

# A Produção de Learning Analytics e Predição de Desempenho Acadêmico por pesquisadores Brasileiros: Uma Revisão Sistemática da Literatura

Carlos Artur Guimarães<sup>1</sup>, Isabel Nunes<sup>1</sup>, Andressa K. Pires<sup>1</sup>, Eric Alencar<sup>1</sup>

<sup>1</sup>Instituto Metrôpole Digital - Universidade Federal do Rio Grande do Norte (UFRN)  
Av. Sen. Salgado Filho, 3000 - Lagoa Nova, CEP: 59.078-970 - Natal - RN - Brazil

(artur, bel)@imd.ufrn.br, kroeffzinha@ufrn.edu.br, eric.ed123@gmail.com

**Abstract:** *Distance learning courses have increased its participation in Brazilian education. Expansion driven by the use of virtual environments, social inclusion, provision in underserved regions and lower cost of execution. Such courses present problematic such as: high dropout rates, failure and retention of students. Prediction is important to identify such problems and propose solutions for mitigation. This work is a Systematic Literature Review (SLR) on the Prediction of Academic Performance of studies published by Brazilian scientists in the research bases: Scopus, Science Direct, Scielo, IEEE and CEIE. Four articles were found that showed techniques, algorithms, practices and controlled studies.*

**Resumo:** *Os cursos a distância têm aumentado a participação na educação Brasileira. Expansão impulsionada pelo uso de ambientes virtuais, inclusão social, oferta em regiões mal atendidas e menor custo de execução. Tais cursos apresentam como problemáticas: altos índices de evasão, reprovação e retenção dos estudantes. É importante a predição para identificar tais problemas e propor soluções para mitigação. Este trabalho é uma Revisão Sistemática de Literatura (RSL) sobre Predição de Desempenho Acadêmico de estudos publicados por cientistas brasileiros nas bases de pesquisas: Scopus, Science Direct, Scielo, IEEE e CEIE. Foram encontrados 4 artigos que mostraram técnicas, algoritmos, práticas e estudos controlados.*

## 1. Introdução

A oferta de cursos na modalidade a distância na educação brasileira vem crescendo, de acordo com os dados da Associação Brasileira de Educação a Distância (ABED, 2019), no seu senso de 2018, a modalidade de ensino a distância em 2016 tinha 3.734.887 alunos, em 2017 um total de 7.773.828 alunos, um aumento de aproximadamente 108,14% entre esses dois anos.

As instituições de ensino privadas e públicas tiveram um grande salto quanto ao uso de Tecnologias Educacionais (TE) como plataformas de aprendizagem, sistemas administrativos, Ambientes Virtuais de Aprendizagem (AVA) e sistemas colaborativos, com uma adesão dessas tecnologias em cursos presenciais, semipresenciais e totalmente a distância. Segundo a ABED (2019) os cursos semipresenciais apresentaram um aumento de 34,5% no ano de 2018. Essa modalidade de ensino se distingue em 3

dimensões: (I) curso inicialmente regulamentado como presencial, mas com oferta de até 20% da sua carga horária na modalidade a distância, (II) cursos regulamentados inicialmente a distância, mas com carga horária presencial obrigatória e (III) cursos presenciais que utilizam tecnologias para complementar os conteúdos de forma remota.

Zhang et al. (2020) afirmam o grande impacto da pandemia COVID-19 na educação por aumentar a necessidade de ensino remoto. Através de tornar o uso de plataformas de ensino ainda mais importante como estratégia de continuidade de cursos, treinamentos e oferta do ensino regular. Este cenário revela a importância do desenvolvimento de soluções diversas por pesquisadores para garantir a continuidade, durante e após a pandemia causada pelo COVID-19, em questão de desenvolvimento da nossa sociedade e do ser humano de forma integral, seja pelos aspectos cognitivos, desenvolvimento humanístico-científico ou geração de emprego e renda. Em consequência das aulas remotas como estratégia de ensino durante a pandemia do COVID-19, tem crescido o aumento de uso de sistemas de TE, resultando no aumento da geração de dados educacionais. O cenário de uso extensivo de aulas remotas utilizando Tecnologias Educacionais é grande geradora de dados educacionais, propiciando utilizar *Learning Analytics* - LA como evidenciado por Baker, Isotani e Carvalho (2011). Através de LA buscar evasão, retenção e reprovação, possibilitando cursos com melhor desempenho acadêmico dos alunos matriculados.

