

Análise Bibliométrica da Revisão Sistemática da Literatura sobre *Learning Analytics* e *Educational Data Mining* focados na Predição de ocorrências no Sistema Educacional

Eulanda Maria Pedro Daniel

Programa de Pós-Graduação em Informática na Educação (PPGIE) - Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS – Brasil
eulanda.daniel@ufrgs.br

Leandro Krug Wives

Programa de Pós-Graduação em Informática na Educação (PPGIE) - Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS – Brasil
leandro.wives@ufrgs.br

Alexandra Lorandi

Faculdade da Educação, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS – Brasil
alexandra@ufrgs.br

Resumo: A partir de um mapeamento sistemático da literatura sobre *Learning Analytics* e *Educational Data Mining*, verificou-se suas contribuições para a predição de ocorrências no contexto educacional. Usou-se como fonte os repositórios Web of Science e Scopus, tendo sido identificados 189 artigos entre 2018 e 2022. Após a aplicação de critérios de exclusão e de inclusão, foram selecionados e analisados 41 artigos. Os resultados mostram uma tendência evolutiva na utilização dessas temáticas no contexto educacional. Além disso, verificou-se uma variedade de ferramentas e recursos utilizados na aplicação de modelos e no desenvolvimento de soluções preditivas, cujo objetivo é apoiar a produção de novos métodos e a aplicação de estratégias para a melhoria de resultados educacionais.

Palavras-chave: *Learning analytics*; *Educational Data Mining*; predição de ocorrências; análise bibliométrica; sistema educacional

Abstract: This article presents a systematic literature review on Learning Analytics (LA) and Educational Data Mining (EDM), focusing on their contributions to predicting occurrences in the educational context. Web of Science and Scopus repositories were used as a source, and 189 articles were identified between 2018 and 2022. After applying exclusion and inclusion criteria, 41 articles were selected and analyzed. The findings demonstrate a growing trend in the adoption of LA and EDM within the educational domain. Additionally, the study highlights the diverse range of

tools and resources employed in the application of models and the development of predictive solutions, aiming to support the creation of innovative methods and strategies for enhancing educational outcomes.

Keywords: Learning Analytics; Educational Data Mining; occurrence prediction; bibliometric analysis; education system

1. Introdução

Diversas pesquisas têm sido realizadas na área educacional, sobretudo relacionadas com a predição de ocorrências (i.e., desempenho escolar, evasão, abandono, retenção escolar, admissão ao curso), no intuito de apoiar a complexidade deste sistema. Nesse contexto, são empregados diversos métodos para melhor atender às demandas de alunos e educadores nos processos de ensino e aprendizagem (Costa et al., 2013). O aprimoramento da educação é uma necessidade que tem se mostrado global. Os desafios enfrentados pelas instituições educacionais são múltiplos, incluindo a necessidade de aprimorar a eficácia do ensino, a de personalizar a aprendizagem para atender às necessidades individuais dos estudantes e a de aprimorar o gerenciamento acadêmico e administrativo.

No entanto, à medida que o cenário educacional se torna cada vez mais complexo, surge a necessidade de desenvolver estratégias inovadoras para apoiar alunos, professores e gestores. A busca por métodos capazes de antecipar e solucionar questões como desempenho acadêmico e retenção escolar tem impulsionado a pesquisa em *Educational Data Mining* (EDM, do português - Mineração de Dados Educacionais) e *Learning Analytics* (LA, do português - Análise da Aprendizagem), como é salientado por Rigo et al. (2014). Essas disciplinas têm se revelado aliadas na transformação do cenário educacional ao fornecer importantes *insights* a partir da análise de dados educacionais. O propósito é fornecer informações para embasar decisões gerenciais, especialmente no que diz respeito ao aprimoramento de métodos e a implementação de técnicas que respaldam a predição de ocorrências. Segundo Paz e Gazela (2021), com a crescente disponibilidade de dados educacionais, a aplicação dessas técnicas tem se mostrado fundamental para a tomada de decisões em instituições de ensino em todo o mundo.

O foco deste trabalho é conhecer a evolução da produção científica no tema desta pesquisa, visando identificar as métricas, ferramentas e o estado da arte relacionados com a EDM e o LA aplicados no contexto de preditivo sobre ocorrências no sistema educacional. O estudo busca não apenas acompanhar as tendências emergentes nessa área, mas também

contribuir para a compreensão das abordagens mais eficazes para a predição de ocorrências no ambiente educacional, proporcionando avanços e melhorias na educação.

2. Propósito

Vários pesquisadores da área da educação vêm se dedicando a explorar técnicas e métodos semelhantes em EDM e LA, com objetivo de favorecer a melhoria do sistema educacional, quer seja entre alunos, professores ou administradores da instituição.

Segundo Rigo *et al.* (2014), a EDM pode ser definida como um procedimento que emprega algoritmos específicos para identificar padrões em conjuntos de dados educacionais, visando descrever ou antecipar eventos ou comportamentos relevantes para a atividade em questão. Exemplos de sua aplicação incluem a emissão de alertas, o suporte a sistemas de recomendação e a identificação de perfis de estudantes.

