Received: 24 March 2023 | Revised: 8 June 2023 | Accepted: 12 June 2023

DOI: 10.1002/sd.2638

RESEARCH ARTICLE

Sustainable Development WILEY

Using data science for sustainable development in higher education

Walter Leal Filho[1,2] | João Henrique Paulino Pires Eustachio[2] | Andreea Corina Nita (Danila)[3] | Maria Alzira Pimenta Dinis[4,5] | Amanda Lange Salvia[6] | Debby R. E. Cotton[7] | Kamila Frizzo[8] | Laís Viera Trevisan[9] | Thais Dibbern[10]

[1]Department of Natural Sciences, Manchester Metropolitan University, Manchester, UK

[2]European School of Sustainability Science and Research (ESSSR), Hamburg University of Applied Sciences, Hamburg, Germany

[3]Department of Economics, Economic Informatics, and Business Management, Faculty of Economics, Administration and Business Administration, Stefan cel Mare University of Suceava, Suceava, Romania

[4]UFP Energy, Environment and Health Research Unit (FP-ENAS), University Fernando Pessoa (UFP), Porto, Portugal

[5]Fernando Pessoa Research, Innovation and Development Institute (FP-I3ID), University Fernando Pessoa (UFP), Porto, Portugal

[6]Graduate Program in Civil and Environmental Engineering, University of Passo Fundo, Passo Fundo, Brazil

[7]SCION Research Group, Plymouth Marjon University, Plymouth, UK

[8]School of Administration, Federal University of Santa Maria, Santa Maria, Brazil

[9]School of Administration, Federal University of Rio Grande do Sul (UFRGS), Porto Alegre, Brazil

[10]Department of Science and Technology Policy, University of Campinas, Campinas, Brazil

Correspondence
João Henrique Paulino Pires Eustachio, European School of Sustainability Science and Research (ESSSR), Hamburg University of Applied Sciences, Ulmenliet 20, Hamburg 21033, Germany.
Email: jh.eustachio@gmail.com

Funding information
Inter-University Sustainable Development Research Programme; Hamburg University of Applied Sciences

Abstract

Despite the abundance of studies focused on how higher education institutions (HEIs) are implementing sustainable development (SD) in their educational programmes, there is a paucity of interdisciplinary studies exploring the role of technology, such as data science, in an SD context. Further research is thus needed to identify how SD is being deployed in higher education (HE), generating positive externalities for society and the environment. This study aims to address this research gap by exploring various ways in which data science may support university efforts towards SD. The methodology relied on a bibliometric analysis to understand and visualise the connections between data science and SD in HE, as well as reporting on selected case studies showing how data science may be deployed for creating SD impact in HE and in the community. The results from the bibliometric analysis unveil five research strands driving this field, and the case studies exemplify them. This study can be considered innovative since it follows previous research on artificial intelligence and SD. Moreover, the combination of bibliometric analysis and case studies provides an overview of trends, which may be useful to researchers and decision-makers who wish to explore the use of data science for SD in HEIs. Finally, the findings highlight how data science can be used in HEIs, combined with a framework developed to support further research into SD in HE.

KEYWORDS
bibliometrics, case studies, data science, higher education institutions (HEIs), sustainable development (SD)

1 | INTRODUCTION

Data science refers to the study of the generalisable extraction of knowledge from data. It involves the arrangement of knowledge into explanations and predictions that can be used for further studies and frequently includes the statistical analysis of results that verifies their credibility (Cao, 2017; Pedamkar, 2022). Data scientists are often tasked with carrying out studies and require skills in machine learning (ML), mathematics, artificial intelligence (AI), decision-making, and solution derivation (Dhar, 2013). Data science is considered a transdisciplinary field that involves the coordination of various disciplines to study data and produce useful insights. It can include videos, pictures (Cao, 2017; Dhar, 2013), and other materials and often involves different types of information, not just quantitative data.

Data handling has four main components: data exploration, modelling, assessing the models, and deploying the models (Pedamkar, 2022; Sarker, 2022). Data investigation takes up the bulk of the process and involves sampling data and removing unnecessary information. Data is transformed and prepared for use in the future steps of the process (Idreos et al., 2015). Modelling is conducted after data is deemed valid. Data is used with ML algorithms to fit or adapt existing models or create entirely new ones (Pedamkar, 2022). The assessment of the models is then carried out. Inaccuracies and discrepancies are fixed to ensure that the model can be utilised. Should this stage fail, a scientist may revert to the design stage. Viable models are then deployed into the industry, research or the target sector for usage (Baesens, 2014; Najafabadi et al., 2015).