Estudos envolvendo análise de dados educacionais para melhoria de desempenho acadêmico são relativamente poucos [BRASIL et al. 2018]. Assim, é necessário que mais trabalhos possam destacar o horizonte das possibilidades associadas com essa tecnologia, não tão recente, para o contexto das escolas brasileiras. O objetivo deste artigo é apresentar uma Revisão Sistemática de Literatura (RSL) sobre a temática de *Learning Analytics* e Predição de Desempenho Acadêmico produzida por pesquisadores brasileiros. Nela serão apresentados trabalhos exclusivamente realizados por pesquisadores brasileiros publicados em inglês ou português em bases científicas diversas, fundamentando-se como forma de sistematizar a pesquisa de brasileiros da área, de forma a colaborar com o estudo e desenvolvimento de pesquisa relacionadas neste momento histórico tão requerente de inovação e tecnologia educacional, de forma a contribuir com o desenvolvimento pleno da educação durante e pós pandemia da COVID-19. Busca-se entender a predição de desempenho acadêmico e relações a partir do seu uso em ambientes educacionais, visando identificar padrões de aprendizagem. O objetivo da RSL é responder questões relacionadas às abordagens, práticas e métodos utilizados na predição de desempenho acadêmico utilizando LA.

O artigo está estruturado da seguinte forma: a seção 2 descreve o processo realizado para obtenção dos estudos; na seção 3 os trabalhos relacionados; e a seção 4 os resultados e percepções da RSL. Por fim, a seção 5 revela as considerações finais.

## **2. Método da Revisão Sistemática**

Segundo Kitchenham (2004), uma Revisão Sistemática da Literatura (RSL) é utilizada para identificar, avaliar e interpretar toda pesquisa disponível acerca de uma questão de pesquisa, tópico ou fenômeno de interesse. Portanto, estudos utilizados em uma revisão sistemática são chamados de estudos primários, enquanto que a revisão realizada é chamada de estudo secundário.

Para organização desta RSL, foi utilizada a ferramenta online Parsifal<sup>1</sup>, focada no desenvolvimento de revisões sistemáticas, colaborando na sistematização da pesquisa conforme Kitchenham (2004), para condução da pesquisa: Objetivos da pesquisa, Questões da pesquisa, Bases de pesquisa e Critérios de qualidade.

Os objetivos específicos buscam compreender como: Realizar predição de desempenho de aprendizagem usando dados educacionais, identificar perfis de aprendizagem e testar algoritmos de predição para desempenho de aprendizagem. As questões da pesquisa norteadoras estão apresentadas na Tabela 1.

**Tabela 1. Questões de pesquisa.**

<b>Código</b>	<b>Questão de Pesquisa</b>
QP1	O que os trabalhos estão prevendo?
QP2	Quais os benefícios e limitações dos modelos de predição?
QP3	Como a predição é utilizada para diagnósticos de aprendizagem?
QP4	Quais as informações utilizadas para a predição?
QP5	Quais características dos cursos dos estudos de predição?

**Fonte: Elaborado pelos autores.**

Foram escolhidas 6 bases de periódicos em âmbito nacional e internacional. As selecionadas foram: Comissão Especial de Informática na Educação (CEIE)<sup>2</sup>, Institute of Electrical and Electronic Engineers (IEEE Digital Library)<sup>3</sup>, Scopus<sup>4</sup>, Scielo<sup>5</sup> e Science Directory<sup>6</sup>. Para definição de aceite ou negação dos artigos foram definidos os critérios de inclusão empregados para a seleção dos trabalhos conforme Tabela 2.