Por outro lado, a LA é um processo que realiza, segundo Carneiro *et al.* (2018), medição, coleta, análise e apresentação de dados sobre os alunos, contextualizando-os de forma a entender e aprimorar a aprendizagem e o ambiente em que esta ocorre. O objetivo da LA é ajudar professores e estudantes a atingirem seus objetivos de ensino e aprendizagem e entender o sistema na sua totalidade.

Apesar das distinções entre LA e EDM, Faria (2014) destaca uma considerável sobreposição entre as duas áreas, tanto em termos de objetivos dos pesquisadores quanto de métodos e técnicas empregados nas investigações. Enquanto a EDM foca em técnicas e tecnologias, a LA adota uma abordagem mais pragmática e aplicada.

Assim, podemos entender que a LA se baseia em uma diversidade de disciplinas acadêmicas, que vão desde a mineração de dados até conceitos e técnicas da ciência da informação, sociologia, ciência da computação, estatística, psicologia e ciências da aprendizagem (Faria, 2014). Ao contrário de EDM, LA busca compreender sistemas de forma integral e apoiar a tomada de decisões humanas, em vez de buscar respostas automáticas.

O objetivo desta pesquisa é identificar as técnicas e métodos de EDM e LA utilizados na predição de ocorrências no sistema educacional, por meio de um processo de mapeamento bibliográfico sistemático. O intuito é identificar as palavras-chave ou temas principais, bem como sua evolução, contribuindo para a gestão de recursos e a seleção de ferramentas ou técnicas de predição de ocorrências no contexto educacional.

3. Materiais e métodos

Este artigo apresenta um estudo exploratório, no qual a investigação é enquadrada como Revisão Sistemática da Literatura (RSL) baseada em pesquisa bibliográfica e bibliométrica, com abordagem quali-quantitativa. Os procedimentos adotados se baseiam no trabalho de Kitchenham e Charters (2007), onde a RSL é realizada a partir de um processo que envolve a execução das etapas de planejamento, condução e publicação de resultados.

Também é baseado no trabalho de Aria e Cuccurullo (2017), que fornece uma análise estruturada para um grande corpo de informações que infere tendências ao longo do tempo, identificando mudanças nas fronteiras das disciplinas e apresentando um “quadro geral” de pesquisas existentes.

Para automatizar a busca e as análises, foram empregadas as ferramentas Parsifal, RStudio e Bibliometrix. Parsifal é uma plataforma online (<https://parsif.al/>) projetada para facilitar a condução de RSL, seguindo o protocolo de Kitchenham e Charters (2007). Esta plataforma, de acesso livre, engloba todos os passos descritos neste artigo, proporcionando aos pesquisadores a capacidade de colaborar de maneira eficiente.

Ao final, são exibidos gráficos a partir de todos os dados extraídos e gerados relatórios completos sobre a revisão. O RStudio é um ambiente de desenvolvimento de scripts em linguagem R. Já a ferramenta Bibliometrix, por sua vez, foi concebida na linguagem R com o propósito de realizar análises bibliométricas. Ela possibilita a criação de matrizes de dados para a cocitação e investigação da colaboração científica (Aria & Cuccurullo, 2017).

O protocolo utilizado nesta revisão consistiu nas seguintes etapas: definição das questões de pesquisa, definição de palavras-chave, definição da *string* de busca, definição das bases de dados e delimitação de estudos, definição de critérios de inclusão e exclusão e seleção propriamente dita. Todas são descritas a seguir.

Etapa 1 - Definição das questões de pesquisa (QPs): as questões exploratórias definidas norteiam a RSL e estão relacionadas com o objetivo do estudo. São elas: (i) Qual é o propósito de se empregar as técnicas de *data mining* e *learning analytics* na educação? (ii) Quais palavras-chave predominam e qual a interconexão entre elas? (iii) Quais algoritmos são explorados nos trabalhos que aplicam a EDM e LA com foco na previsão de ocorrências nas instituições de ensino? (iv) Quais são as principais métricas utilizadas para avaliação do desempenho das abordagens propostas? (v) Quais as principais ferramentas utilizadas?

(vi) Quais são as principais contribuições e resultados reportados relativos à aplicação das duas técnicas na previsão de ocorrências?

Etapa 2 - Definição das palavras-chave: baseando-se nas QPs, foram definidos os termos que servem de referência para os encontrar estudos relacionados. Nesta pesquisa optou-se por buscar apenas estudos na língua inglesa. Assim, os termos definidos também estão escritos em inglês: *educational data mining, learning analytics, machine learning, predict, model*.

Etapa 3 - Definição da *string* de busca: a estratégia para a formação da principal *string* de busca foi relacionada ao objetivo do estudo, combinando os diferentes termos definidos na etapa 2. Assim, a *string* de busca foi definida como: *((("educational data mining") AND ("learning analytics" OR "machine learning")) AND ("predict" OR "model"))*.