Data science has practical applications across a range of fields, including healthcare and biomedical science, where it may assist with classifying and diagnosing diseases or conditions. During the coronavirus disease (COVID-19) pandemic, for example, data science was used to measure contamination risks in certain places and situations, becoming an essential element in preventing the spread of the disease (Caruso et al., 2022; Van Lissa et al., 2022). In other instances, data science assists with research that can help improve the treatment and the care of patients. This is achieved through data science studies in areas such as genome sequencing and drug development, which can significantly transform the medical industry (Upasana, 2018).

Furthermore, deep neural network learning techniques enable the development of effective solutions for remote sensing image interpretation (Esteva et al., 2017). The models are trained to perform image classification, segmentation, object identification, and tracking. Models based on deep neural networks have been found successful in automated dermatology, where medical images are analysed to generate a proper diagnosis in skin cancer investigations (Esteva et al., 2017; Haenssle et al., 2018). Automated radiology is another field of research that has received increased interest with the development of deep learning, being able to recognise potential cancerous lesions (Al-masni et al., 2018), bone fractures (Lindsey et al., 2018) or detect COVID-19 (Ozturk et al., 2020) through radiology images. Image sensing technology is also used in the aerospace and defence sectors to identify objects from satellite images (Izzo et al., 2019), gather military intelligence, or identify suspicious events (Kamran et al., 2018).

Aside from this, data science has proven extremely useful in the business sector. Its deployment allows businesses and employees to make more informed decisions, improving the quality of the business activity and offering a better output (Vicario & Coleman, 2020). Big data analytics assist in delivering business strategies and making faster and more informed decisions that can improve sales and productivity (Agarwal & Dhar, 2014). Data science can also improve business performance in areas as varied as productivity, financial performance, and market access (Vassakis et al., 2018). Sentiment analysis can give businesses a clear image of clients' emotional tone about a specific product or idea. The tools used for sentiment analysis categorise pieces of text collected from social media platforms as positive, negative, and neutral, thereby determining the reputation of a specific brand or product based on online reviews (Revathy et al., 2022). Moreover, sentiment analysis has been used to assess confidence in the stock market or voters' opinions on different issues and how they relate to the actions and speeches of political candidates (Feldman, 2013; Shaik et al., 2022).

Data science also has use cases in agriculture through developing AI algorithms to optimise yield production or disease detection using data on soil, fertilisers, temperature, and humidity collected from satellites or Internet of Things (IoT)-enabled sensors (Liakos et al., 2018). The automotive industry has benefited from the rise of data science and AI with the development of self-driving cars that are able not only to detect traffic signs, lights, pedestrians, and other vehicles but also to make speed or direction adjustments, depending on road conditions (Sharma et al., 2021; Rahul et al., 2022). The technology used ensures more efficient traffic conditions and fewer car accidents. Moreover, self-driving cars can be designed to be more environmentally friendly (Majji & Baskaran, 2021).

In addition, from the economic/financial perspective, data science models have been used in the financial sector to provide forecasts on stock market prices and bankruptcies (Barboza et al., 2017), to detect fraud activities (Pumsirirat & Yan, 2018), or to assess a client's creditworthiness by analysing information coming from several data sources. Credit default risk can be easily estimated based on AI models that account for different dimensions when banks give loans to private companies or individuals (Mancisidor et al., 2020).

Data science may also be further applied in delivering public policies. The amount of data produced by public organisations is large, and data science can be used to create or optimise policies on public affairs. Its deployment aims to improve resource efficacy and the policies' effectiveness and significance, leading to greater sustainability (Liberatore et al., 2019). Data science projects can contribute to more efficient management of smart cities to solve public problems and implement appropriate solutions to the local infrastructure (Grimaldi & Carrasco-Farré, 2021; Sarker, 2022). Since decision-making is increasingly data-driven, its use, transparency, and accountability have become essential issues in public policy-making (Berman et al., 2018). An example that relies on data science is the correlation between mobility in mass transport and mortality due to COVID-19 (Vega-Villalobos et al., 2022) and the use of data science in the prediction of socio-economic indicators (Machicao et al., 2022).

