ISSN: 2358-8829

ANÁLISE DO PERFIL SOCIOEDUCACIONAL E DESEMPENHO EM REDAÇÃO DOS ESTUDANTES DE JACOBINA-BA: MODELAGEM DE DADOS UTILIZANDO OS MICRODADOS DO ENEM

Namilly Acássia Silva Miranda [1]
Ivo Chaves De França [2]

INTRODUÇÃO

Originalmente, o Exame Nacional do Ensino Médio (ENEM), criado em 1998, surgiu com a finalidade de avaliar a qualidade do ensino recebido na educação básica. Todavia, para além desse intuito, a partir de 2009, houve uma reformulação, consolidando-se como um dos principais mecanismos de seleção para ingresso nas instituições de ensino superior (INEP, 2024; Lucena e Santos, 2020). Todos os anos, milhares de dados são gerados a partir do Enem e disponibilizados anonimamente pelo INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira) para permitir estudos e análises sobre o sistema educacional brasileiro.

Entretanto, por mais que os microdados do ENEM sejam públicos, devido ao seu volume e complexidade, eles acabam sendo difíceis de interpretar, principalmente sem o uso de ferramentas tecnológicas de tratamento e análise de dados adequadas, que ajudem na leitura desses arquivos. Diante disso, surge o seguinte problema de pesquisa: De que maneira o uso de ferramentas de análise de dados pode contribuir na compreensão acerca do perfil socioeducacional e desempenho em redação dos estudantes da região de Jacobina-BA a partir dos dados abertos do ENEM 2024?

Sob tal prisma, este estudo tem como objetivo desenvolver um modelo de dados, a partir dos microdados do ENEM 2024, que seja capaz de identificar o perfil socioeducacional dos estudantes que realizaram a prova em Jacobina-BA e analisar a viabilidade de cruzamento com o desempenho em redação, uma das competências mais decisivas do exame. Dessa forma, esse trabalho pode auxiliar análises mais precisas do contexto educacional da microrregião, apoiando gestores, escolas, educadores, pesquisadores e demais pessoas interessadas no assunto a adquirirem informações que revelem quais são as condições

[1] Formanda do Curso Técnico em Informática do Instituto Federal da Bahia - IFBA, snamilly@gmail.com;
[2] Professor orientador: Mestre em gestão e tecnologias aplicadas à educação pela Universidade do Estado da Bahia - UNEB, ivochaves@gmail.com.

ISSN: 2358-8829

socioeducacionais dos estudantes e como esses fatores se relacionam com o desempenho em redação, área decisiva que avalia o domínio da escrita formal, capacidade argumentativa e pensamento crítico, habilidades essenciais para o ingresso no ensino superior, vida acadêmica e profissional (INEP, 2024).

Nesse sentido, para ir ao encontro de tais objetivos de investigação, o uso de ferramentas open source, como o *Pentaho* PDI e *Metabase*, é indispensável na transformação desse grande volume de dados brutos em informações acessíveis sobre o cenário educacional da microrregião de Jacobina.

METODOLOGIA

Esse trabalho configura-se como um estudo exploratório, com abordagem quantitativa. Realizou-se uma revisão bibliográfica em bases de dados acadêmicas, para embasar teoricamente este trabalho, trazendo aspectos socioeducacionais já observados e abordagens metodológicas que foram empregadas em estudos correlatos. O delineamento metodológico consistiu em desenvolver e aplicar um modelo de análise de dados para mapear as características socioeducacionais dos estudantes que realizaram o ENEM 2024 e seu desempenho na prova de redação. Para isso, foram utilizados os microdados do ENEM 2024, uma base oficial de dados extensa, disponibilizada pelo INEP, que contém elementos anonimizados sobre o perfil socioeconômico, escolar e desempenho dos participantes no exame (INEP, 2024).

Devido ao elevado volume de dados, analisou-se o dicionário dos microdados para selecionar como recorte da pesquisa apenas os estudantes que realizaram o exame em Jacobina-BA, visando atualizar a literatura local. Para isso, foram delimitadas variáveis estratégicas agrupadas em quatro eixos principais: informações pessoais do participante, dados demográficos da localidade, perfil socioeconômico e o desempenho obtido na prova de redação.

Sob esse viés, para compreender o fenômeno tal como se apresenta na realidade, a construção do modelo de dados atualizado para extrair informações da realidade local, se deu de forma processual. Foram utilizadas ferramentas de *Business Intelligence* (BI) de código aberto, como a plataforma *Pentaho Data Integration* (PDI) para o tratamento de dados ETL (Extração, Transformação e Carga), integrado a ferramenta *Metabase*, para visualizar as informações (Metabase, 2024; Pentaho, 2025).

ISSN: 2358-8829

REFERENCIAL TEÓRICO

No cenário contemporâneo, é notável que estamos vivendo em uma época marcada pela geração massiva e ininterrupta de dados. No entanto, o simples acúmulo destes não se traduz em conhecimento. Diante disso, para que esses registros sejam úteis, é indispensável o uso de ferramentas de *Business Intelligence* (BI), para analisar, processar, minerar e transformar grandes volumes de dados em informação.

