

# **Identificação de Perfis Motivacionais em Academias de Ginástica: Um Estudo Baseado em Clusterização e HEXAD**

**Daniel M. Pedrozo<sup>1</sup>, Lucas B. Rodrigues<sup>1</sup>, Gabriel Orlow<sup>1</sup>,  
Murilo de Oliveira<sup>1</sup>, Julia Moreira<sup>1</sup>, Leonardo A. Alves<sup>1</sup>,  
Marcos Vinicios P. de Moraes<sup>1</sup>, Paulo César F. Melo<sup>1</sup>**

<sup>1</sup>Instituto de Informática – Universidade Federal de Goiás (UFG)  
Caixa Postal 131 – 74.690-900 – Goiânia – GO – Brazil

{danielpedrozo, marcos.vinicios2}@discente.ufg.br

{gabrielorlow, muriloguimaraes}@discente.ufg.br

julia.moreira@egresso.ufg.br

{lucasbranda04770, pcfm.inf}@gmail.com

leoalvs@ufg.br

**Abstract.** High dropout rates in gyms are often linked to a mismatch between individual motivations and generic engagement strategies. This paper proposes a personalized gamification approach to improve member retention, using anonymized data analyzed with unsupervised machine learning. We identify user clusters with distinct training frequency, intensity, and goals, and associate them with player profiles from the HEXAD model. The results highlight the potential of data-driven segmentation to support more motivating experiences and reinforce the role of personalization in reducing dropout.

**Keywords** Gamification, Clustering, Gyms, HEXAD, K-Means, User Retention, Personalization, Fitness Technology

**Resumo.** A evasão em academias está frequentemente relacionada ao desalinhamento entre motivações pessoais e estratégias de engajamento. Este trabalho propõe uma abordagem de gamificação personalizada para melhorar a retenção de alunos, utilizando dados anonimizados analisados por técnicas de aprendizado de máquina não supervisionado. Foram identificados agrupamentos de alunos com diferenças em frequência, intensidade e objetivos de treino, associados a perfis de jogadores do modelo HEXAD. Os resultados apontam para o potencial da segmentação baseada em dados para promover experiências mais motivadoras, destacando a importância da personalização no combate à evasão.

**Palavras-Chave** Gamificação, Clusterização, Academias de Ginástica, HEXAD, K-Means, Retenção de Usuários, Personalização

## **1. Introdução**

A retenção de alunos é um desafio recorrente na gestão de academias de ginástica<sup>1</sup>. Muitos alunos abandonam a academia nos primeiros meses, apesar de investimentos em

---

<sup>1</sup>Neste trabalho, o termo “academias” é utilizado para se referir às academias de ginástica e musculação.

infraestrutura e campanhas de engajamento genéricas que tendem a ignorar diferenças de motivação entre usuários [Liz e Andrade 2016].

A gamificação tem sido proposta como forma de aumentar o engajamento em contextos de saúde e bem-estar. Seu potencial é maior quando personalizada em comparação com abordagens *one-size-fits-all* [Altmeyer et al. 2021, Zhao et al. 2020, Yang et al. 2023]. Entre os *frameworks* disponíveis, o modelo *HEXAD* [Tondello et al. 2016] se destaca por ter sido desenvolvido para sistemas gamificados aplicados, combinando motivações intrínsecas e extrínsecas em seis tipos de perfis. A hipótese central é que o alinhamento entre elementos de gamificação e o perfil motivacional de cada usuário pode melhorar o engajamento e a retenção.

Na prática, entretanto, perfis *HEXAD* são usualmente obtidos por questionários específicos, o que é difícil de escalar em ambientes operacionais como academias, onde o fluxo de matrícula precisa ser simples e de baixa fricção. Isso motiva a busca por alternativas de baixo custo operacional que utilizem apenas dados já coletados rotineiramente, sem exigir instrumentos adicionais de avaliação.

Este artigo investiga se a clusterização de variáveis comportamentais e demográficas coletadas na matrícula pode servir como um *proxy* prático para segmentação motivacional. Neste contexto, utilizando dados reais e anonimizados de academias parceiras, propomos um fluxo que: (i) realiza pré-processamento e engenharia de atributos; (ii) aplica e compara algoritmos de aprendizado de máquina não supervisionado para identificar agrupamentos coesos; e (iii) interpreta qualitativamente os *clusters* à luz do modelo *HEXAD*, formulando hipóteses de segmentos motivacionais.