Etapa 4 - Definição das bases de busca e delimitação dos estudos: as bases de busca foram selecionadas tendo em foco a área de Ciência da Computação, bem como as bases compatíveis com a ferramenta de análise bibliométrica Bibliometrix/Biblioshiny (Aria; Cuccurullo, 2017) que são Web of Science (WoS) e Scopus. Nesta etapa, os estudos foram limitados aos anos de 2018 a 2022, resultando em 295 estudos, dos quais 127 oriundos da base WoS e 168 da Scopus.

Etapa 5 - Definição de critérios de inclusão e exclusão: os critérios definidos estão descritos na Tabela 1, o que permitiu a avaliação da qualidade dos estudos analisados em três fases, abaixo descritas.

Tabela 1

Critérios de avaliação dos estudos. Fonte: Adaptado pelos autores.

Critérios de Inclusão	Critérios de Exclusão
<ul style="list-style-type: none">Estudos que retratam o tema;Estudos que tratam da utilização de EDM, LA ou machine learning para a predição no ambiente educacional;Estudos completos.	<ul style="list-style-type: none">Estudos duplicados;Estudos secundários (revisões sistemáticas e similares);Estudos envolvendo MOOCs;Não disponíveis para download;Estudos que não estejam na língua inglesa.

Fase 1 - Remoção dos estudos duplicados: Nesta primeira fase, utilizaram-se as ferramentas RStudio e Parsifal para identificar e remover estudos duplicados. Essa etapa é essencial para garantir a

integridade dos resultados, eliminando possíveis repetições na amostra. Ao final desta fase, foram removidos 106 estudos duplicados, resultando em um conjunto mais limpo e consistente de 189 estudos, dos quais 148 foram provenientes da base Scopus e 41 da WoS.

Fase 2 - Seleção pela leitura do título, resumo e palavras-chave:

Nesta fase, os 189 estudos restantes foram submetidos a uma análise mais detalhada. Foram avaliados o título, o resumo e as palavras-chave de cada estudo para determinar sua relevância em relação ao tema da pesquisa. Dessa fase, foram mantidos 75 estudos que atenderam aos critérios de interesse, sendo 48 provenientes da Scopus e 27 da WoS.

Fase 3 - Seleção pela leitura completa dos estudos: Na terceira e última fase, os 75 estudos selecionados na fase anterior foram lidos na íntegra. O objetivo consistiu em verificar se os estudos eram compatíveis com o escopo e os objetivos da pesquisa, respondendo às Questões de Pesquisa (QPs) previamente definidas. No final 41 estudos foram mantidos, sendo 26 provenientes da Scopus e 15 da WoS, compreendendo o conjunto final de estudos analisados.

Essas três fases de seleção garantiram uma abordagem criteriosa e robusta na escolha das publicações, assegurando que apenas aquelas mais pertinentes e alinhadas aos objetivos da pesquisa fossem incluídas na análise.

A execução das cinco etapas do protocolo foi essencial para assegurar que apenas os estudos mais relevantes e alinhados aos objetivos da pesquisa fossem incluídos na análise. Isso fortaleceu a validade e confiabilidade dos resultados obtidos, conferindo à pesquisa uma base sólida e bem fundamentada. O protocolo adotado foi um elemento-chave para o sucesso e a credibilidade desta investigação.

4. Discussão e análise dos estudos

Esta seção apresenta análises detalhadas dos dados coletados a partir dos 41 estudos selecionados. Não se trata apenas de uma apresentação de estatísticas, mas sim de uma imersão na complexidade e nas nuances das pesquisas examinadas. Por meio de análises detalhadas e visualizações gráficas, buscamos extrair *insights* que respondam às questões de pesquisa propostas.

As figuras e tabelas apresentadas fornecem uma visão abrangente das principais tendências e padrões identificados. A Figura 1 apresenta um panorama das estatísticas essenciais, incluindo o número de referências na

área, a contribuição de autores e o crescimento ao longo dos anos. Para melhor compreensão, a Figura 2 oferece uma representação gráfica da trajetória da produção de estudos ao longo do tempo.

Em paralelo, foi necessário visualizar, por meio do gráfico da Figura 3, os autores que mais produziram esses estudos ao longo do tempo definido. Essas informações favoreceram a observação e extração dos dados necessários para responder às QPs definidas neste estudo.

Figura 1

Visão geral das principais informações dos dados selecionados. Fonte: Adaptado pelos autores, 2023.



Figura 2

Tendência das produções dos estudos. Fonte: Adaptado pelos autores, 2023.

