



COLOQUIO INTERNACIONAL  
DE GESTIÓN UNIVERSITARIA  
URUGUAY 2024

Una nueva gestión para una Universidad en Movimiento

Montevideo, Uruguay

02, 03 y 04 de octubre de 2024



## GESTÃO UNIVERSITÁRIA, BIG DATA E DATA DRIVEN: REFLEXÕES A PARTIR DE UMA REVISÃO NA LITERATURA

**ELISANGELA HOFFMANN**

Universidade do Estado de Mato Grosso

[Elisangela.hoffmann@unemat.br](mailto:Elisangela.hoffmann@unemat.br)

**ANTONIO CEZAR BORNIA**

Universidade Federal de Santa Catarina

[cezar.bornia@ufsc.br](mailto:cezar.bornia@ufsc.br)

### RESUMO

Esta pesquisa tem como objetivo explorar o uso da abordagem Data Driven no contexto das Instituições de Ensino Superior (IES), por meio de uma Revisão Sistemática de Literatura (RSL), na plataforma Web Of Science (WoS). O software VOSviewer foi utilizado para apoio às análises. O corpus da pesquisa foi composto por 36 artigos, publicados a partir de 2014. Os resultados indicam que não há estudos em países da América do Sul, há poucos estudos em áreas importantes para as IES como administração, bem estar acadêmico, pesquisa e ciência aberta. Quase a totalidade dos artigos abordam o uso de dados para gestão do ensino, com modelos prescritivos e preditivos para medir, intervir e melhorar o desempenho acadêmico. Com base nos achados, é realizada uma reflexão no contexto das IES sobre a categorização do aluno em métricas de desempenho, suas implicações práticas e políticas, principalmente pelo uso de algorítmicos que podem levar à discriminação. Também refletimos a escassez de evidências que demonstrem benefícios (ou custos) do Data Driven num contexto de adaptação pós-pandemia e de crises de financiamento, embora, uma a transformação digital crítica possa fortalecer o ensino superior frente a tantos desafios.

**Palavras chave:** Dataficação das IES; Big Data; Gestão Baseada em Dados; Data Driven; Gestão Universitária.

## INTRODUÇÃO

As Instituições de Ensino Superior (IES) estão inseridas em um contexto global de crescente pressão para transformar sua estrutura operacional e governativa, produzir melhores resultados nos processos de ensino-aprendizagem, atuar com eficácia e transparência nos processos de gestão, acomodando novas agendas econômicas e sociais, pressões do gerencialismo corporativo e da Nova Gestão Pública (DANIEL, 2017; RAFFAGHELLI; SANGRÀ, 2023). A forte redução de financiamento governamental em muitos países impeliu as IES a buscar estratégias para sobreviver e manter sua posição num ambiente competitivo, pressionadas para cortar custos e manter suas atividades (MENON, 2014).

Com a alta demanda por ensino superior e a expansão massiva associada dos sistemas de ensino superior na segunda metade do século XX, as IES foram chamadas a adotar métodos de gerenciamento que lhes permitissem identificar e atender às necessidades sua comunidade cada vez maior e mais diversificado (MENON et al., 2014). A busca por medidas quantificáveis de resultado e desempenho, da eficiência e produtividade na pesquisa, no ensino e administração foi um motor-chave da cultura das IES nos últimos vinte anos, incorporando ainda conceitos como competitividade e qualidade, estabelecimento de métricas para comparações e elaboração de rankings de eficiência e performance, inclusive internacionais (POZZI et al., 2019; RAFFAGHELLI; SANGRÀ, 2023; GIBBS, 2020).

Esse movimento se amplifica, nos últimos dez anos, com a dataficação da sociedade e, em consequência, das universidades. A forçada adaptação das IES às tecnologias digitais durante a pandemia da COVID-19, acelerou ainda mais a agenda de modernização e dataficação das IES, que geram grandes quantidades de dados por meio da interação entre usuários e sistemas de gestão da aprendizagem, sistemas de registros dos alunos e sistemas de gestão orçamentária, tornando proeminente questões de como gerir melhor os conjuntos de dados e como dar sentido ao volume de dados e gerar valor a partir deles, inserindo na agenda da gestão universitária termos como Inteligência Artificial (IA), Data Driven, Transformação Digital, Criação de Valor Público e Dados Governamentais Abertos (DGA) (BENMOHAMED et al., 2024; STOJANOV; DANIEL, 2024). Alinhadas a novos termos, novos contextos e problemáticas somaram complexidade a já difícil tarefa de gerir instituições tão diversas quanto as IES. Nesse cenário internacionalizado, as IES estão sob constante pressão, do mercado ou do governo, para quantificar e publicizar seus resultados, consolidar ou melhorar a sua reputação e prestígio, geralmente em termos de qualidade percebida, expressa por meio de índices qualitativos e quantitativos, que incluem processos e procedimentos de garantia de qualidade, avaliação, benchmarking e sistemas de classificações (rankings) universitária (POZZI et al., 2019; RAFFAGHELLI; SANGRÀ, 2023).

A gestão das IES na contemporaneidade perpassa então por aspectos positivos e negativos da dataficação da sociedade, por um lado, a medição e quantificação da realidade proporciona oportunidades para extrair conhecimentos para a gestão do ensino, da pesquisa e da administração, fomentando uma gestão orientada a dados, a eficiência e a produtividade e, por outro, comparações e métricas podem dirigir à conclusões precipitada e equivocadas sobre a finalidade dessas organizações sociais promotoras e difusoras do conhecimento, implicando em desafios de ordem prática e política dentro dessas instituições, que precisam compreender melhor as contribuições e limitações do uso de dados e seus impactos sociais e organizacionais (MENON, 2014; VAN ES; SCHÄFER, 2017; RAFFAGHELLI; SANGRÀ, 2023).

Embora a literatura tenha avançado em investigações sobre o Big Data no ensino superior, como um fenômeno novo no contexto universitário, a aplicação do uso dos dados para resolver questões críticas dentro do setor ainda é limitada (DANIEL, 2017; RAFFAGHELLI; SANGRÀ, 2023; STOJANOV; DANIEL, 2017). Há, portanto, relevância científica em compreender como o Data Driven tem sido abordado na gestão universitária, evidenciando contextos de aplicabilidade, seus resultados e ainda áreas inexploradas que carecem de investigação. Visando contribuir com esse contexto, esse artigo objetiva explorar o uso da abordagem Data Driven (ou orientação por dados) no contexto das Instituições de Ensino Superior (IES), focando principalmente nas práticas do uso de grandes bancos de dados para o direcionamento e tomada de decisão, bem como suas implicações.

Partindo de um breve referencial teórico que aborda o avanço da dataficação, Big Data e abordagens Data Driven nas IES, propomos uma análise reflexiva abordando, dentro dos achados na RSL, as perspectivas dos usuários de dados, da funcionalidade, da construção de modelos aplicados e de estruturas conceituais para seu uso, de reflexões sobre o uso de dados, suas práticas, questões éticas, contribuições e limitações no contexto das IES. Esse estudo realizou uma RSL na plataforma WoS com as strings (((("BIG DATA" OR "DATA MINING") AND ("HIGHER EDUCATION" OR "UNIVERSITY" OR "UNIVERSITIES" OR "COLLEGE" OR "FACULTY") AND ("DATA DRIVEN" OR "DATA-DRIVEN")))). De um total de 120 documentos encontrados, após os procedimentos de filtros e exclusões descritos na metodologia, o corpus do estudo foi composto por 36 Artigos selecionados para a leitura integral e análises. A pesquisa utilizou ainda o software VOSviewer foi utilizado para apoio às análises.

