Analisando a Propagação de Viés de Gênero na Aprendizagem por Transferência Aplicada à Educação

Nathan C. Freitas e Gabriel Alves

Departamento de Estatística e Informática
Universidade Federal Rural de Pernambuco (UFRPE)
Recife – PE – Brasil

{nathan.freitas, gabriel.alves}@ufrpe.br

Abstract. *Transfer Learning (TL) is a powerful technique for developing predictive models in education, but it poses the risk of propagating algorithmic biases. This study investigates the phenomenon of gender bias propagation when applying TL to university dropout data. The methodology consisted of training a "teacher model"on a source domain with known gender disparity and transferring its knowledge to "student models"in distinct fields of knowledge. The results indicate that the bias was largely transferred and, in many cases, amplified, increasing the accuracy disparity between genders, although instances of mitigation were also observed. It is concluded that TL can act as a vector for systemic injustice, making fairness auditing an essential and prerequisite step in the selection of pre-trained models for educational applications.*

Resumo. *A Aprendizagem por Transferência (TL) é uma técnica poderosa para desenvolver modelos preditivos na educação, mas apresenta o risco de propagar vieses algorítmicos. Este trabalho investiga o fenômeno da propagação de viés de gênero ao aplicar a TL em dados de evasão universitária. A metodologia consistiu em treinar um modelo-professor em um domínio com notória disparidade de gênero e transferir seu conhecimento para modelos-aluno em distintas áreas do conhecimento. Os resultados indicam que o viés foi majoritariamente transferido e, em muitos casos, amplificado, aumentando a disparidade de acurácia entre os sexos, embora exceções de mitigação tenham sido observadas. Conclui-se que a TL pode atuar como um vetor de injustiça sistêmica, tornando a auditoria de equidade (fairness) uma etapa pré-requisito e essencial na seleção de modelos pré-treinados para aplicações educacionais..*

1. Introdução

O volume crescente e contínuo de dados digitais transformou-se no componente crucial que impulsiona o avanço exponencial e a ampla aplicação de modelos de inteligência artificial. No campo da educação, esses modelos têm sido cada vez mais empregados para tarefas críticas, como a previsão de desempenho acadêmico e a identificação precoce de estudantes em risco de evasão [Barbosa et al. 2024, Cavalcanti et al. 2024]. Analisar essas métricas interessa ao gestor por possibilitar intervenções pedagógicas personalizadas e gestão de recursos eficiente, promovendo o sucesso e permanência dos estudantes.

DOI: 10.5753/weriaedu.2025.16250

Contudo, a implementação de modelos preditivos em um domínio tão sensível quanto a educação acarreta uma responsabilidade ética fundamental: a garantia de justiça (*fairness*). Modelos de aprendizado de máquina são um reflexo dos dados que são treinados e, se estes dados contêm vieses sociais ou demográficos, a técnica pode não apenas perpetuar, mas também amplificar essas desigualdades [Yapo and Weiss 2018]. Uma decisão algorítmica injusta na educação pode levar a suporte desproporcional, ampliando barreiras e afetando análises dos perfis dos estudantes [Omughelli et al. 2024]. Assim, a construção de modelos deve priorizar não só a acurácia, mas principalmente a equidade.

Neste cenário, a Aprendizagem por Transferência (*Transfer Learning* - TL) emerge como uma abordagem poderosa e amplamente adotada. A técnica, extensivamente documentada em revisões da área [Weiss et al. 2016, Gou et al. 2021], permite que o conhecimento adquirido por um modelo em uma tarefa ou domínio de dados massivos seja reaproveitado para otimizar o aprendizado em um novo contexto, frequentemente com dados mais limitados. Para a educação, onde a coleta de dados anotados pode ser dispendiosa, a TL apresenta um potencial imenso para acelerar o desenvolvimento de modelos robustos. No entanto, essa eficiência introduz uma questão crítica: o modelo-aluno herda o viés do modelo-professor? Para responder a esta pergunta, este trabalho se propõe a investigar se e como essas falhas de equidade são transferidas e, a partir dessa análise, explorar quais intervenções podem ser realizadas para mitigar a propagação do viés antes que ele contamine novos domínios.

