Analítica da Aprendizagem em Jogos Educacionais: Um Mapeamento Sistemático da Literatura

Daniel Irano Valentim, Universidade Federal de São Carlos,
danielirano@estudante.ufscar.br, ⟨https://orcid.org/0000-0003-3958-2389⟩
Delano Medeiros Beder, Universidade Federal de São Carlos,
delano@ufscar.br, ⟨https://orcid.org/0000-0003-3692-2202⟩
Joice Lee Otsuka, Universidade Federal de São Carlos,
joice@ufscar.br, ⟨https://orcid.org/0000-0003-3233-0314⟩

Resumo: Este artigo apresenta um mapeamento sistemático da literatura sobre a aplicação da Analítica da Aprendizagem (*Learning Analytics* – LA) e da Mineração de Dados Educacionais (*Educational Data Mining* – EDM) em jogos educacionais. A busca foi realizada nas bases ACM, IEEE Xplore, Scopus, Springer Link, RENOTE e SBC Open Lib, abrangendo publicações entre 2019 e 2024. Dos 170 estudos encontrados, 11 atenderam aos critérios de inclusão e foram analisados em profundidade. As pesquisas concentram-se em duas vertentes: (1) modelagem preditiva e diagnóstica, voltada à previsão de desempenho e identificação de dificuldades; e (2) análise comportamental e contextual, relacionada às interações e motivação dos jogadores. As técnicas mais frequentes incluem SVM, *Random Forest*, *k-means*, IRT, BKT e NLP. Observou-se escassez de estudos sobre inclusão e acessibilidade, indicando a necessidade de novas investigações que integrem práticas inclusivas e analíticas em jogos educacionais.

Palavras-chave: analítica da aprendizagem; mineração de dados educacionais; jogos educacionais; aprendizagem baseada em jogos.

Learning Analytics in Educational Games: A Systematic Literature Mapping

Abstract: This paper presents a systematic mapping of the literature on the application of Learning Analytics (LA) and Educational Data Mining (EDM) in educational games. The search was conducted in the ACM, IEEE Xplore, Scopus, Springer Link, RENOTE, and SBC Open Lib databases, covering publications from 2019 to 2024. From the 170 studies identified, 11 met the inclusion criteria and were analyzed in depth. The reviewed research focuses on two main strands: (1) predictive and diagnostic modeling, aimed at forecasting performance and identifying learning difficulties; and (2) behavioral and contextual analysis, related to player interactions and motivation. The most frequent techniques include SVM, Random Forest, k-means, IRT, BKT, and NLP. A lack of studies addressing inclusion and accessibility was observed, indicating the need for further research that integrates inclusive and analytical practices within educational games.

Keywords: learning analytics; educational data mining; educational games; game-based learning.

1. Introdução

Nas últimas décadas, os processos de ensino-aprendizagem têm evoluído significativamente, impulsionados pela incorporação de novas metodologias e tecnologias digitais. Um exemplo notável dessa transformação é o uso crescente de jogos educacionais, reconhecidos como ferramentas eficazes para avaliar e aprimorar o aprendizado dos alunos (VERBERT *et al.*, 2012).

Os Jogos Sérios (SG - do inglês, *serious games*), definidos como "jogos cujo objetivo vai além do entretenimento, buscando educar, treinar e informar" (MICHAEL; CHEN, 2005), têm se consolidado como estratégias relevantes para enriquecer a experiência educacional e o desempenho acadêmico (DAOUDI, 2022). Evidências empíricas indicam impactos positivos na aprendizagem, abrangendo o desenvolvimento de habilidades cognitivas e socioemocionais (CLARK *et al.*, 2016; BOYLE *et al.*, 2016).

A aprendizagem baseada em jogos (GBL - do inglês, *game-based learning*) caracteriza-se pela integração entre conteúdo e jogabilidade, possibilitando que os alunos aprendam de forma ativa e contextualizada (PRENSKY, 2003; MAYER, 2020). Contudo, ainda há desafios quanto ao acompanhamento dos processos cognitivos e comportamentais durante o uso de jogos educacionais, demandando ferramentas analíticas capazes de coletar e interpretar dados de forma precisa (BANIHASHEM *et al.*, 2024; CHAUDY; CONNOLLY, 2019).

Nesse contexto, a Analítica da Aprendizagem (LA - do inglês, *learning analytics*) tem se destacado como uma abordagem essencial para compreender e otimizar o ensino-aprendizagem por meio da medição, coleta e análise de dados sobre alunos e seus contextos (SIEMENS; LONG, 2011). Em ambientes de GBL, a LA permite identificar padrões de comportamento, dificuldades e oportunidades de adaptação, favorecendo intervenções pedagógicas mais precisas e personalizadas (BANIHASHEM *et al.*, 2024; KANG *et al.*, 2017).