## **1. BIG DATA E DATA DRIVEN NAS IES**

O paradigma baseado em dados, transformar todas as coisas existentes nesse formato, quantificar e conectar, tornou-se um determinante dos processos e práticas sociais do século XXI, anunciando o Big Data como o motor do progresso tecnológico e social sem precedentes (VAN ES; SCHÄFER, 2017). O Big Data tem diversas abordagens conceituais, mas sumariamente são grandes volumes de dados, heterogêneos e que podem ser acessados rapidamente (volume, a variedade e a velocidade) (STOJANOV; DANIEL, 2024). O Big Data é usado para descrever o que envolve a recolha, processamento e análise de grandes conjuntos de dados e suas práticas estão a mudar a forma como o conhecimento é produzido, principalmente pela atribuição de valor económico no processamento de informação e a monetização do conhecimento (VAN ES; SCHÄFER, 2017). O ritmo exponencial da dataficação das coisas abriu uma espécie de precipício para a compreensão da nova realidade da massificação e uso dos dados, sua rápida difusão em vários setores e seu contexto de uso nas IES, historicamente produtoras e disseminadoras do conhecimento científico (RAFFAGHELLI, JE, SANGRÀ, A., 2023).

Esta situação foi acompanhada por uma falta inicial de crítica social que desafiou a postura otimista sobre o poder dos dados (VAN ES; SCHÄFER, 2017). As IES de fato foram afetadas pela dataficação de formas profundas, principalmente por uma junção de contextos nas últimas duas décadas. A modernização do ensino superior através da transformação digital perpassa por debates sobre o gerencialismo e formas de controle burocrático sobre as duas principais atividades acadêmicas, o ensino e a pesquisa, somados a democratização da ciência pela publicidade e as práticas de dados abertos e, ainda, ao desenvolvimento da indústria da Inteligência Artificial, resultaram nas IES em impactos sobre a própria identidade formativa, o profissionalismo dos acadêmicos, tensões e contradições entre as partes interessadas, que

adensam em complexidade as práticas o uso de dados (RAFFAGHELLI, JE, SANGRÀ, A., 2023).

Questões de produção e uso responsável de dados, de atribuição de significado, ética, privacidade e transparência no tratamento de dados constituem agora o núcleo deste novo paradigma (VAN ES; SCHÄFER, 2017). Inovações catalisadas pela tecnologia e dataficação das IES, com abordagens de aprendizagens flexíveis, personalizadas, avaliações bem informadas, relações estreitas com a comunidade e o mercado de trabalho, resultados abertos e transparentes, ressaltando a aprendizagem, o ensino e gestão baseados em dados tem mobilizado a gestão das IES, que buscam incorporar as novas tendências e tecnologias (RAFFAGHELLI, JE, SANGRÀ, A., 2023). Além disso, o número e a complexidade crescentes dos desafios enfrentados pelas IES irão, inevitavelmente, resultar em tentativas de aprimorar processos e serviços com tentativas de recolher e utilizar evidências sob a forma de dados apropriados e relevantes (MENON et al., 2014).

No ambiente das IES, o Big Data incorpora áreas de investigação como a mineração de dados educacionais, o desenvolvimento de ferramentas para sua análise, as análises de aprendizagem individuais, utilizando volumosas quantidades de dados, que vão desde registros de exames e testes de alunos até registros financeiros e orçamentários da instituição (2017). Certamente a utilização crescente de tecnologias digitais para apoiar a aprendizagem, o ensino e a administração têm gerado uma quantidade significativa de dados, principalmente através do envolvimento de usuários e sistemas de gestão e de aprendizagem e, aproveitando o sucesso da utilização de dados para gestão em outros setores, as IES tem a oportunidade para, aplicando técnicas de Big Data, obter informações valiosas para a tomada de decisão (STOJANOV; DANIEL, 2024).

Na intersecção desses contextos, as IES são pressionadas a aderir movimentos como o Data Driven, a orientação baseada dados, implicando que decisões racionais e eficientes precisam apoiar-se na análise e interpretação de dados e que esforços devem ser envidados na captura, tratamento e atribuição de significado e valor a esses dados. Ao buscar percepções da sua realidade a partir de dados para gerar estratégias e acompanhar a complexidade de seu meio, as universidades têm no Big Data potencial para examinar seus desafios e identificar formas de enfrenta-los a partir da análise de grandes bancos de dados. A cultura ou gestão baseada em dados tem sido difundida como prática social nos setores público e privado, prescrevendo que funcionários e gestores observem as informações provenientes de dados disponíveis para tomar decisões (STOJANOV; DANIEL, 2024; RAFFAGHELLI, JE, SANGRÀ, A., 2023)

A teoria da tomada de decisão defende fortemente o uso de pesquisas sobre os mecanismos utilizados por gestores para decidir, subentendendo que os decisores e formuladores de políticas precisam a selecionar as alternativas "certas", visando tanto a eficiência quanto a eficácia nos resultados entregues (MENON et al., 2014). Dado que a abordagem convencional, a coleta de dados pode ser limitada por de captura, limitando ainda a interpretação, a abordagem baseada em grandes quantidades de dados, permitem eliminar barreiras da inadequação da amostragem, aumentando a confiança na generalização e melhorando o rigor da validade (DANIEL, 2017).

Nesse contexto, os tomadores de decisão se esforçam para tomar decisões informadas, com base em um exame e análise completos de dados relevantes, utilizando-os como insumo no planejamento e na tomada de decisões no ensino superior, o que pode beneficiar diversas áreas de atuação, com várias práticas e perspectivas relacionadas a diferentes tipos de dados (Menon et al., 2014). Também, a intensificação da recolha de dados para a tomada de

decisões no ensino superior está principalmente ligada à disponibilidade de dados provenientes de redes sociais, repositórios de dados online, bibliotecas digitais educativas (GIBS, 2014). Essa abordagem orientada a dados, poderia beneficiar atividades de ensino, aprendizagem, pesquisa e gestão, tratando de novas abordagens e decisões fundamentadas, principalmente com ambientes de pesquisa e ensino mais interligados, flexíveis e colaborativos (CZERNIEWICZ, 2022). A gestão orientada a dados, ou Data Driven, é então um promissor campo que poderia contribuir para o conhecimento dos processos e resultados de aprendizagem e ensino, auxiliando alunos e educadores a tomar decisões melhores decisões sobre aprendizagem e ensino rumo à qualidade e eficácia (DANIEL, 2017).

Por outro lado, o entusiasmo desmedido e a ausência de criticidade sobre os resultados “neutros” e precisos produzidos por análise de grandes bancos de dados foram, progressivamente, cedendo espaço para questionamentos sobre as limitações e os fatores que compõe um conjunto de dados, moldam a análise e geram resultados, e temas como impacto social e formas (VAN ES; SCHÄFER, 2017). Também, as consequências críticas da adoção de plataformas e ferramentas digitais, tendo em conta o acesso dos alunos aos espaços digitais, bem como os desequilíbrios epistêmicos na entrega de conteúdos da IES como fonte de poder para os alunos nas suas diversas culturas (CZERNIEWICZ, 2022). De toda forma, como um fenômeno novo no contexto universitário, a abordagem orientada a dados carece, tal qual a aplicação de Big Data nessas organizações, de estudos sobre suas oportunidades e limitações (Daniel, BK, 2017).

