

Ciência de Dados para Equidade de Gênero em STEM: Fatores Associados à Continuidade de Meninas na Tecnologia

Raimunda Nonata Costa Lima¹, Fernanda Ferreira do Nascimento²,
Livia Maria Araújo de Lima¹, Erica Lima Silva¹,
José Maria Monteiro², Rogério Lopes Vieira Cesar¹

¹ Instituto Federal do Ceará – Campus Iguatu – IFCE
Iguatu – CE – Brasil

² Universidade Federal do Ceará – UFC
Fortaleza – CE – Brasil

limanonata66@gmail.com, nascimentofernandaf@outlook.com

Abstract. *Women remain underrepresented in technology, and this gap begins before higher education. This paper presents a data science approach to characterize factors associated with girls' intention to continue in technology-related educational paths. The study emerges from an initiative led by undergraduate women and researchers in Computing who aim to understand the experiences of girls in technical high school and use evidence to guide future interventions. We analyze survey data from female students enrolled in integrated technical high school programs. The pipeline includes data cleaning, feature engineering, statistical association tests with effect size, multiple-testing correction, interpretable modeling, and thematic analysis of open-ended answers. After removing duplicated responses, the dataset contains 155 valid answers. Results indicate that declared interest in technology, satisfaction with the chosen course, family encouragement, and extracurricular technology activities are associated with the intention to continue in the field.*

Resumo. *A sub-representação feminina em tecnologia é um desafio social que se inicia antes do ensino superior. Este artigo apresenta uma abordagem de ciência de dados para caracterizar fatores associados à intenção de continuidade de meninas em percursos formativos ligados à tecnologia. O estudo nasce de uma iniciativa conduzida por estudantes e pesquisadoras da Computação que buscam compreender experiências de meninas no ensino médio técnico e utilizar evidências para orientar intervenções futuras. Foram analisados dados de questionário aplicado a estudantes do gênero feminino em cursos técnicos integrados ao ensino médio. O pipeline inclui limpeza, engenharia de atributos, testes estatísticos com tamanho de efeito, correção para múltiplos testes, modelagem interpretável e análise temática de respostas abertas. Após remoção de duplicatas, a base contém 155 respostas válidas. Os resultados indicam que interesse por tecnologia, satisfação com o curso, incentivo familiar e participação em atividades extracurriculares estão associados à intenção de continuidade.*

1. Introdução

A baixa participação feminina em *Science, Technology, Engineering and Mathematics* (STEM; em português, Ciência, Tecnologia, Engenharia e Matemática) permanece como um problema social e educacional. Ao longo deste artigo, adota-se a sigla STEM para se referir a esse conjunto de áreas. Na computação, desigualdades de gênero se manifestam em estereótipos, ausência de modelos de referência, baixa percepção de pertencimento e menor continuidade em trajetórias técnicas e profissionais [Wang and Degol 2017; Santos and Marczak 2023]. Esse processo não começa apenas no ensino superior: durante o ensino médio técnico, estudantes já vivenciam experiências que podem fortalecer ou reduzir sua intenção de seguir na área, como incentivo familiar, apoio docente, experiências práticas e percepção de barreiras de gênero [Ribeiro and Maciel 2020a,b; Silva et al. 2023].

Este trabalho nasce no contexto de uma iniciativa formada por estudantes e pesquisadoras do ensino superior em Computação interessadas em compreender os fatores que afetam a aproximação, a permanência e a continuidade de meninas na tecnologia ainda durante o ensino médio. A motivação central é produzir evidências que possam orientar ações práticas de intervenção, acolhimento e incentivo à participação feminina em trajetórias formativas em STEM.

A questão investigada é: *quais fatores educacionais, familiares, institucionais e de percepção de gênero estão associados à intenção de continuidade de meninas em cursos técnicos de tecnologia?* Para respondê-la, o estudo organiza respostas de questionário em dimensões analíticas, constrói atributos derivados, aplica testes estatísticos com tamanho de efeito, explora modelos interpretáveis e analisa respostas abertas. As contribuições são: (i) um pipeline reprodutível para análise de dados educacionais sobre meninas em tecnologia; (ii) a caracterização de fatores associados à continuidade; e (iii) a discussão dos achados como subsídio para intervenções futuras, como oficinas, mentorias, rodas de conversa, visitas técnicas e ações de representatividade feminina.

