

Identificação de Perfis Comportamentais de Estudantes em um MOOC Brasileiro por Meio do Algoritmo *K-mean*

Identification of Behavioral Profiles of Students in a Brazilian MOOC Using the K-means Algorithm

Vanessa Faria de Souza*

Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul (IFRS) –
Campus Vacaria. Zona Rural – Vacaria – RS – Brasil.

*vanessa.souza@vacaria.ifrs.edu.br

Resumo

Os Massive Open Online Courses (MOOCs) ampliam o acesso à formação especializada, mas ainda apresentam padrões heterogêneos de participação e baixos índices de conclusão. Este estudo tem como objetivo identificar padrões de interação de estudantes em um MOOC da área de Química, ofertado por uma plataforma brasileira, e interpretar perfis comportamentais associados ao envolvimento e à certificação. Foram analisados dados secundários de 3.540 estudantes, provenientes de relatórios de perfil, progresso, notas e registros de acesso. Após a análise de correlação, a variável de acesso aos materiais foi retirada do agrupamento por redundância, enquanto a emissão de certificado foi reservada como variável externa. O algoritmo K-means foi aplicado aos atributos tempo de permanência, proporção de atividades concluídas e tentativas em questionários, todos padronizados. A escolha de $k = 2$ considerou o método do cotovelo, índices internos, estabilidade por reamostragem, equilíbrio dos grupos e interpretabilidade. Foram obtidos dois agrupamentos, de maior e menor envolvimento. O cruzamento desses agrupamentos com a emissão de certificado originou quatro perfis interpretativos: Engajados, Estratégicos, Inativos e Instrumentais. Os resultados diferenciam os agrupamentos produzidos pelo algoritmo dos perfis construídos posteriormente e oferecem subsídios para intervenções pedagógicas ajustadas aos padrões de participação observados.

Palavras-chave: Educação a distância. Mineração de dados educacionais. Engajamento discente. Conclusão de curso.

Recebido: 25/03/2026
Aceito: 31/08/2026
Publicado: 04/09/2026
Editores Responsáveis:
Daniel Salvador
Carmelita Portela
Daniela Samira

COMO CITAR ESTE TRABALHO

ABNT: SOUZA, V. F. Identificação de Perfis Comportamentais de Estudantes em um MOOC Brasileiro por Meio do Algoritmo *K-means*. *EaD em Foco*, v. 16, n.1, e2812, 2026. DOI: <https://doi.org/10.18264/eadf.v16i1.2812>

Identification of Behavioral Profiles of Students in a Brazilian MOOC Using the K-means Algorithm

Abstract

Massive Open Online Courses (MOOCs) expand access to specialized education, but still exhibit heterogeneous participation patterns and low completion rates. This study aims to identify interaction patterns among students in a Chemistry MOOC offered by a Brazilian platform and to interpret behavioral profiles associated with engagement and certification. Secondary data from 3,540 students were analyzed, obtained from profile, progress, grade, and access-log reports. After correlation analysis, material-access frequency was removed from clustering because of redundancy, while certificate issuance was retained as an external variable. K-means was applied to standardized attributes representing time in the course, proportion of completed activities, and questionnaire attempts. The choice of $k = 2$ considered the elbow method, internal validity indices, resampling stability, group balance, and interpretability. Two clusters were obtained, representing higher and lower engagement. Crossing these clusters with certificate issuance produced four interpretive profiles: Engaged, Strategic, Inactive, and Instrumental. The results distinguish algorithmically generated clusters from profiles interpreted afterward and provide evidence for pedagogical interventions tailored to observed participation patterns.

Keywords: Distance education. Educational data mining. Student engagement. Course completion.

1. Introdução

Os Massive Open Online Courses (MOOCs) ganharam projeção internacional pela possibilidade de ampliar o acesso a conteúdos produzidos por especialistas, com flexibilidade de tempo, ingresso facilitado e atendimento a públicos numerosos. Em 2018, mais de 900 universidades ofertavam aproximadamente 11.400 cursos nesse formato (Shah, 2018). A expansão não eliminou desafios recorrentes. Entre eles, destacam-se a baixa permanência e as taxas elevadas de não conclusão, que podem alcançar cerca de 90% em determinados contextos (Alraimi; Zo; Ciganeck, 2015).