A possibilidade de propagação de viés via TL levanta uma preocupação significativa, pois um único modelo tendencioso tem o potencial de disseminar injustiça algorítmica em larga escala para diversos sistemas derivados. Diante desta lacuna, o presente trabalho tem como objetivo central investigar o fenômeno da propagação de viés de gênero via Aprendizagem por Transferência em dados educacionais, analisando como um modelo-professor, treinado em um domínio com notória disparidade, transfere essas características tendenciosas para modelos-aluno em distintas áreas do conhecimento.

2. Trabalhos Relacionados

A Aprendizagem por Transferência (TL) opera sobre o desafio fundamental do *covariate shift* — a divergência na distribuição de dados entre domínios de origem e destino, que pode levar a um desempenho subótimo se não for tratada. A técnica é amplamente utilizada para superar desafios de performance, como demonstrado por [Celiberto Jr et al. 2010], que aplicaram a TL para acelerar a convergência em Aprendizagem por Reforço. Contudo, o conhecimento transferido de grandes *datasets* de pré-treinamento, como o ImageNet, podem não serem neutros. Esses vieses, sejam naturais ou sintéticos, podem ser notavelmente persistentes, permanecendo no modelo mesmo após o ajuste fino sobre um conjunto de dados-alvo balanceado [Salman et al. 2022].

No âmbito educacional, a TL também tem sido aplicada com foco na otimização. [Tsiakmaki et al. 2020], por exemplo, obtiveram ganhos de desempenho ao transferir conhecimento entre modelos de redes neurais rasas para diferentes cursos universitários. Contudo, a aplicação de TL em contextos educacionais vai além da mera otimização de performance, adentrando o campo crítico da equidade algorítmica. Soluções de *fair transfer learning* já são exploradas em outras áreas, como em [Coston et al. 2019], que propõem estratégias para cenários onde atributos protegidos estão desalinhados entre

domínios, demonstrando ser possível mitigar vieses em setores como seguros e crédito com impacto mínimo na acurácia.

No centro desta discussão no campo da educação está o trabalho seminal de [Gardner et al. 2023], que investigou o impacto da TL na imparcialidade de modelos de predição de evasão com foco em subgrupos intersecionais. No trabalho foi possível notar uma neutralidade: a transferência direta não agravou nem mitigou sistematicamente as disparidades de desempenho em comparação com modelos treinados localmente. No entanto, o estudo expôs uma vulnerabilidade persistente em subgrupos específicos, independentemente da abordagem. Este cenário sugere que, embora a TL possa não ser a causa primária do viés, ela também não é, inerentemente, a solução, motivando a necessidade de investigações mais profundas sobre os mecanismos de propagação de vieses — foco central do presente trabalho.

3. Método

Este trabalho propõe um modelo para a predição da evasão em cursos de graduação da Universidade Federal Rural de Pernambuco. Para tanto, emprega-se uma abordagem de aprendizagem por transferência sobre um conjunto de dados providos pelo Observatório de dados da Gradução (ODG) vinculado a universidade. A tarefa consiste em uma classificação binária, onde, a partir de atributos discentes coletados em um período acadêmico, o modelo deve inferir se o estudante pertence à classe Evasão (positiva) ou Não Evasão (negativa).

A classe positiva foi definida pelos casos de abandono consolidado. A classe negativa, por outro lado, agrupa os status *FORMADO* e *CURSANDO* para representar o conjunto de discentes não evadidos. Embora o estado final dos alunos *CURSANDO* seja, por definição, incerto, sua inclusão é justificada pelo objetivo do modelo: não se trata de prever um desfecho de carreira, mas sim de classificar o risco com base nos padrões observados até um determinado ponto no tempo. Assim, o modelo aprende a diferenciar os perfis de quem já evadiu com os quem, até o momento, permanecem ou finalizaram o vínculo.

O conjunto de dados do estudo, detalhado na Tabela 1, compreende a variáveis heterogêneas que descrevem a trajetória acadêmica discente. Para a modelagem com redes neurais, foi implementada uma pipeline de pré-processamento que incluiu a padronização das variáveis numéricas (*Z-score*)[Morettin and Bussab 2017] e a binarização de atributos categóricos (*one-hot encoding*) [Tan et al. 2018]. Uma característica marcante do *dataset* é seu severo desbalanceamento de classes. A classe minoritária, correspondente à evasão, possui 50.181 instâncias, em forte contraste com as 613.574 instâncias da classe majoritária (não evasão). A fim de mitigar o potencial viés de aprendizado do modelo em favor da classe majoritária, foi empregada a técnica de subamostragem aleatória (*random undersampling*).

