



**INSTITUTO FEDERAL DO PARANÁ**  
**CAMPUS PINHAIS**  
**CURSO TÉCNICO INTEGRADO EM INFORMÁTICA**

**JORGE HENRIQUE KALLUF DE RAMOS**  
**MIGUEL CORREIA CRUZ RODRIGUES**

**ENELYTICS – UMA ANÁLISE DOS MICRODADOS DO ENEM**  
**2023**

**PINHAIS**  
**2025**

**JORGE HENRIQUE KALLUF DE RAMOS**  
**MIGUEL CORREIA CRUZ RODRIGUES**

**ENELYTICS – UMA ANÁLISE DOS MICRODADOS DO ENEM 2023**

Trabalho apresentado ao Curso Técnico Integrado em Informática do *campus* Pinhais, do Instituto Federal do Paraná, como requisito parcial à aprovação na disciplina de Tópicos em Tecnologia da Informação.

Orientador:  
Prof. Dr. João Paulo Orlando

**PINHAIS**  
**2025**

## RESUMO

Os microdados do ENEM 2023 abrangem milhões de registros de diferentes contextos pessoais e cidades de todo o território nacional brasileiro. Através da análise exploratória desse conjunto de dados, por meio da implementação de algoritmos de aprendizado de máquina – e outras técnicas de ciência de dados – é possível identificar os principais fatores socioeconômicos, demográficos e institucionais que influenciam o desempenho dos estudantes brasileiros. Utilizando a linguagem de programação *Python* e as bibliotecas externas *Pandas*, *Matplotlib* e *Seaborn* é possível visualizar relações entre as variáveis do conjunto de dados. Os modelos supervisionados de *machine learning*, como *Decision Tree*, *Random Forest* e *K-Nearest Neighbors*, serão implementados e treinados através da biblioteca externa *Scikit-learn*, para posteriormente terem suas acurácias comparadas. Dessa forma, a mineração dos microdados contribui para a reflexão crítica sobre o acesso e a qualidade da educação no Brasil, reforçando a importância da equidade no processo de aprendizagem. O ENEM, enquanto uma das principais ferramentas de avaliação educacional do Brasil, reflete em seus resultados as diversas realidades dos estudantes. Por conta da amplitude do exame, acredita-se que, ao término do trabalho, serão evidenciadas desigualdades educacionais em diferentes contextos educacionais.

**Palavras-chave:** ENEM; Dados; Aprendizado de máquina; Modelos supervisionados.

## SUMÁRIO

1. INTRODUÇÃO.....	4
2. DESENVOLVIMENTO.....	6
3. CONSIDERAÇÕES FINAIS.....	7
4. REFERÊNCIAS.....	8

## 1. INTRODUÇÃO

O processo ensino-aprendizagem formal no Brasil ocorre em instituições escolares de modo padronizado por todo o território nacional. Esse processo é essencial para o desenvolvimento pessoal e social dos estudantes, destacando-se o aprimoramento de características intrínsecas ao ser humano, como o raciocínio lógico, por exemplo.

Com a ascensão de novas tecnologias – como uma mudança paradigmática –, cada vez mais o método tradicional de ensino pode parecer antiquado ou desinteressante para aqueles que estão inseridos nessas novas áreas, dificultando a desenvoltura acadêmica dos mesmos. Destacam-se também contextos em que essas mudanças sequer atingem os estudantes, que muitas vezes são menos favorecidos economicamente, acentuando a necessidade de discutir e compreender como essas novas tecnologias são distribuídas, tal como a eficácia dos métodos e ferramentas de ensino disponíveis para os estudantes de nosso país. Fatores como região geográfica e renda per capita costumam realçar ainda mais esses problemas infraestruturais e o desafio da igualdade acerca das oportunidades.