De forma complementar, o estudo busca aproximar a produção de evidências quantitativas de uma leitura situada do contexto educacional. Assim, os resultados não são apresentados apenas como medidas estatísticas isoladas, mas como pistas para compreender quais condições podem fortalecer a permanência das estudantes, quais dimensões podem produzir afastamento e que tipos de ação podem ser priorizados por iniciativas de extensão e formação. Essa orientação é relevante porque programas de incentivo à participação feminina em tecnologia frequentemente precisam decidir onde concentrar esforços: divulgação de carreiras, apoio pedagógico, fortalecimento de redes de mentoria, aproximação com famílias ou criação de experiências práticas. A Tabela 1 resume as perguntas analíticas que orientam o estudo.

O restante do artigo está organizado da seguinte forma. A Seção 2 apresenta trabalhos relacionados. A Seção 3 descreve dados e metodologia. A Seção 4 apresenta resultados. A Seção 5 discute implicações e cuidados metodológicos. A Seção 6 conclui o trabalho.

2. Trabalhos Relacionados

Estudos sobre participação feminina em computação apontam que escolha, permanência e continuidade são influenciadas por fatores individuais, familiares, institucionais e soci-

Tabela 1. Perguntas analíticas e estratégias adotadas

Questão	Foco	Estratégia de análise
QP1	Identificar fatores associados à intenção de continuidade em tecnologia.	Testes de associação, V de Cramer, razão de chances e correção para múltiplos testes.
QP2	Verificar se modelos simples reforçam os padrões observados nos testes estatísticos.	Regressão logística, árvore de decisão rasa e floresta aleatória rasa com validação cruzada.
QP3	Compreender demandas expressas pelas estudantes para melhorar o ensino de tecnologia.	Análise de conteúdo das respostas abertas, com agrupamento temático e interpretação qualitativa.

oculturais. Ribeiro and Maciel [2020a,b] investigaram estudantes do ensino médio integrado em Informática e destacaram influência familiar, experiências prévias com tecnologia, interesse pela área, mercado de trabalho e ações de incentivo à equidade. Silva et al. [2023] observaram que muitas meninas possuem conhecimento limitado sobre profissões em tecnologia, especialmente sobre possibilidades além da programação.

Uma referência central é o mapeamento de Santos and Marczak [2023], que analisou fatores de atração, evasão e permanência de mulheres na computação. Entre os fatores de permanência, destacam-se redes de apoio, representatividade, pertencimento, incentivo de terceiros e experiências positivas na formação. Esses elementos orientam a interpretação dos dados deste estudo, sobretudo na relação entre continuidade, apoio familiar, atividades práticas e modelos femininos de referência.

Outros estudos reforçam que a sub-representação feminina não pode ser explicada apenas por escolhas individuais. Leão et al. [2021] mostram a importância de análises regionais e educacionais para compreender trajetórias em STEM, enquanto Menezes et al. [2024] evidenciam que experiências de mulheres na computação são atravessadas por múltiplas dimensões sociais. Esses trabalhos indicam que intervenções efetivas precisam considerar pertencimento, acessibilidade, reconhecimento e condições institucionais, e não apenas a oferta pontual de conteúdos técnicos.

Na literatura internacional, Wang and Degol [2017] também sintetizam que diferenças de gênero em STEM resultam de uma combinação de fatores motivacionais, escolares, familiares e socioculturais. Essa leitura amplia o enquadramento do problema para além do desempenho individual das estudantes e reforça a necessidade de investigar contextos de apoio, pertencimento e expectativa de futuro.