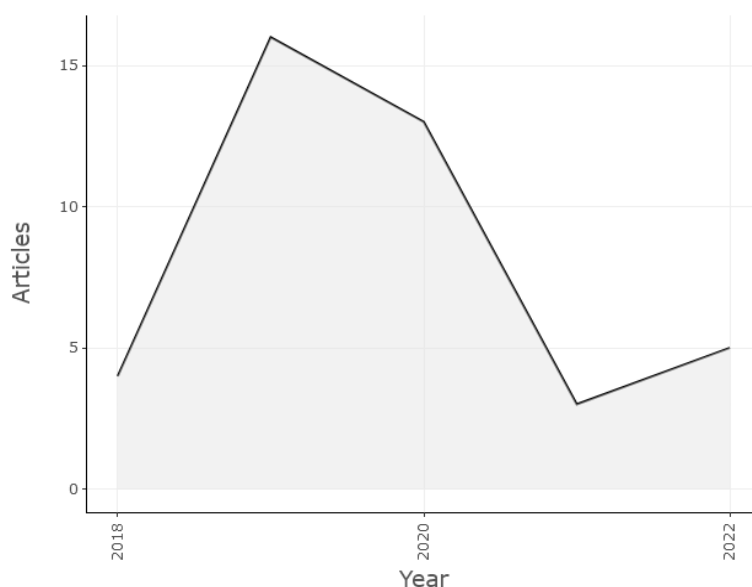
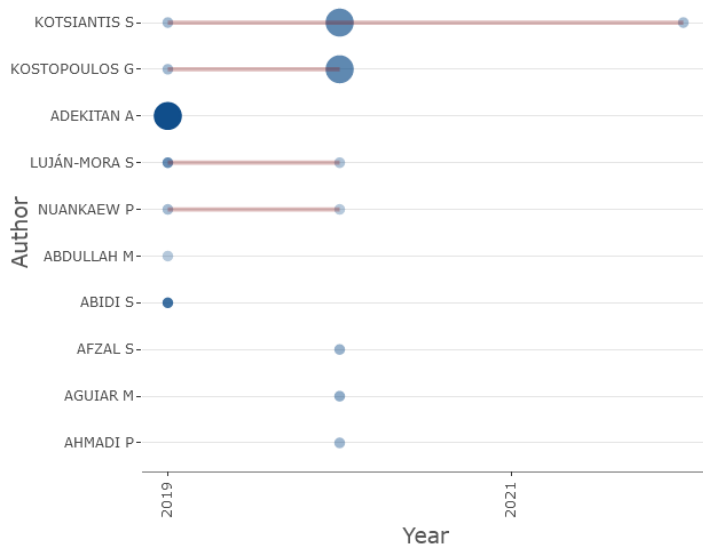


Figura 3

Autores que mais produziram. Fonte: Adaptado pelos autores, 2023.

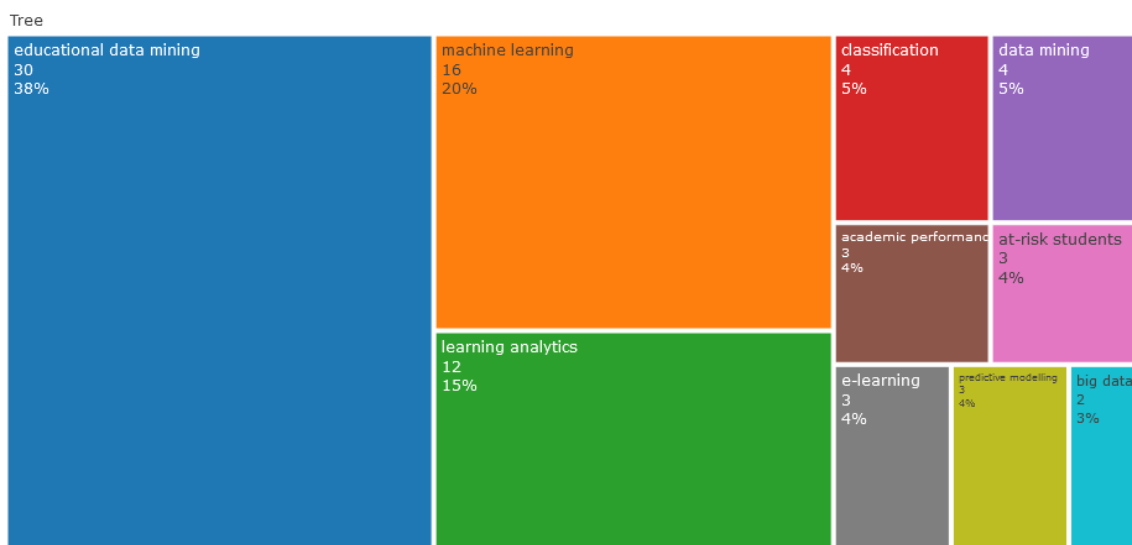


A partir dos dados coletados, ao abordar a primeira questão de pesquisa (QP(i)), ficou evidente que os 41 estudos e os 135 autores selecionados compartilham um objetivo comum: implementar modelos e técnicas para a coleta, tratamento e análise de dados massivos relacionados ao Big Data. Essa abordagem oferece uma rica fonte de informações para a pesquisa de padrões e a interpretação pedagógica, visando aprimorar as práticas de aprendizagem e otimizar o desempenho tanto de professores quanto de alunos, bem como o modelo educacional como um todo.