A mineração de dados educacionais (EDM - do inglês, *educational data mining*) surge como campo complementar, aplicando técnicas computacionais para extrair conhecimento a partir de grandes volumes de dados (ROMERO *et al.*, 2010; BAKER *et al.*, 2011; SILVA; FONSECA, 2017). Enquanto a LA foca na interpretação dos dados, a EDM enfatiza a descoberta de padrões e o desenvolvimento de modelos preditivos (CEREZO *et al.*, 2024). Estudos apontam que a EDM pode ser compreendida como um subconjunto da LA, reforçando o potencial de ambas para aprimorar o aprendizado mediado por jogos (DORMEZIL *et al.*, 2019).

Este mapeamento sistemático busca identificar e discutir como a analítica da aprendizagem e, mais especificamente, técnicas de mineração de dados educacionais, têm sido utilizadas em jogos educacionais.

O presente artigo está estruturado da seguinte maneira: a Seção 2 discute os trabalhos relacionados; a Seção 3 detalha a metodologia empregada no Mapeamento Sistemático da Literatura; a Seção 4 apresenta e analisa os resultados obtidos; a Seção 5 examina as limitações e possíveis ameaças à validade da pesquisa; e, por fim, a Seção 6 traz as discussões e as considerações finais.

2. Trabalhos Relacionados

Nas últimas décadas, a integração entre SG e LA tem ganhado destaque no cenário educacional, em especial pela possibilidade de promover uma aprendizagem mais personalizada e baseada em dados. Essa convergência é abordada por diferentes estudos de revisão sistemática, que analisam como técnicas de EDM e LA têm sido empregadas para potencializar a eficácia dos jogos no contexto de ensino-aprendizagem.

O trabalho de Daoudi (2022) realiza uma revisão sistemática abrangente sobre a aplicação de LA em jogos educacionais, apresentando uma taxonomia multidimensional que agrupa fatores pedagógicos, técnicos e contextuais relevantes à usabilidade desses jogos. Os resultados indicam que a LA pode contribuir significativamente para a avaliação não intrusiva (*stealth assessment*), possibilitando o monitoramento contínuo

do desempenho discente com base em dados interacionais, como *logs* de jogo, métricas comportamentais e sensores em tempo real. Apesar do avanço na área, o estudo destaca desafios recorrentes, como a ausência de padronização nos métodos analíticos e a carência de estudos empíricos voltados à aplicação de técnicas de ciência de dados em contextos educacionais formais.

Corroborando com essa perspectiva, Tlili *et al.* (2021) realizaram uma revisão sistemática voltada exclusivamente para estudos empíricos que aplicam LA em jogos educacionais. A pesquisa identifica que os principais objetivos da aplicação de LA incluem: a modelagem do comportamento dos alunos durante o jogo, o suporte ao design formativo de jogos, a personalização da aprendizagem e a aplicação de avaliações formativas. Além disso, evidenciam-se desafios técnicos, como a dificuldade de integração das abordagens analíticas ao design dos jogos, e éticos, relacionados à coleta e uso dos dados dos estudantes. Os autores concluem que, embora a área tenha crescido, ainda existe uma lacuna para orientar a aplicação de LA de maneira eficaz nesses ambientes.

No contexto nacional, o estudo de Nascimento *et al.* (2021) apresenta um mapeamento sistemático voltado à aplicação da Aprendizagem Baseada em Jogos (GBL) para o ensino de matemática. Embora não tenha como foco direto a LA, o trabalho identifica o uso de técnicas estatísticas, técnicas de mineração de dados, assim como regressão logística, análise de variância, árvores de decisão e redes bayesianas nos jogos analisados, evidenciando o potencial dessas abordagens para apoiar a avaliação e a personalização da aprendizagem. Contudo, uma importante lacuna apontada refere-se à ausência de mais estudos focados no tema.

Ampliando a discussão para a população de estudantes com deficiências, Conde e Rodríguez-Sedano (2024) realizaram uma revisão sistemática específica sobre a aplicação da LA para estudantes com IDD em contextos educacionais diversos, não restritos a jogos. Os autores identificaram 28 estudos entre 364 analisados que de fato aplicavam técnicas de LA para esse público. A principal contribuição do estudo está em evidenciar que a LA é viável no contexto inclusivo, desde que adaptada às necessidades cognitivas e comportamentais desses estudantes, promovendo personalização, intervenções pedagógicas mais precisas e acompanhamento contínuo. Embora o foco não tenha sido o uso em jogos educacionais, seus resultados reforçam o potencial da LA para apoiar práticas inclusivas também nesse tipo de ambiente.