De acordo com Andrade e Teixeira (2013):

O efeito-território é compreendido na literatura sociológica como os benefícios ou prejuízos socioeconômicos que acometem alguns grupos sociais em função da sua localização no espaço social das cidades. A hipótese sociológica a respeito do efeito-território não pressupõe uma ação determinista do espaço sobre as relações sociais, mas investiga as inter-relações entre as características dos espaços (tais como infraestrutura urbana, vizinhança, oferta de serviços) e as características dos grupos sociais (perfil do grupo e a natureza das suas interações internas e externas).

Esse conceito pode corroborar para a compreensão de algumas disparidades educacionais ao relacionarmos a infraestrutura urbana com a infraestrutura institucional das escolas numa determinada região, e posteriormente comparando com outros contextos.

Essas várias nuances acerca das ferramentas de ensino disponíveis – tal como o método utilizado pelo Estado para adotá-las – evidenciam fragilidades e desigualdades que sempre caracterizaram o meio de ensino de nosso país (SOUZA & GUIMARÃES, 2020), comprometendo, dessa forma, o rendimento acadêmico de alunos de diferentes regiões do Brasil.

A pesquisa e compreensão dessas raízes – que podem, ou não, justificar o desempenho acadêmico – é um processo demorado e complexo, visto que é

necessário compreender exceções, contextos diferentes, incentivos estatais, políticas de educação, dentre muitos outros fatores que influenciam nesse meio. Nesse contexto, destaca-se o Exame Nacional do Ensino Médio (ENEM), visto que ele é o principal avaliador de conhecimentos gerais e, consequentemente, desempenho escolar.

O ENEM – utilizado desde 2009 como mecanismo de acesso ao ensino superior – avalia o desempenho escolar dos estudantes ao término da educação básica (INEP). As 180 questões objetivas de quatro áreas do conhecimento distintas evidenciam afinidades dos vestibulandos que desejam ingressar em alguma instituição de ensino superior. Anualmente, o INEP disponibiliza os microdados do ENEM realizado no ano anterior, onde estão contidas todas as informações referentes ao questionário socioeconômico, e também as notas atingidas em cada uma das quatro áreas.

Nesse sentido, os registros obtidos através desses microdados auxiliam na identificação de indicadores de qualidade de ensino e outros problemas infraestruturais da educação brasileira. Esses indicadores e problemas se relacionam com o desempenho desses estudantes, visto que, conforme o desempenho no ENEM aumenta ou diminui, os fatores mais importantes são modificados (FEIJÓ & FRANÇA, 2021 *apud* MORAES & PERES, 2022).

Esses fatores realçam as necessidades estatais de implementar políticas educacionais que atendam as necessidades dos estudantes, de modo a incentivar e influenciar positivamente o rendimento acadêmico, visto que as políticas atuais não se demonstram muito efetivas. É importante compreender que “simplesmente acrescentar recursos financeiros na educação não afeta positivamente o desempenho dos alunos” (ARAUJO *et al.*, 2020).

Ademais, Moraes e Peres (2024) afirmam que:

identificar a diferença de perfil entre alunos de alto e baixo desempenho pode auxiliar na tomada de decisão para elaborar políticas públicas a fim de suprir lacunas, especialmente relacionadas a fatores socioeconômicos, para alunos de baixo desempenho.

De acordo com Freire (1967), a educação deve ser vista como um meio de transformação dos contextos sociais, permitindo uma análise crítica sobre as contradições dessa estrutura. Contudo, ao observarmos a educação no Brasil de uma perspectiva histórica e racial, nos deparamos com desigualdades no acesso à educação de qualidade (LIBÂNEO, 2012 *apud* MELO *et al.*, 2021).

Diante das diversas realidades educacionais no Brasil, torna-se fundamental adotar práticas avaliativas capazes de diagnosticar o desempenho escolar dos estudantes (MENEZES & PAZELLO, 2005 *apud* MELO *et al.*, 2021). No campo das políticas educacionais, tornou-se necessário definir critérios bem fundamentados para avaliar a qualidade do ensino. Para isso, foram selecionadas variáveis objetivas que serviram de base para criar indicadores capazes de medir o desempenho dos estudantes em exames aplicados em larga escala (HARTMAN, 1999 *apud* MELO *et al.*, 2021).