A heterogeneidade do público ajuda a explicar parte dessa dinâmica. Os participantes podem buscar formação integral, atualização de um conteúdo específico, certificação, complementação acadêmica ou apenas uma aproximação inicial com o tema. Estudos baseados em registros de interação mostram que estudantes matriculados no mesmo curso percorrem trajetórias distintas, com diferenças na frequência de acesso, no uso de vídeos e materiais, na realização de avaliações e no tempo de permanência (Kizilcec; Piech; Schneider, 2013; Rodrigues *et al.*, 2016; Cagiltay; Toker; Cagiltay, 2024).

O agrupamento de dados tem sido utilizado para reconhecer padrões recorrentes nessas trajetórias. Ainda assim, a literatura apresenta duas limitações relevantes para esta pesquisa. A primeira é a reduzida quantidade de análises produzidas a partir de MOOCs brasileiros, sobretudo com dados detalhados de interação. A segunda está na forma como os resultados são interpretados: nem sempre se distingue o agrupamento estatístico gerado pelo algoritmo do perfil comportamental construído a partir da leitura pedagógica dos grupos. Também são pouco explicitados os efeitos da escolha das variáveis e do número de agrupamentos sobre as conclusões.

Diante dessa lacuna, formulou-se a seguinte questão: de que maneira os padrões de interação registrados em um MOOC brasileiro podem ser organizados em agrupamentos consistentes e, posteriormente, interpretados em perfis comportamentais relacionados ao envolvimento e à certificação? O objetivo consiste em identificar padrões de interação dos estudantes e interpretar perfis comportamentais a partir da combinação entre os agrupamentos produzidos pelo K-means e a emissão de certificado. Não se pretende

atribuir características psicológicas estáveis aos participantes, mas descrever comportamentos observados em um curso e período específicos.

A contribuição do estudo está na análise de uma base brasileira com 3.540 registros, na avaliação explícita da redundância entre atributos, na comparação de diferentes valores de *k* e na separação entre a etapa de agrupamento e a interpretação dos perfis. As Seções 2 e 3 apresentam os fundamentos conceituais e os trabalhos relacionados. A Seção 4 descreve os dados e o procedimento analítico; a Seção 5 discute os resultados; e a Seção 6 reúne as considerações finais.

2. Perfis Comportamentais e Mapeamento de Estudantes

A expressão perfil de estudante é utilizada em estudos com finalidades distintas. Modelos clássicos, como os de Kolb (1984), Honey e Mumford (1982) e Felder e Silverman (1988), procuram representar preferências ou modos de aprendizagem. Embora sejam relevantes para a história da área, esses modelos não equivalem aos perfis comportamentais derivados de registros digitais. Estilos de aprendizagem remetem a construções pedagógicas ou cognitivas; perfis comportamentais, no sentido adotado neste artigo, correspondem a padrões contextuais de participação observáveis na plataforma.

Essa distinção evita interpretar os agrupamentos como propriedades permanentes dos sujeitos. O fato de um estudante apresentar pouca interação em determinado MOOC não permite concluir que ele se comportará da mesma forma em outro curso. O padrão pode estar relacionado ao tema, ao desenho pedagógico, ao tempo disponível, à familiaridade prévia com o conteúdo ou ao objetivo específico que motivou a matrícula. Assim, o perfil descreve uma configuração de ações registrada em condições delimitadas.

A modelagem de perfis pode integrar informações pessoais, acadêmicas, cognitivas e comportamentais, organizadas por meio de ontologias, redes probabilísticas, portfólios eletrônicos ou técnicas de aprendizagem de máquina (Ferreira-Satler *et al.*, 2012; Hamim; Benabbou; Sael, 2019). Em MOOCs, os registros de acesso permitem examinar a frequência e a sequência das interações, o progresso, o desempenho e o tempo de permanência. Essas variáveis podem apoiar recomendações, identificação de risco e planejamento de intervenções, desde que sejam analisadas sem confundir associação com causalidade.

Entre as técnicas não supervisionadas, o *K-means* organiza os registros em *k* grupos, buscando reduzir as distâncias entre cada observação e o centro do grupo ao qual foi atribuída. O método é simples e escalável, mas exige variáveis numéricas, é sensível à escala, a valores extremos e à inicialização, além de pressupor a escolha prévia de *k* (Jain, 2010). Por isso, o número de agrupamentos não deve ser definido apenas pela aparência de um gráfico. Medidas de coesão e separação, estabilidade por reamostragem, tamanho dos grupos e interpretabilidade substantiva precisam ser avaliados em conjunto. O coeficiente de silhueta, por exemplo, compara a proximidade de cada observação com seu próprio grupo e com o grupo vizinho (Rousseeuw, 1987).