Diferentemente dos estudos analisados, que tendem em enfatizar aspectos técnicos, modelagens preditivas ou análises comportamentais em contextos amplos de aprendizagem baseada em jogos, o presente trabalho propõe analisar a aplicação de técnicas de LA em especial as de mineração de dados, para apoiar a análise de dados de jogos educacionais.

3. Metodologia

Com o intuito de conhecer o estado da arte das pesquisas envolvendo analítica de aprendizagem utilizando técnicas de mineração de dados no apoio à aprendizagem baseada em jogos, foi conduzido um mapeamento sistemático da literatura, seguindo as diretrizes apresentadas por Petersen *et al.* (2008). Todas as etapas do mapeamento sistemático da literatura foram planejadas e conduzidas pelo primeiro autor deste artigo, com revisão e validação realizadas pelos dois últimos autores. Em casos de dúvida sobre a inclusão de um estudo, houve discussão até o consenso.

3.1. Planejamento e condução

O objetivo do presente trabalho foi identificar com qual objetivo a analítica de aprendizagem foi utilizada em cada estudo (para quê?), e entender como a mineração de dados tem sido aplicada nesses estudos.

Para tal, foram definidas as seguintes Questões de Pesquisa (QP):

- **QP1:** Com qual objetivo a LA foi utilizada em cada estudo?
- **QP2:** Como a mineração de dados tem sido aplicada nesses casos?

Para abranger de forma mais clara o objetivo do presente estudo, e para contemplar a QP2, a mesma foi dividida em 3 subquestões (SQP):

- **SQP1:** Quais são os tipos de dados coletados nesses estudos?
- **SQP2:** Quais técnicas de mineração de dados estão sendo utilizadas na analítica de aprendizagem em jogos educacionais?
- **SQP3:** Qual é o tamanho das amostras utilizadas nesses estudos?

A definição das questões de pesquisa seguiu a estrutura conceitual do modelo PICOC (População, Intervenção, Contexto, Comparação e Outcome), adaptada para o contexto da LA e EDM em jogos educacionais.

A coleta de dados foi realizada nas bases de dados *ACM Digital Library, IEEE Xplore Digital Library, RENOTE, SBC Open Lib, Scopus e Springer Link*, onde a *string* de busca utilizada abrangeu trabalhos publicados nos últimos cinco anos, em português e inglês. A string de busca foi elaborada originalmente em inglês, idioma predominante nas bases internacionais, conforme apresentado na Tabela 1. Para as bases nacionais (RENOTE e *SBC Open Lib*), foi desenvolvida uma versão equivalente em português, preservando o mesmo conjunto de termos e operadores.

Tabela 1. String de busca.

Base de Dados	String de busca utilizada
ACM Digital Library, IEEE Digital Library, Springer Link e Scopus	(game* AND (serious OR ”educat*”OR learning) AND (”data mining”) AND (”learning analytics”OR ”game analytics”))

As bases foram selecionadas por sua relevância na área de Informática na Educação e pela ampla cobertura de publicações nacionais e internacionais relacionadas à Analítica da Aprendizagem (LA), Mineração de Dados Educacionais (EDM) e Jogos Educacionais (GBL). O recorte temporal de cinco anos foi definido com o objetivo de refletir o estado mais recente da área, considerando a rápida evolução das pesquisas e o surgimento de novas abordagens e tecnologias nesse período.

Após a aplicação da *string* de busca nas bases de dados mencionadas, foram encontrados 177 trabalhos. A base de dados *ACM Digital Library* retornou 6 trabalhos, as bases de dados *IEEE Digital Library* e *Scopus* retornaram 34 trabalhos cada uma, a base de dados RENOTE retornou 1 trabalho e a base *SBC Open Lib* retornou 3 trabalhos, por fim a base *Springer Link* retornou 99 trabalhos. Após a inserção na ferramenta *Parsifal*, foram encontrados 7 trabalhos duplicados, resultando em um total de 170 trabalhos.

Foram aplicados os seguintes Critérios de Inclusão (CI):

- **CI1:** Artigos completos em português ou inglês, publicados depois de 2018
- **CI2:** Documentos que abordam mineração de dados para analítica de aprendizagem em jogos educacionais

Em contrapartida, foram aplicados os seguintes Critérios de Exclusão (CE):

- **CE1:** Trabalhos secundários
- **CE2:** Documentos não encontrados na íntegra
- **CE3:** Documentos duplicados

O presente trabalho aplicou os critérios de inclusão e exclusão em duas etapas. Na primeira etapa, foi realizada a análise dos títulos e resumos dos estudos, resultando na seleção de 36 estudos considerados relevantes e no descarte de 134 estudos.