No campo da educação, o BI pode viabilizar a transformação de dados brutos como os microdados do ENEM em informações estratégicas, possibilitando identificar fatores que muitas vezes permanecem invisíveis e que podem estar impactando diretamente no ensino-aprendizagem e consequentemente no desempenho dos estudantes que prestam a prova do ENEM para ingressar no ensino superior.

O estudo de Medeiros, França e Souza (2024) destacou a eficiência do PDI e do Metabase no tratamento dos microdados do ENEM. Enquanto o PDI realiza o ETL, o Metabase complementa o processo criando gráficos e painéis interativos de forma intuitiva, transformando dados complexos em informações acessíveis.

Segundo Medeiros, França e Souza (2024), que realizaram uma análise exploratória, no período entre 2021 à 2023 na cidade de Jacobina, quando a estrutura dos microdados do ENEM ainda permitia cruzamento, existe uma correlação direta entre cor/raça, renda e tipo de escola no desempenho final dos estudantes. Candidatos autodeclarados pretos, pardos e indígenas frequentemente apresentam médias inferiores aos brancos, enquanto o aumento da renda familiar está intrinsecamente ligado a melhores resultados em todas as áreas do conhecimento. No que se refere à rede de ensino, a disparidade entre o desempenho de estudantes de escolas públicas e privadas ainda evidencia um cenário de desigualdades.

Além disso, um estudo de revisão sistemática da literatura de Dultra, Júnior, Fernandes (2023) entre os anos 2014 e 2021 sobre os principais fatores que podem influenciar no desempenho dos estudantes na prova do ENEM, aponta como resultados, que entre os fatores mais relevantes, estão questões socioeconômicas, como: renda familiar, sexo, cor/raça, nível de escolaridade e ocupação dos pais.

RESULTADOS E DISCUSSÃO

A análise da amostra foi realizada com base nos microdados do ENEM 2024, referentes ao município de Jacobina–BA, os quais foram identificados 3.565 inscritos. A utilização de

ISSN: 2358-8829

ferramentas de BI, como o PDI e o *Metabase*, mostrou-se eficaz no tratamento, organização e visualização do grande volume de dados. Entretanto, foram identificadas algumas limitações no processo de análise, principalmente referente à estrutura dos microdados, que em 2024 foram disponibilizados em duas bases de dados distintas, uma para PARTICIPANTES_2024 e outra para os RESULTADOS_2024, sem a presença de uma chave de ligação comum que possibilite o cruzamento entre os dados socioeconômicos e as notas de redação. Essa medida segundo o INEP (2024) se deu em conformidade com a Lei Geral de Proteção de Dados (LGPD), visando garantir o anonimato dos participantes.

Apesar das limitações, os dados corroboram com os objetivos da pesquisa ao traçar o perfil dos participantes. Identificou-se maior participação feminina (62,41%) e de jovens entre 17 e 19 anos (55,59%), sinalizando que o exame atrai sobretudo concluintes ou egressos do ensino médio, majoritariamente da rede pública (84,57%). Quanto à cor/raça, a maioria se autodeclarou parda (55%) ou preta (23,9%), o que evidencia a predominância de participantes negros em alinhamento com a demografia de Jacobina-BA. Sob a ótica socioeconômica, sobressai uma vulnerabilidade expressiva: 50,01% possuem renda de até R$ 1.412,00 e 16,55% informaram não possuir renda.

Por fim, no aspecto da inclusão digital, notou-se que embora 86,65% disponham de conexão Wi-Fi, um contingente de 2.290 estudantes de baixa renda não tem acesso a computador ou notebook em casa, o que demonstra uma forte relação entre a condição de renda e o acesso às ferramentas tecnológicas essenciais para o preparo do exame.

No quesito ocupação e escolaridade dos responsáveis, os dados revelam um contexto familiar caracterizado por uma forte representatividade de atividades que não exigem formação de nível superior (Grupos 1[3] e 2[4]) e predominância de níveis até o Ensino Médio completo, com presença também de percursos marcados pela não conclusão da educação básica, o que evidencia a desigualdade estrutural presente no acesso a ocupações de maior remuneração, refletindo limitações no acesso ao ensino superior e à qualificação profissional, que consequentemente pode influenciar as oportunidades educacionais dos estudantes.

Ademais, quanto ao desempenho em redação a análise foi realizada com cautela e em decorrência da omissão de dados do alto quantitativo de participantes (1.375) sobre o tipo e localização da escola que frequentaram no ensino médio, não foi possível fazer cruzamentos

[3] Grupo 1: Lavrador, agricultor sem empregados, bóia fria, criador de animais (gado, porcos, galinhas, ovelhas, cavalos etc.), apicultor, pescador, lenhador, seringueiro, extrativista.
[4] Grupo 2: Diarista, empregado doméstico, cuidador de idosos, babá, cozinheiro (em casas particulares), motorista particular, jardineiro, faxineiro de empresas e prédios, vigilante, porteiro, carteiro, office-boy, vendedor, caixa, atendente de loja, auxiliar administrativo, recepcionista, servente de pedreiro, repositor de mercadoria.