## 2. METODOLOGIA

Tranfield et al. (2003) argumentam que as revisões de literatura propiciam ao pesquisador mapear e avaliar o território intelectual relevante e, nesse sentido, as Revisões Sistemáticas de Literatura (RSL) se diferem por adotar um processo replicável, científico, transparente e detalhada, uma trilha de auditoria das decisões dos revisores e dos procedimentos utilizados. Com características metodológicas desejáveis, Tranfield et al. (2003) propõem uma lista de etapas para a revisão sistemática, associadas a três estágios de revisão: Estágio I – Planejando a revisão (identificação da necessidade, preparação de uma proposta, desenvolvimento de um protocolo); Estágio II – Conduzindo uma revisão (identificação da pesquisa, seleção dos estudos, avaliação da qualidade do estudo, extração de dados e monitoramento do progresso, síntese de dados); Estágio III – Relatório e divulgação (relatório e recomendações, colocando as evidências em prática).

Na fase inicial, no Planejamento da revisão sistemática, foi necessária uma ampla pesquisa (“Big Data” AND “Data Driven” AND “Universities”) para a familiarização com os termos, escopo da literatura e palavras-chave. Ao planejar uma revisão sistemática, um estudo do escopo visa definir uma discussão conceitual do problema de pesquisa, a identificação da necessidade da revisão, um processo de exploração, descoberta e desenvolvimento. No planejamento da revisão sistemática, o escopo foi definido de acordo com o objetivo da pesquisa, explorar o uso da abordagem Data Driven no contexto das Instituições de Ensino Superior (IES). O escopo delimita-se nos estudos sobre Data Driven em IES, incluindo temas relacionados como Big Data e Data Mining. A partir da definição do escopo e da literatura, com a identificação de palavras-chave e termos de busca, são decididas as strings de pesquisa. As strings de pesquisa foram adaptadas do modelo sugerido por Daniel (2024) em uma RSL sobre as aplicações de Big Data no ensino superior, alterando o foco da pesquisa para Data Driven. Pesquisamos (((“BIG DATA” OR “DATA MINING”) AND (“HIGHER EDUCATION” OR “UNIVERSITY” OR “UNIVERSITIES” OR “COLLEGE” OR

"FACULTY") AND ("DATA DRIVEN" OR "DATA-DRIVEN"))), no campo Resumo da Base de dados Web Of Science (WoS). A WoS é a principal plataforma mundial multidisciplinar de pesquisa que contém conjuntos de dados para estudos analíticos, em diversas áreas do conhecimento, sendo utilizado por cientistas globais de diferentes maneiras para responder a questões científicas (LI et al., 2018).

Para a seleção e análise dos artigos, o foco foi o contexto da utilização de Data Driven nas IES, particularmente interessado em como os dados de Big Data estão sendo utilizados para a tomada de decisão de processos e procedimentos nessas organizações, ainda quais suas consequências e se há evidências da sua contribuição para melhores práticas e melhores resultados. Na segunda etapa proposta por Tranfield et al. (2003), a condução da revisão é realizada. A busca (maio de 2024) resultou em 120 entradas, não foram encontradas entradas duplicadas. Foram selecionados apenas documentos do tipo Artigos (77). Não foram aplicados outros filtros. Para a triagem dos artigos, inicialmente, foi realizada a leitura dos títulos e dos resumos para determinar sua relevância para as questões de pesquisa e se os artigos seriam utilizados para análise.

Foram excluídos artigos (1) com publicação excluída da base de dados, (1) com foco na alimentação dos alunos, (3) com foco nas políticas regulatórias de profissão, (1) que trata de implicações do Big Data para economia, (6) onde o foco é o desenvolvimento e exercício da profissão médica, (1) que estuda os níveis de ensino fundamental e médio, (14) cujos objetos de estudo não são assuntos inerentes às universidades e, (2) com foco em disciplinas acadêmicas que tratam de Big Data, restando 43 artigos para leitura integral. Após a leitura integral dos artigos, todos os artigos que tratavam dos contextos de Data Driven nas IES foram selecionados para a análise, não diferenciando estudos empíricos ou bibliográficos, contudo foram excluídos documentos (6) que não apresentaram aderência com o objetivo da pesquisa.

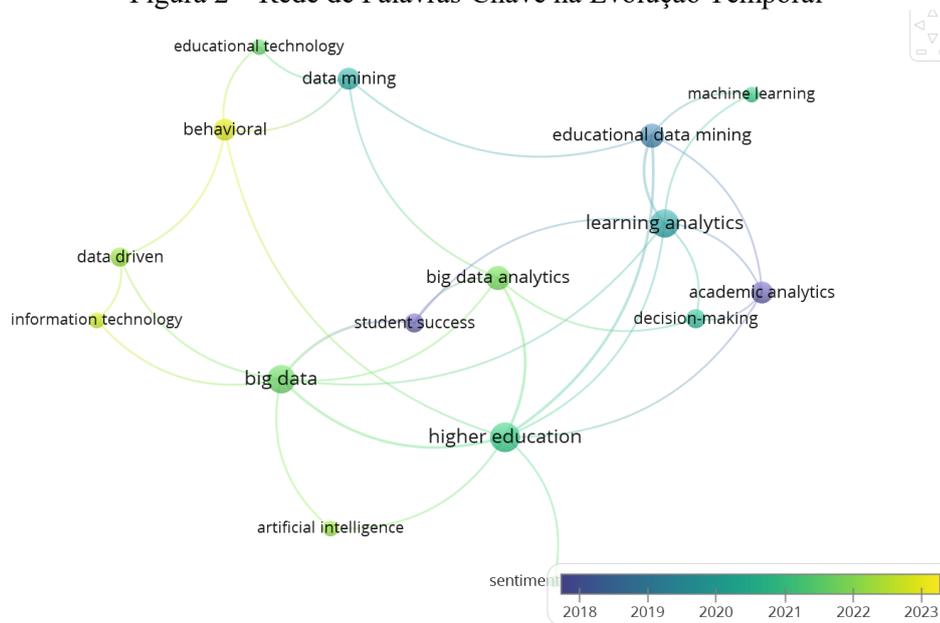
Dessa forma, 36 artigos compuseram o corpus da pesquisa. Os artigos foram analisados sob duas perspectivas: bibliométrica e de conteúdo. Para a análise bibliométrica foram coletados dados quanto a informações de citação, bibliográficas, resumo e palavras-chave e informações quanto às referências. Os dados foram analisados por meio do software VosViewer. Tal software, disponibilizado pela Universidade de Leiden, propicia a construção e visualização de redes bibliométricas, que podem incluir periódicos, pesquisadores ou publicações individuais e podem ser construídas com base em citações, acoplamento bibliográfico, cocitação ou relações de coautoria. Quanto à análise de conteúdo, foi realizada por meio da leitura integral e categorização dos artigos quanto a seus objetivos e abordagens de Data Driven nas IES. O último estágio da revisão sistemática proposto por Tranfield et al. (2003) pressupõe a elaboração de um relatório dos achados durante as fases anteriores, a apresentação dos resultados da pesquisa.

### **3. RESULTADOS**

Anteriormente à leitura completa dos artigos, foram realizadas análises preliminares permitindo uma visualização prévia do estudo. Da amostra selecionada para leitura integral, é possível observar que as publicações sobre o tema Data Driven iniciaram em 2015 (1), em 2016 (1), em 2018 (3), em 2019 (4), 2020 (5), 2021 (6), 2022 (12), 2023 (9) e 2024 (2), com maior frequência de publicações de estudos nos anos de 2022 e 2023, como demonstrado na Figura 1. Tal fato indica que há maior concentração de estudos nos anos mais próximos de 2021, demonstrando que o tema abordado é atual.

