

## USO DE DADOS ABERTOS DO ENEM NA CONSTRUÇÃO DE MODELOS DE ANÁLISE EDUCACIONAL: EVIDÊNCIAS DA MICRORREGIÃO DE JACOBINA-BA

Ketlly Almeida Medeiros<sup>1</sup>

Raíssa Farias Souza<sup>2</sup>

Ivo Chaves De França<sup>3</sup>

### INTRODUÇÃO

O Exame Nacional do Ensino Médio (ENEM) funciona como um dos principais caminhos de acesso à universidade, disponibilizando dados que permitem analisar o desempenho acadêmico dos candidatos em relação a fatores socioeconômicos, evidenciando as raízes das desigualdades na educação.

Apesar da disponibilidade desses dados no Portal do INEP, são raros os trabalhos que utilizam modelagens analíticas de forma sistemática e eficaz, capazes de transformar essas informações em ferramentas diagnósticas úteis para gestores educacionais. Os dados, embora estruturados, exigem tratamento e modelagem para possibilitar cruzamentos legíveis para quem não é da área de ciência de dados.

Este trabalho objetiva desenvolver um processo simplificado, usando ferramentas de modelagem e análise de dados, para extrair informações consolidadas que evidenciem diferenças de desempenho relacionadas à origem da escola (pública ou privada), raça/cor e desempenho nas áreas do conhecimento avaliadas pelo ENEM. Foram adotadas técnicas de Business Intelligence (BI) com ferramentas open source, como Pentaho PDI e Metabase, permitindo um processo de extração, transformação e carga de dados (ETL). O BI compreende um conjunto de soluções e técnicas de tratamento de dados históricos que apoia na consolidação e análise das informações (Barbieri, 2011).

Tanto o Metabase quanto o Pentaho PDI transformam grandes volumes de dados em painéis e relatórios interativos, tornando a visualização mais clara e acessível. Isso auxilia gestores educacionais e estudiosos de políticas públicas na tomada de decisões

---

<sup>1</sup> Formando do Curso de Técnico em Informática do Instituto Federal da Bahia - IFBA, [Ketllyalmeida453@gmail.com](mailto:Ketllyalmeida453@gmail.com) ;

<sup>2</sup> Formando do Curso de Técnico em Informática do Instituto Federal da Bahia - IFBA, [Fariassouzaraiissa@gmail.com](mailto:Fariassouzaraiissa@gmail.com) ;

<sup>3</sup> Mestre em gestão e tecnologias aplicadas à educação pela Universidade do Estado da Bahia - UNEB, [ivochaves@gmail.com](mailto:ivochaves@gmail.com) ;



estratégicas, transformando dados em suporte essencial para a melhoria do sistema educacional.

Estudos, como o da UERJ (Faillace, Britto e Costa, 2022), utilizaram análise exploratória e clusterização nos microdados do ENEM-2019, identificando perfis socioeconômicos distintos e evidenciando o impacto das condições socioeconômicas no desempenho estudantil. Outros trabalhos, como Silva et al. (2020) da UFJF, aplicaram técnicas de mineração de dados para identificar fatores que afetam o desempenho e evidenciar disparidades sociais.

Pesquisas indicam que identidade racial, renda familiar, formação dos pais, instituição de ensino e localização da residência impactam decisivamente o desempenho no ENEM (USP; UNIFESP, 2024). Relatórios como o de Carvalhaes et al. (2023) mostram que a disparidade racial nas médias do ENEM aumentou 44% na rede pública entre 2013 e 2021, e que a pandemia da COVID-19 agravou desigualdades existentes, diminuindo a participação de estudantes de baixa renda e aumentando diferenças de desempenho.

Neste estudo, a modelagem foi aplicada aos dados do ENEM de estudantes que realizaram a prova em Jacobina-BA, cidade pólo da microrregião. A pesquisa bibliográfica revelou que, entre 2021 e 2023, há grandes desigualdades de desempenho entre zonas rurais e urbanas, redes pública e privada, raça e renda familiar. Analisar essas condições é crucial para identificar barreiras e promover a igualdade, permitindo direcionar ações mais eficazes para superá-las.

O trabalho apresenta detalhadamente a metodologia, materiais e técnicas utilizadas, discute os resultados à luz de estudos correlatos e, por fim, traz considerações finais com sugestões para projetos futuros.