A implementação da pipeline de modelagem foi realizada em linguagem Python. A biblioteca *Scikit-learn* [Pedregosa et al. 2011] foi empregada para as tarefas de pré-processamento e para a estruturação da validação cruzada. Para a construção dos modelos de redes neurais, optou-se pela biblioteca *Keras* [Chollet et al. 2015], dada a sua flexibilidade na manipulação e congelamento de camadas, que foi a técnica utilizada para a aplicação de aprendizagem por transferência [Iman et al. 2023]. A otimização da arqui-

Tabela 1. Atributos utilizado da base de dados de alunos.

Atributo	Tipo de dado	Descrição
É funcionário	Booleano	Se o discente se trata de um funcionário da instituição
Semestre de Ingresso	Inteiro	Em qual semestre o discente ingressou na instituição
Duração do vinculo	Inteiro	Respectivo período do discente
Duração do curso	Inteiro	Duração ideal de conclusão do curso
Ingresso Regular	Booleano	Se o discente ingressou no curso através de forma regular
É assistido	Booleano	Discente precisa de alguma assistência nas aulas
Total de aprovações acumulado	Inteiro	Quantidade de aprovações acumulada que o discente possui no respectivo período
Total de reprovações acumulado	Inteiro	Quantidade de reprovações acumulada que o discente possui no respectivo período
Total de cancelamento acumulado	Inteiro	Quantidade de cancelamentos acumulado que o discente possui no respectivo período
Total de outros acumulado	Inteiro	Quantidade de demais estados acumulado que o discente possui no respectivo período
Gênero	Masculino e Feminino	Gênero do discente
Raça	Branco, Preto, Pardo, Indígena, Amarelo, Remanescente Quilombo, Outros e Não informado	Raça do discente
Estado Civil	Solteiro, Casado, Separado, Outros e Não informado	Estado civil do discente
Forma de Ingresso	Vestibular, SISU, Processo de seleção, Diplomado e Outros	Forma de ingresso do discente
Turno	Matutino, Vespertino, Noturno, Integral, Especial e Indefinido	Turno do curso
Tipo de curso	Regular, Especial e EAD	Modalidade do curso

tetura e dos parâmetros de treinamento foi conduzida por meio de uma busca aleatória (*Randomized Search*) [Bergstra and Bengio 2012] de 500 candidados, cujos espaço de busca e melhores parâmetros encontrados são apresentados na Tabela 2. A robustez dos resultados foi assegurada pela utilização de validação cruzada com 5 partições (*5-fold cross-validation*).

O procedimento para a seleção do modelo-professor e a definição dos domínios-alvo foi estruturado em etapas sequenciais. Inicialmente, foi treinado e otimizado um modelo preditivo para cada uma das 11 áreas do conhecimento do CNPq, resultando em 11 modelos-candidatos distintos. Cada um desses modelos foi então avaliado segundo os critérios de seleção predefinidos. O modelo que melhor satisfez os critérios foi designado como o modelo-professor, enquanto as 10 áreas restantes foram definidas como os domínios-alvo para a transferência de conhecimento.

A arquitetura do modelo-aluno foi construída a partir da estrutura do modelo-professor, com sua camada de classificação final sendo descartada. Todas as camadas de extração de características restantes tiveram seus pesos congelados, de modo a pre-

Tabela 2. Configuração de hiperparâmetros da rede neural.

Parâmetro	**Valores**
Otimizador	adam, rmsprop, adagrad e sgd
Neurônios	16, 32 e 64
Taxa de Dropout	0.1, 0.15 e 0.2
Números de camadas	2, 3, 4 e 5
Taxa de aprendizado	0.001, 0.0005 e 0.0001
Épocas	5, 10, 15 e 20

servar integralmente o conhecimento aprendido no domínio-fonte. Novas camadas de classificação, com pesos inicializados aleatoriamente, foi adicionada após a última camada do modelo-professor. O treinamento subsequente, realizado exclusivamente com os dados de outra área de conhecimento, ajustou unicamente os pesos destas novas camadas classificadora [Gou et al. 2021]. Essa estrutura é idealmente representada pela Figura 1.