O ENEM representa uma das principais ferramentas de avaliação educacional no Brasil, abrangendo milhões de estudantes anualmente e refletindo, por meio de seus resultados, as diversas realidades educacionais do país. Por conta de sua amplitude, o ENEM se torna uma importante fonte de dados para a análise do processo de aprendizagem em múltiplos contextos socioeconômicos, demográficos e institucionais (INEP, 2009).

No questionário socioeconômico – que é respondido na data em que a inscrição para a mesma é realizada – podem ser encontrados motivos que influenciam o desempenho dos vestibulandos nas provas. Nesse sentido, torna-se importante a mineração desses dados para a identificação de indicadores de desempenho acadêmico, assim como para estabelecer relações entre diferentes tipos de dados. Por meio de trabalhos realizados neste escopo, foram analisadas informações importantes, como as de que a renda familiar baixa, a escolaridade dos pais de nível primário e a quantidade alta de pessoas que moram com os estudantes são atributos que diminuem o desempenho do aluno (SILVA, MORINO & SATO, 2014 *apud* LIMA *et al.*, 2019). Ainda nesse contexto, foi observado que as condições de infraestrutura e a complexidade das instituições usualmente impactam positivamente no desempenho dos vestibulandos, apresentando melhores resultados em testes padronizados (ALVES & SOARES, 2013 *apud* MEDEIROS & NETO, 2024).

O cruzamento dessas variáveis pode revelar relações significativas entre aspectos socioeconômicos, demográficos, institucionais e o desempenho dos estudantes. Ao identificar esses padrões, por meio da análise de dados utilizando técnicas de *Machine Learning*, torna-se necessário refletir de forma crítica sobre os diversos fatores que interferem no desempenho dos estudantes, como também sobre as condições de acesso à educação no país. Essa necessidade se justifica

pelas desigualdades “entre os estudantes, entre as escolas, entre classes de uma dada escola e entre as regiões em que se localizam as escolas” (TORRES, BICHIR, GOMES & CARPIM, 2006 *apud* MELO *et al.*, 2021). Com base nesses padrões e reflexões, é possível compreender as desigualdades sociais existentes no país para que se amplie o conhecimento sobre como diferentes condições de estudo podem se refletir no desempenho acadêmico.