Foram realizadas análises de redes por meio do software VOSViewer, iniciando com as palavras-chave, conforme apresentado na Figura 2. Pode-se observar que as palavras “Big Data”, “Higher Education”, “Learning analytics” e “Educational Data mining” tem maior incidência. Outra observação importante é que o foco nas palavras-chave dos estudos migrou de “academic analytics” para, a partir do ano de 2022, “Data Driven” e “Artificial Intelligence”.

Figura 2 – Rede de Palavras-Chave na Evolução Temporal



Fonte: Autores (2024)

Tal mudança converge com a literatura, pois o foco dos estudos sobre a abordagem em dados em IES muda de como coletar dados para como dar sentido e gerar valor a partir deles (Raffaghelli, JE, Sangrà, A., 2023). Quanto ao país onde a pesquisa foi aplicada ou, na ausência de aplicação, de vinculação dos pesquisadores, a China (15) apresentou a maior quantidade de trabalhos, seguida dos Estados Unidos (4), África do Sul (2), Alemanha (2), Arábia Saudita (2), Austrália (1), Canadá (1), Coreia do Sul (1), Hong Kong (1), Malásia (1), México (1), Nova Zelândia (1), Paquistão (1), Reino Unido (1), República Tcheca (1), Turquia (1). De fato, a literatura aponta que a concentração geográfica da produção científica em tecnologia tem predominância de países desenvolvidos (Bozkurt, 2020).

Quanto ao conteúdo, os artigos foram classificados em cinco grandes grupos, conforme seus objetivos e métodos para investigar Data Driven nas IES de estudos: que buscam investigar pontos de vista de usuários de dados para a tomada de decisão (3 Artigos); que buscam compreender a funcionalidade dos dados para modelos preditivos e prescritivos (4 Artigos); que propõe, geralmente por meio de mineração de dados, modelos preditivos de desempenho acadêmico (5), prescritivos para melhorar o aprendizado (8) e modelos avaliativos (3); estudos que propõe estruturas conceituais para o uso de dados em IES (2), para sistemas de alerta precoce (1), para o uso de dados para gestão financeira das IES (1) e para avaliar a saúde dos alunos (1) e; estudos teóricos que discutem sobre a prática de big data e Data Driven nas IES (8). Cada grupo foi analisado separadamente. Por fim, uma reflexão sobre os achados nas pesquisas completa a sessão.

#### 4.1 PERSPECTIVA DE USUÁRIOS DE DADOS

Sobre a percepção dos usuários de dados das IES, de forma resumida, buscaram compreender melhor a utilização da capacidade de análise de big data para a tomada de decisões baseada em dados, a fim de obter melhor desempenho das IES.

Na Malásia, Ashaari et al. (2021) concluíram que esse é um campo a ser trabalhado e ainda pouco implementado, seja pela desestruturação dos sistemas de informação ou pela cultura organizacional pouco difundida, mesmo na alta direção dessas organizações e, para que esse processo tenha sucesso, a gestão de topo deve primeiro liderar pelo exemplo, comprometendo-se com as suas intenções sérias e instituindo planos e esforços para tornar uma realidade organizacional o Data Driven. Na mesma linha, mas sem focar na alta administração, Aseeri e Kang (2023) investigaram a relação entre cultura organizacional e a implementação de sistemas baseado em dados para a tomada de decisão na Arábia Saudita. Concluíram que é necessário que os decisores tracem estratégias para implementação desses sistemas, juntamente com a formação de pessoal. Já o estudo de Jones (2019) traz o diferencial em pesquisar a percepção de usuários de sistemas de Data Driven em plena funcionalidade. Nesse caso, o estudo buscou analisar a opinião de usuários decisores, principalmente conselheiros profissionais em uma instituição de ensino superior pública, sobre o uso de um sistema de aconselhamento prescritivos e preditivos. O autor conclui que os entrevistados rejeitaram as ferramentas devido a preocupações de usabilidade, desconforto moral e uma crença de que o uso de medidas preditivas violava um princípio ético profissional para desenvolver uma compreensão abrangente dos seus orientados.

#### 4.2 PERSPECTIVA DA FUNCIONALIDADE DOS DADOS

Nessa perspectiva, quatro artigos buscam fazer a análise da funcionalidade técnica dos dados em sistemas preditivos e prescritivos de aprendizagem. Primeiro, Gao, Dong e Chen (2024) analisam o modelo de plataforma educacional baseada em Big Data já em funcionamento em IES da China, buscando compreender seu desenvolvimento e demanda de dados para essas organizações. Concluem que ao aproveitar a engenharia web e a tecnologia multimídia facilita o compartilhamento de recursos educacionais entre os estudantes, estabelecendo uma plataforma educacional abrangente orientada por big data adaptada às IES. Também, funções como recuperação de informações educacionais universitárias existentes e elementos gráficos podem melhorar a utilização dos dados pelo seu valor inerente, além da abordagem de inteligência artificial recolher e analisar respostas de estudantes e educadores de forma online.

Por outro lado, ao analisar modelos de Data Driven voltados para análise de comportamento e aprendizagem de alunos, FounG e Chen (2022) utilizaram técnicas de Big Data para verificar os modelos de curso e progressão de alunos comparando o progresso entre os diferentes grupos, para decidir o design do curso por decisões baseadas em dados. Ao usar os dados, verificaram o uso da linguagem e as habilidades de citação dos alunos melhoram. Segundo os autores, os resultados indicam que abordagem Data Driven pode ser útil para escrever estudos de progressão, especialmente na era pós-covid (FOUNG; CHEN, 2022). Na mesma premissa, BULUT et al. (2023) analisaram a utilidade das avaliações formativas online na previsão do desempenho final dos alunos no curso, argumentando que os dados extraídos de avaliações formativas sejam utilizados na tomada de decisão dos educadores com insights práticos sobre como a aprendizagem dos alunos pode ser melhorada, fornecendo feedback aos alunos. Concluem que os recursos extraídos das avaliações formativas on-line (por exemplo, conclusão, carimbos de data e hora e pontuações) são preditores fortes e significativos do desempenho final dos alunos no curso, mas a pontuações das avaliações formativas são o preditor mais forte do desempenho dos alunos. Também Park, Yu e Jo

(2016) empregaram a abordagem Data Driven para, a partir de agrupamento de mineração de dados educacionais de comportamento online e do banco de dados de cursos da instituição, extrair características de atividade comuns e padrões de comportamento semelhantes dos alunos de 612 cursos em uma grande IES na Coreia do Sul, verificando quais sistemas de gestão de aprendizado estavam cumprindo com suas funcionalidades. Concluíram que a abordagem baseada em dados ajuda os líderes acadêmicos a monitorar o status do aprendizado combinado e desenvolver sistemas de forma estratégica, atuando na implementação de intervenção online.

#### 4.3 PERSPECTIVA DA CONSTRUÇÃO DE MODELO PARA O USO DE DADOS

A terceira perspectiva, da construção de modelos para o uso de dados, fortemente apoiados por técnicas de mineração de dados, é a mais extensa, contado com dezesseis artigos. As pesquisas utilizam a mineração de dados para construir modelos baseados em Big Data, conforme descritos no Quadro 1.