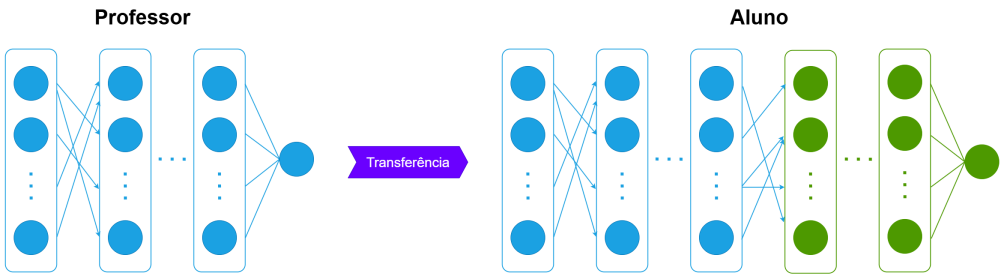

Figura 1. Esquema da arquitetura de rede professor-aluno onde as camadas do professor estão congeladas.

A avaliação do impacto desta transferência foi focada em dois eixos: o desempenho preditivo por meio da acurácia e a equidade do modelo quantificada por meio da disparidade de acurácia, calculada como a diferença absoluta entre a acurácia obtida no subgrupo masculino e no feminino. A comparação dessas métricas entre o modelo-aluno e um modelo de base (treinado do zero no domínio-alvo) permite mensurar o efeito da transferência tanto na performance quanto na propagação do viés.

A fim de garantir que qualquer ganho de performance fosse estritamente atribuível ao conhecimento transferido e não a um aumento na complexidade computacional, a busca de hiperparâmetros do modelo-aluno foi metodologicamente controlada. O processo de busca de hiperparâmetros foi configurado para que o número máximo de camadas e de épocas exploradas não excedesse os respectivos valores do modelo-professor selecionado. Esta abordagem estabelece uma paridade de complexidade máxima, isolando o efeito da transferência de conhecimento como a principal variável sob análise na comparação com o modelo de base.

4. Resultados e Discussão

A primeira etapa dos resultados consistiu na seleção do domínio-fonte (modelo-professor), conforme os critérios de alta acurácia e acentuada disparidade de desempenho entre sexos. Uma análise preliminar da distribuição por área CNPq (Tabela 3) indicou que

as áreas de **Ciências Exatas e da Terra** e **Engenharias** eram as candidatas mais promissoras devido à sua notória assimetria na representação de gênero. O fator decisivo para a seleção, contudo, foi a magnitude do viés de performance. A análise comparativa (Tabela 4) revelou que o modelo treinado com dados de **Engenharias** apresentou uma das maiores discrepância, com a acurácia para um sexo superando o outro em aproximadamente 10 pontos percentuais. Este resultado valida a área de **Engenharias** como um candidato ideal e robusto para investigar a propagação de viés via aprendizagem por transferência.

Tabela 3. Informações de cada amostra por área CNPq.

Área CNPQ	Quantidade Total	Quantidade Feminino	Quantidade Masculino
Ciência Agrárias	134151	52385 (39.0%)	81766 (61.0%)
Ciência Biológicas	70832	46012 (65.0%)	24820 (35.0%)
Ciência da Saúde	62535	33692 (53.9%)	28843 (46.1%)
Ciência Exatas e da Terra	**165454**	**47506 (28.7%)**	**117948 (71.3%)**
Ciência Humanas	94753	48469 (51.2%)	46284 (48.8%)
Ciência Sociais Aplicadas	23339	12088 (51.8%)	11251 (48.2%)
EAD	53706	27105 (50.%)	26601 (49.5%)
Engenharias	**17214**	**5590 (32.5%)**	**11624 (67.5%)**
Linguística, Letras e Artes	19173	12813 (66.8%)	6360 (33.2%)
Matemática	17829	8928 (50.1%)	8901 (49.9%)
Outras	4769	2163 (45.4%)	2606 (54.6%)

Tabela 4. Disparidade absoluta de Acurácia por Sexo por área CNPq

Área CNPQ	Diferença
Ciência Agrárias	0.75%
Ciência Biológicas	4.05%
Ciência da Saúde	4.19%
Ciência Exatas e da Terra	2.85%
Ciência Humanas	2.56%
Ciência Sociais Aplicadas	7.09%
EAD	3.67%
Engenharias	**10.32%**
Linguística, Letras e Artes	**10.37%**
Matemática	9.61%
Outras	1.52%

A análise de desempenho dos modelos-aluno após a transferência de conhecimento revela uma tendência geral de melhoria e convergência na maioria dos domínios-alvo, conforme ilustrado na Figura 2. Contudo, duas áreas emergem como exceções notáveis a essa tendência: **Ciências Sociais Aplicadas** e **Matemática**. Nesses dois domínios específicos, a aplicação da transferência de conhecimento não resultou em um ganho de performance, indicando que o aprendizado transferido não foi eficaz nestes contextos.