A segunda questão de pesquisa (QP (ii)) demandou uma análise bibliométrica para identificar as principais palavras-chave utilizadas pelos autores. Do resultado obtido, foi possível extrair três principais palavras-chave pela maior frequência e por estarem diretamente ligadas ao contexto deste estudo. Conforme apresentado na Figura 4, 38% das palavras-chave representam EDM, 20% *machine learning* e 12% LA.

Figura 4

Árvore das principais palavras-chave. Fonte: Adaptado pelos autores, 2023.

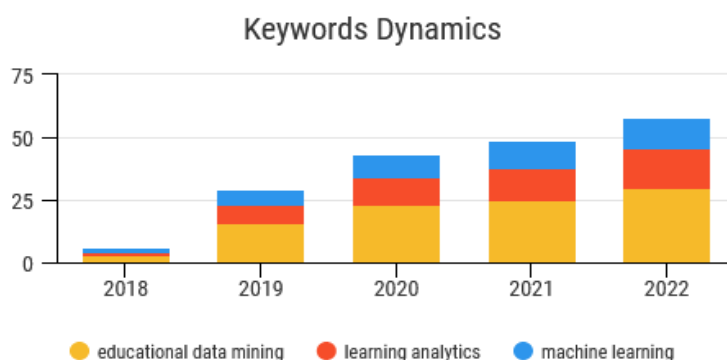


A partir das palavras-chave selecionadas, realizou-se uma medição da proporção dinâmica de forma a perceber a sua evolução ao longo do tempo. Conforme apresentado na Figura 5, os tópicos eram igualmente incipientes em 2018. A partir de 2019 nota-se um crescimento (ou interesse) pelo tema, chegando ao seu auge em 2022.

O mais interessante nessa análise é que a EDM apresenta maior quantidade (sendo a que mais cresce) em todos os anos. Embora esses temas sejam conflitantes em sua definição, eles têm como ponto em comum a maneira com que os dados educacionais são analisados (técnicas analíticas), diferenciando-se basicamente na abordagem em que se coloca cada tipo de problema.

Figura 5

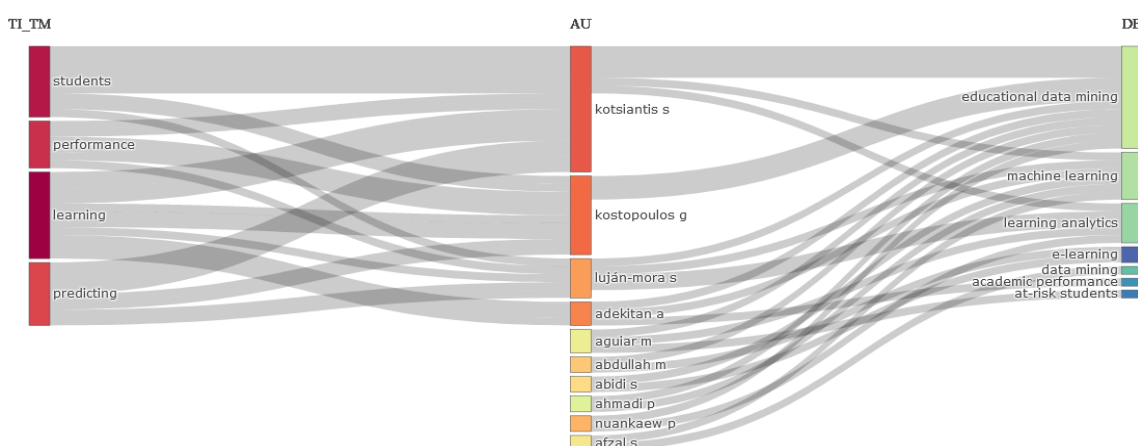
Dinâmica evolutiva das principais palavras-chave. Fonte: Adaptado pelos autores, 2023.



A análise subsequente aborda as questões de pesquisa (QP (iii, iv, v, vi)) por meio de seleção dos autores que mais publicam estudos sobre as três principais palavras-chave selecionadas (EDM, ML e LA). A Figura 6 apresenta uma visualização relacional por meio de um gráfico de Sankey de três pontas, evidenciando as conexões entre os autores e suas palavras-chave, proporcionando insights valiosos para a pesquisa.

Figura 6

Relação entre os autores e suas palavras-chave e palavras que contextualizam o cenário. Fonte: Adaptado pelos autores, 2023



A Figura 6 destaca que 10 autores desempenham um papel central na relação entre as palavras-chave. Dentre esses, quatro autores (AU) concentram-se em temas essenciais (TI_TM) para o contexto do estudo (DE), como desempenho, aprendizado e predição. Esses autores se destacaram na publicação de estudos relacionados aos três tópicos (*performance*, *learning* e *prediction*) e foram objeto de uma análise mais aprofundada. Tais condições foram determinantes para a seleção dos autores que visam responder às QPs.