Na segunda etapa, foi realizada a leitura completa dos 36 estudos selecionados, levando-se em conta os critérios previamente estabelecidos. Como resultado, foram aceitos 11 estudos, considerados relevantes para a contribuição do presente trabalho, enquanto 159 estudos foram rejeitados. A Tabela 2 apresenta uma síntise dos resultados obtidos em cada uma das etapas.

Tabela 2. Síntese dos resultados.

Bases de Dados	Encontrados	Etapa 1: aceitos	Etapa 2: aceitos
ACM Digital Library	6	0	0
IEEE Digital Library	31	12	3
RENOTE	1	0	0
SBC Open Lib	3	2	0
Scopus	30	20	6
Springer Link	99	2	2
Total:	**170**	**36**	**11**

4. Resultados e Discussões

A seguir, são apresentadas as respostas às questões de pesquisa, com base nos trabalhos revisados e aceitos neste estudo. As respostas baseiam-se no número de identificação (ID) dos trabalhos aceitos, listados na Tabela 3.

Tabela 3: Relação dos trabalhos aceitos.

ID	Referência Completa
1	Chen, F., Cui, Y., and Chu, M.-W. (2020). Utilizing game analytics to inform and validate digital game-based assessment with evidence-centered game design: A case study. *International Journal of Artificial Intelligence in Education*, 30:481–503.
2	Alonso-Fernández, C., Martínez-Ortiz, I., Caballero, R., Freire, M., and Fernández-Manjón, B. (2020). Predicting students' knowledge after playing a serious game based on learning analytics data: A case study. *Journal of Computer Assisted Learning*, 36(3):350–358.

Continua na próxima página

Tabela 3 – continuação

ID	Referência Completa
3	Forsyth, C. M., Graesser, A., and Millis, K. (2020). Predicting learning in a multi-component serious game. *Technology, Knowledge and Learning*, 25(2):251–277.
4	Hernández-Lara, A. B., Perera-Lluna, A., and Serradell-López, E. (2019). Applying learning analytics to students' interaction in business simulation games. The usefulness of learning analytics to know what students really learn. *Computers in Human Behavior*, 92:600–612.
5	Amarasinghe, I., Hernández-Leo, D., and Jonsson, A. (2019). Data-informed design parameters for adaptive collaborative scripting in across-spaces learning situations. *User Modeling and User-Adapted Interaction*, 29(4):869–892.
6	Torcate, A. S. and de Oliveira Rodrigues, C. M. (2021). Educational data mining with learning analytics and unsupervised algorithms: analysis and diagnosis in basic education. In *2021 XVI Latin American Conference on Learning Technologies (LACLO)*, pages 67–74. IEEE.
7	Palomo-Duarte, M., Berns, A., Yañez Escolano, A., and Dodero, J.-M. (2019). Clustering analysis of game-based learning: Worth it for all students? *Journal of Gaming & Virtual Worlds*, 11(1):45–66.
8	Gagnon, D. J., Harpstead, E., and Slater, S. (2019). Comparison of off the shelf data mining methodologies in educational game analytics. In *EDM (Workshops)*, pages 38–43.
9	Zapata-Cáceres, M. and Martín-Barroso, E. (2021). Applying game learning analytics to a voluntary video game: Intrinsic motivation, persistence, and rewards in learning to program at an early age. *IEEE Access*, 9:123588–123602.
10	Khorasani, S., Nawaz, S., Syiem, B. V., Wei, J., Pardos, Z. A., Knibbe, J., and Velloso, E. (2023). An empirical evaluation of educational data mining techniques in a dynamic vr application. In *Proceedings of the 35th Australian Computer-Human Interaction Conference*, pages 604–623.
11	Estriegana, R., Garcia-Esteban, S., Rojas, E., and Medina-Merodio, J.-A. (2021). Analytics and predictive models of student's activity in off/on-line learning environments. In *2021 1st Conference on Online Teaching for Mobile Education (OT4ME)*, pages 124–131. IEEE.

4.1. Respondendo QP1 (Qual é o objetivo principal do estudo)

A partir de uma análise cuidadosa dos objetivos de cada estudo, foi possível perceber que podem ser organizados em dois grupos principais, cada um com suas nuances e ênfases particulares. A seguir, são descritos cada grupo:

Grupo 1 (IDs 1, 2, 3, 5, 6, 8, 10 e 11)

Este grupo reúne pesquisas que aplicaram técnicas preditivas, de modelagem e diagnóstico para antecipar ou identificar aspectos dos resultados de aprendizagem. Os estudos buscaram, por meio de dados (como interações, comportamento dos alunos, dados multimodais etc.), prever variáveis importantes como domínio do jogo, aquisição de conhecimento, risco de evasão, dificuldades específicas e até perfis de aprendizagem (por exemplo, aprendizado profundo versus aprendizado superficial).

