

# Predição de Desempenho de Estudantes: Uma Revisão Sistemática de Literatura

Bruno João da Silva<sup>1</sup>, Edson P. Pimentel<sup>1</sup>, Wagner Tanaka Botelho<sup>1</sup>

<sup>1</sup>Universidade Federal do ABC - UFABC - Santo André - SP - Brasil

bruno.joao@aluno.ufabc.edu.br, {edson.pimentel, wagner.tanaka}@ufabc.edu.br

**Abstract.** *Artificial Intelligence in Education has been used to deal with problems of student dropout and academic performance. This article aims to present a Systematic Literature Review (RSL) of works that approach student performance prediction. The work sought to answer the following questions: (a) How are student prediction models used in the teaching-learning process? (b) Which set of variables best explains the prediction of student performance? (c) What does a student performance prediction model look like? The results show that although there are works that present high accuracy values in the prediction, there is still an excellent opportunity to relate and show the benefits for students and teachers.*

**Resumo.** *A Inteligência Artificial (IA) na Educação tem sido utilizada para lidar com problemas de evasão e desempenho acadêmico de estudantes. Este artigo tem por objetivo apresentar uma Revisão Sistemática de Literatura (RSL) de trabalhos que abordam a predição de desempenho de estudantes. O trabalho buscou responder as seguintes questões: (a) Como os modelos de predição de estudante são utilizados no processo de ensino-aprendizagem; (b) Qual o conjunto de variáveis que melhor explica a predição de desempenho dos estudantes; (c) Como é um modelo de predição do desempenho dos estudantes. Os resultados mostram que embora haja trabalhos que apresentam altos valores de acurácia na predição, ainda existe uma grande oportunidade em relacionar e evidenciar quais são os benefícios para os estudantes e/ou professores.*

## 1. Introdução

Pesquisas apontam que nas últimas décadas houve uma grande expansão da educação superior tanto no Brasil como em outros países (DIOGO et al., 2016; MUSSO; HERNÁNDEZ; CASCALLAR, 2020). No entanto, diversos problemas associados ao desempenho e permanência dos estudantes permanecem. Estes problemas são retratados tanto em trabalhos mais antigos como o de Lopes (1999) ou mais recentes como o de Digiampietri, Nakano e Lauretto (2016).

Para lidar com estes problemas de evasão e desempenho estudos como o de Bloom (1984) apresentam o processo de individualização do ensino como uma alternativa para redução de reprovações e abandono. O estudo observou que estudantes com um processo de aprendizagem individualizado obtiveram melhores resultados. Dada a inviabilidade de se ter um professor para cada estudante esta personalização pode ser facilitada pelo uso da Inteligência Artificial na Educação. Nesse sentido, o trabalho de Sekeroglu, Dimililer e Tuncal (2019) mostram que as instituições estão cada vez mais combinando o

modelo tradicional de ensino com as tecnologias de inteligência artificial, com o objetivo de incrementar o resultado no processo de ensino e aprendizagem.

Diversos trabalhos mostram o uso da Inteligência Artificial para tarefas de predição, seja de desempenho ou de evasão dos estudantes, tais como : Zollanvari et al. (2017), Asiah et al. (2019), Hellas et al. (2018) e Alyahyan e Düştegör (2020).

Este artigo tem por objetivo apresentar uma Revisão Sistemática de Literatura (RSL) de trabalhos que abordam a predição de desempenho de estudantes. O trabalho buscou responder as seguintes questões: (a) Como os modelos de predição de estudante são utilizados no processo de ensino-aprendizagem; (b) Qual o conjunto de variáveis que melhor explica a predição de desempenho dos estudantes; (c) Como é um modelo de predição do desempenho dos estudantes. Os resultados mostram que embora exista um alto poder de predição do desempenho dos estudantes, ou seja, os trabalhos apresentam altos valores de acurácia, ainda existe uma grande oportunidade em relacionar e principalmente evidenciar quais são os benefícios para os estudantes e/ou professores.

