

ANÁLISE DA RELAÇÃO ENTRE RENDA E DESEMPENHO NO ENEM: ESTUDO DE CASO BASEADO EM INTELIGÊNCIA ARTIFICIAL PARA DIFERENTES REGIÕES BRASILEIRAS

Tarcísio Alex Almeida de Paula

Aluno - Sistemas de Informação
tarcisioalexalmeida@gmail.com

Ana Clara Oliveira Fernandes

Aluno - Sistemas de Informação
clara.of93@gmail.com

Ana Laura Bastos

Aluno - Sistemas de Informação
anaurabastos890@gmail.com

Hiury Rodrigo da Silva Alves

Aluno - Análise e Desenvolvimento de Sistemas
hiuryalves.contato@gmail.com

Julianny Albuquerque Lima

Aluno - Análise e Desenvolvimento de Sistemas
albuquerque.julianny17@gmail.com

Kauã Silva do Nascimento

Aluno - Análise e Desenvolvimento de Sistemas
kaua.nascimento64@aluno.ce.gov.br

Kaio Gefferson de Almeida Mesquita

Orientador
kaio.mesquita@professor.unifametro.edu.br

Área Temática: Ciências Tecnológicas.

Área de Conhecimento: Inovação e inteligência artificial.

Modalidade: Iniciação Científica.

RESUMO

Introdução: O presente estudo integra um projeto de Iniciação Científica que busca compreender as desigualdades educacionais no Brasil, com foco na relação entre a renda familiar, a localidade e o desempenho dos estudantes no Exame Nacional do Ensino Médio (ENEM). Considerando a importância do ENEM como mecanismo de acesso ao ensino superior



e os contrastes socioeconômicos entre as regiões brasileiras, o trabalho propõe investigar como esses fatores influenciam o rendimento estudantil. Para isso, será realizada uma análise de dados com apoio de modelos de inteligência artificial, visando identificar pontos críticos de influência no desempenho e contribuir com evidências para a formulação de políticas públicas educacionais. A base de dados utilizada para esta análise é proveniente dos microdados do ENEM, disponibilizados pelo INEP por meio do portal de dados abertos: <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/enem>. Como etapa preliminar da investigação, foi elaborada uma matriz de correlação entre variáveis socioeconômicas e as notas dos participantes, com o intuito de explorar possíveis relações estatísticas relevantes. **Objetivo:** Analisar a correlação entre a renda per capita familiar e o desempenho dos estudantes no ENEM 2023, observando as variações dessa relação entre as diferentes regiões do Brasil por intermédio de modelos de inteligência artificial. **Metodologia:** Tipo de estudo: Estudo descritivo, com abordagem quantitativa. Período: Março a maio de 2025. Etapas: (i) Coleta dos dados (Obtenção da base de dados do ENEM 2023 disponibilizada pelo INEP. (ii) Limpeza e tratamento dos dados (Remoção de registros com valores nulos ou inconsistentes; Exclusão de colunas irrelevantes para o objetivo do estudo; Padronização de formatos. (iii) Análise exploratória (Elaboração de uma matriz e gráfico de correlação entre variáveis socioeconômicas e notas dos participantes; A partir dessa análise, observou-se que as variáveis **Q006** - renda familiar e **Q024** - posse de acesso à internet em casa apresentaram as correlações mais significativas com o desempenho dos estudantes, indicando sua relevância no modelo preditivo). (iv) Complementação da base de dados (Inserção do Índice de Desenvolvimento Humano (IDH) correspondente ao estado de cada participante; Identificação se o participante reside em zona rural ou urbana, com base em informações de município e localidade). (v) Equalização das bases (Balanceamento das variáveis para evitar viés no modelo; Normalização ou padronização dos dados numéricos (como renda e notas). (vi) Treinamento do modelo de IA (Escolha e aplicação de algoritmos de regressão para prever as notas dos estudantes com base nos dados tratados; Divisão entre conjunto de treino e teste para avaliação de desempenho do modelo). (vii) Análise dos resultados (Verificação dos fatores com maior influência nas previsões; Interpretação dos padrões encontrados a partir das variáveis socioeconômicas e regionais). Sujeitos envolvidos: Estudantes inscritos no ENEM 2023 com dados válidos sobre rendimento e informações socioeconômicas. Materiais: Utilizou-se a base de microdados do ENEM 2023 disponibilizada no site do GOV.BR. As análises foram



realizadas com o uso da linguagem Python, com apoio das bibliotecas Pandas, NumPy e Seaborn. **Resultados parciais e Discussão:** Até o momento, a equipe concentrou esforços nas etapas iniciais do projeto, realizando o levantamento e a organização da base de dados do ENEM 2023. Foi possível identificar a complexidade e o volume dos dados, o que exigiu um trabalho minucioso de filtragem e limpeza, removendo registros incompletos ou inconsistentes. Também foi iniciada a integração de informações externas, como o IDH dos estados e a classificação das localidades como zonas urbanas ou rurais. A partir dessa experiência, observamos a importância de uma preparação rigorosa dos dados para garantir a qualidade das análises e a eficácia do modelo de IA que será treinado posteriormente. A equipe destaca, ainda, o desenvolvimento de habilidades técnicas e analíticas como parte relevante do processo de pesquisa. **Considerações finais:** O projeto, ainda em fase de desenvolvimento, possibilitou a coleta e o tratamento dos dados do ENEM 2023, além da integração de informações socioeconômicas e regionais, como o IDH e a classificação urbana/rural. Também foi iniciada a análise desses dados, utilizando métodos como a matriz de correlação para verificar quais variáveis exercem maior influência sobre o desempenho dos candidatos. Como contribuição, destaca-se o desenvolvimento de uma base consolidada que permitirá análises mais precisas sobre desigualdades educacionais, bem como de um modelo capaz de prever a nota do candidato com base em dados socioeconômicos e de localidade. A experiência também proporcionou o aprimoramento de habilidades técnicas da equipe, como a análise minuciosa de dados e a construção de modelos preditivos. Como limitação, observa-se a complexidade e o tempo necessário para preparar e equilibrar os dados antes do treinamento do modelo de IA, o que pode impactar o cronograma previsto.

Palavras-chave: Inteligência Artificial. Análise socioeconômica. Análise regional. ENEM. Modelagem.

Referências:

AURELIANO, G. **Machine learning com Python**. São Paulo: Novatec Editora, 2021.

BREIMAN, L. **Random forests**. *Machine Learning*, v. 45, n. 1, p. 5–32, 2001.

CARVALHO, A. C. P. L. F.; FREITAS, A. A. **Inteligência artificial: fundamentos e aplicações**. Rio de Janeiro: LTC, 2014.



HAN, J.; PEI, J.; KAMBER, M. **Data mining: concepts and techniques**. 3. ed. Amsterdam: Elsevier, 2011.

HOSMER, D. W.; LEMESHOW, S.; STURDIVANT, R. X. **Applied logistic regression**. 3. ed. Hoboken: Wiley, 2013.

LIMA, P. D. S. N. *et al.* **Análise de dados do Enade e Enem: uma revisão sistemática da literatura**. Avaliação: Revista da Avaliação da Educação Superior (Campinas), v. 24, p. 89–107, 2019.

NASCIMENTO, M. M.; CAVALCANTI, C.; OSTERMANN, F. **Sucesso escolar em contextos populares: uma análise a partir do Enem**. Estudos em Avaliação Educacional, v. 31, n. 76, p. 134–163, 2020.