3. Trabalhos Relacionados

Os estudos de agrupamento em MOOCs utilizam com maior frequência variáveis relacionadas ao acesso aos recursos, ao progresso, ao desempenho e à duração da participação. Kizilcec, Piech e Schneider (2013) mostraram que a não conclusão não corresponde a um único comportamento e identificaram trajetórias distintas de engajamento. Rodrigues *et al.* (2016), em um MOOC da plataforma Openredu, empregaram *K-means* e reconheceram padrões de estudantes engajados, esporádicos e desengajados. Tan *et al.* (2018) combinaram dados comportamentais, autorrelatos e resultados acadêmicos para caracterizar experiências de aprendizagem em um MOOC de língua inglesa.

Outras pesquisas ampliaram o conjunto de técnicas e escalas de análise. Mareca e Bordel (2019) examinaram perfis na plataforma MiriadaX, enquanto Luna *et al.* (2022) aplicaram descoberta de subgrupos a dados de diferentes MOOCs, buscando regras interpretáveis e não redundantes. Estudos recentes analisaram milhões de matrículas para relacionar características demográficas, recorrência de inscrição e resultados (Cagiltay; Toker; Cagiltay, 2024), além de investigar a influência de características demográficas no engajamento em cursos brasileiros da plataforma Aprenda Mais (Silva *et al.*, 2024). Liu *et al.* (2025) propuseram uma representação que articula dimensões emocionais, comportamentais e de habilidade.

Apesar da diversidade metodológica, parte dos trabalhos concentra-se na descrição dos grupos encontrados e apresenta pouca informação sobre a seleção de atributos, a correlação entre variáveis ou a estabilidade das soluções. Há também diferenças na forma de tratar a conclusão: em alguns casos, ela participa diretamente do agrupamento; em outros, é utilizada para interpretar ou validar grupos formados por comportamentos de interação. Essa decisão é relevante, pois a inclusão da certificação pode fazer o algoritmo separar concluintes e não concluintes antes mesmo que outros padrões sejam examinados.

No contexto brasileiro, a produção permanece menos numerosa e, em geral, está vinculada a plataformas ou cursos específicos. A lacuna não está apenas em aplicar novamente um algoritmo conhecido, mas em tornar explícitas as escolhas que sustentam a interpretação. Este estudo responde a esse ponto ao retirar a emissão de certificado do K-means, avaliar a redundância das variáveis, comparar $k = 2$, $k = 3$ e $k = 4$ e utilizar a certificação somente depois do agrupamento. Desse modo, os perfis resultam de uma etapa interpretativa claramente separada do procedimento estatístico.

4. Metodologia

Trata-se de uma pesquisa quantitativa, exploratória e observacional, baseada em dados secundários gerados durante a utilização de um MOOC. O procedimento foi organizado em quatro etapas: caracterização do curso e da amostra; integração e preparação dos relatórios; seleção dos atributos; e aplicação e validação do agrupamento.

4.1 Contexto, Participantes e Aspectos Éticos

O curso investigado pertence à área de Química, possui carga horária de 15 horas e é ofertado gratuitamente por uma plataforma brasileira de cursos abertos on-line. A identificação da plataforma foi omitida por solicitação da instituição responsável. O MOOC é composto por um módulo introdutório e quatro módulos de conteúdo, com videoaulas, materiais textuais e questionários de múltipla escolha. Ao final, os participantes podem acessar recursos complementares e emitir o certificado quando cumprem os critérios estabelecidos.

Foram analisados 3.540 registros de estudantes que responderam ao questionário de perfil e possuíam informações correspondentes nos relatórios utilizados. Os dados abrangem interações realizadas entre junho de 2019 e agosto de 2020. Como o curso não estabelece prazo fixo de conclusão, a emissão do certificado foi observada até a data de extração. Assim, a ausência de certificado indica que o participante não o havia emitido naquele momento, sem permitir afirmar que abandonou definitivamente o curso.

A pesquisa não foi submetida a Comitê de Ética em Pesquisa. Durante o cadastro, os participantes concordaram com o termo de uso e de autorização para utilização dos dados para fins de pesquisa. Os identificadores disponibilizados aos pesquisadores estavam anonimizados, e os resultados foram tratados de forma agregada. O nome da plataforma e qualquer informação capaz de identificá-la foram preservados conforme a condição estabelecida pela instituição responsável.

4.2. Fonte e Organização dos Dados

A base consolidada foi construída a partir de quatro relatórios: Perfil do Estudante, Conclusão de Atividades, Notas e Logs. Os relatórios foram integrados por identificadores anonimizados. Os registros de acesso foram agregados por estudante e por tipo de recurso. O Quadro 1 apresenta as variáveis consideradas e sua função na análise.