As principais contribuições deste trabalho são três:

- Um procedimento transparente de pré-processamento, engenharia de atributos e clusterização baseado apenas em variáveis de cadastro, sem aplicação de questionários adicionais.
- Um mapeamento teórico, explicitamente exploratório, entre *clusters* comportamentais e perfis do modelo *HEXAD*, oferecendo uma base operacional para estratégias de gamificação personalizadas em academias.
- Uma discussão de limitações, riscos éticos e uma agenda de pesquisa futura, incluindo testes de estabilidade, coleta de rótulos motivacionais fracos e avaliação de possíveis vieses demográficos.

O artigo está organizado da seguinte forma: a Seção 2 apresenta a fundamentação teórica; a Seção 3, a metodologia; a Seção 4, os resultados; a Seção 5, a discussão e limitações; e a Seção 6 traz as conclusões.

## 2. Fundamentação Teórica

A gamificação tem sido amplamente empregada em saúde e bem-estar para aumentar engajamento e adesão. Evidências indicam que abordagens genéricas tendem a produzir efeitos neutros ou de curta duração, enquanto estratégias alinhadas a características individuais apresentam maior potencial de impacto [Altmeyer et al. 2021, Zhao et al. 2020, Yang et al. 2023]. Em academias, essa sensibilidade às diferenças

motivacionais é relevante dada a diversidade de objetivos (emagrecimento, hipertrofia) e hábitos de treino.

Diversos *frameworks* foram propostos para classificar perfis de jogadores, como Bartle [Bartle 1996], Yee [Yee 2006] e *BrainHex* [Nacke et al. 2014]. Embora influentes, esses modelos foram desenhados para jogos digitais de entretenimento puro, o que dificulta sua transferência direta para cenários aplicados, como plataformas de saúde ou academias, focados em comportamentos de longo prazo e objetivos de desempenho concretos.

O modelo *HEXAD*, proposto por Marczewski e validado por Tondello et al. [Tondello et al. 2016], foi desenvolvido especificamente para contextos gamificados que extrapolam o entretenimento. Baseado na Teoria da Autodeterminação (*Self-Determination Theory - SDT*) [Ryan e Deci 2000], define seis perfis motivacionais — *Philanthropist*, *Socialiser*, *Free Spirit*, *Achiever*, *Player* e *Disruptor* — combinando motivações intrínsecas e extrínsecas. Trabalhos como os de Altmeyer et al. [Altmeyer et al. 2021] e Zhao et al. [Zhao et al. 2020] mostram que a personalização de sistemas gamificados com base no *HEXAD* pode melhorar engajamento e satisfação.

Em geral, no entanto, essas aplicações dependem de questionários específicos para estimar o perfil motivacional, o que nem sempre é viável em ambientes de alta rotatividade. Uma linha complementar explora o uso de técnicas de clusterização para inferência de perfis a partir de dados observacionais, com exemplos em análise de clientes em academias [Niu 2025], classificação de perfis em jogos digitais [Bicalho et al. 2019] e plataformas *fitness* gamificadas [Krath et al. 2022]. Contudo, esses estudos raramente conectam explicitamente os segmentos encontrados a modelos motivacionais como o *HEXAD*, especialmente em ambientes físicos de treino.

Dessa forma, permanecem duas lacunas principais: (i) poucos trabalhos combinam clusterização de dados de academias presenciais com uma interpretação motivacional ancorada no *HEXAD*; e (ii) há demanda por soluções de baixo custo operacional que usem apenas dados rotineiramente coletados (idade, sexo, frequência e duração do treino, objetivos declarados), sem questionários extensos ou instrumentação adicional. É nesse espaço que este estudo contribui, aplicando o *K-Means* a registros administrativos de alunos para identificar agrupamentos comportamentais coesos e interpretá-los qualitativamente à luz dos perfis do *HEXAD*.