O artigo está organizado da seguinte forma: a Seção 2 apresenta a metodologia utilizada na RSL; a Seção 3 apresenta e discute os resultados da RSL; e a Seção 5 estabelece as considerações finais.

## **2. Metodologia**

O presente trabalho traz um mapeamento a partir de uma metodologia conhecida como revisão sistemática. Para entender o que é essa abordagem metodológica, o estudo de Cook, Mulrow e Haynes (1997) faz uma comparação entre os dois tipos de revisões a partir de 4 tópicos:

- **Questão:** Para a revisão sistemática é exigido que a questão de pesquisa seja específica, enquanto que na revisão não sistemática a especificidade não é obrigatória;
- **Fonte:** Enquanto que na revisão sistemática a escolha das fontes segue uma metodologia ordenada, na revisão não sistemática não há um protocolo;
- **Seleção:** em uma revisão não sistemática é aceita a seleção aleatória de artigos e os critérios não precisam ser discutidos com o leitor; o inverso é obrigatório na revisão sistemática;
- **Avaliação e Síntese:** os aspectos quantitativos e qualitativos são exigidos na revisão sistemática enquanto que na revisão não sistemática é aceito apenas um dos métodos.

### **2.1. Questões de pesquisas**

Para atender os objetivos do trabalho apresenta-se 3 questões principais destacadas em negrito, vide Tabela 1, compostas por um conjunto de perguntas secundárias que auxiliam na elucidação do objeto de pesquisa. A primeira questão refere-se ao modelo de predição e ao processo de ensino e aprendizagem, a segunda questão tem por objetivo responder quais conjuntos de variáveis trazem um maior poder preditivo e, por fim, a terceira questão refere-se às técnicas de predição encontradas em literatura.

### **2.2. Definição de uma estratégia de busca**

Para a definição de uma string de busca foram utilizados termos que pudessem capturar todas as variações para o mesmo tema de estudo. Com isso a tabela 2 apresenta a string

**Tabela 1. Questões de Pesquisa da RSL**

<b>Código</b>	<b>Questões</b>
<b>P1</b>	<b>Como os modelos de predição de estudante são utilizados no processo de EA?</b>
<b>P1E1</b>	Como os alunos utilizam os modelos de predição?
<b>P1E2</b>	Como os professores utilizam os resultados da predição dos alunos?
<b>P1E3</b>	É possível afirmar que o modelo de predição leva a um maior desempenho dos estudantes?
<b>P2</b>	<b>Qual o conjunto de variáveis que melhor explica a predição de desempenho dos estudantes?</b>
<b>P2E1</b>	Quais conjuntos de variáveis são utilizados para o processo de predição de desempenho dos estudantes?
<b>P2E2</b>	O engajamento (participação) do aluno é avaliado dentro do modelo?
<b>P2E3</b>	Informações socioeconômicas do aluno são utilizadas na modelagem de predição?
<b>P3</b>	<b>Como é um modelo de predição de desempenho dos estudantes?</b>
<b>P3E1</b>	Quais as técnicas de IA são utilizadas para a predição de desempenho dos estudantes?
<b>P3E2</b>	Quais são as métricas de avaliação da acuracidade da predição?
<b>P3E3</b>	Qual é a acurácia encontrada nos modelos de predição de desempenho dos estudantes?

que correlaciona os estudos em relação à predição, sucesso do aluno e inteligência artificial. Vale destacar que a definição do termo de busca foi realizada em inglês para que pudesse trazer os artigos no mesmo idioma, conforme os critérios de inclusão. Na tabela 2 a apresentação da string:

**Tabela 2. Definição de uma estratégia de busca**

<b>Busca String</b>	<b>String</b>
<b>Definição</b>	(( performance OR fail OR successful ) AND ( students OR learners ) AND ( intelligence AND artificial ) OR ( machine AND learning ) OR ( learning AND analytics )))

### 2.3. Critérios de inclusão e exclusão

Os critérios de inclusão e exclusão foram definidos com base na revisão de literatura de autores como Santos et al. (2019) e Alamri e Alharbi (2021). Foram selecionados os seguintes Critérios de Exclusão (CE):

- CE-1: Documentos duplicados;
- CE-2: Documentos com publicação anterior a 2016;
- CE-3: Documentos que não sejam escritos na língua inglesa;
- CE-4: Documentos que tiveram menos de 10 citações;
- CE-5: Documentos que apresentam um número menor do que 10 páginas.