Quadro 1 - Variáveis da base e função no procedimento analítico.

Variável	Definição operacional	Origem	Função analítica
Gênero	Categoria informada no questionário de perfil.	Perfil	Caracterização externa
Idade	Faixa etária informada pelo estudante.	Perfil	Caracterização externa
Escolaridade	Nível de escolaridade informado.	Perfil	Caracterização externa
Motivação	Curiosidade geral, importância para o trabalho ou para os estudos.	Perfil	Caracterização externa
Intenção	Intenção declarada de concluir o curso.	Perfil	Caracterização externa
Nota final	Soma das notas dos quatro questionários, variando de 0 a 40.	Notas	Caracterização externa
Tempo	Diferença, em dias, entre o primeiro e o último registro de atividade.	Progresso	K-means
Proporção concluída	Razão entre atividades concluídas e atividades disponíveis. Varia de 0 a 1; 0,50 corresponde a 50%.	Progresso	K-means
Acessos a textos	Quantidade de acessos aos materiais textuais.	Logs	Caracterização externa
Acessos a vídeos	Quantidade de acessos às videoaulas.	Logs	Caracterização externa
Tentativas em questionários	Quantidade de acessos às tentativas dos quatro questionários.	Logs	K-means
Acessos aos materiais	Soma dos acessos a textos e videoaulas.	Logs	Análise de redundância e caracterização externa
Certificado	Variável binária: 1 para certificado emitido e 0 para não emitido até a extração.	Progresso	Validação externa e interpretação dos perfis

Fonte: elaborado pela autora.

4.3. Pré-processamento e Seleção dos Atributos

Os registros ausentes de tempo foram convertidos em zero quando não havia intervalo entre atividades identificável. A proporção concluída foi mantida na escala de 0 a 1. As contagens de acesso e o tempo foram examinados quanto à distribuição e a valores extremos. Antes do agrupamento, os atributos foram transformados em escores padronizados, de modo que diferenças de escala não determinassem as distâncias.

A seleção das variáveis considerou pertinência conceitual e redundância estatística. Entre os atributos inicialmente avaliados, a proporção de atividades concluídas e o total de acessos aos materiais apresentaram correlação de Pearson de $r = 0,912$. Como ambos descreviam dimensões próximas do avanço no curso, o total de acessos foi retirado do K-means e preservado para caracterização posterior. Uma análise de sensibilidade mostrou que a exclusão dessa variável manteve elevada concordância com a solução completa (índice Rand ajustado = 0,960).

A emissão de certificado também foi excluída do agrupamento. Sua inclusão separava diretamente concluintes e não concluintes e poderia reduzir a capacidade de observar padrões de interação. Gênero, idade, escolaridade, motivação e intenção não foram convertidos arbitrariamente em escalas numéricas. Essas variáveis foram utilizadas somente na descrição posterior. A matriz final do K-means reuniu tempo, proporção concluída e tentativas em questionários.

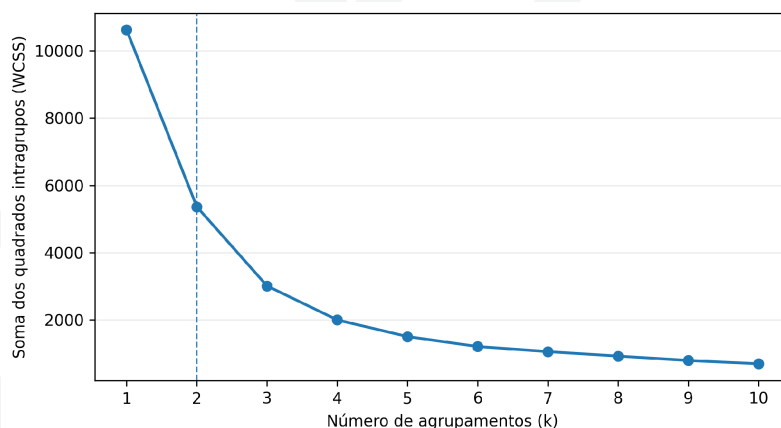
4.3. Aplicação e Validação do *K-means*

O K-means inicia com k centros e atribui cada observação ao centro mais próximo pela distância Euclidiana. Em seguida, os centros são recalculados e o processo é repetido até que as atribuições se estabilizem. Para reduzir a influência da inicialização, foram realizadas 100 inicializações aleatórias para cada valor de k .