Este artigo também dialoga com trabalhos que aplicam ciência de dados a problemas sociais. Silva and Zárate [2025] e Alves et al. [2025] utilizam bases públicas de saúde, pré-processamento, seleção de atributos e modelos interpretáveis para caracterizar populações em contextos de saúde pública e saúde mental. Silva et al. [2024] analisam representação de gênero nas indústrias do livro e da música a partir de dados. Em comum, esses trabalhos mostram que métodos de ciência de dados podem produzir diagnósticos interpretáveis sobre problemas sociais. Diferentemente deles, este estudo aplica essa perspectiva ao contexto educacional de meninas no ensino médio técnico em tecnologia.

3. Materiais e Métodos

O estudo possui caráter exploratório-descritivo e abordagem mista, combinando análise quantitativa de respostas estruturadas e análise qualitativa de respostas abertas [Gil 2008; Creswell 2010; Bardin 2011]. A coleta foi realizada por questionário online aplicado a estudantes do gênero feminino matriculadas em cursos técnicos integrados ao ensino médio no eixo tecnológico, incluindo Informática, Desenvolvimento de Sistemas e Redes de Computadores. A amostra foi não probabilística, por acessibilidade, alcançada por contatos institucionais e divulgação em diferentes escolas públicas estaduais e federais. As participantes eram provenientes de escolas estaduais e de Instituto Federal localizados no interior do Ceará, em instituições que ofertam cursos técnicos integrados ao ensino médio, em tempo integral, na área de tecnologia. Essa contextualização é relevante porque situa o estudo em um cenário específico de formação técnica pública, sem identificar individualmente escolas ou participantes. Por esse motivo, o artigo não compara instituições individualmente nem estima efeitos regionais ou institucionais de forma controlada. Como a coleta envolveu instituições distintas e não havia uma lista nominal consolidada de todas as alunas elegíveis, não foi estimada taxa de resposta populacional; os resultados são tratados como evidências exploratórias para apoiar diagnóstico e planejamento de ações educacionais.

Todas as participantes foram informadas sobre os objetivos do estudo e concordaram com o Termo de Consentimento Livre e Esclarecido. A análise foi conduzida sobre base anonimizada, sem divulgação de nomes, e-mails ou dados individualmente identificáveis, em conformidade com princípios éticos da Resolução CNS nº 466/12 [Conselho Nacional de Saúde 2012].

3.1. Instrumento, atributos e pipeline analítico

O questionário foi organizado em blocos temáticos para explicitar as classes de questões exploradas e melhorar a reprodutibilidade (Tabela 2). As respostas de múltipla escolha foram transformadas em variáveis binárias. Por exemplo, a pergunta sobre áreas de interesse originou atributos como *interesse_tecnologia*, *interesse_engenharia*, *interesse_ciencias*, *interesse_matematica* e *interesse_artes*; a pergunta sobre atividades extracurriculares originou atributos como *atividade_programacao*, *atividade_informatica*, *atividade_design*, *atividade_games*, *atividade_circuitos* e *alguma_atividade*. A variável-alvo foi a resposta à pergunta: “Você pretende seguir na área e fazer uma graduação em Tecnologia?”.

O pipeline incluiu: (i) limpeza, anonimização, padronização de colunas e remoção de respostas duplicadas; (ii) caracterização da amostra; (iii) engenharia de atributos; (iv) testes de associação, usando qui-quadrado ou Fisher, V de Cramer como tamanho de efeito, *odds ratio* para variáveis binárias e correção de Benjamini-Hochberg; (v) modelagem exploratória com regressão logística balanceada, árvore de decisão rasa e floresta aleatória rasa; e (vi) análise temática das respostas abertas. A base bruta continha 158 respostas. Após a identificação de três duplicatas por e-mail, a base analítica passou a conter 155 respostas válidas. As variáveis de múltipla escolha foram decompostas em atributos binários, permitindo comparar a intenção de continuidade entre estudantes com e sem cada fator observado. A ANOVA não foi adotada como análise principal porque a variável de interesse é categórica, tornando mais adequados os testes para tabelas de contingência.