**Tabela 2. Critérios de inclusão.**

<b>Código</b>	<b>Descrição</b>
I1	Artigos em português e inglês
I2	Artigos publicados de Janeiro de 2011 a Abril de 2020
I3	Artigos que propõem/relatam um processo, ferramenta, arquitetura, ou diretriz para processo de desenvolvimento de metodologias
I4	Artigos que contém resultados
I5	Artigos realizados por pesquisadores brasileiros

**Fonte: Elaborado pelos autores.**

<sup>1</sup> <https://parsif.al/>

<sup>2</sup> <https://www.br-ie.org/pub>

<sup>3</sup> <http://ieeexplore.ieee.org>

<sup>4</sup> <http://www.scopus.com>

<sup>5</sup> <http://scielo.com.br>

<sup>6</sup> <http://www.sciencedirect.com>

Para definição de aceite ou negação dos artigos foram definidos, também, os critérios de exclusão empregados para a seleção dos trabalhos conforme Tabela 3.

**Tabela 3. Critérios de exclusão.**

<b>Código</b>	<b>Tipo</b>	<b>Descrição</b>
E1	Exclusão	Estudos duplicados
E2	Exclusão	Estudos que não estão disponíveis com o texto na íntegra
E3	Exclusão	Estudos que não são artigos de revista/evento científico
E4	Exclusão	Estudos que não são primários
E5	Exclusão	Estudos sem resumo

**Fonte: Elaborado pelos autores.**

As questões de qualidade foram definidas sendo planejado três possíveis respostas para a avaliação da pontuação: Sim (1.0), Não (0.0) e Parcialmente (0.5).

## **2.1. A *String* e busca dos artigos**

A pesquisa foi realizada a partir da inserção de informações no Parsifal, na seção de condução do estudo, através da utilização das *strings*, foram realizadas as consultas em português com a adaptação da versão em inglês - conforme Tabela 4 -, nas bases e revista selecionadas para geração dos resultados e exportação dos arquivos BibTeX<sup>7</sup>.

**Tabela 4. *Strings* utilizadas nas consultas.**

<b>Identificador</b>	<b>Idioma</b>	<b><i>String</i> de busca</b>
S1	Português	("análise de dados educacionais" OR "mineração de dados educacionais") AND ("performance de aprendizagem" OR "performance acadêmica" OR "sucesso acadêmico" OR "falha acadêmica")
S2	Inglês	("learning analytics" OR "educational data mining" OR "learning prediction") AND ("learning performance" OR "academic performance" OR "academic success" OR "failure academic" OR "academic achievement")

**Fonte: Elaborado pelos autores.**

Os BibTeX foram exportados para a plataforma Parsifal, para a composição dos resultados associados a cada uma das bases. A base da CEIE não realiza exportação de BibTeX, o mesmo foi criado pelos pesquisadores manualmente, sendo exportado para a plataforma Parsifal. Nesta fase da pesquisa foi possível a primeira quantificação do nível de amostragem de resultados. Os quantitativos dessa exportação podem ser analisados na Tabela 5.

<sup>7</sup> <http://www.bibtex.org>

**Tabela 5. Quantitativo dos estudos encontrados.**

ID String de busca	Bases				
	Scopus	CEIE	IEEE	Science Direct	Scielo
S1	0	1	0	0	0
S2	730	2	17	63	2

**Fonte: Elaborado pelos autores.**

## **2.2. Seleção dos Estudos**

Na base da *Scopus* é possível filtrar a pesquisa exclusiva para pesquisadores brasileiros facilitando o estudo, nas demais bases foi feita pesquisa manual. A realização da aceitação dos estudos foi concebida pelos avaliadores através da leitura dos títulos e resumo dos artigos, buscas dos pesquisadores e validação das informações.

O processo de seleção dos estudos pode apresentar como ameaça pequeno grau de falta de precisão quanto à análise das pesquisas, pois tratando-se principalmente da seleção de estudos focados no aspecto da realização por parte de pesquisadores brasileiros, foi necessária, a busca e filtragem de forma manual nas bases de pesquisa. Sendo esse procedimento passível a erros. Tal procedimento foi realizado de forma separada pelos pesquisadores e depois comparados seus resultados.