Foram comparadas as soluções $k = 2$, $k = 3$ e $k = 4$. A avaliação reuniu a curva do método do cotovelo, o coeficiente médio de silhueta, o índice Davies-Bouldin, a proporção do menor grupo e a estabilidade por reamostragem. A estabilidade foi estimada em 100 amostras contendo 80% dos registros, comparando-se as partições pelo índice Rand ajustado, cuja escala varia até 1 e assume valores mais altos quando as classificações são semelhantes (Hubert; Arabie, 1985).

A escolha final não foi baseada em um único indicador. Foram considerados, simultaneamente, a redução da variabilidade intragrupos, a separação, a estabilidade, o equilíbrio entre os agrupamentos e a possibilidade de interpretação coerente com os comportamentos registrados. Após a definição dos grupos, a emissão de certificado foi cruzada com a solução para construir os perfis interpretativos.

Figura 1 - Curva do método do cotovelo para $k = 1$ a 10.



Fonte: elaborado pela autora.

Os dados individuais não podem ser disponibilizados publicamente devido às condições de uso estabelecidas pela plataforma. Informações adicionais sobre os procedimentos e resultados agregados poderão ser solicitadas aos autores, condicionadas às restrições institucionais.

5. Resultados

Trata-se de uma pesquisa quantitativa, exploratória e observacional, baseada em dados secundários gerados durante a utilização de um MOOC. O procedimento foi organizado em quatro etapas: caracterização do curso e da amostra; integração e preparação dos relatórios; seleção dos atributos; e aplicação e validação do agrupamento.

5.1 Caracterização da Amostra

A amostra apresentou ampla diversidade etária e educacional. A faixa de até 19 anos reuniu 38,8% dos registros, seguida por 20 a 24 anos, com 25,5%. O ensino superior incompleto foi a escolaridade mais frequente, com 36,5%. A principal motivação declarada foi a importância do conteúdo para os estudos, indicada por 46,5% dos participantes. Embora 96,0% afirmassem intenção de concluir o curso, 24,7% haviam emitido certificado até a data de extração. A diferença entre intenção e certificação reforça que a matrícula e a declaração inicial não são suficientes para caracterizar a trajetória efetivamente realizada.

Tabela 1 - Descrição da amostra analisada.

Dimensão	Categoria	n
Gênero	Feminino	2.074
	Masculino	1.457
	Outro	9
Idade	Até 19 anos	1.375
	20 a 24 anos	903
	25 a 29 anos	417
	30 a 34 anos	242
	35 a 39 anos	209
	40 a 49 anos	244
	Acima de 50 anos	150
Escolaridade	Ensino fundamental incompleto	385
	Ensino fundamental completo	69
	Ensino médio incompleto	438
	Ensino médio completo	373
	Ensino superior incompleto	1.292
	Ensino superior completo	435
	Pós-graduação	548
Motivação	Curiosidade geral	1.474
	Importante para o trabalho	421
	Importante para os estudos	1.645
Intenção	Não	6
	Não sei	135
	Sim	3.399
Nota final	0	1.602
	1 a 10	387
	11 a 20	116
	21 a 30	181
	31 a 40	1.254
Certificado	Emitido	876
	Não emitido	2.664
Total		3.540

Fonte: elaborado pela autora.

5.1 Seleção e Caracterização dos Agrupamentos

A curva apresentada na Figura 1 mostra redução acentuada da soma dos quadrados intragrupos até $k = 2$ e ganhos progressivamente menores nos valores seguintes. Os índices internos, apresentados na Tabela 2, não apontaram uma solução única. A silhueta e o Davies-Bouldin favoreceram valores maiores de k ; entretanto, $k = 3$ e $k = 4$ produziram grupos com apenas 55 e 52 estudantes, respectivamente, equivalentes a aproximadamente 1,5% da amostra. Esses grupos eram formados principalmente por participantes com tempo de permanência muito elevado, sem configurar um padrão pedagógico suficientemente distinto para sustentar uma categoria própria.

A solução $k = 2$ apresentou os grupos mais equilibrados, estabilidade média de 0,999 e a interpretação mais parcimoniosa. A diferença entre a silhueta de $k = 2$ e as soluções seguintes foi pequena diante da fragmentação introduzida por $k = 3$ e $k = 4$. Optou-se, portanto, por dois agrupamentos, em concordância com o método do cotovelo e com a finalidade de distinguir níveis gerais de envolvimento antes da utilização da certificação como variável externa.

