

Ciência de dados na educação: contribuições interdisciplinares

Data science in education: interdisciplinary contributions

Patrícia Takaki

Universidade Estadual de Montes Claros (DCC/CCET/UNIMONTES); Programa de Pós-Graduação em Ciência da Informação, Universidade Federal de Santa Catarina (PGCIn/UFSC), Brasil.

E-mail: patricia.takaki@unimontes.br

ORCID: <https://orcid.org/0000-0003-4777-9683>

Moisés Dutra

Programa de Pós-Graduação em Ciência da Informação, Universidade Federal de Santa Catarina (PGCIn/UFSC), Brasil.

E-mail: moises.dutra@ufsc.br

ORCID: <https://orcid.org/0000-0003-1000-5553>

RESUMO

O presente trabalho visa identificar contribuições da Ciência da Informação (CI) e da Ciência da Computação (CC) para atender às necessidades informacionais do contexto educacional. A metodologia adotada é exploratória, descritiva e de natureza aplicada. É utilizada a abordagem qualitativa no tratamento dos dados recuperados pela pesquisa bibliográfica por publicações envolvendo ciência de dados e educação. A pesquisa reuniu aportes teóricos e práticos da CI e da CC capazes de endereçar os principais desafios informacionais da educação relatados na literatura. Os resultados preliminares obtidos constituem-se de conexões de conceitos e técnicas das duas áreas que propõem soluções para um conjunto de necessidades informacionais da educação. Mineração de dados, lógica fuzzy, computação natural, mineração de texto, data warehouse, ontologias, web semântica e recuperação da informação foram as contribuições que se mais se evidenciaram quando analisadas as demandas de alunos, professores e gestores. A conclusão alcançada é que arti-

Como citar: Takaki, P.; & Dutra, M. (2022). Ciência de dados na educação: contribuições interdisciplinares. En T.M.R. Dias (Ed.), *Informação, Dados e Tecnologia. Advanced Notes in Information Science, volume 2* (pp. 149-160). Tallinn, Estonia: ColNes Publishing. DOI: 10.47909/anis.978-9916-9760-3-6.94.

Copyright: © 2022, The author(s). This is an open access work distributed under the terms of the CC BY-NC 4.0 license which permits copying and redistributing the material in any medium or format, adapting, transforming and building upon the material as long as the license terms are followed.

cular conhecimentos interdisciplinares da CC e da CI na proposição de abordagens dirigidas a dados no contexto educacional não só é possível como também necessário. Estas contribuições se revelaram capazes de lidar com o desafio de apoiar a educação de diferentes formas, viabilizando soluções ainda mais robustas teórica e metodologicamente.

Palavras-chave: ciência de dados; educação; aprendizado de máquina; ciência da informação; interdisciplinaridade.

ABSTRACT

The present paper aims to identify the contributions of Information Science (IS) and Computer Science (CS) to meet the informational needs of the educational context. The methodology used is exploratory, descriptive, and applied. A qualitative approach is used to process data retrieved through bibliographic research on data science applied to education. The study brought together the theoretical and practical contributions of IS and CS techniques capable of addressing the main informational challenges of educational environments reported in the literature. The preliminary results obtained consist of connections between concepts and techniques from the two knowledge areas that propose solutions for informational needs in education. Data mining, fuzzy logic, natural computing, text mining, data warehouse, ontologies, semantic web, and information retrieval were the most evident contributions identified when analyzing the demands of students, teachers, and managers. The conclusion reached is that articulating interdisciplinary knowledge of CS and IS in proposing data-driven approaches in the educational context is possible and necessary. These contributions proved capable of dealing with the challenge of supporting education in different ways by enabling even more robust theoretical and methodological solutions.

Keywords: data science; education; machine learning; information science; interdisciplinarity

INTRODUÇÃO

UM CRESCENTE volume de dados tem sido gerado nos contextos educacionais, visto que esses são cada vez mais perpassados por tecnologias. Estes dados carecem de aportes teóricos e metodológicos interdisciplinares para fazer frente às constantes transformações em curso na educação contemporânea.