Os estudos referentes aos IDs 1, 2 e 3 concentraram-se em prever habilidades ou a aquisição de conhecimento com base em dados comportamentais e características identificáveis durante a interação com o jogo ou com o ambiente de aprendizagem.

Já outros estudos (por exemplo, ID 6 e ID 8) utilizaram modelos para identificar dificuldades em conteúdos específicos ou prever comportamentos como, por exemplo, a desistência.

Há também investigações que aplicaram técnicas para estimar a dificuldade dos desafios e a maestria dos alunos (ID 10) ou identificar quais atividades estão mais relacionadas ao sucesso dos alunos no jogo (ID 11).

O estudo referente ao ID 5, por exemplo, utilizou a análise preditiva para antecipar a participação futura em atividades colaborativas, permitindo adaptações dinâmicas dos *scripts* de aprendizagem.

Grupo 2 (IDs 4, 7 e 9)

Neste grupo, os estudos se voltaram para a identificação e análise do comportamento dos alunos dentro de contextos ou ambientes educacionais específicos. A ênfase recaiu sobre como as interações, padrões comunicacionais e fatores motivacionais se manifestaram em cenários particulares – seja em cursos de línguas, jogos de simulação de negócios ou jogos de programação – e como esses aspectos se relacionam com os resultados de aprendizagem.

O estudo referente ao ID 4 investigou fóruns de discussão online de alunos em um jogo de simulação de negócios, buscando compreender quais conteúdos e padrões de comunicação tiveram maior impacto nos resultados de aprendizagem.

O estudo referente ao ID 7 teve foco na identificação do comportamento de alunos em um curso de alemão como língua estrangeira, mostrando como características comportamentais podem ser estudadas em ambientes de ensino de idiomas.

E o estudo referente ao ID 9 aplicou *Game Learning Analytics* em jogos de programação para crianças, explorando fatores como motivação, persistência e recompensas, destacando como esses elementos são cruciais em contextos voltados a públicos específicos (neste caso, crianças).

Cada grupo, portanto, ressaltou uma vertente distinta dentro do campo de analítica de aprendizagem.

4.2. Respondendo QP2 (Como a mineração de dados tem sido aplicada nesses casos?)

Para responder à QP2, foram formuladas três subquestões, conforme apresentado previamente, as quais serão respondidas a seguir.

4.2.1. Respondendo à subquestão SQP1 (Quais são os tipos de dados coletados nesses estudos?)

A análise dos tipos de dados coletados em cada estudo revela uma correspondência significativa tanto com os objetivos individuais dos trabalhos quanto com os grupos que foram previamente definidos.

Os dados do Grupo 1 são predominantemente quantitativos e orientados para medir o desempenho e o comportamento dos alunos durante a interação com o jogo ou plataforma. Essa natureza dos dados é coerente com os objetivos de prever competências, identificar padrões de aprendizagem e diagnosticar dificuldades, sustentando abordagens

de modelagem preditiva e análises estatísticas.

Os dados do Grupo 2 têm um caráter mais contextual e qualitativo, muitas vezes envolvendo análise textual, demográfica e de participação em fóruns. Essa natureza dos dados é compatível com os objetivos de identificar padrões de comunicação, erros específicos e fatores motivacionais em ambientes particulares, contribuindo para uma compreensão mais aprofundada do comportamento dos alunos em contextos específicos.

Após analisar os dados que foram coletados pelos estudos de todos os grupos (1 e 2), é possível mensurar que na maioria dos estudos os dados comportamentais (interações, tempo de jogo) estão presentes nos estudos relacionados aos IDs 1, 5, 8 e 9 e principalmente os dados de desempenho (acertos e erros, progresso no jogo) estão presentes nos estudos relacionados aos IDs 1, 2, 3, 4, 6 e 9. E dados multivariados como, posição, feição dos jogadores, movimentos, erros específicos de um determinado cenário, foram coletados em estudos específicos, relacionados aos IDs 10 e 11, conforme apresentado, a seguir, na Figura 1.

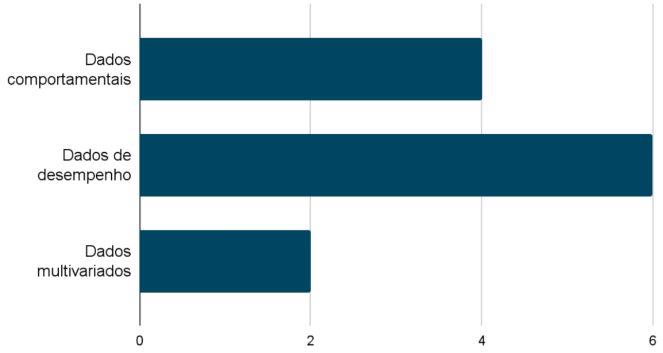

Figura 1. Resumo dos tipos de dados coletados nos estudos.

