

PERFIS DE ENGAJAMENTO E VULNERABILIDADE ACADÊMICA NO ENSINO SUPERIOR: RECOMENDAÇÕES CONTRAFACTUAIS PARA ACOMPANHAMENTO ESTUDANTIL¹

ENGAGEMENT PROFILES AND ACADEMIC VULNERABILITY IN HIGHER EDUCATION: COUNTERFACTUAL RECOMMENDATIONS FOR STUDENT MONITORING

PERFILES DE COMPROMISO Y VULNERABILIDAD ACADÉMICA EN LA EDUCACIÓN SUPERIOR: RECOMENDACIONES CONTRAFÁCTICAS PARA EL SEGUIMIENTO ESTUDIANTIL

Luiz Aurélio Batista Neto²
Rodrigo Miranda Feitosa³
Dayse Yanne Caldas Siqueira Sousa⁴
Omar Andres Carmona Cortes⁵
Alexandre César Muniz Oliveira⁶

RESUMO: Este artigo analisou de que modo perfis de engajamento estudantil, obtidos por clusterização não supervisionada, podem apoiar a identificação de vulnerabilidade acadêmica e a formulação de recomendações contrafactuais interpretáveis para o acompanhamento educacional. O estudo adotou abordagem quantitativa, aplicada e exploratória, utilizando dados secundários do *Open University Learning Analytics Dataset* (OULAD). A metodologia envolveu pré-processamento, seleção de variáveis comportamentais do ambiente virtual de aprendizagem, clusterização por *K-Means*, avaliação da qualidade e estabilidade dos agrupamentos e geração de recomendações contrafactuais restritas a variáveis acionáveis. Atributos demográficos e desempenho final foram excluídos da clusterização, sendo reservados à interpretação e validação externa dos perfis. Os resultados evidenciaram um gradiente de engajamento, com perfis de baixa, média e alta participação. Grupos de baixo engajamento concentraram maiores proporções de desistência e reprovação, indicando associação com vulnerabilidade acadêmica. As recomendações contrafactuais permitiram simular transições para perfis de maior participação e mostraram-se viáveis na maior parte das transições analisadas. Conclui-se que a abordagem proposta pode apoiar a tomada de decisão educacional, desde que suas recomendações sejam compreendidas como subsídios estatísticos auditáveis, e não como prescrições automáticas ou inferências causais.

Palavras-chave: Explicações contrafactuais. Clusterização educacional. Engajamento estudantil. Analítica da aprendizagem.

¹Trabalho com apoio parcial da CAPES (Código 001), CNPq (Processo nº 420224/2025-4), FAPEMA e IFMA.

²Discente do Programa de Pós-graduação Doutorado em Ciência da Computação - Associação UFMA/UFPI.

³Discente do Programa de Pós-graduação Doutorado em Ciência da Computação - Associação UFMA/UFPI.

⁴Discente do Programa de Pós-graduação Doutorado em Ciência da Computação - Associação UFMA/UFPI.

⁵Doutor em Ciência da Computação (USP) do Instituto Federal de Educação, Ciência e Tecnologia do Maranhão (IFMA).

⁶Doutor em Computação Aplicada (INPE) da Universidade Federal do Maranhão (UFMA).

ABSTRACT: This article analyzed how student engagement profiles, obtained through unsupervised clustering, can support the identification of academic vulnerability and the formulation of interpretable counterfactual recommendations for educational monitoring. The study adopted a quantitative, applied, and exploratory approach using secondary data from the Open University Learning Analytics Dataset (OULAD). The methodology involved preprocessing, selection of behavioral variables from the virtual learning environment, K-Means clustering, assessment of cluster quality and stability, and generation of counterfactual recommendations restricted to actionable variables. Demographic attributes and final performance were excluded from clustering and reserved for profile interpretation and external validation. The results revealed an engagement gradient, with low, medium, and high participation profiles. Low-engagement groups concentrated higher proportions of withdrawal and failure, indicating an association with academic vulnerability. The counterfactual recommendations made it possible to simulate transitions toward higher-participation profiles and proved feasible in most analyzed transitions. The study concludes that the proposed approach can support educational decision-making, provided its recommendations are understood as auditable statistical inputs rather than automatic prescriptions or causal inferences.

Keywords: Counterfactual explanations. Educational clustering. Student engagement. Learning analytics.

RESUMEN: Este artículo analizó de qué modo los perfiles de compromiso estudiantil, obtenidos mediante agrupamiento no supervisado, pueden apoyar la identificación de vulnerabilidad académica y la formulación de recomendaciones contrafactuales interpretables para el seguimiento educativo. El estudio adoptó un enfoque cuantitativo, aplicado y exploratorio, utilizando datos secundarios del Open University Learning Analytics Dataset (OULAD). La metodología incluyó preprocesamiento, selección de variables conductuales del entorno virtual de aprendizaje, clusterización mediante K-Means, evaluación de la calidad y estabilidad de los agrupamientos y generación de recomendaciones contrafactuales restringidas a variables accionables. Los atributos demográficos y el desempeño final no se utilizaron en la formación de los agrupamientos, reservándose para la interpretación y validación externa de los perfiles. Los resultados evidenciaron un gradiente de compromiso, con perfiles de baja, media y alta participación. Los grupos de bajo compromiso concentraron mayores proporciones de abandono y reprobación, indicando una asociación con la vulnerabilidad académica. Las recomendaciones contrafactuales permitieron simular transiciones hacia perfiles de mayor participación y se mostraron viables en la mayoría de las transiciones analizadas. Se concluye que el enfoque propuesto puede apoyar la toma de decisiones educativas, siempre que sus recomendaciones se comprendan como insumos estadísticos auditables, y no como prescripciones automáticas o inferencias causales.

Palabras clave: Explicaciones contrafactuales. Clusterización educativa. Compromiso estudiantil. Analítica del aprendizaje.