### 3. Metodologia

Adotamos uma abordagem de aprendizado de máquina não supervisionado estruturada em quatro etapas: (i) definição do conjunto de dados; (ii) pré-processamento e engenharia de atributos; (iii) seleção do modelo e clusterização; e (iv) interpretação motivacional qualitativa à luz do modelo *HEXAD*<sup>2</sup>.

#### 3.1. Conjunto de Dados

O conjunto de dados é composto por 11.388 registros de alunos de academias parceiras da empresa Pacto Soluções Tecnológicas, coletados entre outubro de 2024 e julho de

<sup>2</sup>Código-fonte e scripts estão disponíveis em: [https://github.com/CEIA-PACTO/fitness\\_cluster](https://github.com/CEIA-PACTO/fitness_cluster).

2025. O uso dos dados para fins de pesquisa segue acordo formal de cooperação. O processamento atende à Lei Geral de Proteção de Dados Pessoais (LGPD – Lei nº 13.709/2018), baseando-se no tratamento de dados anonimizados para fins científicos (art. 7º, inciso IV e art. 11, inciso II).

A anonimização foi realizada previamente pela empresa, com remoção de identificadores diretos antes do acesso pela equipe de pesquisa. Não foram utilizados dados sensíveis no sentido legal. Os dados foram armazenados em repositórios seguros, com acesso restrito e documentação dos procedimentos de extração e análise.

O *dataset* original contém 15 atributos brutos organizados em três categorias (demografia/antropometria, comportamento de treino e metadados operacionais). A Tabela 1 resume essas variáveis.

**Tabela 1. Resumo descritivo das variáveis utilizadas na análise**

Categoría	Descrição das Variáveis
<b>Demografia/Antropometria</b>	
Idade	14 a 68 anos; média = $32,4 \pm 9,2$ anos
Altura	142 a 210 cm
Peso	45 a 150 kg; 57,3% ausente
Sexo biológico	Masculino ou Feminino
<b>Comportamento de Treino</b>	
Frequência semanal	Número de treinos por semana (1–6)
Tempo por sessão	Duração média de cada sessão (30–120 minutos)
Nível de experiência	Iniciante, Intermediário ou Avançado
Condição atual	5 categorias, incluindo “Nunca treinei”
Objetivo principal	Emagrecimento, Hipertrofia ou Força
<b>Metadados Operacionais</b>	
Data de cadastro	Marca temporal associada ao registro
Identificadores internos	<i>client ID, request ID</i> (não usados no modelo)

### 3.2. Pré-processamento e Engenharia de Atributos

As variáveis demográficas e comportamentais apresentam alta completude, com a principal exceção sendo o peso (57,3% de valores ausentes). Para mitigar essa ausência, adotamos imputação hierárquica com substituição pela mediana, estratégia escolhida pela robustez a *outliers*. Os grupos de imputação foram formados por: objetivo de treino, sexo biológico e faixa de altura (definida por quartis estratificados por sexo). A idade foi avaliada, mas apresentou baixa correlação com o peso e não foi usada na estratificação.

Na engenharia de atributos, derivamos indicadores relevantes para a clusterização. A Tabela 2 sintetiza as principais variáveis criadas.

Variáveis contínuas (IMC, Idade, Altura, Peso) foram padronizadas com *StandardScaler* (média zero, desvio padrão um). Variáveis binárias e ordinais foram mantidas em sua escala natural, e o *scaler* foi armazenado para uso consistente em dados futuros.

**Tabela 2. Variáveis derivadas e transformações aplicadas na engenharia de atributos**

Atributo Derivado	Descrição
Índice de Massa Corporal (IMC)	Peso dividido pela altura ao quadrado ( $\text{kg}/\text{m}^2$ ).
Intensidade semanal de treino	Categoria baseada em minutos semanais: Baixa ( $<150$ ), Moderada (150–300), Alta (300–450), Muito alta ( $>450$ ).
Indicadores binários ( <i>dummies</i> )	Peso informado/imputado, Sexo biológico, Objetivos de treino e Condições específicas.
Codificação ordinal	Nível de experiência: 0 = Iniciante; 1 = Intermediário; 2 = Avançado.