4.3. Respondendo à subquestão SQP2 (Quais técnicas de mineração de dados estão sendo utilizadas na analítica de aprendizagem em jogos educacionais?)

Na literatura especializada são encontradas diferentes taxonomias para caracterizar as tarefas de mineração de dados. Fayyad *et al.* (1996) apresentam uma taxonomia em dois níveis. No primeiro nível, as tarefas de mineração de dados são divididas em **preditivas** e **descritivas**.

Tarefas **preditivas** usam os valores dos atributos descritivos para predizer valores futuros ou desconhecidos de outros atributos de interesse. Já as tarefas **descritivas** têm o objetivo de encontrar padrões que descrevem os dados de maneira que o ser humano possa interpretar.

No segundo nível da taxonomia (FAYYAD *et al.*, 1996), as tarefas preditivas e descritivas são especializadas. No conjunto de tarefas **preditivas**, os autores inserem: (a) classificação e (b) regressão. Já no conjunto de tarefas **descritivas**, as especializações: (a) agrupamento (*clusterização*), (b) sumarização, (c) modelagem de dependências e (d) detecção de desvios (*outliers*) são incluídas pelos autores.

Na tentativa em prover uma resposta para a subquestão SQP2, a análise dos estudos selecionados revelou uma variedade de técnicas de mineração de dados empregadas na analítica de aprendizagem em jogos educacionais. As técnicas identificadas abrangem desde algoritmos supervisionados e métodos estatísticos até abordagens não supervisionadas e de processamento de dados textuais.

A seguir, apresenta-se uma lista organizada em ordem alfabética, destacando as técnicas utilizadas e os respectivos estudos em que foram aplicadas.

- Análise de Correlação: (ID 3)
- Árvores de Decisão: (ID 2)
- *Bayesian Knowledge Tracing e Item Response Theory* (IRT): (ID 10)
- *Knowledge Discovery in Databases* (KDD): (ID 6)
- *Linear Discriminant Analysis* (LDA): (ID 11)
- Máquinas de Vetores de Suporte (SVM): (IDs 1, 5 e 8)
- Modelagem Estatística: (ID 3)
- *Naive Bayes*: (IDs 8 e 11)
- PCA (Redução de Dimensionalidade): (ID 9)
- *Random Forest*: (IDs 5, 8 e 11)
- Regressão Linear: (ID 2)
- Regressão Logística: (ID 8)
- Técnica de *Clustering: k-means* (IDs 6 e 7)
- *Generalized Linear Model* (GLM): (ID 8)

Figura 2 apresenta a frequência dos métodos e técnicas utilizados nos estudos analisados. Observa-se a predominância da utilização de técnicas relacionadas a tarefas preditivas (classificação e/ou regressão) de mineração de dados.

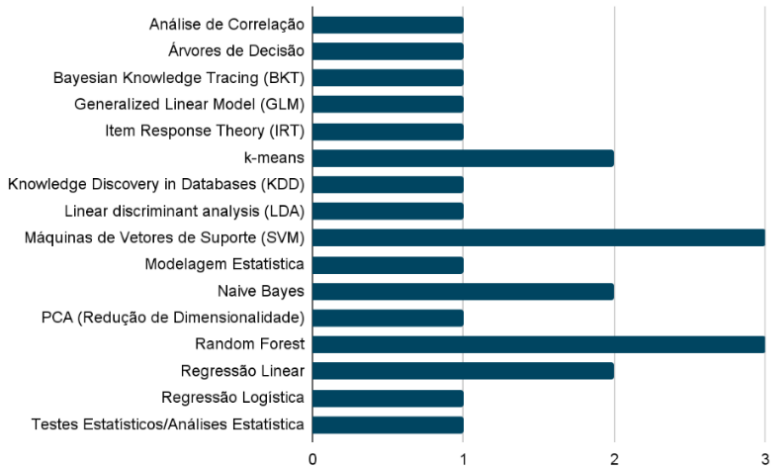

Figura 2. Frequência de utilização dos métodos e técnicas nos estudos.

Os algoritmos *Support Vector Machine* (SVM) e *Random Forest*, ambos geralmente utilizados para tarefas preditivas (classificação e/ou regressão), foram os mais utilizados, aparecendo em três estudos cada.

O algoritmo de clusterização *k-means*, assim como os algoritmos Regressão Linear e *Naive Bayes* (ambos também geralmente utilizados para tarefas preditivas), foram identificados em dois trabalhos.