I. INTRODUÇÃO

A expansão dos ambientes virtuais de aprendizagem e das plataformas educacionais digitais ampliou a disponibilidade de dados sobre trajetórias, interações e padrões de

participação dos estudantes. Registros de acesso, atividades realizadas, uso de recursos didáticos e interações no ambiente virtual passaram a fornecer evidências relevantes para compreender processos de aprendizagem em larga escala. Contudo, a simples disponibilidade desses dados não garante sua utilidade pedagógica. O problema central enfrentado neste estudo consiste em transformar rastros digitais de participação estudantil em informações interpretáveis, confiáveis e úteis ao acompanhamento educacional, de modo que os perfis identificados não se limitem à descrição estatística dos agrupamentos, mas possam apoiar a compreensão da vulnerabilidade acadêmica (Siemens, 2013; Romero; Ventura, 2020).

Entre as estratégias utilizadas para explorar dados educacionais, a clusterização não supervisionada tem sido amplamente empregada para identificar perfis de estudantes com base em padrões não rotulados. Ao reconhecer estruturas latentes nos dados sem depender de rótulos previamente definidos, essa abordagem permite agrupar estudantes com comportamentos semelhantes de engajamento e participação. Embora útil para descrever perfis, esse tipo de modelagem tende a permanecer essencialmente descritivo, oferecendo pouca orientação sobre como diferenças entre grupos se manifestam em termos de ações observáveis. Em contextos de educação a distância, bases como o *Open University Learning Analytics Dataset* (OULAD) favorecem esse tipo de investigação por reunirem registros acadêmicos e interações em ambiente virtual em uma estrutura pública e documentada (Kuzilek; Hlosta; Zdrahal, 2017). Estudos sobre perfilamento estudantil também indicam que agrupamentos derivados de padrões comportamentais podem revelar diferenças relevantes em engajamento, ritmo de aprendizagem e participação acadêmica (Araka et al., 2022; Mojarad et al., 2018).

Apesar de sua utilidade descritiva, a identificação de perfis educacionais não é, por si só, suficiente para produzir conhecimento acionável. Métricas internas podem avaliar a qualidade estatística dos agrupamentos, enquanto validações externas pós-hoc podem relacioná-los a indicadores como aprovação, reprovação ou evasão. Ainda assim, permanece uma lacuna central na literatura: estudos de clusterização educacional raramente indicam quais mudanças mínimas, plausíveis e operacionalizáveis nas variáveis comportamentais poderiam aproximar um estudante de um perfil de maior engajamento. Essa limitação restringe a utilidade prática dos agrupamentos no acompanhamento educacional, uma vez que a simples caracterização de um perfil de baixo engajamento não esclarece quais dimensões comportamentais são efetivamente modificáveis nem quais ajustes seriam suficientes para promover transições entre perfis.

Nesse contexto, a Inteligência Artificial Explicável (*Explainable Artificial Intelligence - XAI*) tem sido utilizada para aumentar a interpretabilidade de modelos computacionais. Métodos como SHAP e LIME são amplamente aplicados para estimar a contribuição de atributos em decisões de modelos preditivos, inclusive em cenários educacionais (Ribeiro; Singh; Guestrin, 2016; Lundberg; Lee, 2017; Khosravi et al., 2022). Entretanto, essas abordagens permanecem predominantemente diagnósticas, pois explicam a influência das variáveis sem indicar intervenções mínimas capazes de alterar um resultado ou aproximar uma instância de outro perfil. Nesse sentido, as explicações contrafactuais emergem como alternativa, ao responderem à questão de quais modificações seriam necessárias para que uma instância alcançasse um novo estado ou classificação (Wachter; Mittelstadt; Russell, 2018).

Diante dessa lacuna, este artigo tem como objetivo analisar como perfis de engajamento estudantil, obtidos por clusterização não supervisionada, podem apoiar a identificação de vulnerabilidade acadêmica e a formulação de recomendações contrafactuais interpretáveis para o acompanhamento educacional. Especificamente, investiga-se se agrupamentos derivados de variáveis comportamentais do ambiente virtual podem revelar estruturas educacionalmente significativas e subsidiar simulações de transição entre perfis, restritas a dimensões potencialmente acionáveis do engajamento estudantil. O estudo utiliza o *Open University Learning Analytics Dataset* (OULAD), com geração de perfis por meio do algoritmo *K-Means* e produção de recomendações voltadas à transição entre agrupamentos. A formação dos perfis considera exclusivamente variáveis comportamentais de interação no ambiente virtual, mantendo atributos demográficos, notas e resultados finais fora da etapa de clusterização, preservando a natureza não supervisionada da análise e reduzindo risco de vazamento de informação.

A contribuição deste trabalho reside na integração entre a clusterização educacional e recomendações contrafactuais restritas ao espaço de ação, permitindo reinterpretar agrupamentos não apenas como descrições de perfis, mas também como estruturas potencialmente acionáveis na simulação de transições comportamentais. Essa integração busca ampliar a utilidade analítica dos clusters no contexto educacional, oferecendo indicações estatísticas sobre mudanças plausíveis sem assumir caráter prescritivo ou causal. Além desta introdução, o artigo apresenta a fundamentação teórica e os trabalhos relacionados, descreve a metodologia adotada, expõe os resultados, discute suas implicações e limitações e, por fim, apresenta as conclusões e direções futuras de pesquisa.

2. REFERENCIAL TEÓRICO E TRABALHOS RELACIONADOS

2.1 ANALÍTICA DA APRENDIZAGEM E PERFIS EDUCACIONAIS POR CLUSTERIZAÇÃO

A Analítica da Aprendizagem constitui um campo interdisciplinar dedicado à coleta, análise e comunicação de dados educacionais para compreender e aprimorar os processos de aprendizagem. Em ambientes virtuais, registros de acesso, interações e regularidade de participação permitem identificar padrões de engajamento estudantil quando analisados por procedimentos interpretáveis e metodologicamente consistentes (Siemens, 2013; Baker; Inventado, 2014; Romero; Ventura, 2020). Evidências recentes indicam ainda que esses registros podem subsidiar análises sobre desempenho, permanência e vulnerabilidade acadêmica em cursos mediados por tecnologia (Aljohani; Fayoumi; Hassan, 2019; Waheed et al., 2020).