Tabela 2. Blocos temáticos do questionário e exemplos de atributos analisados

Bloco	Questões e atributos observados/derivados
Perfil	Ano escolar, curso, escola e concordância com o TCLE.
Interesses formativos	Interesse por tecnologia, engenharia, ciências, matemática e artes.
Contexto familiar	Atuação familiar em tecnologia, discussão familiar sobre STEM e incentivo para seguir carreira em tecnologia.
Ambiente escolar	Qualidade percebida do ensino, incentivo docente, grupos de meninas e acesso a recursos/laboratórios.
Experiências práticas	Informática, programação, design digital, games, eletrônica/circuitos, feiras, competições e indicador de alguma atividade extracurricular.
Representatividade	Presença de figuras femininas inspiradoras na mídia, escola ou comunidade.
Identificação com o curso	Escolha correta, arrependimento, falta de identificação, expectativa frustrada ou escolha por ausência de opção.
Barreiras e apoio	Percepção de prejuízo por ser mulher e dificuldade de pedir ajuda a professores ou colegas.
Desfecho e res-posta aberta	Intenção de continuidade na tecnologia e sugestões para melhorar o ensino de tecnologia.

As análises foram conduzidas em Python/Jupyter, com *pandas*, *numpy*, *scipy*, *scikit-learn*, *matplotlib* e *wordcloud*. Para favorecer a reprodutibilidade, o conjunto de dados anonimizado, o dicionário de variáveis e os notebooks utilizados nas análises estão disponíveis publicamente no repositório do projeto: <https://github.com/FernandaNascimento26/Equidade-STEM-DS4SG>. Antes da disponibilização, os dados foram revisados para remoção de informações potencialmente identificáveis, incluindo instituições específicas e respostas abertas sensíveis.

A etapa de preparação dos dados foi documentada para manter rastreabilidade entre as perguntas originais e os atributos analisados. Esse cuidado é particularmente importante em questionários educacionais, nos quais uma mesma resposta pode indicar dimensões diferentes do fenômeno. Por exemplo, a participação em feiras e competições pode ser interpretada tanto como experiência extracurricular quanto como contato com redes de apoio e oportunidades institucionais. Por isso, as variáveis derivadas foram utilizadas de forma parcimoniosa, preservando a interpretação substantiva de cada atributo.

4. Resultados

4.1. Caracterização da amostra e fatores associados

A base final contém 155 respostas válidas após remoção de duplicatas. A maior parte das participantes estava no 2º ano do ensino médio (69 respostas), seguida pelo 1º ano (63) e pelo 3º ano (23). Em relação aos cursos, 122 estudantes estavam em curso integrado de Informática, 31 em Redes e duas respostas em outras denominações. Quanto à intenção de continuidade, 50 participantes afirmaram que pretendem seguir na área e cursar graduação em tecnologia, enquanto 105 responderam negativamente.

A Figura 1 sintetiza a proporção de estudantes que declararam intenção de continuidade segundo fatores selecionados. As maiores diferenças aparecem para interesse declarado por tecnologia, satisfação com o curso, incentivo familiar e participação em atividades extracurriculares.