Tabela 2* - Indicadores utilizados na comparação do número de agrupamentos.

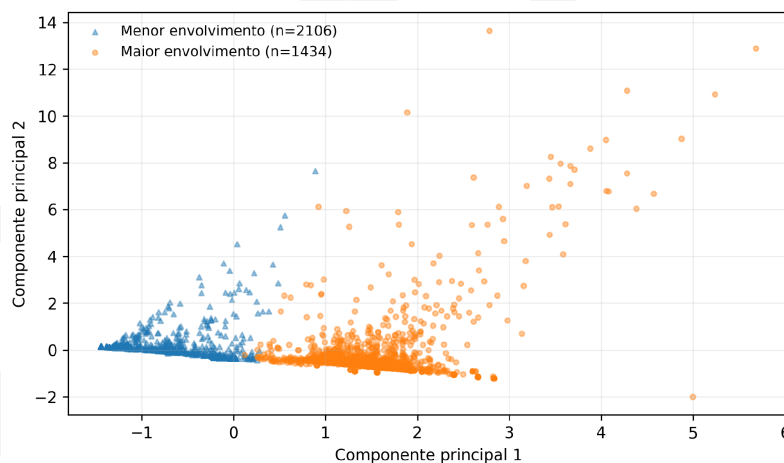
k	WCSS	Silhueta	Davies-Bouldin	Menor grupo	Estabilidade (ARI)
2	5.364,35	0,611	0,689	40,5%	0,999
3	3.011,37	0,636	0,540	1,6%	0,996
4	2.000,14	0,665	0,495	1,5%	0,997

Fonte: elaborado pela autora.

*Valores maiores de silhueta e ARI e menores de Davies-Bouldin são preferíveis; os indicadores foram interpretados em conjunto.

O agrupamento de maior envolvimento reuniu 1.434 estudantes. Em média, foram concluídas 93,1% das atividades, registradas 44,4 tentativas em questionários e 14,6 dias entre o primeiro e o último acesso. O agrupamento de menor envolvimento reuniu 2.106 estudantes, com 8,8% das atividades concluídas, 9,8 tentativas e permanência média de 2,5 dias. Embora os acessos aos materiais não tenham participado do K-means, a média foi de 64,8 no primeiro grupo e 11,5 no segundo, o que confirma externamente a diferença de intensidade entre as trajetórias.

Figura 2* - Projeção dos dois agrupamentos nas duas primeiras componentes principais.



Fonte: elaborado pela autora.

*As duas componentes explicam 90,2% da variabilidade dos três atributos utilizados no K-means.

5.3. Interpretação dos Perfis Comportamentais

Os dois agrupamentos descrevem níveis gerais de envolvimento, mas não equivalem aos quatro perfis. A interpretação foi realizada em uma etapa posterior, pelo cruzamento entre o agrupamento e a emissão do certificado. Dessa combinação resultaram: Engajados, Estratégicos, Inativos e Instrumentais. A Tabela 3 apresenta a regra de classificação e os principais indicadores.

Tabela 3 - Caracterização dos perfis interpretativos.

Perfil	Regra	n	Tempo (dias)	Conclusão	Tentativas	Acessos a materiais	Nota final
Engajados	Maior envolvimento + certificado	855	13,3	97%	44,0	65,2	36,6
Estratégicos	Maior envolvimento + sem certificado	579	16,6	87%	45,0	64,2	33,4
Inativos	Menor envolvimento + sem certificado	2.085	2,5	9%	9,5	11,5	2,5
Instrumentais	Menor envolvimento + certificado	21	0,7	27%	32,4	5,5	33,0

Fonte: elaborado pela autora.

Os Engajados correspondem a participantes que exploraram amplamente os recursos, concluíram quase todas as atividades e emitiram o certificado. O desempenho médio de 36,6 pontos, em uma escala de 0 a 40, indica participação acompanhada de aproveitamento elevado. Esse perfil se aproxima dos grupos de maior engajamento identificados por Rodrigues *et al.* (2016), mas a certificação foi utilizada apenas para interpretação externa.

Os Estratégicos apresentaram intensidade semelhante à dos Engajados, com média de 45 tentativas e 64,2 acessos aos materiais, porém não emitiram certificado até a extração. A denominação não pressupõe abandono: os estudantes podem ter buscado conteúdos específicos, considerado suficiente o aprendizado alcançado ou concluído o curso posteriormente. A permanência média mais longa reforça que a ausência de certificado não equivale automaticamente a desinteresse.