Neste cenário educacional de big data (Daniel, 2017; Fisher *et al.*, 2020), onde as informações em suportes digitais são intensamente produzidas e consumidas por alunos, professores e gestores, o papel interdisciplinar da Ciência da Informação (CI) (Araújo, 2009; Borko, 1968; Saracevic, 1996) assume uma posição de destaque. A CI apresenta-se enquanto ciência capaz de catalisar soluções interdisciplinares para as

demandas informacionais da sociedade na contemporaneidade (Coneglian *et al.*, 2021), alcançando naturalmente a educação.

Abordagens educacionais centradas em dados que articulem contribuições de diferentes ciências, notadamente a CI, a Ciência da Computação (CC) e a Educação, têm o potencial de apoiar o processo de ensino-aprendizagem de várias formas. Elas se fazem indispensáveis para a evolução da educação e da própria sociedade da informação (Angeli *et al.*, 2017; Kitchin, 2014; Romero & Ventura, 2020).

O processo de ensino e aprendizagem apoiado em técnicas da inteligência artificial e da ciência de dados pode melhorar a aprendizagem (presencial ou a distância) dos alunos, apoiar a ação docente e fornecer subsídios informacionais para o processo de tomada de decisão dos gestores (Aldowah *et al.*, 2019). A maior parte da pesquisa recente envolvendo dados na educação tem se dedicado ao comportamento e ao desempenho dos alunos (Baig *et al.*, 2020), porém, outras vertentes também são investigadas.

Uma massiva quantidade de dados tem sido utilizada em ambientes inteligentes de aprendizagem autorregulados (Gambo & Shakir, 2021), na avaliação de alunos, professores, métodos etc. (Rodrigues *et al.*, 2018), em predições de desempenho acadêmico (Costa *et al.*, 2017; Karlos *et al.*, 2020; Tomasevic *et al.*, 2020), em projetos de universidades inteligentes (Artífice *et al.*, 2021), dentre outras aplicações.

Um universo de dados, informações e conhecimentos de diferentes ciências se fazem necessários para a concepção de soluções computacionais efetivas para as demandas educacionais. Objetos, processos e conhecimentos envolvidos no ambiente educacional são representações da “informação como coisa” uma vez que têm a qualidade de transmitir conhecimentos ou comunicar informações; de serem instrutivos e, portanto, informativos (Buckland, 1991).

Este estudo inspira-se na histórica definição de CI de Borko (1968), que aponta para o corpo de conhecimentos relacionados a origem, coleção, organização, armazenamento, recuperação, interpretação, transmissão, transformação e utilização da informação. Isto inclui a pesquisa sobre a representação da informação em sistemas naturais e artificiais, o uso de códigos para a transmissão eficiente da mensagem, bem como o estudo do processamento e de técnicas computacionais.

Os resultados parciais ora obtidos integram um estudo em andamento que assume o pressuposto de que é possível e necessário identificar colaboração e complementaridade nas interfaces entre conhecimentos científicos de diferentes áreas, e que a CC, aliada à CI, pode favorecer a busca pela melhoria contínua dos processos educacionais que explora as técnicas baseadas em dados.

OBJETIVOS

“Para articular e organizar os conhecimentos e assim reconhecer e conhecer os problemas do mundo, é necessária a reforma do pensamento. Entretanto, esta reforma é paradigmática, e não programática” (Morin, 2011, p. 33).

A Teoria da Complexidade de Edgar Morin tem suas bases na teoria dos sistemas, na teoria da informação e na cibernética (Morin, 2015). Esta constelação de saberes que integram a epistemologia complexa registra a importância do diálogo entre as ciências, da inter e transdisciplinaridade, capaz de capturar a essência do que é naturalmente complexo (tecido junto).