Many of the examples outlined above have direct or indirect links to the Sustainable Development Goals (SDGs), and it is evident that there are a number of potential ways in which data science might contribute to enhancing research and practice in SD. However, there is a scarcity of studies that have explored how data science has been used to address SD in the context of higher education institutions. Therefore, this study departs from two research questions (RQ):

> **RQ 1:** How is data science contributing to efforts to pursue sustainability in Higher Education Institutions (HEIs)?
> **RQ 2:** Which examples exist and illustrate how the topic is being handled by them?

To achieve the goals, this research uses a bibliometric assessment of the literature to explore the various ways data science may support efforts towards SD. In addition, this research also relies on an expert-driven literature review showing how data science is being used to create a positive impact at HEIs through the discussion of a set of case studies. It investigates the extent to which data science is contributing to sustainability in HE.

This is an innovative study, which follows previous research on artificial intelligence and SD (Leal Filho, Yang, et al., 2022), and provides a timely addition to the literature since it explores the nexus of data science-SD, an emerging area of research that seeks to use data science techniques and technologies to support the SD of communities and societies. This field has become even more relevant over recent months, when the use of AI tools has attracted much media attention, and the use of data science to guide decision-making has come to the fore.

2 | DATA SCIENCE AND SUSTAINABLE DEVELOPMENT

Data science can play an important role in achieving SD, with SD in this context understood to mean the process of using resources in ways that meet the needs of the present without compromising the ability of future generations to meet their needs (Leal Filho, Vasconcelos, et al., 2022; UN, 1987). Data science can offer an important contribution by identifying and quantifying the environmental and social impacts of human activity (Giuliani et al., 2020; Taylor et al., 2021) to inform decision-making and develop innovative solutions for SD (Guo et al., 2022). It can also be used to measure, monitor and evaluate the effectiveness of policies and practices aimed at achieving SD goals (Eustachio et al., 2019). In addition, data science can be used to identify and address gaps and opportunities for better resource allocation and management (Liu et al., 2021; Tutsch et al., 2020). Finally, it can be used to support the development of evidence-based policies and regulations to ensure SD (Leal Filho, Trevisan, et al., 2023).

Some existing research has considered how data science contributes directly or indirectly to advancing the SDGs. There is an ongoing debate in the literature about the importance of data science in supporting scientists to seek high-quality data and adopting relevant statistical techniques to analyse challenges related to SD (Malhotra et al., 2018; Vance & Love, 2021). This aspect is particularly relevant to SD, a context where systems are complex and dynamic, requiring tools that offer an understanding of how it might be possible to mitigate negative impacts on ecological and social systems, supporting governance (Leal Filho, Abubakar, et al., 2023). Data science can also enhance data-based decision-making processes in order to achieve the United Nations SDGs (Eustachio et al., 2019; Guo et al., 2020; Leal Filho, Trevisan, et al., 2023). Table 1 outlines some of the contributions of data science to SD.

Data science can also support the scaling of SD solutions, making them available to a variety of stakeholders.

Operationally, the creation of centralised data repositories is driving growth across all sectors of society (Linkov et al., 2018). Data is essential for society and the economy because of its potential for innovation in the public and private sectors (Hofman & Rajagopal, 2014), associated with creating environmental, social, and economic value. In this regard, big data is increasingly changing how impacts on the environment are measured and mapped, contributing to SD by being used to measure carbon emissions (Hazen et al., 2016; Huang et al., 2017; Seles et al., 2018), to improve social and environmental sustainability in supply chains (Dubey et al., 2019), to expand the informational landscape of smart sustainable cities (Bibri, 2018), and to improve the allocation and utilisation of natural resources (Song et al., 2017), among other applications. Within this context, data-intensive methods are increasingly being employed to answer environmental questions relevant to sustainability (Pennington et al., 2020).

Some other studies address data science for SD by observing, for instance, soil conditions and humidity (Dhyani et al., 2020), estimating energy consumption (Seyedzadeh et al., 2018; Zheng et al., 2019), greenhouse gas emissions (Hamrani et al., 2020), disaster detection and management (Akter & Wamba, 2019; Sublime & Kalinicheva, 2019), urban water management (Goralski & Tan, 2020; Xiang et al., 2021), poverty assessment and monitoring (Alsharkawi et al., 2021; Jean et al., 2016; Leal Filho, Eustachio, et al., 2022), food consumption and sustainability (Abdella et al., 2020), changes in water resources (Senay et al., 2017; Xu et al., 2019), and biodiversity monitoring (Allen et al., 2019). By improving the accuracy of global climate models and climate forecasts, the importance of data science for the SDGs can be highlighted, considering that a persistent challenge for

TABLE 1 Some of the contributions of data science to sustainable development.