### 3.3. Seleção do Modelo e Clusterização

Foram avaliados três algoritmos de clusterização: *K-Means*, *DBSCAN* e *Hierarchical Clustering*. A comparação considerou qualidade interna dos *clusters* e interpretabilidade da solução para profissionais. Utilizamos três métricas internas para validação: *Silhouette Score* (combina coesão e separação), índice de Calinski–Harabasz (razão entre variância entre e intra-*clusters*) e índice de Davies–Bouldin (similaridade média entre *clusters*).

**Tabela 3. Configurações dos algoritmos testados e métricas internas**

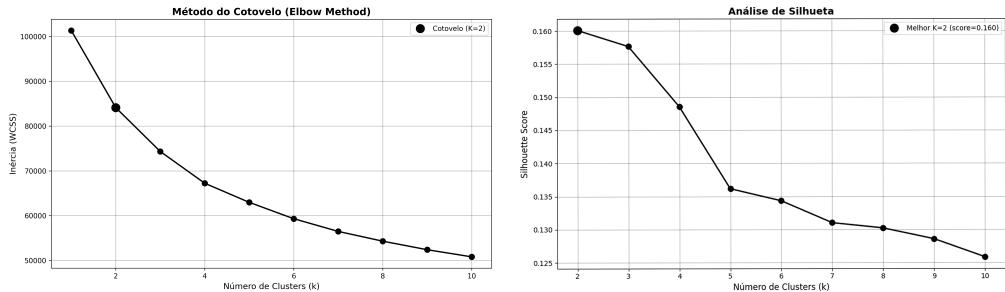
Algoritmo	Configurações	Métricas de Validação
<i>K-Means</i>	<code>n_init = 10</code> <code>max_iter = 300</code> <code>random_state = 42</code>	<i>Silhouette</i> : 0,160 <i>Calinski–Harabasz</i> : 2329,5 <i>Davies–Bouldin</i> : 2,123 <i>Clusters</i> : 2
<i>Hierarchical Clustering</i>	<code>linkage = 'ward'</code>	<i>Silhouette</i> : 0,157 <i>Calinski–Harabasz</i> : 1647,8 <i>Davies–Bouldin</i> : 1,772 <i>Clusters</i> : 2
<i>DBSCAN</i>	<code>eps = 1,5</code> <code>min_samples = 10</code>	<i>Silhouette</i> : 0,079 <i>Calinski–Harabasz</i> : 104,4 <i>Davies–Bouldin</i> : 3,383 <i>Clusters</i> : 3

No *K-Means*, o número de *clusters* ( $K$ ) foi explorado em [1, 10], com `n_init` = 10 e `random_state` = 42. Para o *DBSCAN* e o *Hierarchical Clustering*, realizamos testes e selecionamos as soluções que apresentavam o melhor desempenho nas métricas internas para comparação.

A Tabela 3 resume configurações e resultados médios das métricas para as soluções selecionadas, com o *K-Means* em  $K = 2$  apresentando o melhor balanço.

A escolha de  $K = 2$  para o *K-Means* foi guiada pelo *Método do Cotovelo (Elbow Method)* e pelo *Silhouette Score*, que apontaram o ponto de inflexão e o valor máximo, respectivamente, nesse número de agrupamentos (Figura 1). O *K-Means* convergiu,

em média, após 12 iterações, produzindo dois grupos de tamanho equilibrado: 5.728 alunos (50,3%) e 5.660 alunos (49,7%). Dez execuções independentes com diferentes `random_state` resultaram em coeficiente de variação do *Silhouette* inferior a 0,05, indicando alta estabilidade.



**Figura 1. Resultados para a Escolha de  $K$ : Gráfico do Método do Cotovelo (WCSS, esquerda) e Análise do *Silhouette Score* (direita).**

### 3.4. Procedimento de Interpretação Motivacional

Após o ajuste do *K-Means*, caracterizamos cada *cluster* a partir de médias padronizadas e distribuições de variáveis-chave: frequência semanal de treino, duração das sessões, volume semanal, objetivos declarados e nível de experiência.