Neste contexto, o presente trabalho tem como objetivo estabelecer conexões preliminares entre os saberes de diferentes disciplinas científicas tendo em vista o objetivo de compatibilizar aportes teóricos e práticos para a proposição de soluções informacionais para o contexto educacional.

Como objetivos específicos tem-se a elaboração de conjuntos representativos, ainda que preliminares, de conceitos e técnicas provenientes da CC e da CI que endereçam um conjunto também representativo de necessidades informacionais dos diversos atores presentes no ambiente educacional.

Estes objetivos estão alinhados com o Objetivo de Desenvolvimento Sustentável (ODS) de Número 4 da ONU, que visa a Educação de Qualidade e que tem como um de seus alvos a oferta de “Ambientes de aprendizagem efetivos” (UNESCO, 2017). Naturalmente, bons mecanismos de suporte informacional destes ambientes, incluindo os ambientes virtuais de aprendizagem, devem ser levados a termo pelas comunidades científicas e praticantes das áreas envolvidas, notadamente observando as questões éticas envolvidas (OECD, 2019; UNESCO, 2021).

PROCEDIMENTOS METODOLÓGICOS

A metodologia adotada para a pesquisa proposta é aplicada, de natureza qualitativa, bibliográfica, descritiva e exploratória. Esta proposta de estudo teve como ponto de partida a experiência particular dos autores, atuantes em pesquisas e na docência nas áreas da CC, CI e Educação. Ela permitiu prospectar a construção de um conjunto representativo, ainda que preliminar, de estratégias informacionais e computacionais para atender às necessidades educacionais.

Para subsidiar a validação deste conjunto de conceitos, métodos e técnicas foram buscadas na literatura científica recente pesquisas que

envolvessem soluções educacionais dirigidas a dados. Artigos de revisão (Baig *et al.*, 2020; Daniel, 2017; Rodrigues *et al.*, 2018; Romero & Ventura, 2020) desempenharam papel importante no fornecimento de visões consolidadas nas pesquisas nesta área, ainda que em formatos diferentes de organização dos seus achados. Deste *corpus* científico foram coletados e organizados os resultados a seguir apresentados.

RESULTADOS

Partindo dos três eixos científicos visados no escopo deste estudo, foram realizadas pesquisas bibliográficas em portais de buscas acadêmicas. As publicações científicas analisadas envolveram necessariamente o contexto educacional e abordagens alcançadas pelo conceito de big data, em suas várias formas de apresentação.

No eixo computacional foram reunidos os principais conceitos, métodos e técnicas inseridos no escopo da ciência de dados e diretamente relacionados com os tipos de sistemas e soluções demandados pela educação.

No eixo informacional foram reunidas as principais ferramentas, metodologias e temáticas endereçadas pela Ciência da Informação com vistas a suportar soluções aplicáveis no ambiente educacional.

As possibilidades de conexões entre estes dois eixos e o eixo educacional são apresentadas na subseção 4.3 e constituem a principal contribuição deste trabalho. Os itens a seguir não são formalmente caracterizados por restrições de espaço.

Contribuições da Ciência da Computação (CC)

A CC é responsável por fornecer teorias, métodos e técnicas para que serviços computacionais envolvendo processamento, armazenamento e comunicação de dados sejam oferecidos. Naturalmente, esta área é responsável por caracterizar a ciência de dados, temática deste estudo.