De 815 artigos foram aceitos 4 para o estudo, são eles: [1] Barros et al. (2019); [2] Buschetto Macarini et al. (2019); [3] Etemadpour et al. (2020) e [4] Reyes et al. (2019). Os artigos [1], [2] e [4] foram indexados da Scopus, o [3] foi indexado através da CEIE.

## **2.3. Critérios de Qualidade**

Para os critérios de qualidade foram utilizados 5 perguntas de pesquisa para avaliação dos artigos. As respostas continham os seguintes pesos para composição da nota: Sim (1.0), Parcialmente (0.5) e Não (0.0).

As perguntas de qualidade e as notas de avaliação foram sistematizadas de acordo com o título do artigo, para enumeração e comparativo dos estudos. Dos 4 estudos avaliados, 3 obtiveram nota máxima (5.0) e 1 estudo teve nota (4.0), assim foi possível perceber um grande grau de coerência com os estudos selecionados para busca das respostas às questões norteadoras da pesquisa, conforme perguntas e Tabela 6:

- QA1: O estudo apresenta método, técnica ou ferramenta a auxiliar na abordagem?
- QA2: O estudo apresenta objetivos e/ou questões de pesquisa bem definidos e embasados pela literatura?
- QA3: O estudo avaliado apresenta algum tipo de experimento controlado ou estudo de caso para avaliação da abordagem proposta?
- QA4: O estudo avaliado apresenta uma breve comparação dos seus resultados com resultados de outros trabalhos relacionados?
- QA5: O estudo avaliado apresenta resultados de um contexto real de ensino?

**Tabela 6. Nota de avaliação dos artigos conforme critérios de qualidade.**

<b>Título</b>	<b>QA1</b>	<b>QA2</b>	<b>QA3</b>	<b>QA4</b>	<b>QA5</b>	<b>Total</b>
Role of absence in academic success: an analysis using visualization tools	1.0	1.0	1.0	0	1.0	4.0
Predictive models for imbalanced data: A school dropout perspective	1.0	1.0	1.0	1.0	1.0	5.0
Predicting students success in blended learning-Evaluating different interactions inside learning management systems	1.0	1.0	1.0	1.0	1.0	5.0
Predição de sucesso acadêmico de estudantes: uma análise sobre a demanda por uma abordagem baseada em transfer learning	1.0	1.0	1.0	1.0	1.0	5.0

**Fonte: Elaborado pelos autores.**

### **3. Trabalhos Relacionados**

Nesta seção, os trabalhos apresentados se relacionam com o artigo pela temática e abordagem explorada. Santos et al. (2016) apresentam em uma análise de trabalhos sobre temática de Mineração de Dados Educacionais e Desempenho Acadêmico estudos do período de 2012 a 2016 de pesquisadores brasileiros a partir de buscas no *Google Scholar*. O trabalho busca diversos modelos de trabalhos como conferências, periódicos, teses, dissertações e TCC's. O trabalho apresenta um conjunto de resultados relacionado a fonte das bases de dados, quantitativo de uso dos algoritmos, ferramentas utilizadas e os tipos de análise. Diferenciando do artigo apresentado por ser uma RSL focada em artigos em bases, sendo um processo factível de validação posteriores, além de um escopo quanto modelos preditivos aplicados na predição de desempenho acadêmico e filtrar em um recorte histórico de 2011 a abril de 2020. Os artigos são similares no foco na produção de pesquisadores brasileiros e temática geral.