Element	Implications
Increased awareness and understanding of sustainability issues	Data science can support the raising of awareness and understanding of sustainability issues, and, by doing so, can encourage individuals and communities to take action.
Cost efficiency	Data science can be used to identify the most efficient and cost-effective ways of using resources and energy.
Inclusiveness	Data science can inclusively engage various groups and be of benefit to governments, non-governmental organisations, communities, and also individuals.
Sharing best practices	Data science can significantly facilitate the sharing of successful experiences and good practices, especially via online platforms.
Leveraging partnerships	Data science can facilitate collaboration between governments, non-government organisations, and communities and can accelerate the adoption and implementation of successful sustainability actions
Predictive models	Data science can be used to develop predictive models that can be used to forecast the impacts of different sustainability policies, programmes, and initiatives, and to identify areas of potential risk and vulnerability.

Source: The authors.

the achievement of the SDGs has been a lack of data to assess progress towards each goal and varying capacities among nations to conduct these assessments (Guo, Liang, et al., 2021). Hence, AI-supported research can be used to help map, implement, and report progress on individual SDGs and their related targets, playing an increasingly supporting and coordinating role in SD research (Leal Filho, Yang, et al., 2022).

Open data also has a key role to play here: stakeholders are increasingly requesting access to organisations' sustainability information for research and evaluation purposes. This is made possible through open data processes (Helbig et al., 2021), by which data are made available by organisations, companies, and individuals to access and reuse without restrictions (OECD, 2019; Open Data Institute, 2016), supporting society's development and the formulation of public policies through innovation. Making data accessible to diverse stakeholders, such as researchers, entrepreneurs, business leaders, representatives of public services, non-governmental organisations (NGOs), philanthropic institutions, community groups, and citizens can contribute to the political decision-making process (Kapoor et al., 2015), improve service delivery, build new businesses, and hold aid recipients accountable for resource allocation (Open Data Institute, 2016). This open data movement is similar to Open Source or Open Access (Braunschweig et al., 2012; Kapoor et al., 2015).

Open Government Data Platforms (OGDPs) are web portals that contain data and search and data analysis tools (Marjanovic & Cecez-Kecmanovic, 2020), allowing for increased transparency, participation, and innovation throughout society (Braunschweig et al., 2012). One example is the European Data Portal (EDP), created by the European Commission in 2015 and available in 25 languages. EDP brings together data sets from 35 European countries on agriculture, fisheries, forestry, and food; economics and finance; education, culture and sport; energy; environment; government and public sector; health; international issues; justice, legal system, and public security; population and society; regions and city; scientific technology; and transport. Other examples of portals that provide SD-related data published by the central government, local authorities, and public agencies are data.gov.uk (United Kingdom), data.gov.it (Italy), and govdata.de (Germany). Additionally, other platforms publish research and data regarding global issues, such as the Disclosure Project Action – CDP (data.cdp.net), the Open Data Watch (opendatawatch.com), the Our World in Data (ourworldindata.org), and the Sustainable Data Platform (sustainable-data-platform.org). All of these sites can provide useful data for analysis using data science techniques and technologies. From this perspective, using and promoting open data in SD processes based on cooperation between national and local governments, donors, global institutions, civil society, academy, and industry can make reaching the SDGs a reality by maximising the positive environmental, social, and economic impacts (Open Data Institute, 2016).

Despite the evident importance of data science in overcoming the challenges related to SD, the research field is relatively new, creating room to understand the big picture of how researchers are discussing the use of data science and connecting it to SD in the literature. A comprehensive analysis of SD issues depends on integrating several data types and sources, including dynamic georeferenced datasets and ecological, economic, and social processes based on multi-scale, multi-temporal, and multi-source data (Pitts et al., 2020). In this regard, data science has been seen as a 'strategic highland' in the new data-intensive era (Guo et al., 2020). There is also a lack of literature on data science as related to SD in HE. Some researchers have started to explore possible ways that technologies may support university efforts towards SD—for example, by using ML or other techniques in the education 4.0 context, to mention a few (Costa et al., 2022; González-Pérez & Ramírez-Montoya, 2022). However, to the best of the authors' knowledge, no previous studies have explored in depth the role of data science for SD focusing on HE.