São várias as formas de agrupamento das ferramentas, técnicas e tarefas desenvolvidas no âmbito computacional. Embora simplificada, uma organização possível das principais abordagens da CC presentes nas soluções educacionais é dada a seguir:

- MD: Mineração de Dados (com abordagens supervisionadas ou não)
 - Classificação/Predição
 - Agrupamento
 - Mineração de Regras de Associação
 - Identificação de padrões sequenciais
- LF: Lógica Fuzzy (de forma isolada ou hibridizada)

- CN: Computação Natural
 - Algoritmos Genéticos
 - Algoritmos baseados em colônias de formigas
 - Redes Neurais Artificiais
- MT: Mineração de Texto e Processamento de Linguagem Natural
- WS: Web Semântica e Linked Data
- DW: Técnicas de Data Warehouse

Contribuições da Ciência da Informação (CI)

No âmbito da CI, há uma intensa e intrínseca proximidade de suas contribuições, dado que seu objeto de pesquisa é a informação, com aquelas provenientes da CC, que também se dedica ao tratamento dos dados, em específico no formato digital.

Esta complementaridade de contribuições entre a CI e a CC seguramente não é possível de ser capturada em uma classificação como esta, que pretende ser meramente didática. Esta simplificação não deve ser interpretada como vinculações mutuamente excludentes entre estas ciências.

- SOC: Organização do Conhecimento (Ontologias, Tesouros, Dicionários Controlados etc.)
- EMI: Estudos Métricos da Informação (Webometria, Infometria, Cientometria, Bibliometria e Altmetria)
- CoInfo: Competência em Informação
- UI&UX: Estudos de usuários da informação (interface e experiência do usuário)
- RI: Recuperação da Informação

Demandas educacionais

O contexto educacional é complexo e suas demandas são dinâmicas, epistemologicamente situadas e socialmente (re)construídas. Questões educacionais naturalmente envolvem diferentes recursos humanos, materiais e teórico-metodológicos em um diálogo constante com o processo de ensino-aprendizagem. Para apresentá-las de uma forma organizada, optou-se por considerar os usuários finais aos quais as demandas são vinculadas, o que claramente é uma simplificação da realidade. São eles: alunos, professores e gestores educacionais.

- Demandas dos alunos:
 - Mecanismos de alertas (de prazos, de desempenho, preditivos de reprovação etc.)

- Recomendação de conteúdos
 - Conforme perfil (traço de personalidade, estilo de aprendizagem, objetivos educacionais etc.)
 - Conforme dúvida ou dificuldade encontrada
 - Conforme desempenho em atividades e avaliações
- Sistemas Tutores Inteligentes (intervenções análogas às de tutores humanos)
- Aprendizagem Colaborativa (interação/avaliação entre os pares, compartilhamento de informações etc.)
- Gestão da Aprendizagem (planejamento de atividades, direcionamento de estudos, painel de indicadores etc.)
- Promoção de estratégias de engajamento (por gamificação, por auto percepção motivacional etc.)
- Feedbacks personalizados (imediatos e proativos)

As figuras das conexões entre demandas educacionais e contribuições da CC e da CI utilizam as siglas definidas nas seções 4.1 e 4.2 ao lado de cada item. A figura 1 apresenta as conexões das demandas dos alunos com as contribuições da CC e da CI.

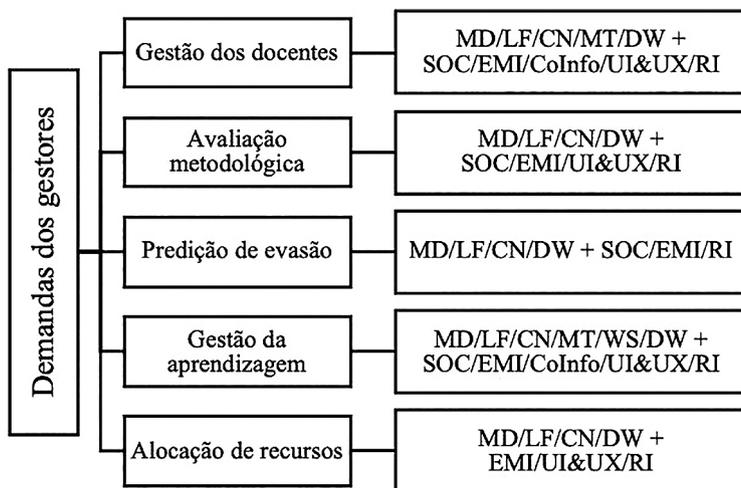


Figura 1. Demandas educacionais dos alunos. Fonte: Dados da pesquisa, 2021.