Para os Critérios de Inclusão (CI) foram definidos os seguintes pontos:

- CI-1: Documentos disponíveis publicamente para leitura (Open Source);

- CI-2: Documentos correlacionados a ciência da computação ou aplicados na área de educação;
- CI-3: Documentos de aplicação ou revisão em predição de desempenho de estudantes.

## 2.4. Avaliação Quantitativa

Os artigos foram submetidos a 4 critérios que resultaram em uma avaliação binária com resultados de 0 ou 1, sendo 0 para ausência da informação e 1 para a existência da mesma. Os critérios são listados a seguir e o resultado da aplicação dos critérios consta nas tabelas 3 e 4:

- A1 - Objetivo/Contribuição: identificação do objetivo da pesquisa e contribuições pretendidas ao longo do projeto.
- A2 - Técnica Utilizada: Descrição da técnica utilizada e sua motivação para a escolha.
- A3 - Resultados: Apresentação clara dos resultados e discussão com o leitor.
- A4 - Evidência da contribuição para o aluno ou professor: O artigo utiliza o modelo de predição desenvolvido para evidenciar quais são as contribuições fornecidas para aluno e ou professor.

## 3. Resultados

Os resultados serão apresentados em três blocos. O primeiro abordará o processo de extração dos dados, o segundo deverá tratar da avaliação dos artigos selecionados e o terceiro bloco irá responder as questões levantadas nesse estudo.

### 3.1. Extração de dados

Dois bancos de dados foram selecionados para capturar os dados pertinentes à esta pesquisa, o Web of Science e o Scopus. A string definida foi buscada através dos títulos, resumos e palavras chaves e foram obtidos, respectivamente, 241 e 279 artigos.

O primeiro passo foi conhecer o mapa de palavras deste artigo, conforme a figura 1, que evidencia as principais palavras chaves destes artigos as quais destacam-se: data mining, machine learning, performance prediction. Os termos são os mesmos listados na busca da string, reforçando a coerência do estudo.

Na sequência procurou-se remover de forma automática a duplicidade de artigos com o uso da ferramenta Mendley, excluindo 108 destes e outros 12 artigos foram excluídos através de um processo manual. O segundo critério de exclusão se deu pela quantidade mínima de 10 citações em cada artigo, resultando na exclusão de mais 96 documentos. Por fim, outros 85 artigos foram retirados por não atenderem aos requisitos de tamanho do documento (49) e avaliação das áreas de pesquisa (36).

Em andamento ao processo de exclusão procurou-se avaliar a relação entre o título dos artigos e seus objetivos, eliminando outros 159 estudos. A exclusão de mais 25 artigos baseou-se na avaliação dos resumos que não estavam dentro dos critérios propostos.

Os 35 artigos que restaram foram submetidos à pesquisa no google scholar de forma manual. O critério de exclusão exigiu que o documento apresenta-se pelo menos 2 citações para os artigos publicados em 2021, ao menos 3 citações para artigos publicados em 2020 e o mínimo de 10 citações para documentos publicados anteriormente à 2020. Obtivemos 22 artigos para a revisão sistemática que serão objetos deste estudo.