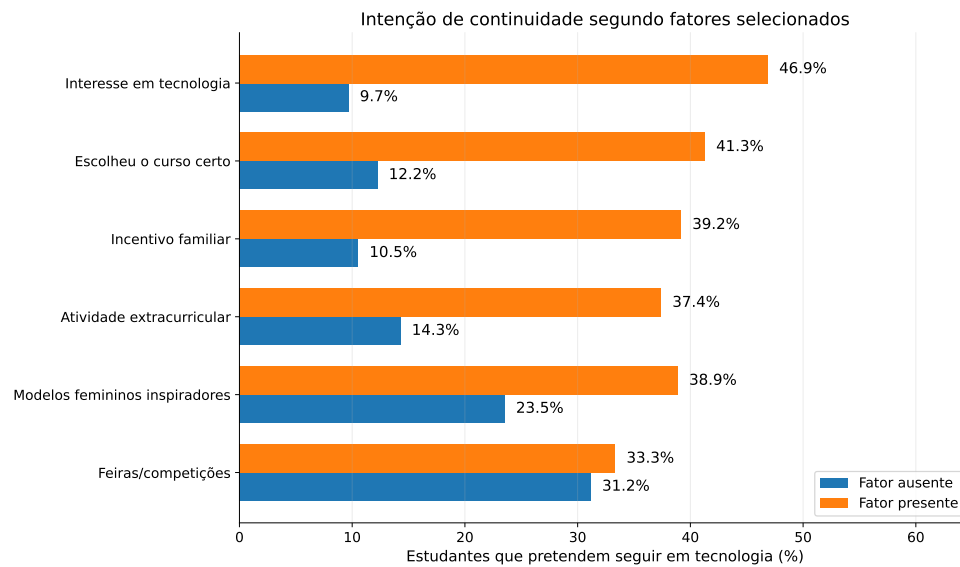


Figura 1. Proporção de estudantes que declararam intenção de continuidade na tecnologia segundo fatores selecionados.

A Tabela 3 apresenta os principais fatores associados à intenção de continuidade. O interesse por tecnologia apresentou a associação mais forte: entre estudantes que declararam esse interesse, 46,8% pretendem seguir na área, contra 9,8% entre as demais (OR=8,07). A satisfação com a escolha do curso também foi relevante: 41,1% das estudantes que afirmaram ter escolhido o curso certo pretendem continuar, contra 12,5% entre as que indicaram arrependimento, falta de identificação, expectativa frustrada ou escolha por ausência de opção. Incentivo familiar e participação em atividades extracurriculares também mantiveram associação após correção para múltiplos testes.

Tabela 3. Fatores associados à intenção de continuidade em tecnologia

Fator	Com	Sem	<i>q</i>	<i>V</i>	OR
Interesse por tecnologia	46,8%	9,8%	< 0,001	0,386	8,07
Escolheu o curso certo	41,1%	12,5%	0,004	0,283	4,89
Incentivo familiar	39,0%	10,8%	0,009	0,257	5,27
Atividade extracurricular	37,5%	14,3%	0,039	0,208	3,60
Figuras femininas inspiradoras	38,6%	23,9%	0,172	0,156	2,01
Atividade em informática	37,3%	22,6%	0,185	0,148	2,03

Nota: “Com” e “Sem” indicam o percentual de continuidade em cada grupo. *q* é o valor-p ajustado por Benjamini-Hochberg; *V* é o *V* de Cramer; OR é a razão de chances.

Esses resultados reforçam que a continuidade não depende apenas de motivação individual, mas de uma combinação de identificação com a área, apoio familiar e vivências práticas. O resultado sobre incentivo familiar converge com estudos que apontam redes próximas como fator de atração e permanência na computação [Ribeiro and Maciel 2020a; Santos and Marczak 2023]. A associação com atividades extracurriculares sugere que oficinas, feiras, clubes e projetos podem aproximar as estudantes de possibilidades concretas de formação e carreira.

4.2. Barreiras de gênero, modelagem e respostas abertas

Embora a percepção de prejuízo por ser mulher não tenha apresentado associação direta significativa com a intenção de continuidade, ela se mostrou associada à dificuldade de

pedir ajuda. Entre as estudantes que relataram ter se sentido prejudicadas por serem mulheres, 38 de 43 também relataram dificuldade para pedir ajuda a professores ou colegas ($p < 0,001$; $V = 0,309$). Também houve associação entre sentir-se prejudicada por ser mulher e perceber menor incentivo docente às meninas em tecnologia ($p = 0,0215$; $V = 0,185$). Esses achados sugerem que barreiras de gênero podem aparecer como redução de segurança, acolhimento e busca por suporte no cotidiano do curso.

A modelagem teve caráter exploratório e interpretável. A Tabela 4 mostra os resultados médios em validação cruzada estratificada com cinco partições. Foram utilizados modelos de baixa complexidade: regressão logística com pesos balanceados, árvore de decisão com profundidade limitada e floresta aleatória rasa. A avaliação considerou acurácia, acurácia balanceada, F1 e ROC-AUC, de modo a observar o desempenho geral e o comportamento em uma base com classes desbalanceadas. Os modelos reforçam os testes de associação: regressão logística e árvore rasa atribuíram maior relevância a interesse por tecnologia, incentivo familiar, satisfação com o curso e experiências extracurriculares. Assim, a modelagem é usada como evidência complementar, não como ferramenta pronta de classificação.