- Demandas dos professores
 - Mecanismos de alertas (preditivos de reprovação, comportamento anormal, prazos etc)
 - Recomendação de recursos e conteúdos pedagógicos (curadoria)
 - Apoio à avaliação dos alunos (correção automática de respostas)

- Agrupamento de alunos (por domínio de conteúdos, por perfis similares ou complementares, preferências de formatos de mídias etc.)
- Identificação de padrões (por *learning pathways*, regras de associação “se-então”, estilos de aprendizagem etc.)
- Detecção de similaridade e plágio
- Descoberta de padrões sequenciais em comportamentos dos alunos
- Gestão de indicadores de desempenho individual e coletivo dos alunos

As conexões destas demandas dos professores com as contribuições da CC e da CI são apresentadas na Figura 2.

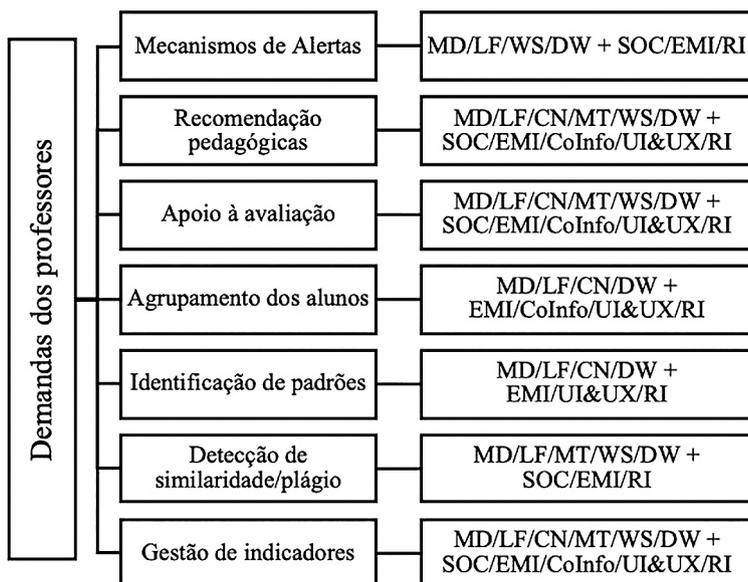


Figura 2. Demandas educacionais dos professores. **Fonte:** Dados da pesquisa, 2021.

- Demandas dos gestores
 - Gestão dos docentes (monitoramento)
 - por presença remota/presencial
 - pelo desempenho em indicadores
 - pela percepção dos alunos
 - por cumprimento de prazos/metras
 - Avaliação metodológica institucional (modelos pedagógicos) e indicadores
 - Predição de evasão e análises de causalidade
 - Gestão da aprendizagem (indicadores chave de qualidade educacional)
 - Alocação de recursos (humanos e materiais) de forma otimizada

As conexões das demandas dos gestores da educação com as contribuições da CC e da CI são apresentadas na Figura 3.