## **METODOLOGIA**

A pesquisa fundamenta-se em um estudo exploratório de natureza bibliográfica, empregando uma abordagem quantitativa. Para isso, são ponderados os microdados do ENEM, que foram disponibilizados pelo INEP. O período de avaliação compreende-se os anos de 2021 a 2023, em virtude da ausência de dados referentes a 2024 no início da coleta para o estudo. A cidade de Jacobina-BA foi anteposto como âmbito de análise por sua relevância como polo regional para a execução do exame, recebendo participantes de diversas áreas da microrregião, incluindo zonas urbanas e rurais.



A fase inicial envolveu uma revisão bibliográfica detalhada. Foram consultados artigos científicos, relatórios institucionais, documentos técnicos e dissertações com o desígnio de identificar os fatores que afetam o desempenho no ENEM e conhecer as abordagens metodológicas já empregadas em estudos similares. Essa etapa também serviu para fundamentar teoricamente o conceito de desigualdade educacional, bem como seus efeitos no acesso à formação superior.

Em seguida, a pesquisa consistiu na análise exploratória dos microdados do ENEM. A partir da verificação da estrutura das bases de dados, foram selecionadas as variáveis de interesse, abrangendo a rede de ensino de origem (pública ou privada), renda familiar, raça/cor, localização geográfica e o desempenho nas diversas áreas do conhecimento.

A etapa de tratamento e organização dos dados foi realizada com ferramentas de Business Intelligence (BI) de código aberto. Para o processo de ETL, foi empregado o Pentaho Data Integration (PDI), operado pelo ambiente gráfico Spoon, o que permitiu estruturar os dados para análises futuras.

Por fim, os dados tratados foram processados no Metabase, uma ferramenta de visualização de dados. Isso permitiu a criação de painéis interativos e relatórios dinâmicos, o que tornou mais fácil a identificação de padrões e desigualdades no desempenho dos participantes. Essa abordagem forneceu a base para interpretações e comparações com outras pesquisas em políticas públicas e educação.

## RESULTADOS E DISCUSSÃO

A análise dos microdados do ENEM entre 2021 e 2023, referentes a Jacobina-BA, partiu de mais de 10 milhões de registros nacionais, filtrando cerca de 7 mil da cidade para a análise. Mesmo com a redução, os dados confirmam tendências sobre o impacto das questões socioeconômicas no desempenho escolar. Houve aumento no número de registros ao longo do tempo, com breve queda em 2021, possivelmente devido a fatores externos como a pandemia.

Em relação à renda, observou-se que quanto maior a renda declarada, melhores as médias em todas as áreas de conhecimento, enquanto estudantes sem renda obtiveram as notas mais baixas. Esse padrão confirma estudos da USP e Unifesp (2021) e relatórios da UERJ (2022) sobre disparidades socioeconômicas e raciais no ENEM. Também houve diferença entre alunos de escolas públicas e privadas, estudantes de



privadas apresentaram desempenho superior, seguidos por federais, enquanto os da rede estadual tiveram as menores médias, refletindo desigualdades educacionais regionais.

O uso do Pentaho e Metabase mostrou-se eficaz para transformar grandes volumes de dados em informações acessíveis, permitindo cruzamentos e análise de séries históricas, inclusive de indicadores locais. Apesar de limitações na qualidade dos microdados, foi possível delinear o perfil dos estudantes e identificar padrões de desigualdade.

As análises indicaram que candidatos autodeclarados pretos, pardos e indígenas tiveram médias inferiores aos brancos; mulheres tiveram melhor desempenho em linguagens e redação, enquanto homens se destacaram em matemática; e estudantes de áreas urbanas obtiveram médias superiores aos da zona rural, evidenciando a influência da infraestrutura e acesso a recursos educacionais.

De modo geral, os resultados comprovam que ferramentas de Business Intelligence aplicadas aos microdados do ENEM contribuem para compreender desigualdades no desempenho escolar, oferecendo evidências locais que podem subsidiar políticas públicas mais direcionadas e efetivas.

## CONSIDERAÇÕES FINAIS

A análise dos dados do ENEM para o município de Jacobina-BA mostrou a persistência de disparidades significativas no rendimento escolar, atreladas a aspectos socioeconômicos, raciais, de gênero, geográficos e ao tipo de instituição de ensino. Os resultados confirmam padrões já observados em estudos nacionais, mas trazem também um olhar regionalizado, que contribui para a compreensão das especificidades locais. Esse recorte reforça que desigualdades amplamente conhecidas em escala nacional também se manifestam no interior do estado, o que exige políticas públicas focalizadas e sensíveis às realidades microrregionais.