Tabela 4. Desempenho exploratório dos modelos interpretáveis

Modelo	Acurácia	Acur. bal.	F1	ROC-AUC
Regressão logística	0,703	0,723	0,630	0,754
Árvore de decisão rasa	0,742	0,741	0,649	0,743
Floresta aleatória rasa	0,755	0,746	0,656	0,778

A leitura dos modelos foi feita de modo complementar aos testes estatísticos. A árvore de decisão rasa, por exemplo, privilegia divisões baseadas em interesse por tecnologia, incentivo familiar e atividade extracurricular, produzindo regras simples e interpretáveis. A regressão logística, por sua vez, permite observar o sentido das associações sem transformar o estudo em uma proposta de predição individual. Essa escolha metodológica é importante no contexto educacional: o objetivo não é rotular estudantes, mas identificar fatores que podem orientar ações coletivas de apoio, escuta e fortalecimento de trajetórias.

As respostas abertas foram lidas integralmente e agrupadas em categorias temáticas por análise de conteúdo [Bardin 2011]. As categorias foram construídas de forma indutiva a partir da recorrência de temas nas respostas, como infraestrutura, práticas pedagógicas, apoio docente, representatividade e orientação de carreira. Uma mesma resposta pôde contribuir para mais de uma categoria quando mencionava múltiplas demandas. A Tabela 5 mostra que as demandas mais frequentes envolveram infraestrutura, didática e apoio docente, representatividade, aulas práticas e conteúdos técnicos. Esses achados complementam os resultados quantitativos ao mostrar que permanência e continuidade dependem de condições materiais, experiências concretas de aprendizagem e pertencimento simbólico.

5. Discussão

Os resultados apontam que a intenção de continuidade em tecnologia é multifatorial. O interesse por tecnologia aparece como fator central, mas não deve ser interpretado como característica puramente individual: interesses são construídos em experiências famili-

Tabela 5. Categorias temáticas nas respostas abertas

Categoria	n	Exemplos de demandas
Infraestrutura e recursos	26	Laboratórios, computadores, equipamentos, kits de robótica e componentes eletrônicos.
Didática e apoio docente	24	Melhor explicação, apoio dos professores e maior abertura para pedir ajuda.
Representatividade e incentivo feminino	20	Professoras, mulheres de referência, debates sobre mulheres na tecnologia e acolhimento.
Aulas práticas e projetos	17	Aulas práticas, oficinas, projetos e atividades em laboratório.
Programação e conteúdo técnico	17	Mais programação, aprofundamento técnico, softwares e desenvolvimento de sistemas.
Outras demandas	15	Organização, comunicação, carga de atividades e melhorias diversas.
Orientação de carreira e vivências	6	Palestras, visitas técnicas, contato com profissionais e informações sobre mercado.

ares, escolares e institucionais [Wang and Degol 2017; Santos and Marczak 2023]. A satisfação com o curso e o incentivo familiar indicam a importância de orientação vocacional, acolhimento nos primeiros anos e aproximação das famílias com possibilidades de formação e carreira.

A participação em atividades extracurriculares reforça a relevância de oficinas, clubes, feiras, projetos e grupos de apoio. Esses espaços podem funcionar como ambientes de experimentação e pertencimento, especialmente quando incluem práticas colaborativas e modelos femininos. As barreiras de gênero apareceram de forma relevante na relação com dificuldade de pedir ajuda e percepção de menor incentivo docente, sugerindo que parte do problema está em dimensões menos visíveis da experiência escolar, como segurança psicológica, acolhimento e confiança para expor dúvidas.

Como implicação prática, os achados podem orientar ações conduzidas por meninas e mulheres do ensino superior junto a estudantes do ensino médio: oficinas de programação e robótica, rodas de conversa com mulheres da área, visitas técnicas, aproximação com ambientes universitários, orientação vocacional, ações com famílias e sensibilização docente. Assim, os dados não são tratados apenas como descrição do problema, mas como insumos para desenho de intervenções educacionais baseadas em evidências. A Tabela 6 sintetiza possíveis desdobramentos práticos.