**Tabela 3. Avaliação dos artigos selecionados - parte 1**

Artigo	Referência	Título	A1	A2	A3	A4	Total
AR01	(ZOHAIR; MAHMOUD, 2019)	Prediction of Student's performance by modelling small dataset size	1	1	1	0	3
AR02	(ADNAN et al., 2021)	Predicting at-Risk Students at Different Percentages of Course Length for Early Intervention Using Machine Learning Models	1	1	1	1	4
AR03	(AKÇAPINAR; ALTUN; AŞKAR, 2019)	Using learning analytics to develop early-warning system for at-risk students	1	1	1	1	4
AR04	(ALMASRI; CELEBI; ALKHAWAL- DEH, 2019)	EMT: Ensemble Meta-Based Tree Model for Predicting Student Performance	1	0	1	0	2
AR05	(ALSHANQITI; NAMOUN, 2020)	Predicting Student Performance and Its Influential Factors Using Hybrid Regression and Multi-Label Classification	1	1	0	0	2
AR06	(AZCONA; HSIAO; SMEATON, 2019)	Detecting students-at-risk in computer programming classes with learning analytics from students' digital footprints	1	1	0	1	3
AR07	(BUJANG et al., 2021)	Multiclass Prediction Model for Student Grade Prediction Using Machine Learning	1	1	1	0	3
AR08	(DEO et al., 2020)	Modern Artificial Intelligence Model Development for Undergraduate Student Performance Prediction: An Investigation on Engineering Mathematics Courses	1	1	1	0	3
AR09	(GAŠEVIĆ et al., 2016)	Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success	1	1	0	0	2
AR10	(IMRAN et al., 2019)	Student Academic Performance Prediction using Supervised Learning Techniques	1	1	1	0	3
AR11	(KHAN et al., 2021)	An artificial intelligence approach to monitor student performance and devise preventive measures	1	1	1	1	4

**Tabela 4. Avaliação dos artigos selecionados - parte 2**

<b>Artigo</b>	<b>Referência</b>	<b>Título</b>	<b>A1</b>	<b>A2</b>	<b>A3</b>	<b>A4</b>	<b>Total</b>
AR12	(LAGUS et al., 2018)	Transfer-Learning Methods in Programming Course Outcome Prediction	1	1	1	0	3
AR13	(MARBOUTI; DIEFES-DUX; MADHAVAN, 2016)	Models for early prediction of at-risk students in a course using standards-based grading	1	1	1	0	3
AR14	(MUSSO; HERNÁNDEZ; CASCALLAR, 2020)	Predicting key educational outcomes in academic trajectories: a machine-learning approach	1	1	0	0	2
AR15	(ZOLLANVARI et al., 2017)	Predicting Students' GPA and Developing Intervention Strategies Based on Self-Regulatory Learning Behaviors	1	1	1	0	3
AR16	(POLYZOU; KARYPIS, 2016)	Grade prediction with models specific to students and courses	1	1	1	0	3
AR17	(POPESCU; LEON, 2018)	Predicting Academic Performance Based on Learner Traces in a Social Learning Environment	1	1	1	0	3
AR18	(WAKELAM et al., 2020)	The potential for student performance prediction in small cohorts with minimal available attributes	1	1	1	1	4
AR19	(WANG et al., 2020)	Student Performance Prediction with Short-Term Sequential Campus Behaviors	1	1	1	0	3
AR20	(WEI et al., 2020)	Predicting Student Performance in Interactive Online Question Pools Using Mouse Interaction Features	1	1	1	0	3
AR21	(YU; WU; LIU, 2019)	Predicting Learning Outcomes with MOOC Clickstreams	1	1	1	0	3
AR22	(ZHAO et al., 2020)	Academic Performance Prediction Based on Multisource, Multifeature Behavioral Data	1	1	1	0	3

desempenho dos alunos, ainda que os trabalhos que abordaram esse ponto demonstrassem ganhos significativos para a aprovação dos alunos. Vale-se destacar que apenas 5 dos 22 artigos estudados demonstraram a existência dessa relação. Adicionalmente, os trabalhos que discutem sobre este tema trazem argumentação e referências para indicar as predições nos primeiros anos graduação para um melhor desempenho na disciplina e no processo total de graduação.