Os Inativos constituíram o perfil mais numeroso. A baixa proporção de atividades concluídas, o pequeno número de tentativas e o tempo reduzido indicam contato inicial seguido de pouca continuidade. Entre eles, 96,2% haviam declarado intenção de concluir o MOOC, mostrando que a intenção autorrelatada não se converte necessariamente em ação sustentada. Esse contraste pode orientar notificações precoces, apoio à organização do estudo e revisão dos recursos iniciais.

Os Instrumentais reuniram apenas 21 participantes e devem ser interpretados com cautela. Eles emitiram certificado em menos de um dia, realizaram em média 32,4 tentativas nos questionários e acessaram poucos materiais. O padrão sugere concentração nas avaliações e uso pragmático do curso para certificação. Por se tratar de um subgrupo pequeno, sua recorrência precisa ser examinada em outros MOOCs antes de qualquer generalização.

Os resultados demonstram que a conclusão isolada não descreve adequadamente o comportamento. Engajados e Instrumentais emitiram certificado, mas percorreram trajetórias muito diferentes. Da mesma forma, Estratégicos e Inativos não o emitiram, embora apresentem níveis opostos de envolvimento. Essa distinção reforça a utilidade de combinar registros de interação e desfechos externos, sem permitir que o desfecho determine previamente os grupos.

6. Conclusões

O estudo identificou dois padrões gerais de interação em um MOOC brasileiro: maior e menor envolvimento. A seleção de $k = 2$ foi sustentada pelo método do cotovelo, pela estabilidade, pelo equilíbrio dos grupos e pela interpretação substantiva, mesmo que índices internos isolados apresentassem valores ligeiramente superiores para soluções mais fragmentadas. A análise também mostrou que a emissão de certificado não deve compor o agrupamento quando se pretende examinar o comportamento que antecede esse desfecho.

O cruzamento posterior dos dois agrupamentos com a certificação permitiu interpretar quatro perfis: Engajados, Estratégicos, Inativos e Instrumentais. A principal contribuição está em explicitar que esses perfis não são quatro clusters produzidos pelo K-means. São categorias analíticas construídas a partir de dois grupos comportamentais e de uma variável externa. Essa separação reduz a inconsistência entre o procedimento estatístico e a interpretação pedagógica.

Do ponto de vista aplicado, estudantes com menor envolvimento podem receber apoio nas etapas iniciais, lembretes e orientações de estudo. Para os Estratégicos, trilhas flexíveis e certificações parciais podem reconhecer objetivos de aprendizagem específicos. O padrão Instrumental sugere a necessidade de revisar critérios de certificação e diversificar as atividades avaliativas, evitando que tentativas repetidas substituam o contato com os conteúdos. Ações dessa natureza devem ser testadas antes de sua adoção em larga escala.

As limitações incluem a análise de um único curso, a ausência de prazo fixo para conclusão e a impossibilidade de verificar se estudantes sem certificado o emitiram após a extração. Os perfis são dependentes do contexto e não devem ser tratados como características permanentes. Estudos futuros podem replicar o procedimento em outras áreas, acompanhar trajetórias ao longo do tempo e avaliar se intervenções orientadas pelos perfis produzem melhora no envolvimento e na aprendizagem.

Biodados e contatos dos autores



SOUZA, V. F. de. é doutora em Informática na Educação pelo Programa de Pós-Graduação em Informática na Educação (PPGIE) da Universidade Federal do Rio Grande do Sul (UFRGS). Mestre em Informática pelo Programa de Pós-Graduação em Informática (PPGI) da Universidade Tecnológica Federal do Paraná (UTFPR), na área de Computação Aplicada, com ênfase em Engenharia de Software. Possui Especialização em Educação Especial Inclusiva, com ênfase em Tecnologias Assistivas, pela Universidade Estadual do Norte do Paraná (UENP). É bacharela em Sistemas de Informação e licenciada em Computação pela UENP, formações realizadas concomitantemente. Também é licenciada em Matemática por meio do Programa de Formação Pedagógica da UTFPR. Atua no Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul (IFRS) – Campus Vacaria. Seus interesses de pesquisa incluem Inteligência Artificial na Educação, Mineração de Dados Educacionais, Educação a Distância e Tecnologias Educacionais.

ORCID: <https://orcid.org/0000-0002-8804-2680>

E-mail: vanessa.souza@vacaria.ifrs.edu.br

Agradecimentos

Agradeço ao Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul, por ter oportunizado a realização desta pesquisa.