A utilização de ferramentas de *Business Intelligence* mostrou-se eficiente na transformação de grande volumes de dados em informações acessíveis, possibilitando a identificação de padrões e disparidades educacionais. Essa experiência reforça o potencial do uso de softwares livres, como Pentaho e o Metabase, para auxiliar gestores e pesquisadores a elaborar diagnósticos mais precisos e estratégias de enfrentamento das desigualdades.



Apesar das limitações decorrentes da qualidade e completude dos microdados, o estudo cumpriu seus objetivos, oferecendo evidências importantes sobre como as condições sociais e o ambiente escolar impactam o desempenho dos alunos. Como perspectiva futura, recomenda-se a ampliação da análise para incluir períodos mais longos e novas variáveis, como acesso à internet, participação em programas sociais e condições de infraestrutura escolar.

Dessa forma, a pesquisa poderá contribuir ainda mais para a formulação de políticas públicas que promovem maior equidade no acesso à educação de qualidade e no desempenho acadêmico dos estudantes.

Palavras-chave: Enem; Análise De Dados, Metabase, Ensino Médio.

## REFERÊNCIAS

BARBIERI, Carlos. **BI2 - Business Intelligence: modelagem & qualidade**. Rio de Janeiro: Elsevier, 2011. 392 p.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **Microdados do ENEM**. Brasília: INEP, 2025. Disponível em: <https://www.gov.br/inep>. Acesso em: 16 ago. 2025.

CARVALHAES, Flavio; KLITZKE, Melina; CASTRO, Daniel; BARTHOLO, Tiago. **Oportunidades educacionais de estudantes concluintes do ensino médio: desempenho médio no ENEM entre 2013 e 2021**. LaPOpE / NIED – UFRJ, com apoio do Instituto Unibanco, 2023. Disponível em: <https://www.institutounibanco.org.br/aprendizagem-em-foco/2023/11/oportunidades-educacionais-de-estudantes-concluintes-do-ensino-medio-desempenho-medio-no-enem-entre-2013-e-2021/>. Acesso em: 18 ago. 2025.

FAILLACE, Helena Ferreira Paraiso; BRITTO, Isadora Lopes Maldonado; COSTA, Fernanda da Serra. **O desempenho dos alunos do ensino médio no ENEM 2019 e a desigualdade social: um estudo utilizando análise exploratória e técnicas de agrupamento**. Cadernos do IME – Série Estatística, Universidade do Estado do Rio de Janeiro, v. 49, n. 2, p. 1-16, 2022. Acesso em: 14 ago. 2025.

HITACHI VANTARA. **Pentaho Data Integration (PDI) – User Guide**. 2023. Disponível em: <https://help.hitachivantara.com>. Acesso em: 16 ago. 2025.

METABASE. **Metabase Documentation**. 2024. Disponível em: <https://www.metabase.com/docs>. Acesso em: 16 ago. 2025.

SILVA, Vinicius Alberto Alves da; MORENO, Lorenza Leão Oliveira; GONÇALVES, Luciana Brugiolo; SOARES, Stênio São Rosário Furtado; SOUZA JÚNIOR, Robson



Rocha. **Identificação de desigualdades sociais a partir do desempenho dos alunos do Ensino Médio no ENEM 2019 utilizando mineração de dados.** In: Simpósio Brasileiro de Informática na Educação (SBIE), 31., 2020. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2020. p. 1715-1724. Acesso em: 14 ago. 2025.

**SANTOS, Cíntia Velasco; FONTOURA, Helena Amaral da (orgs.).** *Processos formativos e desigualdades sociais: a educação entre políticas e práticas.* Rio de Janeiro: UERJ, 2022. Disponível em: <https://www.cepuerj.uerj.br/cursos2.php?ano=2022&curso=E02660&tipo=cursos>. Acesso em: 2 set. 2025.

**SENA, Elisa Thomé; HELENE, Otaviano.** **Estudantes negros e de baixa renda têm desempenho prejudicado no Enem, constata estudo.** *Jornal da USP*, São Paulo, 15 mar. 2024. Disponível em: <https://jornal.usp.br/diversidade/estudantes-negros-e-de-baixa-renda-tem-desempenho-prejudicado-no-enem-constata-estudo/>. Acesso em: 14 set. 2025.

**SANTOS, Cíntia Velasco; FONTOURA, Helena Amaral da.** **Processos formativos e desigualdades sociais.** Rio de Janeiro: UERJ, 2022. Disponível em: <https://www.cepuerj.uerj.br/cursos2.php?ano=2022&curso=E02660&tipo=cursos>. Acesso em: 2 set. 2025.