3 | METHODS

This study aims to review the various means via which data science may contribute to efforts to pursue SD in HE and has deployed two main methods. The first method adopted was a bibliometric analysis to explore the possible connections between terms related to data science, SD and HEIs. The second method deployed was an expert-driven literature review, resulting in 10 case studies that illustrate the relevant discussion in the field. The choice of bibliometric analysis was based on the fact that it provides a succinct overview of the literature on data science and how it is being deployed in an SD context. This has been complemented by an expert-driven literature review, resulting in the identification of selected case studies, where the topic was analysed and synthesised to exemplify some of the current trends regarding SD in HEIs.

In order to collect data for the bibliometric analysis, a search string was created in three blocks of terms. The first block relates to terms commonly adopted to discuss data science issues such as big data, ML, and data visualisation. The second group, in turn, embraces terms related to SD, such as SDGs and the 2030 Agenda, and the last group covers the context of HE, with terms such as HEIs and universities. The authors aimed to find papers containing these three groups of terms, whether they appear in the title, abstract, or keywords. Table 2 illustrates the search string used and the number of peer-reviewed documents found.

The search for documents was performed on the Scopus database, which is considered one of the most relevant and biggest academic databases, covering more than 23,452 peer-reviewed indexed journals and 77.8 million records from several subject fields such as life sciences, social sciences, physical sciences, and health sciences (Scopus, 2022). After applying the search string formulated (see Table 2), the Scopus database initially returned 405 documents on 28th August 2022. All the titles and abstracts were checked, and after a screening process, 164 papers were dropped due to their lack of connection to the research goal of this study, remaining 241 adherent papers written in English and published from 2007 to 2022, in order to proceed with the analysis.

The bibliometric assessment was based on the co-occurrence of terms. The VOSviewer (2022) software was selected to perform the analysis of the 241 chosen documents since it was designed to construct and visualise bibliometric networks. The results are shown through a network graph where the size of the bubbles represents the frequency of a term, and the stronger the connections are between two bubbles, the higher the probability of two terms co-occur and, therefore, becoming a thematic cluster represented by the colours of a network graph (van Eck & Waltman, 2011, 2017). The results of the bibliometric analysis are presented in Figure 1.

In addition, and in order to illustrate the connection between data science and SD in HEIs, the authors judiciously selected a set of 10 case studies based on their relevance to the SD topic. This approach was based on previous studies published in the SD field (Leal Filho, Levesque, et al., 2022) and has proven helpful since it can enhance the discussion of the bibliometric analysis by highlighting the multidimensional aspects related to the connection of data science, sustainability, and HE.

4 | RESULTS AND DISCUSSION

This section is divided into two subsections. The first one discusses the broad theoretical background by providing an overview of what researchers cover in data science and SD in HEIs. The second subsection brings 10 representative case studies that contributed to delving into the impacts of data science in SD in HEIs.

4.1 | Bibliometric analysis

Figure 1 shows that data science has been regularly discussed in relation to SD, HE, and other related technologies such as AI, ML, and data mining. This vibrant field is evidenced by the five thematic clusters represented in Figure 1, which shows at least five research strands that researchers are driving in the SD research field, evidencing how data science may support efforts towards SD from several perspectives, as discussed below.

The red cluster reports on the techniques commonly used for data science, embracing terms such as decision trees, predictive models, data mining, and ML. Papers that belong to this specific cluster report on how data science contributes to implementing SD in the various activities undertaken at HEIs, especially in teaching and research practices. For example, some authors argue about the relevance of technology in sustainability teaching in HE (Bonini, 2020). Other studies focus on how the leading technologies related to data science, for example, the creation of new algorithms and statistical

TABLE 2 Search criteria and number of publications in the Scopus database.