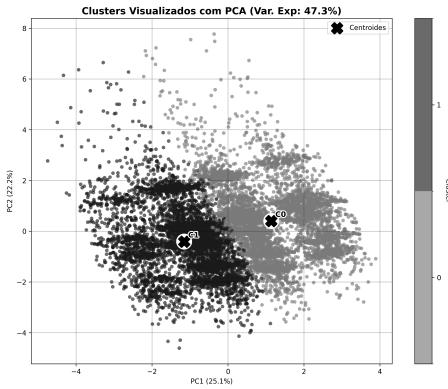
A interpretação motivacional seguiu duas etapas: (i) descrição do comportamento típico de cada grupo; e (ii) comparação desses perfis às descrições teóricas dos tipos do *HEXAD* [Tondello et al. 2016]. \*\*A lógica adotada foi ancorar a intensidade de esforço e o foco em desempenho ao eixo *Achiever*: padrões de alta frequência/volume e foco em metas de performance (hipertrofia) foram associados a uma orientação *Conquistador (Achiever-like)*, que valoriza progresso mensurável e maestria. Em contraste, padrões com menor volume e foco em resultados gerais (emagrecimento), sugerindo maior sensibilidade a barreiras, foram associados a um perfil *Jogador (Player-like)*, compatível com maior responsividade a recompensas extrínsecas e incentivos de curto prazo.\*\*

Esse mapeamento é estritamente qualitativo, servindo como rótulo de segmento para orientar hipóteses de gamificação. Os termos *Achiever-like* e *Player-like* são usados para designar segmentos, e não como diagnósticos definitivos de perfis motivacionais individuais.

## 4. Resultados

A aplicação do algoritmo *K-Means* com  $K = 2$  resultou na formação de dois *clusters* com tamanhos bastante similares: 5.728 alunos (50,3%) em um grupo e 5.660 (49,7%) no outro, permitindo comparações balanceadas. Para visualizar essa separação, a projeção das variáveis padronizadas nos dois primeiros componentes principais via *PCA* (Figura 2) mostra agrupamentos com centróides distintos e sobreposição limitada, sugerindo diferenças sistemáticas nas combinações de atributos de treino e perfil.

A Tabela 4 resume os principais atributos dos grupos, revelando diferenças em frequência de treino, volume semanal, intensidade, objetivos declarados e características demográficas/antropométricas.



**Figura 2.** Visualização dos *clusters* via PCA. A distribuição espacial indica separação significativa entre os agrupamentos.

Com base na interpretação motivacional descrita na Seção 3, os padrões de alto volume e foco em hipertrofia do Cluster 0 são compatíveis com uma orientação *Conquistador (Achiever-like)*, associada a esforço elevado e valorização de metas de progresso mensurável. Já os padrões de menor volume e maior foco em emagrecimento do Cluster 1 indicam um perfil mais próximo de um *Jogador (Player-like)*, compatível com engajamento mais moderado e maior responsividade a recompensas extrínsecas e incentivos acessíveis.

Essas associações são exploratórias e não derivam de rótulos motivacionais observados, mas de uma leitura teórica dos padrões de treino. Os termos *Achiever-like* e *Player-like* funcionam, neste contexto, como rótulos de segmento úteis para o desenho de estratégias de gamificação diferenciadas, e não como diagnósticos individuais de perfil *HEXAD*.

## 5. Discussão e Limitações

Os resultados indicam que um conjunto restrito de variáveis coletadas na matrícula é suficiente para revelar dois segmentos comportamentalmente distintos em academias de ginástica. Interpretados à luz do *HEXAD*, esses segmentos sugerem uma distinção pragmática entre usuários com comportamentos mais orientados a desempenho e maestria (*Achiever-like*) e usuários com padrões de engajamento mais moderados, potencialmente mais sensíveis a recompensas acessíveis e incentivos de curto prazo (*Player-like*). Essa distinção oferece uma base inicial para o desenho de estratégias de gamificação diferenciadas.

A principal limitação é que o vínculo entre os *clusters* identificados e os perfis do *HEXAD* é qualitativo e não supervisionado. Não foram coletados, nesta fase, rótulos motivacionais por meio de questionários *HEXAD*, o que impede quantificar o alinhamento entre segmentos e orientações motivacionais reais. Em linha com o rigor metodológico e sugestões da literatura, um próximo passo natural é aplicar um questionário *HEXAD* reduzido em uma subamostra de usuários, permitindo calcular métricas como ARI, NMI e *precision@k* para avaliar a correspondência entre clusterização e perfis auto-relatados.