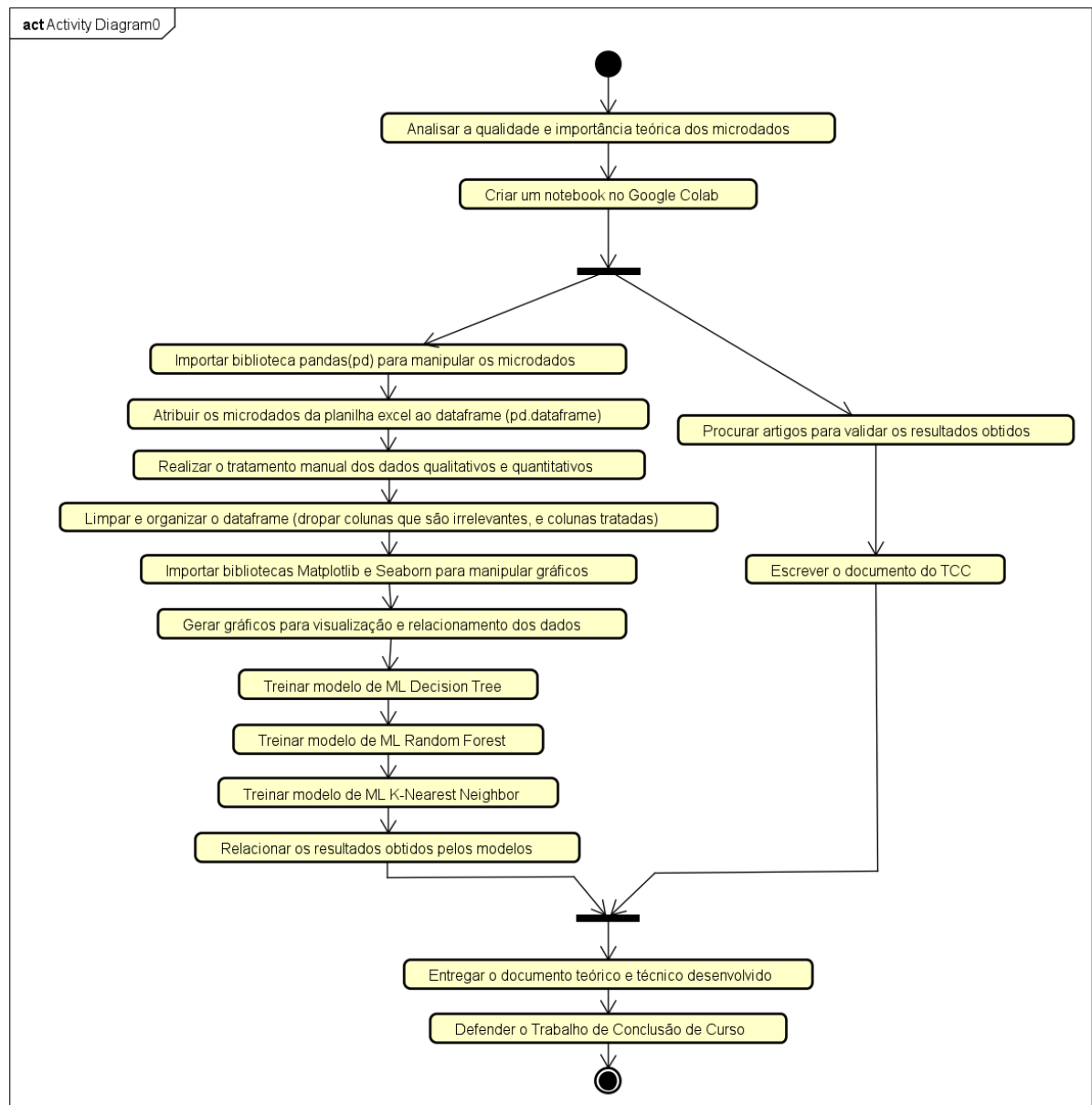
Nesse sentido, esse trabalho tem como objetivo realizar uma análise dos microdados do ENEM para, dessa maneira, evidenciar desigualdades educacionais e compreender a eficácia das políticas de educação existentes atualmente no Brasil.