Nesse contexto, o Open University Learning Analytics Dataset (OULAD) tornou-se uma base de referência para estudos em Analítica da Aprendizagem por reunir dados anonimizados de estudantes da Open University, incluindo informações acadêmicas, demográficas e avaliativas, além de registros de interação no ambiente virtual (Kuzilek; Hlostá; Zdrahal, 2017). Sua estrutura relacional favorece a análise de trajetórias educacionais em cursos a distância e permite separar variáveis utilizadas na formação de perfis daquelas destinadas à validação externa, preservando a natureza não supervisionada da análise e evitando vazamento de informação.

Entre as estratégias analíticas mais utilizadas nesse contexto, a clusterização não supervisionada destaca-se por permitir a identificação de perfis estudantis com base em padrões comportamentais não rotulados, como regularidade de acesso, ritmo de estudo e interação com recursos digitais. Embora eficaz para revelar estruturas latentes nos dados, essa abordagem tende a produzir representações essencialmente descritivas, com limitada capacidade de orientar ações educacionais concretas. Estudos indicam que agrupamentos derivados desses padrões podem revelar diferenças relevantes em participação, permanência e desempenho acadêmico quando analisados posteriormente à luz de indicadores educacionais externos (Araka et al., 2022; Mojarad et al., 2018).

Entre os métodos disponíveis, o *K-Means* permanece amplamente utilizado em dados educacionais tabulares devido à sua simplicidade, eficiência computacional e interpretabilidade baseada em centroides, favorecendo a caracterização de perfis segundo padrões médios de participação e interação no ambiente virtual. Contudo, sua aplicação requer validação rigorosa,

incluindo métricas internas de qualidade, estabilidade das partições e validação externa pós-hoc, de modo a assegurar consistência estatística e relevância educacional dos agrupamentos. Ainda assim, mesmo quando bem validados, os clusters não oferecem, por si só, suporte para ação educacional, uma vez que descrevem perfis sem indicar como uma instância pode transitar entre eles.

Essa limitação torna-se evidente ao se observar que a interpretação baseada em centroides não responde à questão de quais mudanças mínimas e concretas poderiam aproximar um estudante de um agrupamento mais favorável. Assim, embora indicadores externos como aprovação, reprovação ou evasão possam ser utilizados em validações pós-hoc, eles não devem integrar o processo de formação dos clusters, preservando sua natureza não supervisionada. O resultado é um cenário em que os agrupamentos são estatisticamente consistentes, mas ainda insuficientes para produzir conhecimento acionável no contexto educacional.

Nesse cenário, torna-se necessário recorrer a abordagens complementares capazes de traduzir diferenças entre agrupamentos em formas interpretáveis de transição comportamental.

2.2 EXPLICABILIDADE, ATRIBUIÇÃO E RECOMENDAÇÕES CONTRAFACTUAIS

6

Partindo dessa necessidade de complementar a clusterização com mecanismos de interpretação acionável, a Inteligência Artificial Explicável tem sido mobilizada para ampliar a transparência, a auditabilidade e a interpretabilidade de modelos computacionais em contextos educacionais, especialmente na Analítica da Aprendizagem, em que evidências algorítmicas podem apoiar o acompanhamento estudantil e a identificação de vulnerabilidades acadêmicas. Métodos pós-hoc, como SHAP e LIME, são amplamente utilizados para estimar a contribuição de atributos em decisões algorítmicas, oferecendo explicações baseadas na importância relativa das variáveis (Ribeiro; Singh; Guestrin, 2016; Lundberg; Lee, 2017). Embora úteis para identificar fatores associados ao desempenho, à evasão e à permanência, essas técnicas permanecem predominantemente diagnósticas, pois não indicam quais mudanças mínimas poderiam reposicionar uma instância em direção a outro perfil. Revisões recentes sobre explicações contrafactuais em educação indicam crescimento do interesse por abordagens interpretáveis capazes de apoiar decisões educacionais, embora persistam lacunas quanto à validação pedagógica, à plausibilidade das recomendações e ao uso em cenários não supervisionados (Buñay-Guisñán; Lara; Romero, 2026).

Nesse contexto, as explicações contrafactuais oferecem uma alternativa mais alinhada ao problema de ação educacional, ao indicar quais alterações seriam necessárias para que uma instância alcançasse um resultado ou perfil distinto (Wachter; Mittelstadt; Russell, 2018). Diferentemente dos métodos de atribuição, elas propõem modificações específicas nas variáveis de entrada e devem ser avaliadas por critérios como validade, proximidade, diversidade, plausibilidade e factibilidade, conforme sistematizado na literatura recente sobre explicações contrafactuais (Mothilal; Sharma; Tan, 2020; Guidotti, 2024). No âmbito do recourse algorítmico, tais recomendações incorporam restrições de ação e custos associados às mudanças propostas, buscando garantir coerência com o domínio de aplicação e evitar sugestões inviáveis ou incompatíveis com o contexto real (Ustun; Spangher; Liu, 2019; Karimi et al., 2022).

No domínio educacional, essas restrições são centrais. Recomendações contrafactuais não devem incidir sobre atributos imutáveis, dados demográficos, histórico acadêmico consolidado, notas já obtidas ou resultados finais, mas concentrar-se em variáveis comportamentais potencialmente modificáveis, como regularidade de acesso, participação em atividades e interação com recursos educacionais. Assim, neste estudo, as explicações contrafactuais são tratadas como recomendações estatísticas de apoio à interpretação dos perfis de engajamento, complementando a clusterização ao traduzir diferenças entre agrupamentos em possibilidades de transição sob restrições realistas, sem caráter prescritivo ou inferência causal.

2.3 EXPLICAÇÕES CONTRAFACTUAIS PARA AGRUPAMENTOS E LACUNA DO ESTUDO

As explicações contrafactuais têm sido desenvolvidas predominantemente em cenários supervisionados, nos quais indicam alterações mínimas capazes de modificar a classe predita de uma instância. Sua aplicação à clusterização exige, contudo, uma reinterpretação conceitual, pois os agrupamentos não decorrem de rótulos prévios, mas de estruturas latentes identificadas nos dados. Nesse contexto, a mudança contrafactual corresponde à transição entre perfis, sem implicar causalidade, prescrição automática ou melhoria acadêmica.