Brasil et al. (2018), em sua RSL, expõe o uso de *Learning Analytics* em Ambientes Virtuais de Aprendizagem (AVA) brasileiros, o trabalho possui como objetivos compreender a relação da análise de aprendizagem abordados, instituições, abordagem, tecnologia e que ambientes virtuais são utilizados. O trabalho utilizou como fonte para buscas o repositório de pesquisa SCOPUS e CEIE. Foram selecionados 22 estudos para a análise, servindo de diagnóstico quanto uso de LA em AVA's nas instituições brasileiras e suas características relacionadas aos objetos da pesquisa. Diferenciando-se deste artigo quanto a temática de predição que mesmo por utilizar AV para geração e análise dos dados, possui o diferencial da pesquisa quanto os modelos preditivos para mitigação de reprovação, evasão e retenção nos processos de desempenho acadêmico.

### **4. Resultados e discussões**

O trabalho de Reyes et al. (2019) conduz uma análise que destaca a existência de divergências entre distribuições de dados gerados em contextos educacionais diferentes,

e envolvidos na predição de insucesso acadêmico. Os dados analisados pertencem a um curso superior de modalidade a distância. Durante a análise, de fato foram encontradas as divergências apontadas em períodos letivos diferentes, o que pode ser descrito sob a ótica do fenômeno conhecido como *Covariate Shift*.

Já Buschetto Macarini et al. (2019) introduzem um estudo comparativo para encontrar a melhor combinação de variáveis e algoritmos de classificação para definir estudantes em risco de forma prévia. A investigação é realizada sobre treze conjuntos de dados, os quais representam diferentes aspectos das interações dos estudantes na plataforma Moodle. Diferente de outros trabalhos que buscam identificar estudantes em zona de risco, este busca categorizar a contagem de *logs* do AVA, para de acordo com Garrison identificar uma experiência mediada por computador que seja de boa qualidade. Isso conforme três aspectos: cognitivo, social e presença de ensino.

Barros et al. (2019) expõe uma abordagem para encontrar resultados mais precisos na busca por definir estudantes com tendência a persistir e ou evadir nas instituições de ensino. É relatado que muitos trabalhos encontram nas pesquisas um fenômeno onde os dados ficam mascarados, devido a falta de balanceamento. Aqui são usadas as técnicas de árvore de decisão, redes neurais e *balanced bagging*.

Etemadpour et al. (2020) apresentam uma ferramenta que combina modelos de visualização de dados e modelos preditivos para compreensão do desempenho acadêmico, pesquisando tópicos específicos como notas, tarefas e presenças, os quais podem servir de indicadores para mensuração de desempenho acadêmico. Foram desenvolvidos modelos preditivos com base nos semestres anteriores, combinados com layouts interativos que destacam notas, frequências e escolaridade dos seus pais. Esse trabalho tem como objetivo principal responder como a predição de desempenho acadêmico vem sendo trabalhada por pesquisadores brasileiros. Foram elaboradas cinco questões norteadoras para a elucidação desse questionamento conforme Tabela 1.

O resultado da RSL após a seleção foi de 4 artigos, buscando responder às corresponder os objetivos da pesquisa em seus pontos positivos, negativos e aspectos gerais da pesquisa. Na avaliação os pesquisadores utilizaram *Data Extraction* do Parsifal para extração, organização e sistematização das informações dos artigos.

A QP1, Barros et al. (2019) utilizam notas de semestres anteriores para analisar o desempenho acadêmico, a existência de relação entre fatores externos (gênero sexual, escolaridade dos pais e aspectos sócio econômicos) e desempenho dos estudantes. Buschetto Macarini et al. (2019), é analisado como os indícios dos dados provenientes de contexto educacionais distintos na tarefa de predição de insucesso acadêmico de estudantes. Etemadpour et al. (2020) analisam a capacidade de diagnóstico na primeira semana do curso e baixa eficiência nas semanas seguintes. Reyes et al. (2019) realizam a predição de sucesso acadêmico de estudantes. Das respostas dos estudos analisadas, Barros et al. (2019) utilizam notas de semestres. é uma abordagem negativa para a pesquisa pretendida, pois caracterizamos como tardio para recuperação de alunos no primeiro semestre. Em Etemadpour et al. (2020) é identificado que a primeira semana é muito importante para desenvolvimento da aprendizagem do aluno, com características de abandono, repetência e reprovação quando seguido de baixa desempenho acadêmico nas semanas seguintes. Tal conceito apresentado é importante para a pesquisa por explicitar em períodos menores de tempo.