Antes de executar intervenções, os achados também sugerem a necessidade de diálogo com as próprias estudantes e com as equipes pedagógicas. Ações baseadas em dados não devem substituir a escuta qualitativa, mas apoiar decisões mais informadas. Desse modo, uma escola com baixa participação em atividades extracurriculares poderia priorizar oficinas e projetos; outra, com relatos de dificuldade para pedir ajuda, poderia investir em mentoria, acolhimento e formação docente. Essa leitura contextual evita que os resultados sejam tratados como receita única e reforça seu papel como instrumento de planejamento.

Quanto ao escopo, os resultados não pretendem produzir estimativas populacionais generalizáveis, pois a amostra foi obtida por acessibilidade e a coleta ocorreu em instituições distintas, sem controle do universo total de alunas convidadas. Os dados são autorrelatados e os modelos foram treinados com amostra reduzida; portanto, devem

Tabela 6. Relação entre achados e possíveis intervenções educacionais

Achado	Interpretação	Possível ação
Interesse por tecnologia associado à continuidade	A aproximação com a área precisa ocorrer antes da decisão pelo ensino superior.	Oficinas introdutórias, desafios práticos, clubes de programação e atividades de experimentação.
Incentivo familiar relevante	A família pode validar ou desencorajar escolhas formativas.	Rodas de conversa com responsáveis e materiais simples sobre carreiras em tecnologia.
Satisfação com o curso associada à continuidade	A experiência no curso influencia pertencimento e expectativa de futuro.	Acolhimento nos primeiros anos, orientação sobre trilhas profissionais e acompanhamento pedagógico.
Atividades extracurriculares associadas à continuidade	Vivências práticas ampliam repertório e confiança.	Feiras, visitas técnicas, projetos de curta duração e participação em eventos.
Dificuldade de pedir ajuda relacionada a barreiras de gênero	Barreiras podem aparecer como insegurança e menor busca por suporte.	Mentorias, grupos de apoio, sensibilização docente e espaços seguros para dúvidas.

ser lidos como apoio à interpretação, e não como classificação definitiva. Essa natureza exploratória é coerente com o objetivo diagnóstico do estudo. Também não foram realizadas análises de equidade algorítmica entre subgrupos, pois a base disponível não continha volume suficiente nem variáveis contextuais adequadas para estimar métricas como paridade demográfica ou igualdade de oportunidade de forma responsável. Essa escolha evita interpretações instáveis em uma amostra pequena, mas aponta uma direção relevante para coletas futuras.

6. Conclusão

Este artigo apresentou uma abordagem de ciência de dados para caracterizar fatores associados à intenção de continuidade de meninas em cursos técnicos de tecnologia. A partir de 155 respostas válidas, foram aplicadas etapas de limpeza, engenharia de atributos, análise estatística, modelagem interpretável e análise temática de respostas abertas.

Os achados indicam que interesse por tecnologia, satisfação com o curso, incentivo familiar e participação em atividades extracurriculares estão associados à intenção de seguir na área. Além disso, a percepção de prejuízo por ser mulher mostrou associação com a dificuldade de pedir ajuda, indicando que barreiras de gênero podem afetar a experiência cotidiana das estudantes mesmo quando não aparecem diretamente como intenção declarada de desistência.

Como continuidade, pretende-se usar os achados para planejar oficinas, mentorias, rodas de conversa, visitas técnicas e ações de aproximação entre meninas do ensino superior e alunas do ensino médio técnico. Uma possibilidade é organizar ciclos de intervenção em que cada ação seja associada a uma dimensão observada nos dados: pertencimento, orientação de carreira, apoio pedagógico, contato com modelos femininos e experiências práticas. Em seguida, novas coletas poderão avaliar se as intervenções alteram percepções de confiança, interesse e intenção de continuidade. Trabalhos futuros também poderão ampliar a amostra, incorporar variáveis contextuais adicionais, avaliar diferenças entre subgrupos e explorar técnicas de mineração de texto para aprofundar a análise semântica das respostas abertas. Ao transformar dados educacionais em

evidências para ação, este trabalho reforça o potencial da ciência de dados na construção de trajetórias mais inclusivas para meninas e mulheres em STEM.