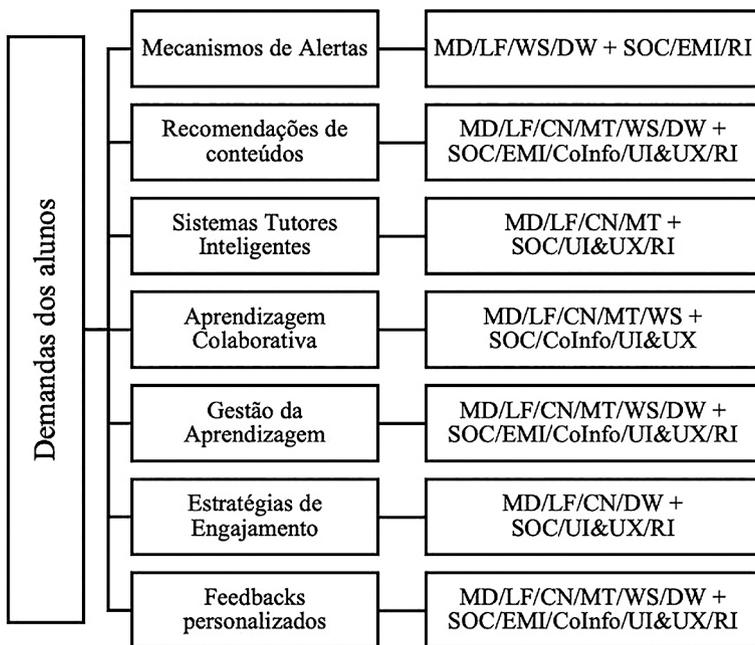


Figura 3. Demandas educacionais dos gestores. Fonte: Dados da pesquisa, 2021.

Em síntese, MD, LF, CN, MT, DW, SOC, WS e RI foram as contribuições da CC e da CI mais recorrentes nas proposições de soluções para as demandas educacionais.

CONSIDERAÇÕES FINAIS

Os resultados obtidos vão ao encontro dos objetivos propostos e têm o potencial de contribuir com futuros estudos que visem melhorar práticas educacionais com técnicas orientadas a dados da CC e da CI.

O paradigma da complexidade foi evidenciado nas conexões interdisciplinares neste recente e dinâmico campo de estudos que se forma na interseção entre educação e as ciências da computação e da informação.

Conclui-se ainda que as contribuições interdisciplinares identificadas neste estudo são representativas das práticas científicas em curso na interface entre as áreas citadas, prospectando trabalhos futuros que visem replicar e expandir os resultados publicados.

CONFLITOS DE INTERESSE

Os autores declaram não haver conflito de interesses.

DECLARAÇÃO DE CONTRIBUIÇÃO

Concepção, supervisão, metodologia, curadoria dos dados, escrita do original e revisão e edição do texto: Patrícia Takaki.

Supervisão e revisão e edição do texto: Moisés Dutra.

DECLARAÇÃO DE CONSENTIMENTO DE DADOS

Nenhum dado foi gerado durante o desenvolvimento deste estudo. 

REFERÊNCIAS

- ALDOWAH, H., AL-SAMARRAIE, H., & FAUZY, W. M. (2019). Educational Data Mining and Learning Analytics for 21st Century higher education: A review and synthesis. *Telematics and Informatics*, 37, 13-49. <https://doi.org/doi:10.1016/j.tele.2019.01.007>
- ANGELI, C., HOWARD, S. K., MA, J., YANG, J., & KIRSCHNER, P. A. (2017). Data mining in educational technology classroom research: Can it make a contribution? *Computers & Education*, 113(113), 226-242. <https://doi.org/10.1016/j.compedu.2017.05.021>
- ARAÚJO, C. A. Á. (2009). Correntes teóricas da ciência da informação. *Ciência Da Informação*, 38(3), 192-204. <https://doi.org/10.1590/s0100-19652009000300013>
- ARTÍFICE, A. F. V. P., SARRAIPA, J., & JARDIM-GONCALVES, R. (2021). Improvement of Student Attention Monitoring Supported by Precision Sensing in Learning Management Systems. In I. Dei (Ed.), *Computer-Mediated Communication*. <https://doi.org/10.5772/intechopen.98764>
- BAIG, M. I., SHUIB, L., & YADEGARIDEHKORDI, E. (2020). Big data in education: a state of the art, limitations, and future research directions. *International Journal of Educational Technology in Higher Education*, 17(1). <https://doi.org/10.1186/s41239-020-00223-0>
- BORKO, H. (1968). Information Science: What is it? *American Documentation*, 19(1), 3-5. <https://doi.org/10.1002/asi.5090190103>
- BUCKLAND, M. K. (1991). Information as thing. *Journal of the American Society for Information Science*, 42(5), 351-360. [https://doi.org/10.1002/\(SICI\)1097-4571\(199106\)42:5<351::AID-A-SI5>3.0.CO;2-3](https://doi.org/10.1002/(SICI)1097-4571(199106)42:5<351::AID-A-SI5>3.0.CO;2-3)