A QP2, Barros et al. (2019) destacam os benefícios do uso dos algoritmos de Máquinas de Vetores de Suporte e Floresta Aleatória por serem mais precisos nos modelos de predição. Os algoritmos de Regressão Linear, MLP e Árvore de Decisão apresentam resultados um pouco inferiores. Buschetto Macarini et al. (2019) em sua pesquisa apresenta a riqueza do número de variáveis para modelos de predição de bons resultados. Para desenvolvimento dos testes é utilizada a biblioteca *Scikit Learn*<sup>8</sup>. [3] Etemadpour et al. (2020) informam que os alguns algoritmos possuem limitação para aplicação em recortes temporais diferentes, nas primeiras semanas possuem melhor desempenho e nas posteriores mostram baixa eficiência. Reyes et al. (2019), relatam que o índice de generalização de modelos preditivos para contexto distintos é positivo. As informações apresentadas nos artigos quanto aos modelos de predição são de grande contribuição ao estudo proposto por considerar aspectos úteis entre os destacados para a pesquisa a ser desenvolvida em estudos posteriores, mas por ser uma aspecto secundário, não é necessário para a pesquisa.

Em QP3, Barros et al. (2019) fazem análise de predição para semestres e variáveis para buscar relação com notas, criando um algoritmo possível de identificar comportamentos de risco semelhante em turmas futuras. Foi mostrada as seguintes relações no trabalho: notas e frequência, notas e sexos, notas e nível educacional dos pais e populações e sexos dos alunos. Buschetto Macarini et al. (2019), seleciona no mínimo duas variáveis (períodos letivos e disciplinas), criando um *setup* de teste e validação. Reyes et al. (2019) faz a predição através de um conjunto de variáveis registradas pelo sistema Moodle. Barros et al. (2019) e Etemadpour et al. (2020) apresentam contribuições não aderentes ao estudo, por serem apresentarem limitadas a ambientes específicos, mas não contribuem para os objetivos da pesquisa apresentada. A QP4 apresenta em: [1] Gênero, semestre do curso, assunto, nota e frequência.

Implementaram uma variedade de modelos de aprendizado de máquina que preveem o desempenho dos alunos com base em sua frequência (taxa de ausência) e empregaram uma técnica de visualização para verificar a precisão de diferentes modelos. [2] *Logs* do Moodle (*logins*, discussão em fóruns, postagem, visualizações, submissão de tarefas, tempo de visualização do material, atividades e quantidade de dias distintos que fez login). [3] Cognitivo: atividades (*upload* de: trabalho, arquivos, texto, trabalho de casa, página de acesso), pesquisa, presença no sistema, url, *wiki*, arquivo e laboratório virtual; Ensino: comentários sobre os trabalhos enviados; Social: fórum e votações. [4] Variáveis do Moodle, como: *login*, fórum, *chat*, *post's*, recursos visualizados e tempo de uso. Os estudos apresentam ricas contribuições da pesquisa. A limitação é ponderar quanto o número de variáveis e testes controlados.

Em QP5, Barros et al. (2019) realizam estudo em um curso de nível superior na área de exatas no período de 2009 a 2017. Buschetto Macarini et al. (2019) desenvolvem seus estudos em cursos de nível superior a distância. Etemadpour et al. (2020) apresentam um estudo sobre ensino híbrido na graduação. Em Reyes et al. (2019) é realizado o estudo em cursos de nível superior, na modalidade totalmente EAD. Os artigos anteriores apresentam características quanto os ambientes dos cursos em que realizam predição de aprendizagem acadêmica sendo ricos para o estudo.