**P2: Qual o conjunto de variáveis que melhor explica a predição de desempenho dos estudantes?** Identificou-se que não existe um único grupo que pode ser utilizado para trazer a melhor explicação de desempenho dos alunos e sim, combinações de grupos de variáveis que apresentam melhores poderes preditivos para aquele conjunto de dados. Os conjuntos de variáveis mais utilizados para a predição foram o desempenho histórico do aluno e as características demográficas.

**P2E1: Quais conjuntos de variáveis são utilizados para o processo de predição de desempenho dos estudantes?** Foram encontrados 7 conjuntos de variáveis que estão listadas a seguir e a quantidade dos artigos que contemplam cada um deles: Histórico de desempenho (16 artigos); Características demográficas (15 artigos); Interações Ambiente Virtual Aprendizagem (9 artigos); Participação em atividades durante o curso (8 artigos); Características socioeconômicas (5 artigos); Interação social online (3 artigos); Interação através do mouse (1 artigo).

**P2E2: O engajamento (participação) do aluno é avaliado dentro do modelo?** O engajamento apareceu na terceira posição entre o mais frequentes para a predição de desempenho dos estudantes. Os artigos que mensuram o engajamento do aluno são baseados nas interações do aluno com o ambiente virtual de aprendizagem através de discussão em fórum, quantidade de logs no período, quantidade de participações em chats e tempo em que o aluno ficou logado.

**P2E3: Informações socioeconômicas do aluno são utilizadas na modelagem de predição?** Apenas 6 trabalhos, dos 22 selecionados, abordaram os atributos relacionados às características socioeconômicas. Vale destacar que embora haja atributos escolhidos para avaliar o modelo, durante o processo de seleção de variável os atributos de características socioeconômicas podem não ter sido selecionados.

**P3: Como é um modelo de predição de desempenho dos estudantes?** Os modelos de predições utilizados nos artigos selecionados apresentaram diversas técnicas, não sendo possível definir qual é a melhor técnica e como são desenvolvidos para obter a melhor predição. O que se observou é que cada disciplina, objetivo e processo de ensino-aprendizagem apresentaram uma modelagem própria para aumentar a previsibilidade.

**P3E1: Quais as técnicas de IA são utilizadas para predição de desempenho dos estudantes?** Por meio dos 22 documentos foi possível localizar 23 variações de técnicas para predição de desempenho dos alunos, sendo que um único artigo pôde utilizar-se de uma ou mais técnicas e com isso justifica-se a quantidade de variáveis superior a quantidade de artigos. As técnicas mais observadas neste conjunto de artigos foram: Floresta aleatória (11 artigos); Support Vector Machines (11 artigos); K-Vizinhos mais próximos (9 artigos); Naive Bayes (8 artigos); Árvore de decisão (8 artigos); MLP-ANN (6 artigos); Regressão Logística (6 artigos).

**P3E2: Quais são as métricas de avaliação da acuracidade da predição?** Em

relação as métricas identificou-se que não existe uma variação entre as medidas de desempenho dos algoritmos. A principal modificação está em relação aos resultados do modelo; se há como resultado um modelo de classificação com resposta aprovado ou reprovado, ou se o modelo apresenta como predição a nota do aluno. De qualquer forma, destacam-se as seguintes métricas: Acurácia; Precisão; Sensibilidade; Especificidade; F-Measure; Erro quadrático médio; Erro médio absoluto; Curva ROC.

**P3E3: Qual é a acurácia encontrada nos modelos de predição de desempenho dos estudantes?** As acurácias encontradas apresentam valores elevados. Os trabalhos, no geral, apresentaram acurácias maiores do que 80 por cento e vale destacar que o momento de predição e a quantidade de informações influenciaram diretamente no valor de acurácia. Ou seja, quanto mais tempo o aluno estiver cursado a disciplina maior o poder de discriminação.