Essa adaptação a modelos não supervisionados ainda envolve desafios relacionados à interpretação dos agrupamentos, à estabilidade das explicações e à plausibilidade das instâncias geradas. Estudos recentes, como Vardakas et al. (2025) e Karra et al. (2025), avançam nessa

direção ao investigar contrafactuais para K-Means, modelos gaussianos de mistura e estratégias baseadas na equivalência entre clusterização e classificação. Ainda assim, permanece em aberto como incorporar restrições de domínio e avaliar a utilidade interpretativa das transições entre grupos.

No domínio da Analítica da Aprendizagem, essa questão torna-se especialmente relevante, pois os clusters frequentemente representam perfis de engajamento, participação e vulnerabilidade acadêmica. Assim, a transição entre agrupamentos deve ser compreendida como hipótese estatística de mudança comportamental, e não como simples deslocamento geométrico no espaço de características. Para que essas recomendações sejam educacionalmente úteis, devem respeitar restrições de imutabilidade, plausibilidade e acionabilidade, concentrando-se apenas em dimensões potencialmente modificáveis do comportamento estudantil.

Diante disso, a lacuna deste estudo reside na integração ainda limitada entre clusterização educacional, explicações contrafactuais para agrupamentos e mecanismos formais de restrição de ação. Embora a literatura em Analítica da Aprendizagem utilize clusterização para descrever perfis e relacioná-los a indicadores acadêmicos, ainda são escassas as abordagens que traduzem diferenças entre perfis em recomendações contrafactuais plausíveis, restritas e acionáveis. Este artigo busca contribuir nessa direção ao ampliar a interpretabilidade dos agrupamentos educacionais por meio de recomendações estatísticas de transição entre perfis, preservando seu caráter não causal, não prescritivo e orientado ao suporte à decisão.

3. METODOLOGIA

3.1 TIPO DE ESTUDO, BASE DE DADOS E RECORTE ANALÍTICO

Dando continuidade ao arcabouço teórico apresentado nas Seções 2.1 a 2.3, este estudo operacionaliza uma arquitetura analítica integrada para interpretação de perfis de engajamento estudantil, combinando clusterização não supervisionada, validação externa e explicações contrafactuais sob restrições de ação. Essa arquitetura é estruturada em três camadas: (i) construção de perfis a partir de variáveis comportamentais, (ii) validação estatística e educacional pós-hoc dos agrupamentos e (iii) geração de recomendações contrafactuais para simulação de transições entre perfis, respeitando restrições de imutabilidade e plausibilidade. Trata-se de um estudo quantitativo, aplicado e exploratório, no qual a componente quantitativa

decorre da implementação computacional do pipeline e o caráter exploratório resulta da ausência de rótulos supervisionados na formação dos agrupamentos.

A base de dados utilizada foi o *Open University Learning Analytics Dataset* (OULAD), conjunto público e anonimizado amplamente empregado em Analítica da Aprendizagem por integrar informações acadêmicas, demográficas, avaliativas e registros de interação em ambiente virtual (Kuzilek; Hlostá; Zdrahal, 2017). Em coerência com a arquitetura proposta, os dados foram organizados em uma matriz centrada no engajamento comportamental, enquanto variáveis acadêmicas e contextuais foram reservadas exclusivamente para validação externa pós-hoc.

A unidade de análise corresponde ao estudante em uma apresentação de módulo, isto é, uma oferta específica de disciplina em determinado período acadêmico (ex.: 2013B, 2014B, 2013J). Assim, cada instância representa a participação do estudante em um contexto específico, preservando variações temporais e institucionais. O recorte empírico concentrou-se nessas três apresentações, sendo duas utilizadas como análise principal e uma como verificação de consistência, evitando a suposição de homogeneidade entre contextos distintos.

O corpus final foi composto por 21.333 instâncias, correspondentes a 19.720 estudantes únicos, distribuídas em sete módulos e 15 combinações módulo-apresentação. Essa organização preserva a heterogeneidade contextual do conjunto de dados e evita tratá-lo como uma coleção homogênea de observações independentes.

A Figura 1 sintetiza a arquitetura metodológica do estudo, evidenciando o fluxo integrado entre seleção da base, recorte por apresentação, pré-processamento, construção da matriz de engajamento, clusterização, validação dos agrupamentos e geração de recomendações contrafactuais. O pipeline explicita três componentes funcionais: (i) variáveis usadas na formação dos clusters (engajamento comportamental), (ii) variáveis destinadas à validação externa pós-hoc (indicadores acadêmicos e contextuais) e (iii) variáveis elegíveis para geração de explicações contrafactuais sob restrições de ação.

Figura 1 – Fluxo metodológico do estudo



Fonte: Elaborada pelos autores, 2026.

3.2 VARIÁVEIS, PRÉ-PROCESSAMENTO E PREVENÇÃO DE VAZAMENTO DE INFORMAÇÃO

A construção da matriz analítica seguiu o princípio de separação estrita entre variáveis utilizadas na clusterização, variáveis reservadas à validação externa e variáveis elegíveis para geração de recomendações contrafactuais, em coerência com a arquitetura metodológica apresentada na Seção 3.1. Essa separação foi fundamental para preservar a natureza não supervisionada do estudo e evitar o vazamento de informações provenientes de desfechos acadêmicos na formação dos agrupamentos.

O espaço de entrada do *K-Means* foi definido por cinco dimensões comportamentais: participação ativa, volume de interações, regularidade de acesso, atividade passiva e atividade social. Essas variáveis representam padrões observáveis de engajamento no ambiente virtual e constituem exclusivamente o espaço de formação dos clusters, sem incorporação de notas, desempenho ou atributos demográficos. Tais variáveis foram mantidas apenas para validação externa pós-hoc e análise interpretativa dos perfis.

O pré-processamento envolveu consolidação dos logs de interação, tratamento de valores ausentes por imputação via mediana e escalonamento robusto das variáveis. Valores extremos foram preservados, uma vez que podem refletir comportamentos reais de engajamento, sendo apenas mitigado o impacto desses valores na geometria do espaço por meio da padronização.

De forma coerente com a etapa posterior de explicações contrafactuais (Seção 3.4), a definição do espaço de variáveis também delimitou o conjunto de atributos elegíveis para

intervenção simulada, restringindo-se às dimensões comportamentais acionáveis e excluindo variáveis imutáveis, acadêmicas e de resultado final.