Embora todos os estudos tenham sido analisados na íntegra, os quatro autores são os que mais se destacaram na publicação de estudos relacionados com os três tópicos. Assim sendo, para cada autor, fez-se um levantamento do número de publicações correspondentes, conforme apresentado na Tabela 2.

Um dado de análise interessante encontrado nesta pesquisa é que alguns desses autores aparecem como coautores entre eles e/ou com outros autores não citados, e essas publicações foram contabilizadas para cada coautor e autor correspondente. Essa ação faz com que a quantidade

frequente dos estudos seja reduzida a 7, culminando numa análise exploratória da RSL, condensada na Tabela 3.

Tabela 2

Distribuição da frequência das publicações.

Autor	Qtd. inicial	Qtd. frequente
Adekitan A	1	1
Kostopoulos G	3	2
Kotsiantis S	4	2
Luján-Mora S	2	2
Total Geral	10	7

Tabela 3

Análise exploratória dos estudos dos autores selecionados. Fonte: Adaptado pelos autores, 2023.

Artigo	Propósito	Algoritmos	Palavras-Chave	Métricas	Ferramentas	Contribuições e Resultados
Gkantzis, A. F., Kotsiantis, S., Panagiotakopoulos, C. T., & Verykios, V. S. (2022)	Entender os principais determinantes do abandono, identificar com precisão alunos com probabilidade de abandono e recomendar intervenções para reduzi-la	Redes Bayesianas, Naive Bayes, Multilayer Perceptron, SMO, KNN, C4.5, JRIP e Random Forest	EDM; big data; ML; distance learning; LA	Área sob a Curva ROC (AUC), Precisão e Revocação	Hadoop, Weka e scikit-learn (Python)	A contribuição deste trabalho são as previsões oportunas do desgaste dos alunos para as semanas de um ano letivo e o desenvolvimento de uma ferramenta preditiva front-end e amigável para cada semana, fornecendo suporte tangível aos professores e autoavaliação aos alunos. Seguindo o fluxo de trabalho do estudo, uma ferramenta de previsão foi desenvolvida para cada categoria- alvo, auxiliando a atenção dos tutores para a tendência de reprovação dos alunos. Além disso, os alunos têm a oportunidade de identificar sua possível reprovação em cada semana acadêmica e aplicar habilidades de aprendizagem autorregulada. O reconhecimento semanal preciso do desgaste do aluno desde o início do ano letivo tem o potencial de minimizar a falta de contato pessoal no ensino a distância e maximizar intervenções personalizadas e serviços de aprendizado pessoal.
Karlos, Kostopoulos e Kotsiantis (2020)	Implementar um algoritmo de semi-regressão para prever as notas de alunos de graduação nos exames finais de um curso on-line que explora três visões de recursos independentes e naturalmente formadas	KNN, Regressão Linear típica, Gradient Boosting com funções de perda (ls e huber), regressor MLP	EDM; student grade prediction; semi-regression; early prognosis; interpretation; COREG algorithm	Erro absoluto médio (MAE)	Scikit-learn (Python) e Framework COREG	Contribui com uma perspectiva diferente para os métodos de visualização única existentes, explorando totalmente o potencial de diferentes subconjuntos de recursos, estendendo a estrutura COREG para a configuração de múltiplas visualizações. Além disso, aponta a importância de atributos específicos que influenciam fortemente o atributo alvo. Os resultados experimentais demonstram que o prognóstico precoce de alunos em risco de reprovação pode ser alcançado com precisão em comparação aos modelos supervisionados, mesmo para uma pequena quantidade de dados coletados inicialmente nos dois primeiros semestres.

Tabela 4 (cont.)

Análise exploratória dos estudos dos autores selecionados. Fonte: Adaptado pelos autores, 2023.

Artigo	Propósito	Algoritmos	Palavras-Chave	Métricas	Ferramentas	Contribuições e Resultados
Tsiakmaki, M., Kostopoulos, G., Kotsiantis, S., & Ragos, O. (2020)	Propor uma metodologia de transferência de aprendizagem para prever o desempenho de alunos no ensino superior	Redes Neurais profundas	Transfer learning; deep learning; EDM; student performance prediction	Média de acurácia	Keras (Python)	Contribui com uma proposta de um método de aprendizagem por transferência para a tarefa de prever o desempenho do aluno em cursos de graduação. O aprendizado por transferência permite realizar o treinamento de uma rede profunda usando o conjunto de dados de um curso anterior (curso de origem) e reutilizá-lo como ponto de partida para um conjunto de dados de um novo curso relacionado (curso alvo). Além disso, é possível ajustar ainda mais o modelo reaproveitado. As descobertas do estudo provaram que um desempenho razoável foi alcançado na maioria dos casos, enquanto o método proposto supera com folga o modelo de linha de base. Os resultados são encorajadores e devem ser validados por amostras maiores de cursos de diferentes departamentos e programas.
Kostopoulos, G., Kotsiantis, S., Fazakis, N., Koutsonikos, G., & Pierrakeas, C. (2019)	Examinar a eficácia de um algoritmo SSL proposto para tarefas de regressão no campo educacional, em particular para previsão de notas em exames finais de um curso online a distância	Regressão Linear; SMOreg; KNN; M5 Rules; Tree M5; Random Forest	EDM; distance learning; grade prediction; semi supervised Regression; MSSRA	MAE, erro absoluto relativo, erro quadrático médio, coeficiente de correlação de Pearson	WEKA	Um algoritmo SSR, denominado MSSRA, é proposto para previsão de notas de alunos em exames finais de um curso a distância. Eles usaram uma série de atributos relacionados a trabalhos escritos, presença ou ausência nas sessões de contato opcionais agendadas e interação dos alunos dentro do LMS do curso ao longo do ano letivo e vários experimentos foram realizados para avaliar a eficácia do algoritmo em comparação com a regressão familiar métodos. Os resultados verificaram não só a superioridade do algoritmo MSSRA, mas também indicaram que uma previsão precoce da nota dos alunos é viável, permitindo assim que os professores façam intervenções personalizadas e eficazes para os alunos de baixo desempenho e, finalmente, melhorem seu desempenho acadêmico.