### 3.4. Síntese e oportunidades identificadas

O trabalho tem os respectivos insights:

- A maior parte dos trabalhos utilizam o desempenho histórico do aluno através das notas dos cursos anteriores ou notas dos exames para definir o desempenho do estudante;
- O segundo maior conjunto de atributos utilizado é o de dados demográficos como idade, sexo e características gerais;
- Os atributos de interação social apresentam relevância para discriminar o desempenho dos alunos;
- Os atributos de comportamento de aprendizagem têm alta usabilidade, inclusive em aspectos online através de ambientes virtuais de aprendizagem ou até mesmo offline como acesso as bibliotecas;
- A ferramenta de predição é um modelo poderoso para identificar o progresso do aluno;
- Os primeiros anos dos alunos de graduação são cruciais para o desempenho acadêmico e profissional, dando maior relevância para intervenções neste período;
- Oportunidade para desenvolver um modelo que possa ser aplicado em diferentes disciplinas e que possa ter a comprovação na contribuição para o desempenho acadêmico do aluno.

Dessa forma, uma das principais oportunidades identificadas dentro da revisão sistemática é que poucos trabalhos exploram a evidência da contribuição para o estudante e/ou professor. Os documentos apresentaram um grande esforço para aumentar a acurácia e os medidores de previsibilidade dos algoritmos. Ainda há um grande espaço para o desenvolvimento desses modelos de predição e uma pesquisa de como seria a melhor forma de implementá-los.

### 3.5. Ameaças a validade

O processo de revisão sistemática exige uma metodologia muito rígida em relação a transparência de cada etapa. Neste sentido, o presente trabalho esforçou-se para documentar todos os procedimentos que foram seguidos ao longo do projeto. Entretanto, como ameaça à validade da pesquisa existe a preocupação do processo de seleção dos artigos. O

primeiro ponto de ameaça identificado foi a inclusão do critério para que o artigo selecionado tivesse 10 citações ou mais, através dos bancos de dados e na sequência analisado a citação no google, o que poderia implicar na exclusão de artigos que fossem mais novos e ainda não possuíssem uma divulgação maior. Entretanto, ao avaliar os anos de publicação foi possível validar que existe um volume considerável para todos os anos entre 2016 e 2021 como era o desejado inicialmente.

A segunda preocupação é referente ao processo de classificação dos artigos baseados na leitura do título e, posterior, do resumo. Os artigos que não pareciam ter uma proximidade com o objetivo desta pesquisa precisaram ser excluídos. A exclusão ocorreu sob o olhar de somente um avaliador, o que pode implicar eventualmente em falha operacional. Para reduzir a probabilidade de erros adotou-se a releitura dos artigos em dias diferentes.

#### **4. Considerações Finais**

Os estudos demonstraram que o número de alunos nas universidades tem um crescimento muito forte nos últimos anos. Esse crescimento ocorre não apenas no cenário brasileiro, mas existe uma tendência mundial de crescimento. Entretanto, mesmo apresentando um indicador positivo ainda existe uma preocupação com a qualidade e permanência desses alunos durante o curso. Estudos demonstram que existe a preocupação em melhorar a qualidade de ensino e aprendizagem, bem como alavancar os resultados acadêmicos.

Neste sentido o trabalho identificou como oportunidade o processo de individualização do ensino, surgindo como uma alternativa mais factível com o aparecimento das novas tecnologias de inteligência artificial. Abordando essas tecnologias, existe uma área de grande expansão que foi abordada ao longo deste trabalho que é a mineração de dados educacionais (EDM). O objetivo deste trabalho foi aplicar uma revisão sistemática nesta área para o desenvolvimento de modelos de predição de desempenho dos alunos.

Para tanto, o trabalho seguiu a metodologia de revisão sistemática que consiste, basicamente, em uma questão de pesquisa que deverá guiar a uma definição de termos e busca dentro dos bancos de dados. Os meios pelos quais resultaram na definição dos 22 artigos foram evidenciado ao longo deste trabalho.