Referências Bibliográficas

- ALRAIMI, K. M.; ZO, H.; CIGANEK, A. P. Understanding the MOOCs continuance: the role of openness and reputation. **Computers & Education**, v. 80, p. 28–38, 2015. DOI: <https://doi.org/10.1016/j.compedu.2014.08.006>
- CAGILTAY, N. E.; TOKER, S.; CAGILTAY, K. Exploring MOOC learners' behavioural patterns considering age, gender and number of course enrolments: insights for improving educational opportunities. **Open Praxis**, v. 16, n. 1, p. 70–81, 2024. DOI: <https://doi.org/10.55982/openpraxis.16.1.543>
- FELDER, R. M.; SILVERMAN, L. K. Learning and teaching styles in engineering education. **Engineering Education**, v. 78, n. 7, p. 674–681, 1988.
- FERREIRA-SATLER, M. *et al.* Fuzzy ontologies-based user profiles applied to enhance e-learning activities. **Soft Computing**, v. 16, n. 7, p. 1129–1141, 2012. DOI: <https://doi.org/10.1007/s00500-011-0788-y>
- HAMIM, T.; BENABBOU, F.; SAEL, N. Student profile modeling: an overview model. In: INTERNATIONAL CONFERENCE ON SMART CITY APPLICATIONS, 4., 2019. **Proceedings [...]**. New York: ACM, 2019. p. 1–9. DOI: <https://doi.org/10.1145/3368756.3369075>
- HONEY, P.; MUMFORD, A. **The manual of learning styles**. Maidenhead: Peter Honey, 1982.
- HUBERT, L.; ARABIE, P. Comparing partitions. **Journal of Classification**, v. 2, n. 1, p. 193–218, 1985. DOI: <https://doi.org/10.1007/BF01908075>
- JAIN, A. K. Data clustering: 50 years beyond K-means. **Pattern Recognition Letters**, v. 31, n. 8, p. 651–666, 2010. DOI: <https://doi.org/10.1016/j.patrec.2009.09.011>
- KIZILCEC, R. F.; PIECH, C.; SCHNEIDER, E. Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In: INTERNATIONAL CONFERENCE ON LEARNING ANALYTICS AND KNOWLEDGE, 3., 2013, Leuven. **Proceedings [...]**. New York: ACM, 2013. p. 170–179. DOI: <https://doi.org/10.1145/2460296.2460330>
- KOLB, D. A. **Experiential learning: experience as the source of learning and development**. Englewood Cliffs: Prentice Hall, 1984.
- LIU, J.; *et al.* Research on learner “emotion-behavior-ability” characteristics based on MOOC online education user profiles. **Information Processing & Management**, v. 62, n. 3, art. 104026, 2025. DOI: <https://doi.org/10.1016/j.ipm.2024.104026>
- LUNA, J. M.; *et al.* Subgroup discovery in MOOCs: a big data application for describing different types of learners. **Interactive Learning Environments**, v. 30, n. 1, p. 127–145, 2022. DOI: <https://doi.org/10.1080/10494820.2019.1643742>
- MARECA, P.; BORDEL, B. Students profiles and their behavior in MOOC platforms: MiriadaX platform. In: IBERIAN CONFERENCE ON INFORMATION SYSTEMS AND TECHNOLOGIES, 14., 2019, Coimbra. **Proceedings [...]**. Piscataway: IEEE, 2019. p. 1–6. DOI: <https://doi.org/10.23919/CISTI.2019.8760696>
- RODRIGUES, R. L.; *et al.* Discovery engagement patterns MOOCs through cluster analysis. **IEEE Latin America Transactions**, v. 14, n. 9, p. 4129–4135, 2016. DOI: <https://doi.org/10.1109/TLA.2016.7785943>
- 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: [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- SHAH, D. **By the numbers: MOOCs in 2018**. Class Central, 2018. Disponível em: <https://www.classcentral.com/report/mooc-stats-2018/>. Acesso em: 03 jan 2026.

SILVA, J. M. C. da; *et al.* Learner engagement and demographic influences in Brazilian massive open online courses: Aprenda Mais platform case study. **Analytics**, v. 3, n. 2, p. 178–193, 2024. DOI: 10.3390/analytics3020010

TAN, Y.; *et al.* Learning profiles, behaviors and outcomes: investigating international students' learning experience in an English MOOC. In: INTERNATIONAL SYMPOSIUM ON EDUCATIONAL TECHNOLOGY, 2018. **Proceedings [...]**. Piscataway: IEEE, 2018. p. 214–218. DOI: <https://doi.org/10.1109/ISET.2018.00055>

PRELIMINARY