---

<sup>8</sup> <https://scikit-learn.org/stable/>



## 5. Conclusões

Estudos brasileiros tendo como temática a predição de aprendizagem são poucos conforme a quantidade de artigos aprovados nesta RSL. A temática da análise de dados educacionais também é uma área relativamente nova, o que contribui para essa pouca disponibilidade de conteúdo. Apesar dessa carência, a temática em questão é passível de trazer grandes contribuições para o contexto educacional.

Os artigos utilizados na pesquisa apresentam contextos diversos onde determinadas técnicas e algoritmos não apresentam os mesmos resultados. Essas diferenças nos resultados podem ser traduzidas pela mudança dos contextos dos dados onde as técnicas foram aplicadas, como por exemplo o uso de soluções de balanceamento de dados. Ademais, os trabalhos pretendem com esses ajustes contornar o fato de não existir uma solução genérica o suficiente para todos os casos.

Esses estudos apresentam-se, mesmo que dentro de um pequeno número de trabalhos encontrados, de grande relevância, pois a predição do desempenho estudantil mostra-se um tema de muita importância para docentes. Por meio da análise preditiva torna-se possível a tomada de decisão quanto às medidas a serem aplicadas para os casos de baixo desempenho acadêmico. Trazendo assim a possibilidade de respostas para evitar reprovações e a evasão.

Uma observação a considerar como possível ameaça ao estudo, são possíveis erros humanos na condução da pesquisa, como característica a ser levada em consideração em revisões sistemáticas da literatura de forma geral.

Por fim, como pretensão de próximos trabalhos, fica destacada aqui a necessidade do aprofundamento no estudo de modelos preditivos, isto é, a elaboração de pesquisas testando a eficácia dos modelos disponíveis em ambientes de teste reais e simulados pelo grupo de pesquisadores. Assim sendo possível fundamentar uma pesquisa e resultados posteriores de forma a contribuir a área de LA como forma de colaboração científica e ao campo educacional.

## 6. Referências

- ABED – Associação Brasileira de Educação a Distância (2019) Censo EAD.BR: relatório analítico da aprendizagem a distância no Brasil 2018 – [livro eletrônico] Curitiba: InterSaberes.
- Baker, R.; Isotani, S.; Carvalho, A. (2011) Mineração de Dados Educacionais: Oportunidades para o Brasil. Revista Brasileira de Informática na Educação, v. 19, n. 02, p. 03, 24 ago.
- Barros, T. M. et al. (2019) Predictive Models for Imbalanced Data: A School Dropout Perspective. Education Sciences, v. 9, n. 4, p. 275, dez.
- Brasil, P. et al. (2018) Uma Revisão Sistemática sobre o uso de Learning Analytics em ambientes virtuais de aprendizagem brasileiros. In: Ctrl+ E Congresso sobre Tecnologias Educacionais, Fortaleza-CE.

- Macarini B., L. A. et al. (2019) Predicting Students Success in Blended Learning—Evaluating Different Interactions Inside Learning Management Systems. *Applied Sciences*, v. 9, n. 24, p. 5523, jan.
- Etemadpour, R. et al. (2020) Role of absence in academic success: an analysis using visualization tools. *Smart Learning Environments*, v. 7, n. 1, p. 2, 7 jan.
- Kitchenham, B. (2007) Guidelines for performing Systematic Literature Reviews in Software Engineering, Version 2.3, EBSE Technical Report EBSE-2007-01, Keele University and University of Durham.
- Reyes, D. A. G. D. L. et al. (2019) Predição de sucesso acadêmico de estudantes: uma análise sobre a demanda por uma abordagem baseada em transfer learning. *Revista Brasileira de Informática na Educação*, v. 27, n. 01, p. 01, 1 jan.
- Santos, R. et al. (2016) Análise de Trabalhos Sobre a Aplicação de Técnicas de Mineração de Dados Educacionais na Previsão de Desempenho Acadêmico. *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*, v. 5, n. 1, p. 960, 10 nov.
- Zhang, W. et al. (2020) Suspending Classes Without Stopping Learning: China's Education Emergency Management Policy in the COVID-19 Outbreak. *Journal of Risk and Financial Management*, v. 13, n. 3, p. 55, mar.