As demais técnicas, como Árvores de Decisão, Regressão Logística, *Bayesian Knowledge Tracing* (BKT), *Linear Discriminant Analysis* (LDA), Testes Estatísticos, Processamento de Linguagem Natural, PCA, *Generalized Linear Model* (GLM), *Item Response Theory* (IRT), Análise de Correlação, Modelagem Estatística e *Knowledge Discovery in Databases* (KDD), foram cada uma mencionada em apenas um estudo.

Ressalta-se que as quatro primeiras técnicas listadas no último parágrafo também são geralmente utilizadas no contexto de tarefas preditivas – fato que evidencia ainda mais a predominância da utilização, nos estudos analisados, de técnicas relacionadas a tarefas preditivas de mineração de dados.

4.4. Respondendo à subquestão SQP3 (Qual é o tamanho das amostras utilizadas nesses estudos?)

Após a análise individual de cada estudo, foi possível constatar que alguns estudos (IDs 6 e 10) trabalharam com amostras reduzidas (0–100 participantes), enquanto o estudo referente ao ID 2, por exemplo, se concentrou em amostras intermediárias (201–300 participantes) e alguns empregaram amostras mais robustas (IDs 8 e 9, com mais de 1000 participantes). Ademais, observa-se que o estudo referente ao ID 11 não especificou o tamanho da amostra utilizada, conforme apresentado na Tabela 4.

Tabela 4. Agrupamento pelo tamanho da amostra.

ID	Tamanho da amostra
6 e 10	0 - 100
3 e 7	101 - 200
2	201 - 300
4 e 5	301 - 400
1	401 - 500
8 e 9	> 1.000
11	Não especificado

5. Limitações e ameaças à validade

Entre as limitações deste estudo, destaca-se o número reduzido de estudos selecionados, o que pode ter restringido a abrangência da análise e limitado uma visão mais completa do estado da arte sobre a aplicação da LA e da EDM em jogos educacionais. A variação nos termos utilizados pelos autores e a predominância de publicações em língua inglesa podem ter levado à exclusão de estudos potencialmente relevantes em outros idiomas ou na literatura cinzenta. Além disso, não foi realizada validação cruzada entre avaliadores na etapa de codificação, o que pode introduzir vieses na categorização dos dados. Contudo, essas limitações não comprometem a validade do mapeamento, mas apontam oportunidades de aprimoramento metodológico em pesquisas futuras, que podem ampliar o período de busca e considerar outras bases de dados, de modo a oferecer uma perspectiva mais abrangente e atualizada.

6. Discussão e considerações finais

O mapeamento sistemático apresentado neste estudo permitiu uma visão abrangente do estado da arte sobre a aplicação da Analítica da Aprendizagem (LA) e da Mineração de Dados Educacionais (EDM) no contexto da Aprendizagem Baseada em Jogos (GBL). A partir da análise dos 11 estudos selecionados, foi possível identificar como essas técnicas têm sido utilizadas para aprimorar os processos de ensino-aprendizagem, além de evidenciar os principais desafios e oportunidades emergentes na área.

Os estudos analisados foram organizados em dois grandes grupos, conforme detalhado nas seções anteriores. Observou-se uma predominância de abordagens voltadas à modelagem preditiva e à análise de desempenho discente, com ênfase em variáveis como aquisição de conhecimento, persistência e padrões de interação nos jogos. Contudo, identificou-se uma lacuna significativa quanto à validação dessas abordagens em contextos educacionais inclusivos. Além disso, muitos trabalhos priorizam aspectos técnicos, sem considerar mecanismos de apoio ao docente, como o uso do *feedback* gerado ou a adaptação das ferramentas ao contexto escolar.

O estudo conduzido se destaca por organizar os estudos com base nos objetivos de uso da LA, nos tipos de dados coletados, nas técnicas de EDM aplicadas e nos tamanhos das amostras utilizadas. Os resultados evidenciam uma ausência de estudos voltados a estudantes com deficiência, o que aponta para uma lacuna significativa na literatura.

Embora o mapeamento não tenha se restringido a estudos voltados a estudantes com deficiência, os resultados evidenciam uma lacuna importante quanto à aplicação de técnicas de LA e EDM em jogos educacionais inclusivos. Ao destacar essa ausência, esta pesquisa se diferencia ao propor uma agenda crítica de investigação, um aspecto ainda pouco abordado na literatura atual.

Como direções futuras, recomenda-se o desenvolvimento e a avaliação de ferramentas de LA voltadas especificamente a jogos educacionais inclusivos, bem como o aprimoramento de interfaces e visualizações acessíveis que permitam aos educadores interpretar e utilizar os dados de forma autônoma e significativa.

Referências

BAKER, R.; ISOTANI, S.; CARVALHO, A. Mineraçao de dados educacionais: Oportunidades para o brasil. **Revista Brasileira de informática na educação**, v. 19, n. 02, p. 03, 2011.

