

ANTEPROJETO DE TCC - BSI

	Nome	E-mail	Telefone
ORIENTADOR	Vanderson		
COORIENTADOR			
ALUNO	Lucas Garcia de Souza		997990256
ALUNO	Pedro Emanuel Facchera Pereira		33 98400-0026
FACILITADOR	Ricardo Tedesco	ricardotedescodasilva@gmail.com	

TEMA:

A influência da escolaridade e de fatores socioeconômicos na renda dos trabalhadores do Brasil

PROBLEMA REAL :

O ponto de partida deste trabalho é um desafio que afeta milhões de brasileiros: compreender em que medida o esforço em estudar e se qualificar realmente se traduz em melhores oportunidades no mercado de trabalho e em salários mais altos. Imagine, por exemplo, um pai de família que precisa sustentar a casa ou um jovem recém-formado que está prestes a ingressar na vida profissional. Ambos desejam saber se sua escolaridade e suas competências lhes garantirão um futuro melhor, com estabilidade e remuneração digna.

No Brasil, existe a percepção difundida de que quanto mais anos de estudo, maior o salário. Contudo, na prática, pessoas com o mesmo nível de escolaridade frequentemente apresentam rendas muito diferentes. Essa discrepância ocorre porque a empregabilidade e os rendimentos não dependem apenas da formação, mas também de fatores como região, setor econômico, idade, gênero e condições do mercado de trabalho.

Esse quadro gera incerteza para trabalhadores e estudantes, que muitas vezes não conseguem prever suas perspectivas salariais ou tomar decisões informadas sobre educação e carreira.

PROBLEMA CIENTÍFICO :

É possível estimar, a renda mensal esperada de um trabalhador no Brasil com base em seu nível de escolaridade e em características socioeconômicas (como idade, gênero, setor e região)?

HIPÓTESE :

H1: Se utilizarmos um modelo de inteligência artificial treinado com dados da PNAD Contínua que considerem escolaridade e características socioeconômicas, então será possível prever de a renda mensal esperada de indivíduos no Brasil.

VARIÁVEL INDEPENDENTE :

O uso de um modelo de inteligência artificial que, a partir da escolaridade e de características socioeconômicas.

VARIÁVEL DEPENDENTE :

O valor da renda mensal **esperada** de um indivíduo.

OBJETIVO GERAL :

Comparar as previsões do modelo com os dados reais da PNAD Contínua e estatísticas oficiais para verificar se é possível prever de forma confiável a renda média de um trabalhador.

OBJETIVOS ESPECÍFICOS :

- Revisar a literatura sobre a relação entre escolaridade, renda e fatores socioeconômicos no Brasil e em países em desenvolvimento.
- Selecionar e preparar variáveis relevantes da PNAD Contínua para modelagem preditiva de renda e empregabilidade.
- Implementar e comparar algoritmos de aprendizado de máquina (ex.: regressão, Random Forest, Gradient Boosting, XGBoost, redes neurais) para previsão da renda.
- Avaliar o desempenho dos modelos a partir de métricas de erro (MAE, RMSE) e de explicabilidade (SHAP values).
- Analisar como variáveis socioeconômicas e regionais modulam a relação entre escolaridade e renda no Brasil.
- Identificar limitações do modelo em diferentes faixas salariais e grupos populacionais.
- Discutir implicações acadêmicas e práticas, destacando a utilidade dos resultados para estudantes, trabalhadores e formuladores de políticas públicas.
- Desenvolver e avaliar um modelo de inteligência artificial capaz de prever a renda mensal esperada e a taxa de empregabilidade de trabalhadores no Brasil, a partir de seu nível de escolaridade e características socioeconômicas.

INTRODUÇÃO :

A relação entre escolaridade e renda tem sido amplamente documentada na literatura econômica e social, sendo reconhecida como um dos principais determinantes da empregabilidade e dos salários. Em linhas gerais, quanto maior o nível de instrução, maiores tendem a ser as oportunidades de inserção no mercado de trabalho e a remuneração média associada (Mullainathan & Spiess, 2017; Neri, 2008). Essa percepção consolidada, presente

tanto em estudos acadêmicos quanto no senso comum, tem orientado investimentos em educação, políticas públicas de qualificação profissional e decisões individuais de carreira.

Entretanto, a realidade brasileira mostra que essa relação, embora positiva, não é linear nem uniforme. Evidências empíricas recentes indicam que indivíduos com níveis semelhantes de escolaridade podem apresentar rendimentos bastante distintos em função de fatores como região geográfica, setor econômico, idade, gênero e mesmo a formalização do vínculo de trabalho. Herrera et al. (2023), utilizando a PNAD Contínua e técnicas de aprendizado de máquina interpretável (valores SHAP), demonstram que, embora a educação seja a variável mais relevante para explicar a renda, características como idade, gênero e região modulam fortemente esse resultado. De forma complementar, Santolin e Oliveira (2023) analisam os efeitos da pandemia e confirmam que fatores de vulnerabilidade — como menor escolaridade, gênero feminino, raça/cor e residência em regiões metropolitanas — aumentaram significativamente a probabilidade de queda de renda e de ingresso em situação de pobreza.