Do ponto de vista metodológico, este estudo emprega três métricas internas (*Silhouette*, Calinski–Harabasz e Davies–Bouldin) e uma análise preliminar de

**Tabela 4. Comparação das características dos clusters identificados**

Característica	Cluster 0	Cluster 1
<b>Tamanho da Amostra</b>	5.728 (50,3%)	5.660 (49,7%)
<b>Demografia</b>		
Idade média (anos)	30,7 ± 10,7	33,6 ± 11,3
Faixa etária principal	18–34 anos (66,6%)	25–44 anos (58,5%)
<b>Características Físicas</b>		
Altura média (cm)	168,0	167,6
Peso médio (kg)	75,6	75,6
IMC médio	26,7 (Sobrepeso)	26,9 (Sobrepeso)
<b>Comportamento Fitness</b>		
Dias de treino/semana	5,0	3,5
Tempo de treino (min)	83,4	54,5
Minutos semanais totais	416,5	181,1
Objetivo principal	Hipertrofia (53,4%)	Emagrecimento (47,7%)
Experiência predominante	Intermediário (44,2%)	Intermediário (45,1%)
Intensidade predominante	Alta (62,5%)	Moderada (74,4%)
<b>Outros Indicadores</b>		
Alta frequência (5+ dias)	87,1%	20,7%
Sessões longas (90+ min)	43,6%	5,7%
Peso informado	44,2%	41,1%

estabilidade via múltiplas inicializações do *K-Means*. Contudo, para maior robustez, ainda não foram conduzidos testes mais sofisticados de robustez, como procedimentos baseados em *bootstrap* ou *jackknife* com índices de similaridade (por exemplo, Jaccard) entre partições. A incorporação sistemática desses testes em trabalhos futuros pode fornecer evidências mais fortes de que os segmentos não são artefatos de particularidades da amostra ou da parametrização.

Outra limitação importante é o escopo temporal e a ausência de desfechos comportamentais de longo prazo. A análise utiliza apenas informações de matrícula e características declaradas de treino, sem séries temporais de uso efetivo das academias, histórico de presença ou dados de cancelamento. Portanto, qualquer inferência sobre impacto direto na evasão é, por enquanto, hipotética. A contribuição atual é definir uma segmentação interpretável e hipóteses de design para estratégias gamificadas; a avaliação de efeitos sobre engajamento e retenção exige estudos longitudinais e/ou experimentais adicionais.

A generalização dos resultados também é limitada pelo contexto geográfico e operacional. Os registros são provenientes de academias que utilizam uma mesma plataforma de gestão, em um país específico, o que pode refletir características culturais, econômicas, ou de modelo de negócio particulares. A replicação da metodologia em outras regiões, faixas de preço e perfis de academias é necessária antes de se recomendar a adoção ampla de segmentos e mecânicas específicas.

Por fim, a utilização de variáveis demográficas e antropométricas em processos de

segmentação levanta questões éticas e de justiça. Embora o conjunto de dados tenha sido previamente anonimizado e não envolva dados sensíveis no sentido legal, é importante monitorar, em aplicações futuras, como os *clusters* se distribuem por idade, sexo e outras características, evitando interpretações estigmatizantes ou intervenções que reforcem desigualdades. Estratégias gamificadas diferenciadas devem ser analisadas quanto a possíveis efeitos colaterais indesejados, em consonância com a LGPD e boas práticas de proteção de dados.

Em síntese, os resultados apresentados devem ser entendidos como um primeiro passo em uma agenda de pesquisa que combina segmentação comportamental, teoria motivacional e experimentação controlada para avançar rumo à personalização responsável de sistemas gamificados em academias.

## 6. Conclusão

Este estudo apresentou uma abordagem baseada em clusterização para segmentar usuários de academias de ginástica a partir de dados coletados no momento da matrícula, com o objetivo de apoiar estratégias de gamificação personalizada. Utilizando o algoritmo *K-Means* e métricas internas de validação, identificamos dois grupos comportamentais coesos e de tamanhos similares, que diferem de forma consistente em frequência de treino, volume semanal, intensidade e objetivos declarados.