Uso de Inteligência Artificial

A Inteligência Artificial foi utilizada como apoio à revisão textual, organização de seções e geração inicial de trechos de código; todas as saídas foram revisadas, verificadas e ajustadas pelas autoras.

Referências

- Alves, G. E. F., Couto, P. G., and Zárate, L. E. (2025). Descrevendo o perfil de pessoas com depressão nas regiões sudeste e centro-oeste do Brasil por meio de mineração de dados. In *Anais Estendidos do XL Simpósio Brasileiro de Banco de Dados*, pages 324–333, Fortaleza/CE. Sociedade Brasileira de Computação.
- Bardin, L. (2011). *Análise de Conteúdo*. Edições 70, Lisboa.
- Conselho Nacional de Saúde (2012). Resolução nº 466, de 12 de dezembro de 2012. Acesso em: 04 mar. 2025.
- Creswell, J. W. (2010). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Sage Publications.
- Gil, A. C. (2008). *Métodos e técnicas de pesquisa social*. Atlas.
- Leão, J., Araújo, A., Souza, B. M. F., Araújo, J. F., Ferreira, K., Barros, M., Lauschner, T., and Nakamura, F. G. (2021). Uma análise regional sobre o perfil de estudantes concluintes em cursos STEM através do enade. In *Anais do XV Women in Information Technology (WIT)*, pages 200–209. Sociedade Brasileira de Computação.
- Menezes, N., Mendes, C. M., Corrêa, J. C., Rocha, T. A., and Mota, M. P. (2024). Além do gênero: Explorando as múltiplas perspectivas de mulheres na computação. In *Anais do XVIII Women in Information Technology (WIT)*, pages 104–114, Brasília/DF. Sociedade Brasileira de Computação.
- Ribeiro, K. S. F. M. and Maciel, C. (2020a). Fatores de influência na escolha pela continuidade da carreira em computação pelas estudantes de ensino médio técnico em informática. In *Anais do XIV Women in Information Technology (WIT)*, pages 40–49. Sociedade Brasileira de Computação.
- Ribeiro, K. S. F. M. and Maciel, C. (2020b). Meninas e identidade profissional: percepções das estudantes de ensino médio integrado em informática sobre a área de computação. *Revista Brasileira de Informática na Educação*, 11(1):309–316.
- Santos, N. D. d. and Marczak, S. (2023). Fatores de atração, evasão e permanência de mulheres nas Áreas da computação. In *Anais do XVII Women in Information Technology (WIT)*, pages 136–147, João Pessoa/PB. Sociedade Brasileira de Computação.
- Silva, A. G., Prado, R. M., Moro, M. M., and Araújo, A. (2023). Autopercepção de meninas do ensino básico em relação às carreiras de STEM. In *Anais do XVII Women in Information Technology (WIT)*, pages 91–102, João Pessoa/PB. Sociedade Brasileira de Computação.
- Silva, A. L. and Zárate, L. E. (2025). Caracterização da população brasileira que sofre de catarata-hipertensão: Um estudo baseado na base de dados PNS 2019. In *Anais Estendidos do XL Simpósio Brasileiro de Banco de Dados*, pages 272–281, Fortaleza/CE. Sociedade Brasileira de Computação.
- Silva, M. O., Oliveira, G. P., and Moro, M. M. (2024). Data insights on gender representation: Analyzing the book and music industries. In *Anais Estendidos do XXXIX Simpósio Brasileiro de Banco de Dados*, pages 338–347, Florianópolis/SC. Sociedade Brasileira de Computação.
- Wang, M.-T. and Degol, J. L. (2017). Gender gap in science, technology, engineering, and mathematics (STEM): Current knowledge, implications for practice, policy, and future directions. *Educational Psychology Review*, 29(1):119–140.