Quadro 1 – Perspectiva da construção de Modelos para uso de Dados

Autores	Cla.	Modelo e Dados utilizados
SAQR.; FORS; NOURI, 2018	Preditivo de desempenho acadêmico	prever desempenho acadêmico por meio de dados de rede social, na interação aluno professor
RAMASWAMI, G. et al., 2019		Construir modelo preditivo de desempenho acadêmico, com dados de atividades realizadas pelos alunos durante o curso, notas da avaliação dos alunos, notas finais
PAN, L. et al., 2020		Preditivo de desempenho acadêmico, com dados pontuações de avaliação dos alunos ao longo de um período de ensino
YAGCI, 2022		modelo baseado em algoritmos de aprendizado de máquina para prever as notas dos exames finais de alunos de graduação, usando dados de desempenho acadêmico
ZHOU et al., 2023		Preditivo de desempenho acadêmico, registros de frequência às aulas, alimentação e banho, utilizando redes diárias de pares de estudantes universitários e seus comportamentos
BEARD, D. et al., 2019	Prescritivos para melhorar o aprendizado	Modelo de utilização de corpora linguísticos (grandes coleções de textos reais) para embasar e melhorar o ensino gramatical, dados de uso gramatical.
TAN, 2022		Algoritmo para sugerir recursos de ensino e atividades personalizadas para cada aluno, com base na análise dos dados e nos padrões identificados
WANG, 2022		Desenvolver um algoritmo de ensino de tradução de inglês para IES que utilize técnicas de aprendizado profundo orientado por dados, com diversos dados dos cursos
SONG, Y. et al., 2022		propor um modelo de curso mais adequado aos alunos, com dados de realização de tarefas, presença na sala e resultado de exames
MA; ZHENG; XU, 2023		recomendação inteligente para indicação de melhores práticas e cursos baseado no comportamento, com dados de comentários de alunos participantes de cursos
XU, 2023		empregando redes neurais profundas desenvolve um algoritmo de ensino de tradução de inglês IES, com dados de deficiências de aprendizagem e a escrita dos alunos
CHEN; XUE; CUI, 2024		modelo de aprendizagem do aluno por meio de salas de aula inteligente, com dados de alfabetização informacional dos estudantes universitários, integrando-os aos recursos da biblioteca universitária
QJ, 2021	sistema de estrutura de serviço inteligente de aprendizado móvel, com dados do usuário; de atividade de aprendizagem, de avaliação e de uso do aplicativo	
OKOYE, K. et al., 2020	Avaliativo	Modelo de avaliação da qualidade do ensino, usando dados de termos mais frequentes usados pelos alunos para descrever os professores e o número de comentários que fazem referência ao gênero e às diferenças dos professores
WANG; WANG, 2023		avaliar o efeito da educação musical universitária na saúde mental dos estudantes, usando dados de avaliação, combinando as características do aconselhamento psicológico dos estudantes universitários
NIE.; LUO; YU, 2023		Sistema de avaliação automática do efeito educacional usando algoritmos digitais inteligentes, com dados de padrões comportamentais

Fonte: Autores (2024)

É possível verificar no Quadro 1 que são propostos modelos prescritivos de cursos, práticas, recursos e atividades personalizadas, baseados em mineração de dados sobre o comportamento dos estudantes nas suas interações com os componentes do curso, tais como dados de avaliação e exames, dados de comentários online, da interação com os professores e uso de aplicativos e espaços, tais como bibliotecas e plataformas, todos com o objetivo de melhorar o aprendizado e a performance dos alunos. Já os modelos preditivos, que usam dados para prever o desempenho acadêmico dos alunos, utilizam dados como interação de alunos e professores, ocupação de espaços físicos das IES, pontuações de avaliações durante o período de aula e relacionamento com pares, buscando antever o problema e encontrar causas, possibilitando a interferência do gestor para melhorar os resultados desses alunos. Quanto aos modelos de avaliação, dos dados são utilizados para verificar efeito, seja do processo educacional ou na saúde mental com intervenções já realizadas.

#### 4.4 PERSPECTIVA CONCEITUAL SOBRE O USO DE DADOS

Nessa perspectiva, são discutidos modelos conceituais sobre o uso de dados para as IES. As estruturas conceituais encontradas nessa sessão da RSL, de forma geral, utilizam estudos bibliográficos e estudos de caso, diferenciando-se das perspectivas de construção de modelos, principalmente por sua abordagem mais ampla, focando em conceitos do que deveria ser considerado em sistemas de uso dos dados, não em conceitos de programação para a construção desses.

Buscando propor uma abordagem prescritiva direcionada a dados para otimizar a educação e o aprendizado dos alunos, acompanhar o progresso dos alunos e fornecer orientação individualizada, Fan e Zhao (2023) concluem que, com o apoio das tecnologias de Big Data e a disponibilidade de dados, como os de avaliação e feedback, a educação baseada em dados explora a trajetória e as leis dos pensamentos e comportamentos, muda a atribuição da orientação para resultados para a função de previsão de dados, desde a teoria até a prática aprofundada, o que abre uma ideia totalmente nova para a pesquisa da educação. Em outra abordagem para o uso de dados, a preditiva, Cele (2021) propõe uma estrutura orientada a dados integrados para um sistema de alerta precoce que ajude identificar alunos em risco de abandono ou mau desempenho, e possibilitar intervenções oportunas e direcionadas que podem ajudar em seu sucesso, ressaltando a importância da ética na implementação e o cuidado à proteção da privacidade do aluno e ao uso responsável dos dados. Já Piety (2019), por meio de uma revisão bibliográfica sobre o uso de dados na educação, propõe um modelo conceitual de como os dados poderiam ser transformados em conhecimento, um modelo que se centra em quatro tipos de fatores ou elementos: componentes técnicos, infraestruturas, capacidade e práticas num contexto político e sistêmico. O estudo conclui ainda que, na educação superior, as principais abordagens se baseiam na mineração de dados para análise de aprendizagem.

Único entre os artigos analisados, Jin (2018) propõe uma estrutura conceitual para uso de dados na gestão financeira das IES, combinado Data Warehouse, mineração de dados financeiros e tecnologia de análise, o Data Driven é adotado para estabelecer o orçamento e modelo financeiro da universidade. O estudo aponta que novas tecnologias são aplicadas à construção e gerenciamento a fim de tornar a gestão financeira universitária mais científica. Também único entre os estudos, Deng et al. (2022) propõe, após a análise de pesquisa relevantes e a literatura sobre educação física, ciência da computação, pedagogia, gestão e outras disciplinas, um modelo conceitual que poderia usar dados de Big Data para apresentar um sistema de governança inteligente baseado que gerencie a saúde física de estudantes

universitários. A pesquisa aponta que, entre os motivos para o declínio contínuo da saúde física dos estudantes, estão a implementação deficiente de políticas esportivas escolares e que, com o uso de dados, pode-se reunir o papel dos gestores, da sociedade, da família e dos alunos para criar um impacto interligado para promover a saúde física dos alunos.

#### 4.5 PERSPECTIVA DAS PRÁTICAS SOBRE O USO DE DADOS

Na perspectiva de práticas sobre o uso de dados nas IES, estudos teóricos, geralmente provenientes de metodologias que incluem revisões de literatura, buscam apontar o atual estado do uso de dados em sistemas, seus contextos e consequências.