A performance insatisfatória observada nos modelos de **Ciências Sociais Aplicadas** e **Matemática** duas hipóteses. A primeira, e mais provável, reside no conceito de transferência negativa. É plausível que a distribuição de atributos e o perfil dos discentes de Engenharia (domínio-fonte) sejam substancialmente distintos dos perfis encontrados nesses dois domínios-alvo. Tal divergência pode tornar o conhecimento pré-treinado inadequado ou até mesmo prejudicial para o aprendizado. Uma segunda hipótese, de natureza metodológica, é que o regime de treinamento de 5 épocas, embora suficiente para as demais áreas, pode ter sido uma limitação para esses casos específicos. É possível que estes modelos requeressem um treinamento mais extenso para superar platôs de aprendizado e convergir para uma solução mais ótima.

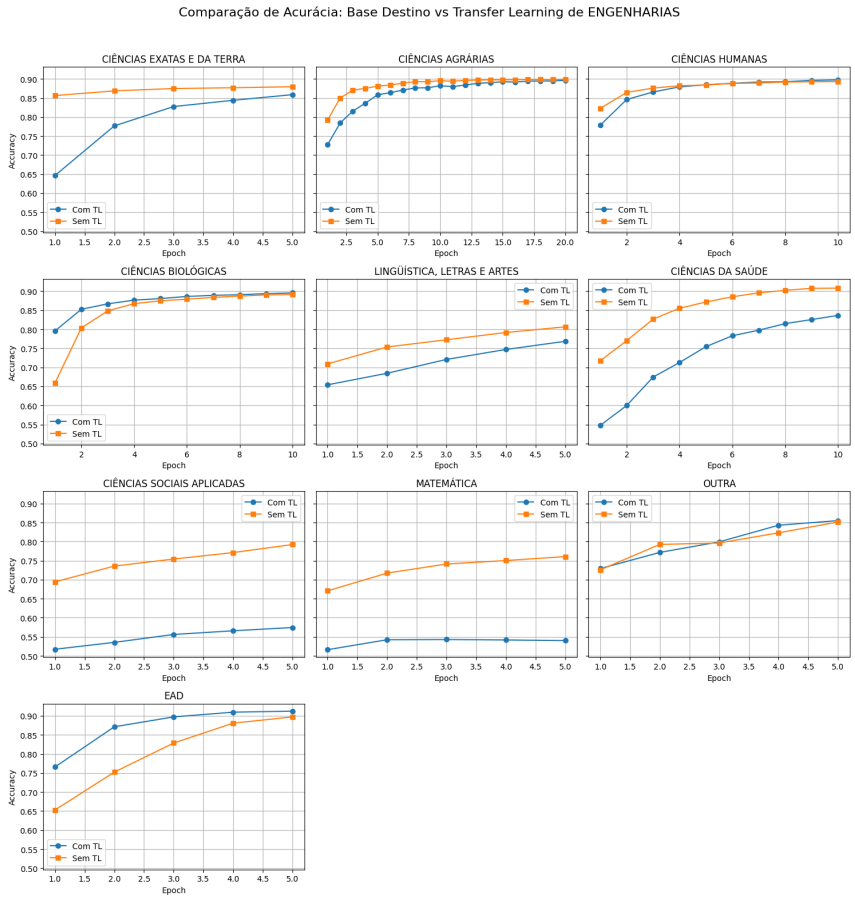

Figura 2. Comparação do desempenho por épocas entre modelos com e sem *transfer learning*

A análise de equidade, avaliada através da disparidade de acurácia entre os sexos, revelou que a transferência de conhecimento teve um impacto direto, e predominantemente negativo, no viés dos modelos-aluno. Conforme demonstrado na Tabela 5, a maioria dos domínios-alvo exibiu uma amplificação do viés após o procedimento, com um aumento na diferença de acurácia entre os grupos masculino e feminino. As únicas exceções expressivas a esta tendência foram as áreas de **Ciências Humanas** e **Matemática**, nas quais foi observada uma mitigação do viés, com a redução da disparidade de performance entre os sexos.