3.3 CLUSTERIZAÇÃO, VALIDAÇÃO E ESTABILIDADE DOS PERFIS

A etapa de clusterização foi realizada com o algoritmo *K-Means*, aplicado separadamente às apresentações 2013B, 2014B e 2013J, conforme definido no recorte analítico. O objetivo foi identificar perfis de engajamento estudantil com base em padrões comportamentais previamente definidos, sem recorrer a rótulos supervisionados.

A definição do número de clusters seguiu critérios combinados de qualidade estatística, estabilidade e interpretabilidade. Foram testadas diferentes soluções de k , sendo adotado $k = 5$ por apresentar melhor equilíbrio entre separação dos grupos, consistência sob reamostragem e viabilidade interpretativa das transições analisadas na etapa contrafactual. O algoritmo foi executado com múltiplas inicializações e controle de reprodutibilidade para reduzir a variabilidade associada à inicialização aleatória.

A avaliação dos agrupamentos incluiu métricas internas de coesão e separação, além de análise de estabilidade via *bootstrap*, comparando partições geradas sob reamostragem. Em continuidade à estratégia metodológica apresentada na Seção 3.2, realizou-se ainda a validação externa pós-hoc, utilizando variáveis acadêmicas não incluídas na clusterização, com o objetivo de avaliar o significado educacional dos perfis sem comprometer o desenho não supervisionado.

Após a formação dos clusters, os perfis foram caracterizados com base nas médias não escalonadas das variáveis comportamentais, permitindo interpretação pedagógica dos agrupamentos. Essa caracterização serviu de base para a etapa seguinte de análise de transições e de geração de explicações contrafactuais.

3.4 GERAÇÃO E AVALIAÇÃO DAS RECOMENDAÇÕES CONTRAFACTUAIS

Em continuidade à estrutura de perfis obtida na etapa anterior, as recomendações contrafactuais foram utilizadas como simulações de transição entre agrupamentos, indicando quais mudanças comportamentais seriam necessárias para aproximar uma instância de um grupo-destino no espaço de clusterização. Diferentemente de abordagens gerais como DiCE (*Diverse Counterfactual Explanations*) e *Growing Spheres* (Laugel et al., 2017), o método adotado combina restrição explícita do espaço de ação, controle de esparsidade, busca por vizinhos e validação em uma arquitetura Proxy-Oracle. Nesse contexto, as recomendações são

interpretadas como instrumentos de análise estrutural das diferenças entre perfis, e não como prescrições pedagógicas.

As transições analisadas seguem o gradiente de engajamento identificado pelos clusters, incluindo movimentos de baixo para intermediário e de intermediário para alto engajamento, considerando as apresentações 2013B, 2014B e 2013J. Todas as transições são interpretadas como aproximações estatísticas no espaço de comportamento, sem implicações causais ou garantia de melhoria acadêmica.

A arquitetura Proxy-Validador foi implementada em duas etapas. Inicialmente, um classificador multiclasse de Regressão Logística, treinado para reproduzir os grupos obtidos pelo K-Means, orientou a busca por modificações comportamentais mínimas e esparsas. Em seguida, um classificador multiclasse HistGradientBoosting atuou como validador supervisionado auxiliar, confirmando a atribuição da instância modificada ao grupo de destino. Assim, a recomendação foi considerada válida apenas quando a transição foi confirmada nesse esquema de dupla verificação, mantendo-se o K-Means como referência para a definição inicial dos agrupamentos.

A elegibilidade para geração contrafactual é restrita aos indivíduos pertencentes ao cluster de origem, e apenas variáveis comportamentais acionáveis são modificadas. Caso não seja possível gerar uma alteração válida sob essas restrições, a instância permanece elegível, mas não compõe o conjunto de recomendações bem-sucedidas. A avaliação das recomendações considera validade, cobertura, proximidade, esparsidade, plausibilidade, factibilidade e estabilidade, permitindo mensurar simultaneamente capacidade de transição, custo de intervenção e robustez das explicações sob reamostragem.

3.5 CONSIDERAÇÕES ÉTICAS E LIMITAÇÕES METODOLÓGICAS

O estudo utiliza dados públicos e anonimizados do OULAD, analisados de forma agregada, sem possibilidade de identificação individual. Identificadores foram usados apenas para integração dos registros, sem participação nas etapas de clusterização ou geração de recomendações. Os agrupamentos são interpretados como padrões estatísticos de engajamento, e não como categorias diagnósticas, enquanto as recomendações contrafactuais correspondem a simulações interpretativas de mudanças no espaço comportamental, sem caráter causal ou prescritivo. Por se tratar de dados observacionais, os resultados têm finalidade exploratória e de apoio interpretativo pós-hoc, não sendo apropriados para inferência causal ou intervenção

direta, devendo ser complementados por validação pedagógica e supervisão humana em contextos reais.

4. RESULTADOS

4.1 AMOSTRA E QUALIDADE DA CLUSTERIZAÇÃO

A análise empírica foi conduzida com dados do OULAD previamente processados, totalizando 21.333 instâncias de 19.720 estudantes distribuídos nas apresentações 2013B, 2014B e 2013J, sendo as duas primeiras o recorte principal e a última utilizada como verificação de consistência. Os agrupamentos obtidos refletem padrões de engajamento comportamental, uma vez que foram formados exclusivamente a partir de variáveis de interação no ambiente virtual. A solução com $k = 5$ apresentou desempenho satisfatório em termos de qualidade e estabilidade das partições.

Esse desempenho foi avaliado por meio de indicadores de separação e coesão, representados por Silhouette, que mensura o grau de adequação de cada instância ao seu cluster (Rousseeuw, 1987), Davies-Bouldin, que expressa a similaridade média entre clusters, sendo valores menores indicativos de melhor separação (Davies; Bouldin, 1979), e Calinski-Harabasz, que relaciona a dispersão intra e interclusters, sendo valores maiores associados a grupos mais bem definidos (Calinski; Harabasz, 1974). Além disso, foram consideradas medidas de estabilidade sob reamostragem, representadas pelo ARI (*Adjusted Rand Index*), índice ajustado de similaridade entre partições (Hubert; Arabie, 1985), e pelo NMI (*Normalized Mutual Information*), medida normalizada de informação compartilhada entre agrupamentos (Strehl; Ghosh, 2002), conforme apresentado na Tabela 1.