As perguntas realizadas foram respondidas na seção de resultados. Nos aspectos educacional, atributos de explicação e técnicas de predição, identificam-se grande similaridade entre os trabalhos. A maioria dos trabalhos são aplicados nas áreas de exatas e as variáveis utilizadas nos modelos de predição eram, em suma, dados demográficos e histórico de desempenho. Embora exista um alto poder de predição do desempenho dos estudantes, ainda existe uma grande oportunidade em relacionar e, principalmente, evidenciar quais são os benefícios para os estudantes e/ou professores. Conclui-se que o presente trabalho alcançou o objetivo de mapear todo o estado da arte referente aos modelos de predição de desempenho do estudante e além disso, conseguiu entender quais são as limitações das pesquisas e oferecer um direcionamento para estudos futuros.

## Referências

- ADNAN, M. et al. Predicting at-risk students at different percentages of course length for early intervention using machine learning models. *Ieee Access*, IEEE, v. 9, p. 7519–7539, 2021.
- AKÇAPINAR, G.; ALTUN, A.; AŞKAR, P. Using learning analytics to develop early-warning system for at-risk students. *International Journal of Educational Technology in Higher Education*, Springer, v. 16, n. 1, p. 1–20, 2019.
- ALAMRI, R.; ALHARBI, B. Explainable student performance prediction models: a systematic review. *IEEE Access*, IEEE, v. 9, p. 33132–33143, 2021.
- ALMASRI, A.; CELEBI, E.; ALKHAWALDEH, R. S. Emt: Ensemble meta-based tree model for predicting student performance. *Scientific Programming*, Hindawi, v. 2019, 2019.
- ALSHANQITI, A.; NAMOUN, A. Predicting student performance and its influential factors using hybrid regression and multi-label classification. *IEEE Access*, IEEE, v. 8, p. 203827–203844, 2020.
- ALYAHYAN, E.; DÜŞTEGÖR, D. Predicting academic success in higher education: literature review and best practices. *International Journal of Educational Technology in Higher Education*, Springer, v. 17, n. 1, p. 1–21, 2020.
- ASIAH, M. et al. A review on predictive modeling technique for student academic performance monitoring. In: EDP SCIENCES. *MATEC Web of Conferences*. [S.l.], 2019. v. 255, p. 03004.
- AZCONA, D.; HSIAO, I.-H.; SMEATON, A. F. Detecting students-at-risk in computer programming classes with learning analytics from students' digital footprints. *User Modeling and User-Adapted Interaction*, Springer, v. 29, n. 4, p. 759–788, 2019.
- BLOOM, B. S. The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational researcher*, Sage Publications Sage CA: Thousand Oaks, CA, v. 13, n. 6, p. 4–16, 1984.
- BUJANG, S. D. A. et al. Multiclass prediction model for student grade prediction using machine learning. *IEEE Access*, IEEE, v. 9, p. 95608–95621, 2021.
- COOK, D. J.; MULROW, C. D.; HAYNES, R. B. Systematic reviews: synthesis of best evidence for clinical decisions. *Annals of internal medicine*, American College of Physicians, v. 126, n. 5, p. 376–380, 1997.
- DEO, R. C. et al. Modern artificial intelligence model development for undergraduate student performance prediction: An investigation on engineering mathematics courses. *IEEE Access*, IEEE, v. 8, p. 136697–136724, 2020.
- DIGIAMPIETRI, L. A.; NAKANO, F.; LAURETTO, M. de S. Mineração de dados para identificação de alunos com alto risco de evasão: Um estudo de caso. *Revista de Graduação USP*, v. 1, n. 1, p. 17–23, 2016.
- DIOGO, M. F. et al. Percepções de coordenadores de curso superior sobre evasão, reprovações e estratégias preventivas. *Avaliação: Revista da Avaliação da Educação Superior (Campinas)*, SciELO Brasil, v. 21, p. 125–151, 2016.