BANIHASHEM, S. K.; DEHGHANZADEH, H.; CLARK, D.; NOROOZI, O.; BIEMANS, H. J. Learning analytics for online game-based learning: A systematic literature review. **Behaviour & Information Technology**, Taylor & Francis, v. 43, n. 12, p. 2689–2716, 2024.

BOYLE, E. A.; HAINEY, T.; CONNOLLY, T. M.; GRAY, G.; EARP, J.; OTT, M.; LIM, T.; NINAUS, M.; RIBEIRO, C.; PEREIRA, J. An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. **Computers & Education**, Elsevier, v. 94, p. 178–192, 2016.

CEREZO, R.; LARA, J.-A.; AZEVEDO, R.; ROMERO, C. Reviewing the differences between learning analytics and educational data mining: Towards educational data science. **Computers in Human Behavior**, Elsevier, v. 154, p. 108155, 2024.

CHAUDY, Y.; CONNOLLY, T. Specification and evaluation of an assessment engine for educational games: Integrating learning analytics and providing an assessment authoring tool. **Entertainment Computing**, Elsevier, v. 30, p. 100294, 2019.

CLARK, D. B.; TANNER-SMITH, E. E.; KILLINGSWORTH, S. S. Digital games, design, and learning: A systematic review and meta-analysis. **Review of educational research**, Sage Publications Sage CA: Los Angeles, CA, v. 86, n. 1, p. 79–122, 2016.
CONDE, M. Á.; RODRÍGUEZ-SEDANO, F. J. Is learning analytics applicable and applied to education of students with intellectual/developmental disabilities? a systematic literature review. **Computers in Human Behavior**, Elsevier, p. 108184, 2024.
DAOUDI, I. Learning analytics for enhancing the usability of serious games in formal education: A systematic literature review and research agenda. **Education and Information Technologies**, Springer, v. 27, n. 8, p. 11237–11266, 2022.
DORMEZIL, S.; KHOSHGOFTAAR, T. M.; ROBINSON-BRYANT, F. Differentiating between educational data mining and learning analytics: A bibliometric approach. In: **EDM (Workshops)**. [S.l.: s.n.], 2019. p. 17–22.
FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P.; UTHURUSAMY, R. (Ed.). **Advances in knowledge discovery and data mining**. USA: American Association for Artificial Intelligence, 1996. ISBN 0262560976.
KANG, J.; LIU, M.; QU, W. Using gameplay data to examine learning behavior patterns in a serious game. **Computers in Human Behavior**, Elsevier, v. 72, p. 757–770, 2017.
MAYER, R. E. Cognitive foundations of game-based learning. The MIT Press, 2020.
MICHAEL, D. R.; CHEN, S. L. **Serious games: Games that educate, train, and inform**. [S.l.]: Muska & Lipman/Premier-Trade, 2005.
NASCIMENTO, J. B.; RODRIGUES, R. L.; ANDRADE, V. L. V. X. D. Aplicações de game learning analytics na abordagem sobre conceitos de matemática. **Revista Novas Tecnologias na Educação**, v. 19, n. 2, p. 51–60, 2021.
PETERSEN, K.; FELDT, R.; MUJTABA, S.; MATTSSON, M. Systematic mapping studies in software engineering. In: BCS LEARNING & DEVELOPMENT. **12th international conference on evaluation and assessment in software engineering (EASE)**. [S.l.], 2008.
PRENSKY, M. Digital game-based learning. **Computers in entertainment (CIE)**, ACM New York, NY, USA, v. 1, n. 1, p. 21–21, 2003.
ROMERO, C.; VENTURA, S.; PECHENIZKIY, M.; BAKER, R. S. **Handbook of educational data mining**. [S.l.]: CRC press, 2010.
SIEMENS, G.; LONG, P. Penetrating the fog: Analytics in learning and education. **EDUCAUSE review**, ERIC, v. 46, n. 5, p. 30, 2011.
SILVA, C.; FONSECA, J. Educational data mining: a literature review. **Europe and MENA Cooperation advances in information and communication technologies**, Springer, p. 87–94, 2017.
TLILI, A.; CHANG, M.; MOON, J.; LIU, Z.; BURGOS, D.; CHEN, N.-S. *et al.* A systematic literature review of empirical studies on learning analytics in educational games. International Journal of Interactive Multimedia and Artificial Intelligence ..., 2021.
VERBERT, K.; MANOUSELIS, N.; DRACHSLER, H.; DUVAL, E. Dataset-driven research to support learning and knowledge analytics. **Journal of Educational Technology & Society**, JSTOR, v. 15, n. 3, p. 133–148, 2012.