No campo da previsão salarial, Pedroso (2023), aplicando modelos de machine learning sobre dados administrativos do CAGED, mostrou que variáveis como ocupação, escolaridade, idade e localidade são determinantes da renda, mas que os modelos apresentam maior acurácia ao prever salários em faixas baixas e médias, revelando a complexidade da distribuição salarial no Brasil. Além disso, análises macroeconômicas, como as de Baltar et al. (2024), evidenciam que a estagnação econômica e o aumento da informalidade nos últimos anos intensificaram desigualdades estruturais, reforçando que a renda não é explicada apenas pela escolaridade individual, mas também por fatores institucionais e setoriais.

Outro aspecto relevante é a dimensão espacial. Estudos recentes têm demonstrado que a desigualdade socioeconômica se manifesta fortemente no território. Chi et al. (2022), em escala global, mostraram que variáveis de contexto espacial — captadas por imagens de satélite e conectividade digital — são capazes de explicar diferenças substanciais em indicadores de riqueza. No Brasil, Castro e Álvarez (2023) utilizaram imagens de satélite e técnicas de transfer learning para prever renda média municipal e observaram que os resultados variam significativamente entre estados, sugerindo que a desigualdade regional afeta diretamente a previsibilidade da renda. De forma ainda mais localizada, Machicao et al. (2022) confirmaram que aspectos visíveis na infraestrutura urbana, captados por imagens do Google Street View, correlacionam-se com diferenças de renda per capita em regiões pobres e semiurbanas.

Diante desse corpo de evidências, emerge a necessidade de aprofundar a investigação sobre até que ponto a escolaridade realmente se traduz em melhores salários no Brasil e como fatores socioeconômicos, institucionais e regionais modulam essa relação. Mais do que um exercício acadêmico, compreender esse problema é fundamental para oferecer subsídios a trabalhadores, estudantes e formuladores de políticas públicas que buscam decisões mais fundamentadas sobre educação, empregabilidade e perspectivas de renda.

JUSTIFICATIVA E RELEVÂNCIA :

A relação entre escolaridade e renda é amplamente reconhecida na literatura econômica, mas a realidade brasileira revela que indivíduos com níveis semelhantes de instrução frequentemente apresentam rendimentos bastante distintos. Evidências empíricas recentes reforçam que fatores como região geográfica, setor econômico, idade, gênero e formalização do vínculo de trabalho modulam significativamente os resultados salariais (Herrera et al., 2023; Santolin & Oliveira, 2023; Baltar et al., 2024). Ao mesmo tempo, estudos aplicados ao mercado de trabalho brasileiro (Pedroso, 2023) mostram que, mesmo com o uso de técnicas modernas de aprendizado de máquina, a previsão de salários é desafiadora, sobretudo em faixas mais altas de renda.

Nesse contexto, investigar em que medida a escolaridade se traduz em melhores salários, considerando simultaneamente características socioeconômicas e regionais, é relevante tanto do ponto de vista acadêmico quanto prático. Academicamente, o trabalho contribui para a literatura sobre retorno da educação e desigualdade no Brasil, aplicando metodologias robustas de machine learning a bases de microdados representativas, como a PNAD Contínua. Do ponto de vista prático, os resultados podem apoiar estudantes, trabalhadores e formuladores de políticas públicas ao oferecer uma visão mais clara e fundamentada sobre perspectivas de renda e empregabilidade.

REFERÊNCIAS :

Kumar, V. (2024). Income prediction using machine learning. *Soft Computing Fusion with Applications*, 1(3), 150–160. <https://doi.org/10.22105/scfa.v1i3.46>

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, 31(2), 87–106. <https://doi.org/10.1257/jep.31.2.87>

Pedroso, K. H. V. (2023). Previsão de renda no Brasil utilizando técnicas de aprendizado de máquina: Uma análise com dados do CAGED. [Trabalho acadêmico de conclusão]. (arquivo analisado previamente).

Santolin, R., & Oliveira, J. R. (2023). Determinantes da renda domiciliar no Brasil durante a pandemia: Uma análise a partir da PNAD Contínua. *Revista de Economia*, 44(2), 45–72. (arquivo: 4104-Texto do artigo-16130-17321-10-20231004.pdf).

Herrera, G. P., Constantino, M., Su, J.-J., & Naranpanawa, A. (2023). The use of ICTs and income distribution in Brazil: A machine learning explanation using SHAP values. *Telecommunications Policy*, 47(3), 102598. <https://doi.org/10.1016/j.telpol.2023.102598>