Iniciamos por Antonopoulou e Dare (2022) que objetivam oferecer um contraponto sobre a neutralidade e o realismo dos dados, analisando as intenções e impactos no acesso e preservação da informação. Nesse estudo, teórico e bastante denso, os autores argumentam que é necessária uma criticidade na percepção baseada em dados, principalmente ao entender uma vida como negócio, como dados, passíveis de categorização e gerenciamento, uma prática neoliberal e antiacadêmica. Focados em questões éticas, Kobis e Mehner (2021) objetivam analisar o uso de mentorias apoiadas por inteligência artificial (IA) no ensino superior. Em suas conclusões argumentam que tais discussões são relevantes pois objetivam aumentar a conscientização para o desenvolvimento ético e o uso de futuros ambientes de mentoria baseados em dados e suportados por inteligência artificial no ensino superior.

Gaftandzhieva et al. (2023) buscam analisar o estado da arte da tomada de decisões baseada em dados em IES, concluindo que as IES devem utilizar ferramentas de análise de dados, incluindo mineração de dados educacionais, análise de aprendizagem e inteligência de negócios, para extrair insights e conhecimento de dados educacionais. Já Prinsloo et al. (2015) realiza um estudo de caso descritivo do estado atual dos dados dos alunos, bem como explora o impacto das fontes de dados existentes e abordagens analíticas. Na mesma linha, Xing et al. (2022) investigam os fatores que influenciam o uso eficaz de dados pelos alunos no ensino superior na era do Big Data, enquanto Waheed et al. (2018) realizaram uma revisão bibliográfica para entender o estado da arte do big data na análise de aprendizagem.

#### 4.6 REFLEXÕES SOBRE DATA DRIVEN NAS IES

Ao buscar estabelecer conexões e reflexões entre pesquisas realizadas em diferentes geografias, contextos históricos, sociais e de desenvolvimento, esse estudo pretende compreender como a abordagem direcionada a dados, ou Data Driven, ou decisão baseada em dados é utilizada nas IES, bem como seus impactos, fatores direcionadores e achados já consolidados na literatura. A busca no WoS retornou, de fato, uma diversidade de estudos, com diferentes abordagens e objetivos. Entretanto, foi possível dividi-los em cinco grupos de análise, conforme apresentado nos tópicos acima. Essa última sessão de resultados perpassa essa divisão, buscando entre elas, apontamentos importantes sobre o uso de dados para decisão nas IES. Inicialmente, é preciso observar que não há estudos de países da América do Sul, ainda que o tema seja atual e necessário para o desenvolvimento e fortalecimento das IES, países desenvolvidos costumam liderar estudos sobre Data Driven e Big Data (BOZKURT, 2020).

Quase a totalidade dos artigos tem estudos ou objetivos voltados para a gestão do ensino e pesquisa. Único entre os artigos analisados, o estudo de Jin (2018) propõe um sistema de gestão financeira universitária e de apoio à decisão orientadas por dados,

argumentando que dados ajudam os gestores a tomar decisões precisas e rápidas e a livrar-se da cegueira do desconhecimento e que, por essa tarefa estar cada vez mais complicada, a gestão do orçamento financeiro orientada ao Data Driven pode contribuir para o pleno uso dos recursos de informação, com informações de orçamentos ex ante, controle de incidentes e avaliação ex post como formas eficazes de tomar decisões científicas.

Outro ponto inicial de atenção foi que apenas o estudo de Tzanova (2020) tratou da problemática de Ciência Aberta, defendendo sua essencialidade para a forma como o conhecimento é criado e disseminado pela sociedade e que o acesso amplo aos resultados da investigação pode ajudar a promover uma cultura de maior educação e literacia científica, o que, por sua vez, pode ter um impacto direto nas políticas públicas. Em contrapartida, argumenta que atual sistema de divulgação do conhecimento científico, baseado em assinatura, defendido pelos editores de periódicos e seu sistema de divulgação com fins lucrativos, não serve ao carácter público da ciência. Embora nenhum outro estudo aborde essa problemática, Raffaghelli e Sangrà (2023) afirmam que a ciência aberta deveria ser uma contrapartida à dataficação da investigação científica e à privatização das métricas, alimentadas através de plataformas adoptadas para gerir revistas de investigação e revisões por pares.

Poucos também são os estudos que expressam a preocupação com o bem estar do corpo académico das IES. Embora três estudos tenham objetivado obter pontos de vista de usuários de dados, apenas o de Jones (2019) se preocupa com esse aspecto, os demais eram focados em como usar dados. Também, dos 36 artigos, apenas dois focaram em modelos preditivos de bem estar, os dois realizados em anos recentes: Wang e Wang (2023) propuseram a construção e aplicação de uma plataforma de análise de big data de educação musical universitária para saúde mental de estudantes universitários e Deng (2022), que propôs um sistema de governança de dados inteligente para detectar e compreender a saúde física de estudantes universitários. Alguns outros estudos até propõe analisar dados de comportamento, mas voltados para modelos preditivos de desempenho académico.

Sobre esses modelos, maior parte dos estudos encontrados na RSL, preocupam-se com modelos prescritivos e preditivos para melhorar o desempenho académico dos estudantes, seja pela oferta de serviços, materiais, cursos e atividades personalizados no primeiro caso, ou no uso de dados para antever o possível resultado em provas finais, exames ou finalização dos cursos, no segundo caso. Inicialmente, nossa reflexão foca nessa categorização de pessoas (alunos), seja para modelos de interferência ou oferta. Categorizar um aluno para qualquer fim tem implicações práticas e políticas no contexto das IES. Embora estudos como o de Park, Yu e Jo (2016) e Fan e Zhao (2023) defendam que a decisão baseada em dados em sistemas prescritivos de aprendizagem ou preditivos de desempenho, por ser mais extensa no número de participantes, mais abrangente no número de sessões e mais requintado no nível de detalhes, tenha vantagens quando comparadas à observação humana, ou como o de Wang (2022) ao concluir que os problemas da instrução tradicional, como conteúdo desatualizados e conhecimento insuficiente, causados por alunos e professores, podem ser superados com métodos de aprendizagem baseada em dados, pois ela fornece uma variedade de recursos de aprendizagem e ajuda a cultivar a capacidade de aprendizagem autônoma dos alunos e a atender suas necessidades individuais de aprendizagem, alguns estudos demonstram a preocupação da capacidade dos sistemas absorver todas as variáveis necessárias para essas intervenções.

Por exemplo, Gao, Dong e Chen (2024) e Cele (2021), argumentam que análise de dados deve considerar e ser utilizada para avaliar o impacto da diversidade demográfica, dos

padrões de acessibilidade e das necessidades dos alunos, e que sejam observados aspectos diversos, que considerem uma compreensão abrangente dos alunos como pessoas no mundo que chegam às universidades com vários problemas complexos da vida que podem interromper suas oportunidades de aprendizagem. Para tanto, a abordagem Data Driven deve promover o progresso educacional, criação de um ambiente de aprendizagem equitativo e acessível, promoção da diversidade na análise de dados, considerações de acessibilidade, percursos de aprendizagem personalizados, concepção curricular inclusiva, atribuição equitativa de recursos, diversidade e inclusão docente, serviços de apoio culturalmente competentes, acesso à tecnologia, formação em competências culturais, envolvimento comunitário, políticas institucionais e um ciclo de feedback contínuo.