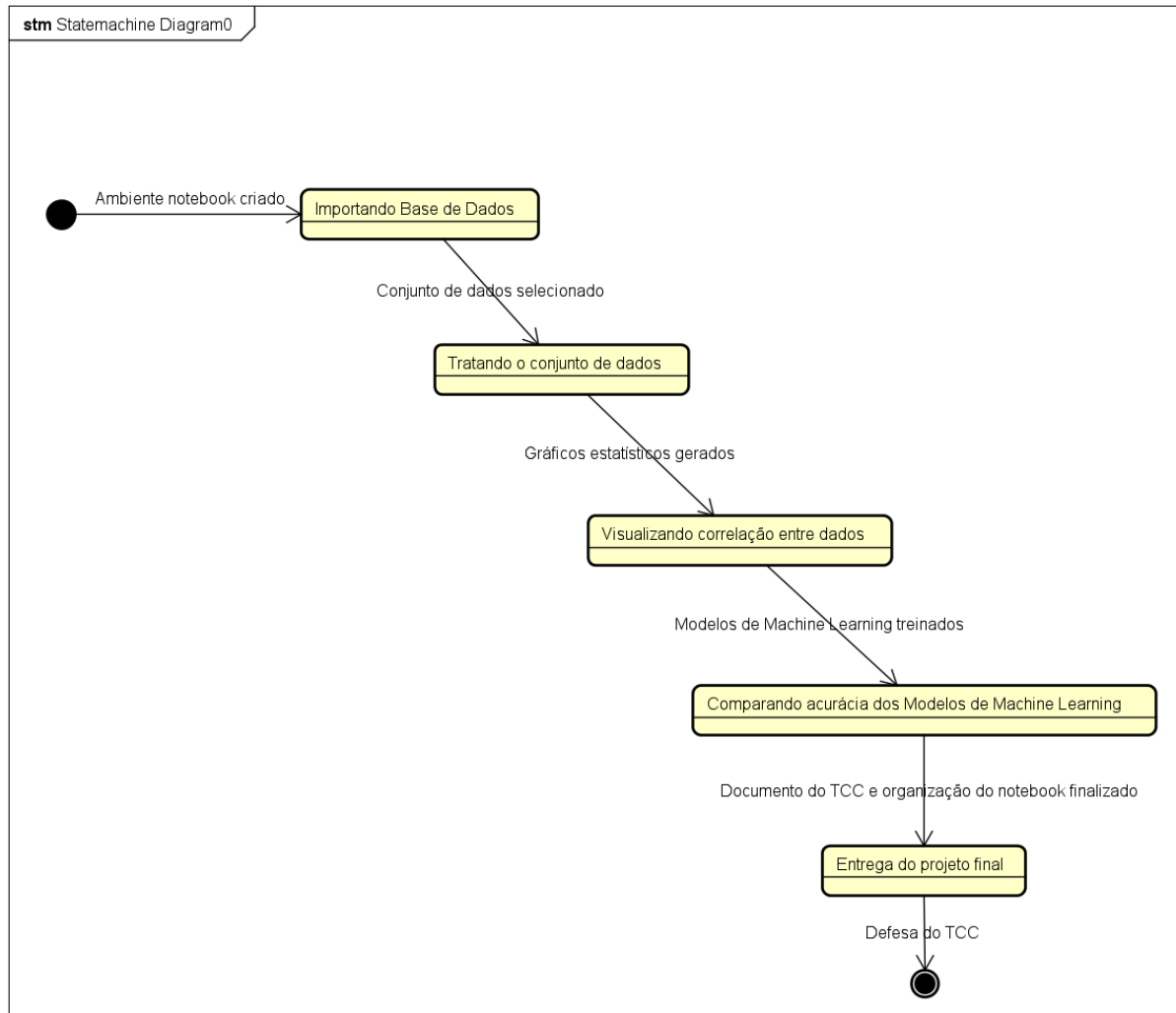
## 2. DESENVOLVIMENTO

Nessa seção encontram-se os diagramas da UML desenvolvidos no decorrer do projeto.