13

Tabela 1 – Qualidade e estabilidade da clusterização por apresentação

Apresentação	Silhouette	Davies-Bouldin	Calinski-Harabasz	ARI médio	NMI médio
2013B	0,583	0,884	3.444,25	0,920	0,878
2014B	0,558	0,827	5.382,12	0,879	0,817
2013J	0,629	0,876	6.263,54	0,890	0,832

Fonte: Elaborada pelos autores, 2026.

De forma geral, os resultados sugerem estabilidade adequada dos grupos, com valores de ARI superiores a 0,870 e NMI superiores a 0,810 nas três apresentações analisadas (Tabela 1).

4.2 PERFIS DE ENGAJAMENTO E VALIDAÇÃO EXTERNA

Com a clusterização validada na subseção anterior, passou-se à identificação dos padrões de engajamento no ambiente virtual de aprendizagem. Embora a solução tenha produzido cinco agrupamentos, a análise concentrou-se nos três perfis mais relevantes para a investigação das transições contrafactuais: baixo, médio e alto engajamento. Esses perfis foram caracterizados com base nas médias das dimensões comportamentais dos grupos (Tabela 2).

Tabela 2 – Perfis predominantes de engajamento identificados a partir das médias comportamentais por grupo.

Apresentação	Perfil	n (%)	Volume de Cliques	Dias ativos	Atividade ativa	Atividade passiva	Atividade social	Classe de engajamento
2013B	G1	66,46%	533,51	38,00	36,06	286,18	138,78	Baixo
	G3	12,34%	2.366,17	129,20	76,23	1.239,12	1.050,82	Médio
	G4	15,46%	3.078,73	99,47	733,77	1.992,90	352,05	Alto
2014B	G0	72,12%	417,34	29,27	71,10	215,01	60,57	Baixo
	G2	11,57%	1.912,28	105,42	120,68	1.115,86	675,74	Médio
	G4	13,39%	2.792,49	99,02	875,83	1.620,87	295,79	Alto
2013J	G0	75,87%	591,31	41,13	48,93	339,61	125,93	Baixo
	G2	10,04%	2.828,08	140,90	110,60	1.536,68	1.180,81	Médio
	G4	11,18%	3.040,16	103,56	739,68	1.934,64	365,84	Alto

Fonte: Elaborada pelos autores, 2026.

Observa-se predominância do perfil de baixo engajamento nas três apresentações, com 66,46% em 2013B, 72,12% em 2014B e 75,87% em 2013J. Esses grupos apresentaram menor volume de cliques, menor regularidade de acesso e menor participação nas dimensões comportamentais analisadas. Em contraste, os perfis médio e alto concentraram estudantes com maior intensidade de engajamento, distinguindo-se, respectivamente, pela maior regularidade e interação social, e pelo maior volume de interações em atividades ativas e passivas.

A validação externa pós-hoc, realizada com a variável `final_result` apenas após a formação dos agrupamentos, indicou maior concentração de desistências e reprovações nos perfis de baixo engajamento. Essa variável corresponde ao resultado acadêmico final do estudante. Nas apresentações principais, a associação entre perfis e desfechos acadêmicos foi estatisticamente significativa e de magnitude moderada: 2013B, $\chi^2 = 1309,42$, $gl = 12$, $p < 0,001$, V de Cramer = 0,3053; e 2014B, $\chi^2 = 2165,44$, $gl = 12$, $p < 0,001$, V de Cramer = 0,3041.

4.3 RECOMENDAÇÕES CONTRAFACTUAIS ENTRE PERFIS DE ENGAJAMENTO

Após a caracterização dos perfis, analisaram-se as recomendações contrafactuais como simulações estatísticas de transição entre perfis de engajamento. Ao todo, foram investigadas seis transições, tendo as apresentações 2013B e 2014B como recorte principal e 2013J como verificação complementar. Para cada transição, foram reportados o número de estudantes elegíveis, as recomendações geradas, a cobertura, a validade, o número de variáveis alteradas e a distância em relação ao perfil-alvo (Tabela 3).

Tabela 3 – Recomendações contrafactuais para transições entre perfis de engajamento.

Apresentação	Transição	Elegíveis	Geradas	Cobertura (%)	Validade (%)	Nº de variáveis alteradas	Distância ao perfil-alvo
2013B	G ₁ →G ₃	55	55	100	100	4	12,86
	G ₃ →G ₄	578	578	100	100	3	2,43
2014B	G ₀ →G ₂	5.628	5.628	100	100	3	1,90
	G ₂ →G ₄	903	903	100	100	3	4,54
2013J	G ₀ →G ₂	6.711	6.711	100	100	4	2,72
	G ₂ →G ₄	888	6	0,7	100	3	1,80

Fonte: Elaborada pelos autores, 2026.

Em termos interpretativos, as transições entre perfis refletem deslocamentos ao longo de um gradiente de engajamento, desde padrões mais baixos até níveis intermediários e elevados de participação no ambiente virtual. Esses resultados devem ser compreendidos como aproximações estatísticas entre agrupamentos, sem caráter prescritivo nem garantia de melhoria acadêmica.

De modo geral, observou-se alta cobertura e validade de 100% na maioria das transições, indicando que foi possível gerar recomendações capazes de reatribuir as instâncias ao perfil-alvo nas apresentações 2013B e 2014B, bem como na transição G₀→G₂ de 2013J. Essas recomendações envolveram alterações em três a quatro variáveis comportamentais, com complexidade moderada, sendo a maior distância ao perfil-alvo identificada na transição G₁→G₃ de 2013B, que exigiu mudanças mais expressivas para reposicionamento no espaço de clusters.

Em contraste, a transição 2013J: G₂→G₄ apresentou baixa cobertura, apesar de manter 100% de validade nas recomendações geradas, sugerindo que a migração para o perfil de maior engajamento demandaria mudanças fora das restrições de plausibilidade adotadas. Esse

resultado reforça o caráter restritivo do método, que evita recomendações quando estas não são compatíveis com a estrutura empírica observada nos dados.