O contraponto é feito na pesquisa de Kobis e Mehner (2021) que aponta um viés algorítmico que, tendenciosos no emparelhamento, pode levar à discriminação, além da falta de transparência nas decisões da IA, questões de privacidade e de desumanização. Sobre os dados, argumentam que estão sempre a serviço de uma ideia, gerando informações que devem ser sempre observadas com cuidado e criticidade. Assim, duas questões emergem sobre a questão da categorização de um aluno para direcionar serviços ou oportunidades: primeiro a ideia de que decisão baseada em dados gera resultados “neutros”, racionais, precisos e livres de vieses; a segunda é que o uso de grandes bancos de dados e processamento por algoritmos inteligentes ou de inteligência artificial conseguiriam superar as limitações da análise humana. Nesse sentido, Prinsloo et al. (2015) faz uma crítica ao pensamento que os dados falam por si, e deve ter cuidado pois se a decisão é baseada em dados, não é sem intenção.

Conseguimos nos amparar na teoria em afirmar que as decisões orientadas a dados já enfrentam questionamentos sobre as limitações e os fatores que compõe um conjunto de dados, moldam a análise e geram resultados, todo resultado é intencional, questões sobre quem projeta e controla os sistemas orientados a dados, quais interesses defendem são essenciais nas discussões, também, temas como impacto social e formas que conteúdos são entregues já são caracterizados como uma forma do exercício de poder (VAN ES; SCHÄFER, 2017; DANIEL, 2024; RAFFAGHELLI; SANGRÀ, 2023; CZERNIEWICZ, 2022).

Para complementar essa questão, o estudo de Jones (2019) é relevante, pois trata dos impactos que esses direcionamos por algoritmos podem ter na vida acadêmica dos alunos. No estudo, os orientadores se preocupam com esse tipo de ranking, argumentando que o tratamento não traz justiça, pois indica que alunos de alto risco possam ser ignorados por precisarem de mais recursos e tempo para sua formação, e que uma otimização de recursos sugere concentrar a atenção nos alunos que tiveram mais chances de sucesso. Além disso, questões como a ética e a proteção à privacidade da comunidade acadêmica são citadas por Cele (2021), também ressonando na teoria com o embasamento de Raffaghelli e Sangrà (2023), que defendem as boas práticas no cruzamento entre a modernização do ensino superior, a ciência aberta, a privacidade de dados e a ética.

Outra questão importante no contexto de recursos das IES, embora estabelecer políticas e cultura baseadas em dados para planejamento, investimento, coordenação e controle, implementando o “Data Driven decision” seja analisado como uma questão estratégica e essencial para a gestão dessas organizações, existem custos associados ao uso de dados de Big Data. Em primeiro lugar, são necessários investimentos na aquisição de infraestruturas tecnológicas e sistemas de big data apropriados (ASHAARI et al., 2021). Armazenar grandes quantidades de dados não gera nenhum valor ou conhecimento às IES, segundo Piety (2019) é necessário que esses dados sejam organizados, filtrados e analisados

para se tornarem informações e depois combinados com a compreensão e experiência das partes interessadas para se tornarem conhecimento acionável.

Nessas condições, Piety (2019) ainda argumenta que a escassez de evidências que demonstrem benefícios (ou custos) dessa abordagem orientada a dados traz questões sobre sua sistematicidade no contexto das IES pois, embora o campo de estudo sobre métodos de investigação e os tipos de dados e informações que estão implicados esteja progredindo, questões sobre impacto na prática permanecem descobertas. Na literatura, Czerniewicz (2022) corrobora ao questionar até que ponto o foco na exploração e transformação de uma cultura de dados pode ser um item crucial em instituições que lidam com situações críticas como a adaptação da educação ao cenário pós-pandemia, ou que tentam sobreviver a grande crise de financiamento, embora, argumente que uma transformação digital remodelada e criticamente reconsiderada possa fortalecer o ensino superior frente a tantos desafios.

Ainda sobre a dataficação, Antonopoulou e Dare (2022) argumentam que a exposição das IES à metrificação intensificada, a processos de gestão cada vez mais orientadas por dados, implica na redução do qualitativa a painéis e pontuações. Kobis e Mehner (2021) argumentam que, por um lado, dados e algoritmos inteligentes proporcionam maior acesso para todos os alunos, atendimento personalizado as necessidades individuais, um processo mais objetivo e imparcial, por outro, admitem a possibilidade de viés algorítmico que, tendenciosos no emparelhamento, podem levar à discriminação, também ressaltam a falta de transparência nas decisões da IA, questões de privacidade e de desumanização.

Para o fim dessa reflexão, mas não pretendendo esgotar os debates sobre o uso de dados nas IES, Raffaghelli e Sangrà (2023) lembram que a discutir sobre abordagem orientada a dados, particularmente no ensino superior, exige a superação de declarações simplistas relativas à inovação e ao desenvolvimento e, ao mesmo tempo, devem evitar críticas performativas à dataficação.

#### **4. CONCLUSÃO**

Essa RSL objetivou explorar o uso da abordagem Data Driven no contexto das Instituições de Ensino Superior (IES). Na busca no WoS e aplicados os critérios de exclusão e seleção, o corpus da pesquisa foi composto de 36 artigos, que foram analisados nas perspectivas dos usuários de dados, da funcionalidade, da construção de modelos aplicados e de estruturas conceituais para seu uso, de reflexões sobre o uso de dados, suas práticas, questões éticas, contribuições e limitações no contexto das IES.

Os resultados demonstraram que não há estudos em países da América do Sul, há poucos estudos em áreas importantes para as IES como administração, bem estar acadêmico, pesquisa e ciência aberta. Quase a totalidade dos artigos abordam o uso de dados para gestão do ensino, com modelos prescritivos e preditivos para medir, intervir e melhorar o desempenho acadêmico. Com base nos achados, propusemos uma reflexão no contexto das IES sobre a categorização do aluno em métricas de desempenho, suas implicações práticas e políticas, principalmente pelo uso de algorítmicos que podem levar à discriminação. Também refletimos a escassez de evidências que demonstrem benefícios (ou custos) do Data Driven num contexto de adaptação pós-pandemia e de crises de financiamento, embora, uma a transformação digital crítica possa fortalecer o ensino superior frente a tantos desafios. Assim, o estudo em questão alcançou seu objetivo inicial de realizar uma exploração bibliográfica reflexiva sobre o uso da abordagem Data Driven no contexto das IES.

As limitações do estudo são pautadas pela necessidade de resumir grandes artigos científicos em poucas linhas de informação, realizadas na perspectiva do entendimento dos autores. Como sugestão, primeiro observamos a necessidade de estudos nessa área na América do Sul e em países em desenvolvimento, depois, estudos que abordem uso da abordagem baseada em dados para a administração das IES e, por último, estudos que visem utilizar o potencial do Big Data para promover o bem estar acadêmico e o desenvolvimento sustentável da educação superior.

## REFERÊNCIAS

ANTONOPOULOU, A.; DARE, E. The Riverine Archive: Nausea and information loss on the neoliberal ship of fools. *Visual Communication*, v. 21, n. 3, p. 418-436, 2022.

ASEERI, M.; KANG, K.. Organizational culture and big data socio-technical systems on strategic decision making: Case of Saudi Arabian higher education. *Education and Information Technologies*, v. 28, n. 7, p. 8999-9024, jan. 2023. doi: <https://doi.org/10.1007/s10639-022-11500-y>.

ASHAARI, M. A. et al. Big data analytics capability for improved performance of higher education institutions in the Era of IR 4.0: A multi-analytical SEM & ANN perspective. *Technological Forecasting and Social Change*, v. 173, p. 121119, dez. 2021. doi: <https://doi.org/10.1016/j.techfore.2021.121119>.

BEARD, D. et al. Rethinking the teaching of grammar from the perspective of corpus linguistics. 2019.