O resultado geral corrobora a hipótese central deste estudo: o viés de

Tabela 5. Disparidade absoluta de acurácia por sexo nos modelos-aluno com transferência de conhecimento a partir de Engenharias.

Área CNPQ	Diferença
Ciência Agrárias	0.68%
Ciência Biológicas	4.43%
Ciência da Saúde	9.51%
Ciência Exatas e da Terra	4.03%
Ciência Humanas	1.67%
Ciência Sociais Aplicadas	21.93%
EAD	3.89%
Linguística, Letras e Artes	11.68%
Matemática	3.27%
Outras	2.02%

representação presente no domínio-fonte (**Engenharias**) foi efetivamente transferido e, na maioria dos casos, amplificado nos domínios-alvo. A aparente mitigação do viés em **Matemática**, contudo, deve ser interpretada com cautela. Como o modelo para esta área demonstrou um baixo desempenho geral, a redução da disparidade pode ser uma consequência de um aprendizado ineficaz, e não uma melhora genuína na equidade. Em contrapartida, o caso de **Ciências Humanas** representa um desfecho positivo e significativo. A hipótese para este sucesso é que a robustez do sua amostra — tanto em volume quanto em equivalência na quantidade de instância por gênero — permitiu que o modelo-aluno aprendesse novos padrões a partir dos dados locais, que foram capazes de sobrepor e mitigar o viés herdado.

Os padrões de propagação de viés observados neste estudo representam uma implicação crítica para a aplicação responsável da aprendizagem por transferência em domínios sensíveis como a educação. Se um modelo preditivo padrão já incorre no risco de perpetuar desigualdades, a transferência de conhecimento pode atuar como um potente mecanismo de amplificação, disseminando o viés de um único modelo-fonte para múltiplos domínios-alvo. Diante disso, a seleção de um modelo-professor não pode ser guiada unicamente por métricas de performance. Torna-se necessária que a auditoria de equidade seja uma etapa fundamental e pré-requisito no processo, garantindo que o modelo base realize classificações equitativas entre os diferentes grupos de amostras antes que seu conhecimento seja utilizado em larga escala.

Os achados deste estudo possuem implicações diretas para a formulação de políticas públicas e diretrizes institucionais sobre o uso de Inteligência Artificial na educação brasileira. A mitigação dos riscos de propagação de viés, aqui demonstrada, demanda uma colaboração sinérgica, como por exemplo o Ministério da Educação (MEC), instituições de ensino superior e desenvolvedores de tecnologia. Um modelo de governança eficaz poderia atribuir a um órgão central, como o MEC, a responsabilidade de auditar e certificar modelos-professores de base que possam ser utilizados nacionalmente. Em paralelo, cada universidade que desenvolva ou adapte seus próprios modelos deveria ser incentivada a seguir tais protocolos ou a estabelecer comitês internos de ética em IA, garantindo a supervisão local.

Conforme investigado no presente estudo, esta análise deve ir além do gênero, abrangendo também atributos protegidos como raça, estado civil e turno do curso. Este método diagnóstico fornece um panorama concreto para que instituições, gestores e pesquisadores possam iniciar a implementação de práticas de IA mais responsáveis, promovendo uma cultura de equidade no desenvolvimento tecnológico educacional.

5. Considerações Finais

Apesar de suas vantagens de desempenho, a transferência de conhecimento pode servir como um meio para a transmissão e ampliação de vieses algorítmicos no contexto educacional. Ao transferir conhecimento a partir de um domínio-fonte com reconhecida disparidade de gênero, verificou-se que a maior parte dos modelos-alvo não apenas herdou, mas intensificou o viés preditivo. O estudo também indicou, entretanto, que tal resultado não é determinístico, como demonstrado por **Ciências Humanas**, onde características específicas do domínio-alvo possibilitaram a mitigação do viés transmitido. A principal consequência deste estudo é um alerta: a seleção de modelos-professores baseada exclusivamente em métricas de acurácia constitui uma prática de alto risco, com potencial para promover injustiça em escala sistêmica.