ISSN: 2358-8829

conclusivos que ilustrasse com precisão as médias da realidade local por dependência administrativa (federal, estadual, privada) ou localização (urbana/rural). Contudo, os números obtidos apontam para médias superiores em instituições privadas e federais em comparação à rede estadual. Destaca-se, ainda, a expressiva taxa de ausência na prova de redação 1.158 ausentes, fenômeno relevante que pode servir de dado para investigações posteriores caso esse alto percentual tenda a persistir.

Diante dos resultados e discussões supracitados, ainda que a estrutura atual das bases tenha impedido correlações diretas entre variáveis socioeducacionais e o desempenho na redação, os achados vão ao encontro dos estudos correlatos de Medeiros, França, Souza (2024) e Dultra, Júnior, Fernandes (2023), que apontam a influência de questões socioeducacionais como renda familiar, nível de escolaridade e ocupação dos pais, sexo, cor/raça, tipo e localização da escola, como alguns dos principais condicionantes no desempenho acadêmico.

CONSIDERAÇÕES FINAIS

Desenvolver e aplicar um modelo de análise de dados, fundamentado em ferramentas de BI de código aberto, possibilitou o mapeamento do perfil socioeducacional dos estudantes da microrregião de Jacobina, Bahia. Conclui-se que o uso dos softwares livres, *Pentaho* PDI e *Metabase*, mostrou-se altamente eficaz no tratamento e visualização de grandes volumes de dados, bem como democratização do acesso a informações acessíveis do contexto educacional local. O modelo de análise de dados permitiu identificar um perfil majoritariamente jovem, feminino, autodeclarado pardo/preto, de baixa renda, oriundo de escola pública, com contexto familiar marcado por responsáveis de baixa escolaridade e com atividades laborais que não exigem formação de nível superior. Além de evidenciar um paradoxo entre alto acesso à internet e ausência de recursos tecnológicos de suporte à aprendizagem. No que tange ao desempenho em redação, a reestruturação das bases dos microdados disponibilizados pelo INEP em 2024 e a omissão de dados de uma quantidade expressiva de participantes sobre o tipo e localização escolar, inviabilizaram cruzamentos estatísticos conclusivos, o que resultou em lacunas informacionais que dificultaram uma análise completa do perfil e desempenho em redação dos participantes. Contudo, entre a quantidade de participantes que consta essa informação, as maiores médias foram observadas em estudantes de instituições federais e privadas, outro dado que se sobressai é a expressiva taxa de ausência no exame.

De modo geral, o modelo proposto demonstra potencial para subsidiar discussões crítico-reflexivas sobre o planejamento de políticas educacionais mediadas por dados e orientadas pelas reais necessidades locais, contribuindo para o debate sobre a cultura digital na educação e apontando caminhos para o avanço no uso de tecnologias digitais na análise de dados acerca do ENEM. Por fim, recomenda-se que pesquisas futuras acompanhem a evolução das bases do INEP e explorem abordagens metodológicas complementares que contornem as lacunas informacionais, aprofundando a compreensão sobre como as desigualdades socioeducacionais podem impactar o desenvolvimento das competências de escrita, argumentação e pensamento crítico dos estudantes.

Palavras-chave: ENEM, Análise de dados, Metabase, PDI.

REFERÊNCIAS

DUTRA, J. F.; JÚNIOR, J. B. F.; FERNANDES, D. Y. de S. Fatores que podem interferir no desempenho de estudantes no ENEM: uma revisão sistemática da literatura. **Revista Brasileira de Informática na Educação**, [S. l.], v. 31, p. 323–351, 2023. DOI: 10.5753/rbie.2023.3087. Disponível em: https://journals-sol.sbc.org.br/index.php/rbie/article/view/3087. Acesso em: 15 jan. 2026.

INEP. INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA . **Microdados do Enem 2024**. Brasília: Inep, 2025. Disponível em: < https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enem>. Acesso em: 16 dez. 2025.

LUCENA, João Paulo Oliveira; SANTOS, Heric Nero Lisboa dos. A relação entre desempenho no Exame Nacional do Ensino Médio e o perfil socioeconômico: um estudo com os microdados de 2016. **Revista de Gestão e Secretariado**, v. 11, n. 2, p. 1-23, 2020. Disponível em: https://ojs.revistagesec.org.br/secretariado/article/view/994/pdf. Acesso em: 15 jan. 2026.

MEDEIROS, K. A; FRANÇA, I. C; SOUZA, R. F. Uso de dados abertos do enem na construção de modelos de análise educacional: evidências da microrregião de Jacobina-ba. Anais XI Congresso Nacional de Educação... Campina Grande: **Realize Editora**, 2025. Disponível em: <https://editorarealize.com.br/artigo/visualizar/133289>. Acesso em: 16 dez. 2025.

METABASE. **Metabase Documentation.** 2024. Disponível em: https://www.metabase.com/docs. Acesso em: 25 jan. 2026.

PENTAHO. **Pentaho Data Integration 11.0**. 2025. Disponível em: <https://docs.pentaho.com/pdia-data-integration>. Acesso em: 25 jan. 2026.