5. DISCUSSÃO E IMPLICAÇÕES PRÁTICAS

5.1 PERFIS DE ENGAJAMENTO E VULNERABILIDADE ACADÊMICA

A associação estatisticamente significativa entre os perfis de engajamento e os desfechos acadêmicos, com magnitude moderada (V de Cramer $\approx 0,30$), reforça a relevância dos agrupamentos, sem indicar relação causal. Os resultados mostram que os perfis obtidos pela clusterização evidenciam diferenças consistentes no engajamento estudantil no ambiente virtual, especialmente quanto à regularidade de acesso, ao volume de interações e à participação em atividades.

Esse resultado está alinhado à literatura de Analítica da Aprendizagem, que utiliza rastros digitais para compreender padrões de participação e vulnerabilidade acadêmica (Siemens, 2013; Romero; Ventura, 2020). Os perfis de baixo engajamento concentraram a maior proporção de estudantes em todas as três apresentações analisadas. Esses grupos também apresentaram menor participação no ambiente virtual, enquanto os perfis médio e alto exibiram níveis mais elevados de interação e regularidade.

A validação externa pós-hoc com o desfecho acadêmico indicou maior concentração de reprovação e evasão nos perfis de baixo engajamento e melhores resultados nos perfis de maior engajamento. Apesar disso, os resultados são apenas associativos e não permitem inferência causal. Assim, os perfis devem ser interpretados como agrupamentos estatísticos de comportamento, e não como categorias diagnósticas.

5.2 RECOMENDAÇÕES CONTRAFACTUAIS E INTERPRETABILIDADE DOS PERFIS

Este estudo amplia a análise dos perfis ao incluir recomendações contrafactuais baseadas nas diferenças entre agrupamentos. Em vez de apenas descrever os grupos, o método busca identificar mudanças comportamentais necessárias para aproximar um estudante de outro perfil. Essa abordagem desloca a análise de uma descrição dos perfis para uma interpretação mais operacional, baseada em possíveis transições entre grupos de engajamento. As recomendações contrafactuais representam alterações em variáveis comportamentais

acionáveis, como regularidade de acesso, volume de interações e participação em atividades no ambiente virtual.

Diferentemente de métodos como SHAP e LIME, que indicam a importância das variáveis (Ribeiro; Singh; Guestrin, 2016; Lundberg; Lee, 2017), as explicações contrafactuais indicam o que precisa mudar para alcançar outro resultado. Esse tipo de abordagem é mais próximo de recomendações interpretáveis, pois considera critérios como proximidade, plausibilidade e factibilidade (Wachter; Mittelstadt; Russell, 2018; Karimi et al., 2022). Na clusterização, não há classes reais, mas reatribuições baseadas em modelos. Por isso, a validade contrafactual foi definida como a mudança de um perfil para outro pelo modelo de referência, sem interpretação causal.

Os resultados mostram alta validade e cobertura na maioria das transições analisadas. No entanto, uma das transições apresentou baixa cobertura, indicando que nem todas as mudanças entre perfis são viáveis sob restrições realistas.

5.3 IMPLICAÇÕES, LIMITAÇÕES E PESQUISAS FUTURAS

Os resultados indicam que perfis de engajamento e recomendações contrafactuais podem apoiar a interpretação de padrões educacionais em nível agregado. Para professores e tutores, os perfis ajudam a identificar estudantes com menor participação no ambiente virtual. Para gestores, oferecem uma visão comparativa entre grupos e contextos de oferta.

A principal contribuição prática reside em explicitar as diferenças entre perfis. Em vez de apenas identificar baixo engajamento, o modelo indica quais dimensões comportamentais estão associadas a diferentes níveis de participação. As recomendações contrafactuais devem ser interpretadas com cautela. Elas não indicam que mudanças comportamentais geram automaticamente melhores resultados acadêmicos. Trata-se de simulações estatísticas, sem caráter causal ou prescritivo. Portanto, não devem ser usadas para decisões automatizadas ou punitivas.

Como limitação, destaca-se o uso de dados históricos e observacionais do OULAD, o que limita a análise a um estudo pós-hoc. Os resultados também dependem das escolhas de modelagem e do recorte das apresentações analisadas. Pesquisas futuras podem ampliar a abordagem por meio da validação com especialistas educacionais, aplicação em outros contextos institucionais e comparação com outros métodos de clusterização e geração de explicações contrafactuais.

6. CONCLUSÃO

Este estudo investigou a relação entre perfis de engajamento estudantil obtidos por meio de clusterização não supervisionada e recomendações contrafactuais no contexto da Analítica da Aprendizagem. Os resultados indicam a presença de um gradiente consistente de engajamento, associado a diferentes níveis de vulnerabilidade acadêmica. A principal contribuição reside na integração entre clusterização e recomendações contrafactuais baseadas em restrições de ação, permitindo simular transições interpretáveis entre perfis de engajamento.

Os achados devem ser interpretados como evidências analíticas derivadas de dados observacionais, com uso pós-hoc, sem implicar inferência causal ou aplicação automatizada. Futuras pesquisas podem ampliar a abordagem por meio de validação externa com especialistas, outros contextos institucionais e comparação com diferentes modelos de clusterização e recourse contrafactual.

REFERÊNCIAS

- ALJOHANI, N. R.; FAYOUMI, A. G.; HASSAN, S.-U. **Predicting at-risk students using clickstream data in the virtual learning environment.** *Sustainability*, v. 11, n. 24, artigo 7238, 2019. DOI: 10.3390/su11247238.
- ARAKA, E.; OBOKO, R.; MAINA, E.; GITONGA, R. **Using educational data mining techniques to identify profiles in self-regulated learning: an empirical evaluation.** *The International Review of Research in Open and Distributed Learning*, v. 23, n. 1, p. 131-162, 2022. DOI: 10.19173/irrod1.v23i1.5401.
- BAKER, R. S.; INVENTADO, P. S. **Educational data mining and learning analytics.** In: LARUSSON, J. A.; WHITE, B. (org.). *Learning analytics: from research to practice*. New York: Springer, 2014. p. 61-75. DOI: 10.1007/978-1-4614-3305-7_4.
- BUÑAY-GUISÑÁN, Pamela; LARA, Juan A.; ROMERO, Cristóbal. **Counterfactual explanations in education: a systematic review.** *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, v. 16, n. 1, e70060, 2026. DOI: 10.1002/widm.70060.
- CALINSKI, T.; HARABASZ, J. **A dendrite method for cluster analysis.** *Communications in Statistics*, v. 3, n. 1, p. 1-27, 1974. DOI: 10.1080/03610927408827101.
- DAVIES, D. L.; BOULDIN, D. W. **A cluster separation measure.** *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. PAMI-1, n. 2, p. 224-227, 1979. DOI: 10.1109/TPAMI.1979.4766909.

