e a qualidade do leite.

A Figura 5B correlaciona os indicadores de mastite CCS (eixo X) e EST (eixo Y) nos estados brasileiros entre 2000 e 2023. O Rio Grande do Sul destacou-se pela estabilidade temporal, mas com dados agrupados em cerca de 600 CCS e 12,2% de EST. Por superar o limite regulatório da IN 76/77 do MAPA (500x10³ células/mL), esse cenário pode acarretar penalizações financeiras aos produtores da região, reforçando a necessidade de políticas de manejo sanitário orientadas por dados.

Por outro lado, o estado de Minas Gerais, o maior produtor de leite nacional, possui mais variações mais amplas, em alguns anos o CSS de 300 e 450, porém com média de 11.9% de EST, enquanto outros anos o CSS varia de 450 a 500 e o EST se mantém próximo a 12.6%. Esses números indicam que o estado de Minas Gerais teve a tendência de apresentar animais mais saudáveis ao longo do tempo e possivelmente com produção mais qualificada.

5. Considerações Finais

O artigo apresenta uma estratégia computacional baseada em engenharia de dados para integrar dados dispersos sobre a cadeia do leite brasileira. A estratégia mostrou-se viável com a criação de um *pipeline, scripts* e *datasets* de dados curados e enriquecidos com proveniência, recursos reprodutíveis em equipamentos de baixo custo. Os autores destacam o ineditismo da estratégia, ressaltando a ausência de trabalhos similares na literatura nacional voltados a esse setor. Os artefatos e *datasets* resultantes desta pesquisa estão disponíveis ao público através da plataforma Zenodo.

Os autores também destacam que o trabalho é apenas um primeiro passo para contribuir com novas pesquisas do setor leiteiro. Verificou-se assimetria regional em relação aos dados, com maior escassez de informações relativas as regiões Norte/Nordeste, o que pode comprometer análises nacionais mais amplas e dificultar comparações inter-regionais. Adicionalmente, verificou-se que alguns indicadores são frequentemente restritos a poucos estados. O uso de técnicas de imputação de dados sintéticos pode atenuar essa lacuna parcialmente. No entanto, eles não oferecem a precisão necessária para estudos mais densos, o que pode acentuar as percepções de desigualdades entre as regiões. Por exemplo, a falta de dados consolidados de partes do

Nordeste contrasta com os crescentes aumentos da produção leiteira de Pernambuco nos últimos anos. Desse modo, dados mais robustos permitirão compreender as dinâmicas de crescimento ou mesmo avaliar o peso da influência de outros fatores sobre a produção, como por exemplo, investimentos em melhoramento genético, infraestrutura logística ou incentivos fiscais.

A ausência de variáveis avançadas nos repositórios brasileiros de dados leiteiros como microclima, genética, doenças e emissões de gases atua como uma limitação que restringe análises mais amplas de produção e sustentabilidade. No entanto, o *pipeline* proposto permite a fácil integração futura dessas informações. Como próximos passos, sugere-se aumentar a granularidade dos parâmetros ao nível de microrregiões ou propriedades ao integrar este projeto com a plataforma *OpenSoils* [Cruz et al, 2019]. Esses refinamentos poderão viabilizar ações extensionistas mais direcionadas e avaliações mais precisas, a exemplo do cálculo do Índice de Temperatura e Umidade (ITU) para mensurar o impacto do estresse térmico nos animais.

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento [001]; também contou com apoio do CNPq - Processos nº [310452/2025-2 e 150731/2024-8] e da FAPERJ - Processo nº [E-26/260.253/2026].

6. Referências

Ahn, H., Sun, K., and Kim, K. (2021). Comparison of missing data imputation methods in time series forecasting. Computers, Materials & Continua, 70(1):767–779.

Barros, M. V. et al. (2022). An analysis of Brazilian raw cow milk production systems and environmental product declarations of whole milk. Journal of Cleaner Production, 367:133067.

Cesarini, L. et al. (2024). Comparison of deep learning models for milk production forecasting at national scale. Computers and Electronics in Agriculture, 221:108933.

Confederação da Agricultura e Pecuária do Brasil (2025). PIB do agronegócio registra crescimento de 6,49% no primeiro trimestre de 2025. https://www.cnabrasil.org.br/. Acesso em: 13 out. 2025.

Cordeiro, M. P., Viana, J. G. A., and Silveira, V. C. P. (2022). Influence of meso-institutions on milk supply chain performance: A case study in Rio Grande Do Sul, Brazil. Agriculture, 12(4):482.

Cruz, S. M. S. et al. OpenSoils: uma plataforma de apoio à Ciência do Solo. In: CONGRESSO BRASILEIRO DE AGROINFORMÁTICA (SBIAGRO), 12., 2019, Indaiatuba. Anais [...]. Indaiatuba: [s. n.], 2019.

de Alencar Nääs, I. et al. (2008). Estimativa de estro em vacas leiteiras utilizando métodos quantitativos preditivos. Ciência Rural, 38:2383–2387.

Dong, X. L. and Srivastava, D. (2013). Big data integration. Proceedings of the VLDB Endowment, 6(11):1188–1189.

Grigol, N. (2025). Por que monitorar os preços do leite e dos lácteos? Cepea. Acesso em: 13 out. 2025.

Groth, P. and Moreau, L. PROV-overview. disponível em https://www.w3.org/TR/prov-overview/

Hott, M. C., Andrade, R. G., and Magalhães Jr., W. C. P. (2019). Distribuição da produção de leite por estados e mesorregiões.

Rajini, A. and Sravani, T. (2025). Machine learning approaches for dairy (milk) quality assurance. In Kumar, A. et al., editors. Springer Nature, Singapore.

Spers, R. G., Wright, J. T. C., and Amedomar, A. D. A. (2013). Scenarios for the milk production chain in Brazil in 2020. Revista de Administração, pages 254–267.

Telles, T. S. et al. (2020). Milk production systems in Southern Brazil. Anais da Academia Brasileira de Ciências, 92(1):e20180852.