Database	Search string	Number of documents	Type of documents	Time frame
Scopus	TITLE-ABS-KEY ([‘Data Science*’ OR ‘Big*Data’ OR ‘Machine Learning’ OR ‘Artificial Intelligence’ OR ‘Data Mining’ OR ‘Data Analytics’ OR ‘Data Visuali*ation’] AND [‘sustainable development’ OR ‘sustainability’ OR ‘2030 Agenda’ OR ‘SDG*’ OR ‘Sustainable Development Goal*’] AND [‘HEI’ OR ‘HEIs’ OR ‘Higher Education Institutions’ OR ‘Higher Education’ OR ‘University’ OR ‘Universities’])	*Initial number of entries*: 405 documents *After the screening process*: 241 selected documents published in English	Articles; Conference Papers; Book Chapters; Reviews.	2007–2022

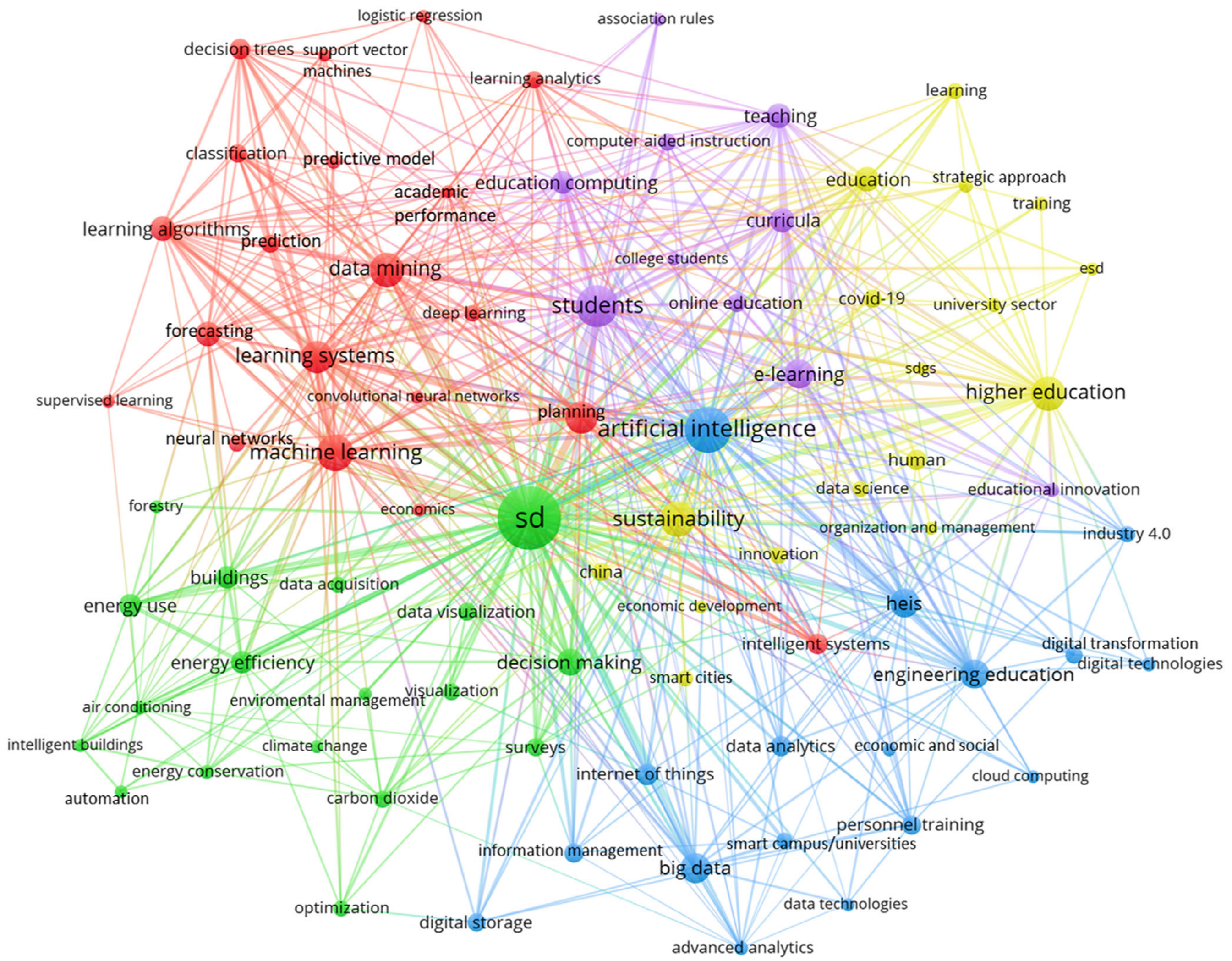

FIGURE 1 Co-occurrence of the terms—data science, SD, and HE. Co-occurrence of the terms—VOSviewer output.

models, are helping in sustainability research by using different ways of monitoring, collecting, and analysing data (Poudyal et al., 2020). Another strand in this cluster uses data science techniques to understand the impact of HE courses and programmes in several dimensions of SD (Arango-Uribe et al., 2022).