A análise concentrou-se no viés de gênero e utilizou dados de uma única instituição, o que requer cautela na generalização dos resultados. Desta forma, torna-se evidente a necessidade de replicação do experimento em contextos multi-institucionais e a inclusão de outros atributos protegidos e suas interseccionalidades. Ademais, a comparação entre diferentes abordagens de Transferência de Aprendizagem, como o ajuste fino (*fine-tuning*), poderia desvendar dinâmicas distintas de propagação de viés. Finalmente, este trabalho propicia o desenvolvimento de métodos que atentem à manutenção da equidade, sendo capazes não apenas de evitar a propagação, mas também de ativamente mitigar vieses durante o processo de aprendizado.

6. Agradecimentos

Os autores agradecem ao Observatório de Dados da Graduação (ODG) pelo fornecimento dos dados e à Gestão Institucional da UFRPE pelo apoio fundamental à condução deste estudo. Este trabalho foi desenvolvido no âmbito das iniciativas do ODG, que visam fomentar a gestão universitária baseada em evidências.

Referências

Barbosa, T., Freitas, N., Cavalcanti, L., da Conceicao Batista, M., Gouveia, R., and Alves, G. (2024). Construção dinâmica de modelos de learning analytics explicáveis e justos aplicados ao acompanhamento de estudantes de graduação. In *Anais do XXXV Simpósio Brasileiro de Informática na Educação*, pages 1918–1930, Porto Alegre, RS, Brasil. SBC.

Bergstra, J. and Bengio, Y. (2012). Random search for hyper-parameter optimization. *The journal of machine learning research*, 13(1):281–305.

Cavalcanti, L., Barros, A., Falcão, T., da Conceicao Batista, M., Cristino, C., and Alves, G. (2024). Avaliando o impacto da mudança do projeto pedagógico de cursos sobre a evasão através da análise de sobrevivência. In *Anais do XXXV Simpósio Brasileiro de Informática na Educação*, pages 2617–2626, Porto Alegre, RS, Brasil. SBC.

Celiberto Jr, L. A., Matsuura, J. P., De Màntaras, R. L., and Bianchi, R. A. (2010). Using transfer learning to speed-up reinforcement learning: a cased-based approach. In *2010 latin american robotics symposium and intelligent robotics meeting*, pages 55–60. IEEE.

Chollet, F. et al. (2015). Keras.

Coston, A., Ramamurthy, K. N., Wei, D., Varshney, K. R., Speakman, S., Mustahsan, Z., and Chakraborty, S. (2019). Fair transfer learning with missing protected attributes. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pages 91–98.

Gardner, J., Yu, R., Nguyen, Q., Brooks, C., and Kizilcec, R. (2023). Cross-institutional transfer learning for educational models: Implications for model performance, fairness, and equity. In *Proceedings of the 2023 acm conference on fairness, accountability, and transparency*, pages 1664–1684.

Gou, J., Yu, B., Maybank, S. J., and Tao, D. (2021). Knowledge distillation: A survey. *International journal of computer vision*, 129(6):1789–1819.

Iman, M., Arabnia, H. R., and Rasheed, K. (2023). A review of deep transfer learning and recent advancements. *Technologies*, 11(2):40.

Morettin, P. A. and Bussab, W. O. (2017). *Estatística básica*. Saraiva Educação SA.

Omughelli, D., Gordon, N., and Al Jaber, T. (2024). Fairness, bias, and ethics in ai: Exploring the factors affecting student performance. *Journal of Intelligent Communication*, 3(2):100–110.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Salman, H., Jain, S., Ilyas, A., Engstrom, L., Wong, E., and Madry, A. (2022). When does bias transfer in transfer learning? *arXiv preprint arXiv:2207.02842*.

Tan, P.-N., Steinbach, M., Karpatne, A., and Kumar, V. (2018). *Introduction to Data Mining (2nd Edition)*. Pearson, 2nd edition.

Tsiakmaki, M., Kostopoulos, G., Kotsiantis, S., and Ragos, O. (2020). Transfer learning from deep neural networks for predicting student performance. *Applied Sciences*, 10(6):2145.

Weiss, K., Khoshgoftaar, T. M., and Wang, D. (2016). A survey of transfer learning. *Journal of Big data*, 3(1):9.

Yapo, A. and Weiss, J. (2018). Ethical implications of bias in machine learning.