BOZKURT, A. Educational technology research patterns in the realm of the digital knowledge age. *Journal of Interactive Media in Education*, n. 1, 2020.

BULUT et al. Standing on the shoulders of giants: Online formative assessments as the foundation for predictive learning analytics models. *British Journal of Educational Technology*, v. 54, n. 1, p. 182-198, 2023. ISSN: 0007-1013 (impresso) | EISSN: 1467-8535 (eletrônico)

CELE, Njabuliso. Big data-driven early alert systems as means of enhancing university student retention and success. *South African Journal of Higher Education*, v. 35, n. 2, p. 59-72, 2021.

CHEN, S., XUE, Y. & CUI, X. Information literacy of college students from library education in smart classrooms: based on big data exploring data mining patterns using Apriori algorithm. *Soft Comput* 28, 3571–3589 (2024). <https://doi.org/10.1007/s00500-023-09621-8>

CZERNIEWICZ, L. (2022). Multi-layered digital inequalities in HEIs: The paradox of the post-digital society. *New visions for higher education towards 2030-Part 2: Transitions: Key topics, key voices*.

DANIEL, B. K.. Big data in higher education: The big picture. In *Big data and learning analytics in higher education* (pp. 19–28), 2017. Springer. [https://doi.org/10.1007/978-3-319-06520-5\\_3](https://doi.org/10.1007/978-3-319-06520-5_3)

DENG, C. et al. Big data-driven intelligent governance of college students' physical health: System and strategy. *Frontiers in Public Health*, v. 10, p. 924025, 2022.

FOUNG, D.; CHEN, J. Tracing writing progression in English for academic purposes: A data-driven possibility in the post-COVID era in Hong Kong. *Frontiers in Education*, v. 7, n. 1, p. 1-14, 2022.

GAO, Y.; DONG, Q.; CHEN, Z.. Leveraging the Synergy of IPv6, Generative AI, and Web Engineering to Create a Big Data-driven Education Platform. *Journal of Web Engineering*, v. 23, n. 2, p. 197-226, abr. 2024. doi: <https://doi.org/10.13052/jwe1540-9589.2321>.

GIBBS, P. The Marketingisation of Higher Education. In: Rider, S., Peters, M.A., Hyvönen, M., Besley, T. (eds) *World Class Universities. Evaluating Education: Normative Systems and Institutional Practices*, 2020. Springer, Singapore. [https://doi.org/10.1007/978-981-15-7598-3\\_13](https://doi.org/10.1007/978-981-15-7598-3_13)

JIN, M. et al. Financial Management and Decision Based on Decision Tree Algorithm. *Wireless Pers Commun* 102, 2869–2884, 2018. <https://doi.org/10.1007/s11277-018-5312-6>

JONES, K. M. L. Advising the whole student: eAdvising analytics and the contextual suppression of advisor values. *Education and Information Technologies*, v. 24, n. 1, p. 437-458, jan. 2019. doi: <https://doi.org/10.1007/s10639-018-0843-3>.

KÖBIS, L.; MEHNER, C. Ethical questions raised by AI-supported mentoring in higher education. *Frontiers in Artificial Intelligence*, v. 4, p. 624050, 2021.

MA, R., ZHENG, M., & XU, J. Visualized Evaluation and Intelligent Recommendation of International Chinese MOOCs Based on Learning Data Mining. *Journal of Technology and Chinese Language Teaching*, 14(2), 1-24, 2023.

MENON, A. et al. Antecedents and consequences of marketing strategy making: a model and a test. *Journal of marketing*, v. 63, n. 2, p. 18-40, 1999.

NIE Y., LUO X. & YU Y., "A Data-Driven Knowledge Discovery Framework for Smart Education Management Using Behavioral Characteristics," in *IEEE Access*, vol. 11, pp. 72562-72574, 2023.

OKOYE, K. et al. Impact of students evaluation of teaching: a text analysis of the teachers qualities by gender. *Int J Educ Technol High Educ* 17, 49 (2020). <https://doi.org/10.1186/s41239-020-00224-z>

PAN, L. et al. Gathering Intelligence on Student Information Behavior Using Data Mining. *Library Trends*, v. 68, n. 4, p. 636-658, 2020.

PARK, Y.; YU, J. H.; JO, I. H. Clustering blended learning courses by online behavior data: A case study in a Korean higher education institute. *The Internet and Higher Education*, v. 29, n. 1, p. 1-11, 2016. doi: <https://doi.org/10.1016/j.iheduc.2015.11.001>

PIETY, P. J. Components, infrastructures, and capacity: The quest for the impact of actionable data use on P-20 educator practice. *Review of Research in Education*, v. 43, n. 1, p. 394-421, Mar. 2019. ISSN: 0091-732X.

QU, J. Research on mobile learning in a teaching information service system based on a big data driven environment. *Educ Inf Technol* 26, 6183–6201 (2021). <https://doi-org/10.1007/s10639-021-10614-z>

RAFFAGHELLI, J.E., SANGRÀ, A. Data, Society and the University: Facets of a Complex Problem. In: Raffaghelli, J.E., Sangrà, A. (eds) *Data Cultures in Higher Education . Higher Education Dynamics*, vol 59, 2023. Springer, Cham. [https://doi.org/10.1007/978-3-031-24193-2\\_2](https://doi.org/10.1007/978-3-031-24193-2_2)

RAMASWAMI, G. et al. Usando técnicas de mineração de dados educacionais para aumentar a precisão da previsão do desempenho acadêmico dos alunos. *Information and Learning Sciences* , v. 120, n. 7/8, p. 451-467, 2019.

- SAQR, M.; FORS, U.; NOURI, J.. Using social network analysis to understand online Problem-Based Learning and predict performance. PloS one, v. 13, n. 9, p. e0203590, 2018.
- SONG, Y. et al. Avaliação de um modelo de ensino oral de inglês online usando big data. Mobile Information Systems , v. 2022, n. 1, p. 7934575, 2022.
- TAN, J. Information Analysis of Advanced Mathematics Education-Adaptive Algorithm Based on Big Data. Mathematical Problems in Engineering, v. 2022, n. 1, p. 7796681, 2022.
- TZANOVA, S. Changes in academic libraries in the era of Open Science. Education for Information, v. 36, n. 3, p. 281-299, 2020.
- TRANFIELD, D. et al. Rumo a uma Metodologia para Desenvolver Conhecimento de Gestão Baseado em Evidências por Meio de Revisão Sistemática. British Journal of Management, v. 3, pág. 207–222, conjunto. 2003.
- WANG, X.; WANG, T.. Construction and Application of a Big Data Analysis Platform for College Music Education for College Students' Mental Health. International Journal of Data Warehousing and Mining, v. 19, n. 4, p. 1-16, jan./mar. 2023. DOI: <https://doi.org/10.4018/IJDWM.324060>
- XU, J. Big data-driven english teaching for social media: a neural network-based approach. Evol. Intel. 16, 1589–1597, 2023. <https://doi.org/10.1007/s12065-022-00792-1>
- Yağcı, M. Educational data mining: prediction of students' academic performance using machine learning algorithms. Smart Learn. Environ. 9, 11 (2022). <https://doi.org/10.1186/s40561-022-00192-z>
- ZHOU Y. et al. Daily Peer Relationships and Academic Achievement among College Students: A Social Network Analysis Based on Behavioral Big Data. Sustainability. 2023; 15(22):15762. <https://doi.org/10.3390/su152215762>