GUIDOTTI, Riccardo. **Counterfactual explanations and how to find them: literature review and benchmarking**. *Data Mining and Knowledge Discovery*, v. 38, p. 2770–2824, set. 2024. DOI: 10.1007/s10618-022-00831-6.

HUBERT, Lawrence; ARABIE, Phipps. Comparing partitions. *Journal of Classification*, v. 2, n. 1, p. 193–218, 1985. DOI: 10.1007/BF01908075.

KARIMI, A.-H.; BARTHE, G.; SCHÖLKOPF, B.; VALERA, I. **A survey of algorithmic recourse: contrastive explanations and consequential recommendations**. *ACM Computing Surveys*, v. 55, n. 5, artigo 95, p. 1–29, 2022. DOI: 10.1145/3527848.

KARRA, Antonia; VARDAKAS, Georgios; PITOURA, Evaggelia; LIKAS, Aristidis. **Generating counterfactual explanations for clustering models based on their equivalence to classification models**. In: MAGLOGIANNIS, Ilias et al. (ed.). *Artificial Intelligence Applications and Innovations: AIAI 2025*. Cham: Springer, 2025. p. 85–100. IFIP Advances in Information and Communication Technology, v. 755. DOI: 10.1007/978-3-031-96239-4_7.

KHOSRAVI, H.; SHUM, S. B.; CHEN, G.; CONATI, C.; TSAI, Y.-S.; KAY, J.; KNIGHT, S.; MARTINEZ-MALDONADO, R.; SADIQ, S.; GAŠEVIĆ, D. **Explainable artificial intelligence in education**. *Computers and Education: Artificial Intelligence*, v. 3, artigo 100074, 2022. DOI: 10.1016/j.caeai.2022.100074.

KUZILEK, J.; HLOSTA, M.; ZDRAHAL, Z. **Open University Learning Analytics dataset**. *Scientific Data*, v. 4, artigo 170171, 2017. DOI: 10.1038/sdata.2017.171.

LAUGEL, Thibault; LESOT, Marie-Jeanne; MARSALA, Christophe; RENARD, Xavier; DETYNIECKI, Marcin. **Inverse classification for comparison-based interpretability in machine learning**. In: *2017 IEEE International Conference on Information Reuse and Integration (IRI)*. San Diego: IEEE, 2017. p. 100–107. DOI: 10.1109/IRI.2017.44.

LUNDBERG, Scott M.; LEE, Su-In. A unified approach to interpreting model predictions. In: GUYON, Isabelle et al. (ed.). *Advances in Neural Information Processing Systems 30*. Red Hook: Curran Associates, 2017. p. 4765–4774.

MOJARAD, S.; ESSA, A.; MOJARAD, S.; BAKER, R. S. **Data-driven learner profiling based on clustering student behaviors: learning consistency, pace and effort**. In: NKAMBOU, R.; AZEVEDO, R.; VASSILEVA, J. (org.). *Intelligent Tutoring Systems*. Cham: Springer, 2018. p. 130–139. DOI: 10.1007/978-3-319-91464-0_13.

MOTHILAL, R. K.; SHARMA, A.; TAN, C. **Explaining machine learning classifiers through diverse counterfactual explanations**. In: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. New York: ACM, 2020. p. 607–617. DOI: 10.1145/3351095.3372850.

RIBEIRO, M. T.; SINGH, S.; GUESTIN, C. **“Why should I trust you?”: explaining the predictions of any classifier**. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM, 2016. p. 1135–1144. DOI: 10.1145/2939672.2939778.

ROMERO, C.; VENTURA, S. **Educational data mining and learning analytics: an updated survey.** *WIREs Data Mining and Knowledge Discovery*, v. 10, n. 3, artigo e1355, 2020. DOI: 10.1002/widm.1355.

ROUSSEEUW, P. J. **Silhouettes: a graphical aid to the interpretation and validation of cluster analysis.** *Journal of Computational and Applied Mathematics*, v. 20, p. 53-65, 1987. DOI: 10.1016/0377-0427(87)90125-7.

SIEMENS, G. **Learning analytics: the emergence of a discipline.** *American Behavioral Scientist*, v. 57, n. 10, p. 1380-1400, 2013. DOI: 10.1177/0002764213498851.

STREHL, Alexander; GHOSH, Joydeep. Cluster ensembles: a knowledge reuse framework for combining multiple partitions. **Journal of Machine Learning Research**, v. 3, p. 583-617, 2002.

USTUN, B.; SPANGHER, A.; LIU, Y. **Actionable recourse in linear classification.** In: *Proceedings of the Conference on Fairness, Accountability, and Transparency*. New York: ACM, 2019. p. 10-19. DOI: 10.1145/3287560.3287566.

VARDAKAS, G.; KARRA, A.; PITOURA, E.; LIKAS, A. **Counterfactual explanations for k-means and Gaussian clustering.** In: *2025 IEEE International Conference on Tools with Artificial Intelligence*. IEEE, 2025. p. 977-983. DOI: 10.1109/ICTAI66417.2025.00143.

WACHTER, S.; MITTELSTADT, B.; RUSSELL, C. **Counterfactual explanations without opening the black box: automated decisions and the GDPR.** *Harvard Journal of Law & Technology*, v. 31, n. 2, p. 841-887, 2018. DOI: 10.2139/ssrn.3063289.

WAHEED, H.; HASSAN, S.-U.; ALJOHANI, N. R.; HARDMAN, J.; ALELYANI, S.; NAWAZ, R. **Predicting academic performance of students from VLE big data using deep learning models.** *Computers in Human Behavior*, v. 104, artigo 106189, 2020. DOI: 10.1016/j.chb.2019.106189.