Com base nesses padrões, propusemos uma interpretação motivacional qualitativa ancorada no modelo *HEXAD*, na qual um segmento foi caracterizado como *Achiever-like* (mais orientado a desempenho, progresso e rotinas de alta carga) e o outro como *Player-like* (com engajamento mais moderado e potencial sensibilidade a recompensas extrínsecas e incentivos acessíveis). Essa leitura deve ser vista como um arcabouço de design para estratégias diferenciadas, e não como tipificação definitiva de usuários.

Do ponto de vista prático, a contribuição central é propor um procedimento de baixo custo operacional, baseado exclusivamente em registros administrativos já coletados, que pode servir como base para concepção de intervenções gamificadas segmentadas em academias. Em vez de afirmar impacto direto sobre evasão, o trabalho organiza evidências e hipóteses que podem orientar experimentos futuros com mecânicas personalizadas.

As limitações discutidas na Seção 5 delineiam uma agenda clara para trabalhos futuros. Em termos metodológicos, destacam-se: (i) a incorporação de testes de estabilidade mais robustos (por exemplo, *bootstrap* e índices de similaridade entre partições); e (ii) a coleta de rótulos motivacionais fracos, por meio de um questionário *HEXAD* reduzido aplicado a uma subamostra, permitindo avaliar quantitativamente a correspondência entre clusterização e perfis auto-relatados. Em termos aplicados, próximos passos incluem: (iii) o desenho e a avaliação experimental de mecânicas de gamificação específicas para cada segmento, em colaboração com academias parceiras; e (iv) o monitoramento cuidadoso de efeitos sobre engajamento, retenção e possíveis vieses demográficos, em conformidade com a LGPD.

Ao combinar clusterização, teoria motivacional e avaliação empírica planejada, este trabalho busca contribuir para o desenvolvimento de abordagens de personalização em gamificação que sejam operacionalmente viáveis e sensíveis às particularidades do contexto de academias de ginástica.

## Referências

- Altmeyer, M., Schubhan, M., Krüger, A., e Lessel, P. (2021). A long-term investigation on the effects of (personalized) gamification on course participation in a gym. *arXiv preprint arXiv:2107.12597*.
- Bartle, R. (1996). Hearts, clubs, diamonds, spades: Players who suit muds.
- Bicalho, L. F., Baffa, A., e Feijó, B. (2019). A game analytics model to identify player profiles in singleplayer games. In *Proceedings of SBGames 2019*. SBGames.
- Krath, J., Palmquist, A., Jedel, I., Barbopoulos, I., Helmefalk, M., e Munkvold, R. I. (2022). Does behaviour match user typologies? an exploratory cluster analysis of behavioural data from a gamified fitness platform. In *GamiFIN*, pages 105–114.
- Liz, C. M. d. e Andrade, A. (2016). Análise qualitativa dos motivos de adesão e desistência da musculação em academias. *Revista Brasileira de Ciências do Esporte*, 38(3):267–274.
- Nacke, L. E., Bateman, C., e Mandryk, R. L. (2014). Brainhex: A neurobiological gamer typology survey. *Entertainment computing*, 5(1):55–62.
- Niu, Z. (2025). Customer segmentation and management strategy optimization for gyms using k-means clustering.
- Ryan, R. M. e Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist*, 55(1):68.
- Tondello, G. F., Wehbe, R. R., Diamond, L., Busch, M., Marczewski, A., e Nacke, L. E. (2016). The gamification user types hexad scale. In *Proceedings of the 2016 annual symposium on computer-human interaction in play*, pages 229–243.
- Yang, Y., Goh, K. Y., Teo, H. H., e Tan, S. S. (2023). Compete with me? the impact of online gamified competition on exercise behavior. *Journal of the Association for Information Systems*, 24(3):912–935.
- Yee, N. (2006). Motivations for play in online games. *CyberPsychology & behavior*, 9(6):772–775.
- Zhao, Z., Arya, A., Orji, R., Chan, G., et al. (2020). Effects of a personalized fitness recommender system using gamification and continuous player modeling: system design and long-term validation study. *JMIR serious games*, 8(4):e19968.