**Figura 1 - Diagrama de Atividades do ENELYTICS.**

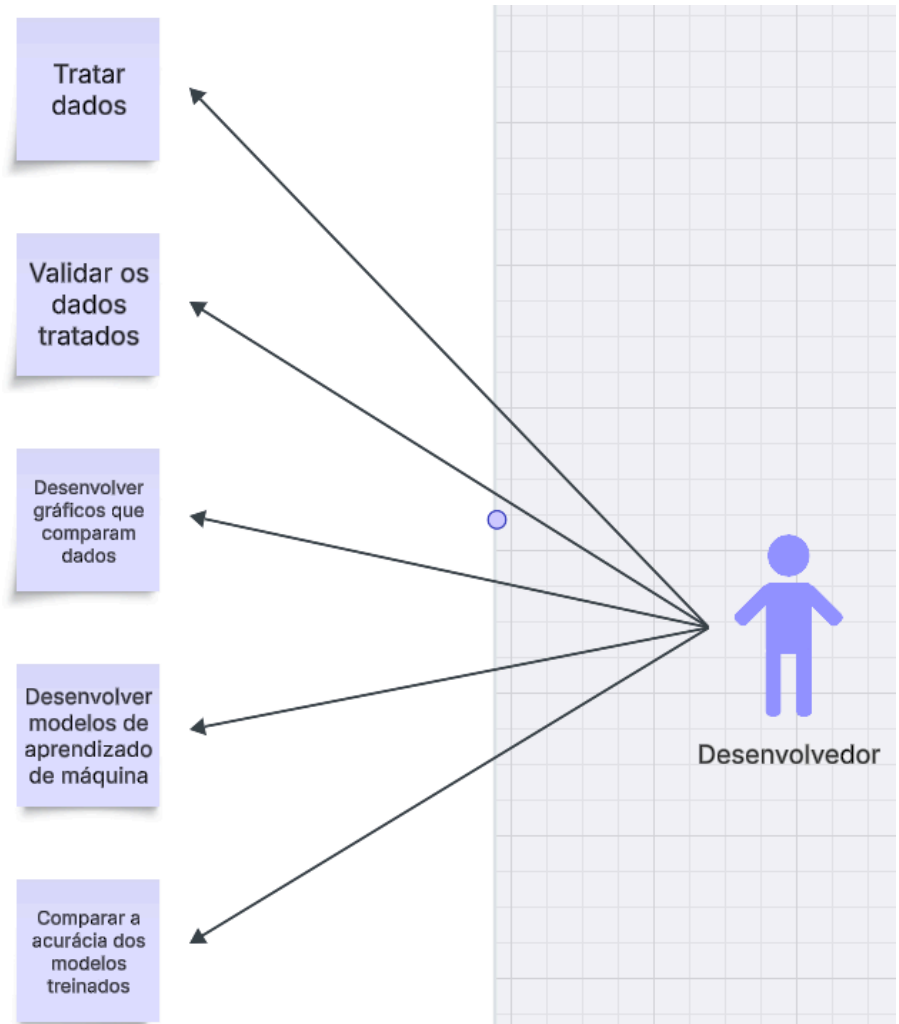


**Fonte: Os Autores (2025).**

**Figura 2 - Diagrama de Estado do ENELYTICS.**

**Fonte: Os Autores (2025).**

**Figura 3: Diagrama de Casos de Uso do ENELYTICS.**



**Fonte: Os Autores (2025).**

### 3. CONSIDERAÇÕES FINAIS

Este estudo analisou, por meio do desenvolvimento de algoritmos de aprendizado de máquina, os microdados do ENEM-2023, evidenciando as desigualdades educacionais presentes na realidade acadêmica brasileira.

Após o tratamento do conjunto de dados selecionado, foram implementados e treinados os seguintes modelos de *machine learning*: *Decision Tree*, *Random Forest* e *K-Nearest Neighbors*. Como resultado, foram obtidos em todos os modelos uma acurácia superior a 90%, superando o que era esperado durante o desenvolvimento do projeto. Dessa forma, tornou-se possível concluir que, de fato, o contexto no qual o vestibulando está inserido reflete diretamente em seu resultado naquele teste, corroborando com o propósito da pesquisa de evidenciar fragilidades educacionais.

Embora a taxa de acerto seja satisfatória, ao lidar com centenas de milhares de registros – referentes ao número de inscritos na prova – torna-se possível perceber que a proporção de erros é considerável. Isso, porém, não desestabiliza a tese defendida nesta pesquisa. Acredita-se que essa proporção de erros se dê pelo conjunto de dados selecionado para o treinamento dos modelos. Nesse sentido, recomenda-se que seja investido mais tempo para a pesquisa de artigos e estudos que validem outros indicadores de qualidade de ensino no contexto acadêmico, a fim de diminuir a quantidade de predições erráticas.