- CONEGLIAN, C. S., VALENTIM, M. L. P., & SANTAREM SEGUNDO, J. E. (2021). Muli e interdisciplinaridade entre a Ciência da Informação e a Ciência da Computação no âmbito da Web Semântica. *Informação & Sociedade: Estudos*, 31(1), 1-18. <https://doi.org/10.22478/ufpb.1809-4783.2021v31n1.52120>
- COSTA, E. B., FONSECA, B., SANTANA, M. A., DE ARAÚJO, F. F., & REGO, J. (2017). Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. *Computers in Human Behavior*, 73(73), 247-256. <https://doi.org/10.1016/j.chb.2017.01.047>
- DANIEL, B. K. (2017). Big Data in Higher Education: The Big Picture. In B. K. Daniel (Ed.), *Big Data and Learning Analytics in Higher Education* (pp. 19-28). https://doi.org/10.1007/978-3-319-06520-5_3
- FISCHER, C., PARDOS, Z. A., BAKER, R. S., WILLIAMS, J. J., SMYTH, P., YU, R., ... WARSCHAUER, M. (2020). Mining Big Data in Education: Affordances and Challenges. *Review of Research in Education*, 44(1), 130-160. <https://doi.org/10.3102/0091732x20903304>
- GAMBO, Y., & SHAKIR, M. Z. (2021). WIP: Model of Self-Regulated Smart Learning Environment. *2021 IEEE World Conference on Engineering Education (EDUNINE)*, 1-4. <https://doi.org/10.1109/edunine51952.2021.9429090>
- 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. <https://doi.org/10.3390/app10238413>
- KITCHIN, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1), 205395171452848. <https://doi.org/10.1177/2053951714528481>
- MORIN, E. (2011). *Os sete saberes necessários a educação do futuro* (2nd ed.; C. Eleonora & Jeanne Sawaya, Trans.). Brasília/DF: Unesco.
- MORIN, E. (2015). *Introdução ao pensamento complexo* (5th ed.; E. Lisboa, Trans.). Porto Alegre: Sulina.
- OECD. (2019). *Recommendation of the Council on Artificial Intelligence - OECD Legal Instruments 0449*. Retrieved from Organisation for Economic Co-operation and Development website. Obtido de: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>
- RODRIGUES, M. W., ISOTANI, S., & ZÁRATE, L. E. (2018). Educational Data Mining: A review of evaluation process in the e-learning. *Telematics and Informatics*, 35(6), 1701-1717. <https://doi.org/10.1016/j.tele.2018.04.015>
- ROMERO, C., & VENTURA, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery*, 10(3), 1-21. <https://doi.org/10.1002/widm.1355>

- SARACEVIC, T. (1996). Ciência da informação: origem, evolução e relações. *Perspectivas Em Ciência Da Informação*, 1(1), 41-62. Obtido de <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/235/22>
- TOMASEVIC, N., GVOZDENOVIC, N., & VRANES, S. (2020). An overview and comparison of supervised data mining techniques for student exam performance prediction. *Computers & Education*, 143(143), 103676. <https://doi.org/10.1016/j.compedu.2019.103676>
- UNESCO. (2017). *Ensure quality education for all: Sustainable development goal 4; ten targets*. In UNESDOC. Obtido de <https://unesdoc.unesco.org/ark:/48223/pf0000259784>
- UNESCO. (2021). *AI and education-Guidance for policymakers*. Obtido de <https://unesdoc.unesco.org/ark:/48223/pf0000376709>