Tabela 5 (cont.)

Análise exploratória dos estudos dos autores selecionados. Fonte: Adaptado pelos autores, 2023.

Artigo	Propósito	Algoritmos	Palavras-Chave	Métricas	Ferramentas	Contribuições e Resultados
Villegas-Ch, Palacios-Pacheco e Luján-Mora (2020)	Implementar um <i>framework</i> de BI para melhorar os métodos ou atividades que geram aprendizado ao aluno.	Árvore de decisão (J48)	Business intelligence; EDM; learning management systems; LA	Precisão, taxa de erro, revocação e f-measure	Repositório DW e WEKA	Apresenta um método para a implementação de um <i>framework</i> de Business Intelligence (BI) que abrange tratamento, análise e apresentação de dados que podem responder a qualquer problema gerado em uma instituição de ensino. Como resultado, verificou-se que 37% dos alunos trabalham e passam em todos os cursos, sem problemas, e os outros 63% vivem alguma complicação. A análise realizada mostra que 63% dos alunos precisam de pelo menos 19 horas de uso do <i>Learning Management System</i> (LMS) para se familiarizar com sua gestão.
Bueaño-Fernandez, Gil e Luján-Mora (2019)	Prever as notas dos alunos de acordo com diversas características de seu desempenho acadêmico	Árvore de decisão (não mencionado o tipo)	EDM; LA; ML; big data; prediction grades	Precisão e matriz de confusão	WEKA	Apresenta uma análise dos GFs (notas finais) por áreas de conhecimento (p.ex., banco de dados e infraestrutura de rede). Pretende-se assim justificar que as notas numa disciplina podem ser previstas a partir das notas dos alunos nos anos letivos anteriores. Por exemplo, os GFs da disciplina de Certificação de Bancos de Dados podem ser previstos a partir dos GFs das disciplinas Bancos de Dados I, Bancos II e Administração de Bancos de Dados, enquanto os GFs da disciplina Certificação de Redes podem ser previstos a partir dos GFs das disciplinas Redes I e Redes II. Como resultado da pesquisa, as autoridades da universidade aprovaram a alteração do percentual atribuído a cada PG (notas parciais). Dessa forma, foi possível melhorar as notas e o desempenho acadêmico dos alunos do PG3, bem como reduzir a taxa de absenteísmo dos alunos ao final de cada período letivo (PG3).
Adekitan e Noma-Osaghae (2019)	Determinar a extensão da relação entre critérios de admissão usados pelas universidades nigerianas e o	Random Forest, Tree Ensemble; Decision Tree, Naïves Bayes, Logistic Regression e Resilient Back	Academic Performance; ML; EDM; data mining algorithms; knowledge discovery; Nigerian	AUC, precisão, revocação e F1-score	KNIME e Orange	A análise neste estudo revela até que ponto os requisitos de admissão podem ser usados como preditores de desempenho universitário, tendo a Universidade Covenant, Nigéria, como estudo de caso. Algoritmos e modelos foram criados e treinados usando KNIME e Orange. Análise de regressão mostrou que existe uma relação entre os requisitos de admissão e o desempenho acadêmico do aluno admitido, no primeiro ano, mas essa relação não é muito

Tabela 6 (cont.)

Análise exploratória dos estudos dos autores selecionados. Fonte: Adaptado pelos autores, 2023.

Artigo	Propósito	Algoritmos	Palavras-Chave	Métricas	Ferramentas	Contribuições e Resultados
	desempenho acadêmico do aluno após a primeira sessão acadêmica	Propagation MultiLayer Perceptron	university			forte como revelado pela precisão de outros algoritmos preditivos. Os critérios de seleção para admissão na universidade na Nigéria são baseados principalmente no perfil acadêmico dos candidatos, e o resultado desse estudo, portanto, enfatiza a necessidade de reavaliar esses critérios e de considerar outros fatores não acadêmicos, tais como indicadores de dedicação, motivação e paixão do aluno pelo sucesso.