Ademais, nossos resultados explicitam precariedades infraestruturais da educação brasileira, tal como desigualdades no processo de aprendizagem, facilitando a identificação de setores que necessitam de investimentos para a transformação positiva na vida de milhões de estudantes, assim como o incentivo ao estudo.

#### 4. REFERÊNCIAS

ANDRADE, Luciana Teixeira, SILVEIRA, Leonardo Souza. **Efeito-território: Explorações em torno de um conceito sociológico**. Civitas: Revista De Ciências Sociais, 2013. Disponível em <<https://www.scielo.br/j/civitas/a/JvcZT5bSCHns49P4QDhnHxf/>>. Acesso em: 7 de maio de 2025.

ARAUJO, Juliana Maria de *et al.* **Fatores determinantes do desempenho educacional no Sudeste Brasileiro**. Gestão e Sociedade, 2020. Disponível em <<https://ges.face.ufmg.br/index.php/gestaoesociedade/article/view/2942>>. Acesso em: 6 de maio de 2025.

INEP. **Apresentação do Exame Nacional do Ensino Médio (ENEM)**. Disponível em <<https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enem>>. Acesso em: 5 de maio de 2025.

JALOTO, Alexandre, PRIMI, Ricardo. **Fatores socioeconômicos associados ao desempenho no Enem**. Revista de Administração Pública, 2021. Disponível em <<https://emaberto.inep.gov.br/ojs3/index.php/emaberto/article/view/5002>>. Acesso em: 5 de maio de 2025.

LIMA, Priscila da Silva Neves *et al.* **Análise de dados do Enade e Enem: uma revisão sistemática da literatura**. Avaliação Campinas: Revista Da Avaliação Da Educação Superior, 2018. Disponível em <<https://www.scielo.br/j/aval/a/L4J43gBxhXmjYhT5cX6BTM/>>. Acesso em: 30 de abril de 2025.

MEDEIROS, Hugo Augusto Vasconcelos, NETO, Ruy de Deus e Mello. **Schooling inequality on standardized test scores and extended journey: an analysis of Pernambuco (Brazil) high school public system**. Ensaio: Avaliação E Políticas Públicas Em Educação, 2024. Disponível em <<https://www.scielo.br/j/ensaio/a/mh5zXHmKGjCx9qK3bKR9DtQ/?lang=en>>. Acesso em: 30 de abril de 2025.

MELO, Rafael Oliveira *et al.* **Impacto das variáveis socioeconômicas no desempenho do Enem: uma análise espacial e sociológica**. Revista de Administração Pública, 2021. Disponível em <<https://www.scielo.br/j/rap/a/ZHJFnmsrdgGH8cj6xHHwbKg/>>. Acesso em: 5 de maio de 2025.

MORAES, Caroline Ponce de, PERES, Rodrigo Tosta. **REFLEXÕES SOBRE DIFERENÇAS DE DESEMPENHO NO ENEM: UMA ANÁLISE SOCIOECONÔMICA E ESCOLAR DO SUDESTE DO BRASIL**. Jornal de Políticas Educacionais, 2022. 3-4 p. Disponível em <<https://revistas.ufpr.br/jpe/article/view/85377>>. Acesso em: 6 de maio de 2025.

SOUZA, Marcelo Nogueira de, GUIMARÃES, Lislaine Mara da Silva.  
**VULNERABILIDADE SOCIAL E EXCLUSÃO DIGITAL EM TEMPOS DE PANDEMIA: UMA ANÁLISE DA DESIGUALDADE DE ACESSO À INTERNET NA PERIFERIA DE CURITIBA.** Revista Interinstitucional Artes de Educar, 2020.  
Disponível em <<https://www.e-publicacoes.uerj.br/riae/article/view/51097>>. Acesso em: 7 de maio de 2025.