GAŠEVIĆ, D. et al. Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, Elsevier, v. 28, p. 68–84, 2016.

HELLAS, A. et al. Predicting academic performance: a systematic literature review. In: *Proceedings companion of the 23rd annual ACM conference on innovation and technology in computer science education*. [S.l.: s.n.], 2018. p. 175–199.

IMRAN, M. et al. Student academic performance prediction using supervised learning techniques. *International Journal of Emerging Technologies in Learning*, v. 14, n. 14, 2019.

KHAN, I. et al. An artificial intelligence approach to monitor student performance and devise preventive measures. *Smart Learning Environments*, SpringerOpen, v. 8, n. 1, p. 1–18, 2021.

LAGUS, J. et al. Transfer-learning methods in programming course outcome prediction. *ACM Transactions on Computing Education (TOCE)*, ACM New York, NY, USA, v. 18, n. 4, p. 1–18, 2018.

LOPES, A. Algumas reflexões sobre a questão do alto índice de reprovação nos cursos de cálculo da ufrgs. *Sociedade Brasileira de Matemática. Rio de Janeiro*, n. 26/27, p. 123–146, 1999.

MARBOUTI, F.; DIEFES-DUX, H. A.; MADHAVAN, K. Models for early prediction of at-risk students in a course using standards-based grading. *Computers & Education*, Elsevier, v. 103, p. 1–15, 2016.

MUSSO, M. F.; HERNÁNDEZ, C. F. R.; CASCALLAR, E. C. Predicting key educational outcomes in academic trajectories: a machine-learning approach. *Higher Education*, Springer, v. 80, n. 5, p. 875–894, 2020.

POLYZOU, A.; KARYPIS, G. Grade prediction with models specific to students and courses. *International Journal of Data Science and Analytics*, Springer, v. 2, n. 3, p. 159–171, 2016.

POPESCU, E.; LEON, F. Predicting academic performance based on learner traces in a social learning environment. *IEEE Access*, IEEE, v. 6, p. 72774–72785, 2018.

SANTOS, H. G. d. et al. Machine learning para análises preditivas em saúde: exemplo de aplicação para prever óbito em idosos de são paulo, brasil. *Cadernos de Saúde Pública*, SciELO Public Health, v. 35, p. e00050818, 2019.

SEKEROGLU, B.; DIMILILER, K.; TUNCAL, K. Student performance prediction and classification using machine learning algorithms. In: *Proceedings of the 2019 8th International Conference on Educational and Information Technology*. [S.l.: s.n.], 2019. p. 7–11.

WAKELAM, E. et al. The potential for student performance prediction in small cohorts with minimal available attributes. *British Journal of Educational Technology*, Wiley Online Library, v. 51, n. 2, p. 347–370, 2020.

WANG, X. et al. Student performance prediction with short-term sequential campus behaviors. *Information*, MDPI, v. 11, n. 4, p. 201, 2020.

WEI, H. et al. Predicting student performance in interactive online question pools using mouse interaction features. In: *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*. [S.l.: s.n.], 2020. p. 645–654.

YU, C.-H.; WU, J.; LIU, A.-C. Predicting learning outcomes with mooc clickstreams. *Education sciences*, MDPI, v. 9, n. 2, p. 104, 2019.

ZHAO, L. et al. Academic performance prediction based on multisource, multifeature behavioral data. *IEEE Access*, IEEE, v. 9, p. 5453–5465, 2020.

ZOHAIR, A.; MAHMOUD, L. Prediction of student's performance by modelling small dataset size. *International Journal of Educational Technology in Higher Education*, SpringerOpen, v. 16, n. 1, p. 1–18, 2019.

ZOLLANVARI, A. et al. Predicting students' gpa and developing intervention strategies based on self-regulatory learning behaviors. *IEEE Access*, IEEE, v. 5, p. 23792–23802, 2017.