5. Conclusão

Este estudo exploratório e quali-quantitativo revelou a tendência evolutiva recente dos campos de *Educational Data Mining*, *Learning Analytics* e *Machine Learning* na educação. As análises mostram que a pesquisa na área tem sido cada vez mais ativa, com foco no uso das principais técnicas de computação e inteligência para desenvolver novos métodos que melhorem os resultados educacionais. Os resultados apresentam modelos de predição que detectam diferentes situações no contexto educacional, evidenciando uma comunidade diversa e inclusiva de pesquisadores.

Destaca-se, ainda, a relevância dos resultados encontrados, proporcionando uma visão clara das tendências e direções futuras no uso de técnicas analíticas e de mineração de dados na educação. As implicações práticas desses achados são significativas, podendo orientar educadores, administradores e formuladores de políticas na implementação de intervenções mais eficazes e personalizadas.

Perspectivas futuras incluem a avaliação de outros critérios para a análise que não tenham sido sintetizados, bem como a apresentação de um gráfico relacional entre os autores dos 41 estudos mantidos. Além disso, sugere-se a elaboração de um guia de definição de critérios de qualidade para a seleção de estudos na área educacional, o que poderia contribuir para a padronização e transparência na seleção de trabalhos.

Acreditamos que os insights obtidos neste estudo têm o potencial de impactar positivamente a área da educação, auxiliando na tomada de decisões e na melhoria das práticas educacionais. Ao continuar avançando nesse campo, esperamos ver uma crescente integração dessas técnicas inovadoras, promovendo um ambiente de aprendizado mais eficaz e adaptável às necessidades dos alunos.

6. Referências

- Adekitan, A. I., & Noma-Osaghae, E. (2019). Data mining approach to predicting the performance of first year student in a university using the admission requirements. *Education and Information Technologies*, 24(2), 1527-1543.
- Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959-975.
- Buenaño-Fernández, D., Gil, D., & Luján-Mora, S. (2019). Application of machine learning in predicting performance for computer engineering students: A case study. *Sustainability*, 11(10), 2833.

- Carneiro, J. N., et al. (2018). Aplicação de Learning Analytics para Modelagem do Aluno de acordo com a Taxonomia de Bloom Revisada.
- Conforto, E. C., Amaral, D. C., & Silva, S. L. da. (2011). Roteiro para revisão bibliográfica sistemática: aplicação no desenvolvimento de produtos e gerenciamento de projetos. Trabalho apresentado, 8.
- Costa, E., Aguiar, J., & Magalhães, J. (2013). Sistemas de Recomendação de Recursos Educacionais: conceitos, técnicas e aplicações. *Jornada de Atualização em Informática na Educação*, 1(1).
- Faria, S. M. S. M. L. de. (2014). Educational data mining e learning analytics na melhoria do ensino online. (Tese de Doutorado).
- Gkontzidis, A. F., Kotsiantis, S., Panagiotakopoulos, C. T., & Verykios, V. S. (2022). A predictive analytics framework as a countermeasure for attrition of students. *Interactive Learning Environments*, 30(6), 1028-1043.
- Karlos, S., Kostopoulos, G., & Kotsiantis, S. (2020). Predicting and interpreting students' grades in distance higher education through a semi-regression method. *Applied Sciences*, 10(23), 8413.
- Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. *Technical Report EBSE-2007-01, School of Computer Science and Mathematics, Keele University*.
- Kostopoulos, G., Kotsiantis, S., Fazakis, N., Koutsonikos, G., & Pierrakeas, C. (2019). A semi-supervised regression algorithm for grade prediction of students in distance learning courses. *International Journal on Artificial Intelligence Tools*, 28(04), 1940001.
- Paz, F. J., & Cazella, S. C. (2021). Reflexões sobre um modelo de Analítica Acadêmica em instituições comunitárias de ensino superior: estudo piloto. *Revista on line de Política e Gestão Educacional*, 2178-2192.
- Rigo, S. J., Cambruzzi, W., Barbosa, J. L., & Cazella, S. C. (2014). Aplicações de Mineração de Dados Educacionais e Learning Analytics com foco na evasão escolar: oportunidades e desafios. *Revista Brasileira de Informática na Educação*, 22(01), 132.
- Siemens, G., & Baker, R. S. J. d. (2012). Learning analytics and educational data mining: towards communication and collaboration. In: *Proceedings of the 2nd international conference on learning analytics and knowledge*, 252-254.
- Tsiakmaki, M., Kostopoulos, G., Kotsiantis, S., & Ragos, O. (2020). Transfer learning from deep neural networks for predicting student performance. *Applied Sciences*, 10(6), 2145.
- Villegas-Ch, W., Palacios-Pacheco, X., & Luján-Mora, S. (2020). A business intelligence framework for analyzing educational data. *Sustainability*, 12(14), 5745.