The green cluster is related to how data science can support decision-making at HEIs, especially regarding building efficiency and implementing sustainability in campus operations (Gholami et al., 2020; Soares et al., 2015). Therefore, this cluster considers terms such as buildings, energy use, energy efficiency, energy conservation, intelligent buildings, automation, and energy conservation. For example, Cai et al. (2009) developed an interactive decision support system that could lead university managers to optimise and plan energy management, generating strategies and policies that could contribute to SD challenges. In the same perspective, Cao et al. (2015) studied university facilities that represent several challenges in the perspective of environmental management and decision-making, proposing energy-aware intelligent facility management using data sciences principles and techniques.

Papers that belong to the purple cluster report on issues related to teaching and learning aspects of how data science could contribute to the SD challenges, as well as understanding how data science and the technologies related to it are being deployed in teaching for SD (Mian et al., 2020), including terms such as online education and e-learning. The blue cluster, in turn, is similar to the purple cluster and evidences the perspective of how data science techniques (data analytics, internet of things, big data) are connected to engineering education, specifically. Researchers in this cluster are positing efforts to understand how digital transformation is shaping society and how engineering education should adapt to it, such as guaranteeing competencies related to sustainability, data science, and others connected to the rise of Industry 4.0 (Gürdür Broo et al., 2022; Okoroafor et al., 2022). In addition, another aspect covered in this cluster is related to the need to integrate SD issues into the curricula of information technology and AI engineering programmes (Paulauskaite-Taraseviciene et al., 2022).

Finally, the yellow cluster reports on how HEIs are innovating and adopting data science strategies in education for sustainable development (ESD), as well as managing the other educational systems at a time when the COVID-19 pandemic fostered swift technological changes. These changes kept requiring universities to adapt to the new reality in their several practices, such as research on COVID-19 and possible connections to SD, the adoption of digital learning platforms, and the need to foster education on data science (Bozkurt, 2022; Budaya et al., 2021).

4.2 | Case studies

Organisations collect data from different sources and with different aims, as mentioned before. In the past, storing data from smart devices, email information, social media, and a variety of other sources was too costly. Today, the economic burden has been made cheaper through cloud and data lakes, improving information accuracy, optimising development, or enabling smart decision-making. Technology holds power to change our world, overcoming existing barriers and increasing the universality and speed of teaching and learning (Kwok & Treiblmaier, 2022). The transformation of business through digital technology was considered by the European Green Deal (European Commission (EC), 2019) as a major factor in the achievement of the SDGs (UN, 2015). The examples in Table 3 reveal that data science is highly relevant in terms of the achieved positive impacts involving the SD under work in HEIs, with positive practical implications for the entire society.

The connections between HEIs and SD achievement in companies are clearly affected by the digital challenges surrounding the firms, with implications at the staff level, digital technology, capital, and infrastructure. Yuan et al. (2022) reported having found a statistically significant association between technology and the performance of business incubators in China, which seems to be more relevant than government support itself, showing how important information and communication technology is in supporting SD. From the perspective of SD, digital media technology was found to be fundamental in positively improving education quality, accessibility, inclusion, and

TABLE 3 Examples of studies focused on the relevant impacts of data science in SD in HEIs.

Case title	Focus/country	Country	References
Antecedents of big data analytics adoption: an analysis with future managers in a developing country	Big data analytics by future managers of companies	Brazil	de Moraes et al. (2022)
Climate change and COP26: Are digital technologies and information management part of the problem or the solution? An editorial reflection and call to action	Climate challenges viewed through a technology lens	Worldwide	Dwivedi et al. (2022)
Components of education 4.0 in 21st century skills frameworks: systematic review	Core Education 4.0 strategies	Worldwide	González-Pérez and Ramírez-Montoya (2022)
Digital transformation in Romanian accounting practice and education: impact and perspectives	The potential impact of the digitalisation-based approach to academic education	Romania	Guşe and Mangiuc (2022)
Inclusive education and sustainable development	Policy framework for inclusive education ignored in higher education.	Sri Lanka	Dorabawila et al. (2022)
No one left behind in education: blockchain-based transformation and its potential for social inclusion	Social inclusion of students with limited access to education found unreachable	South Korea	Kwok and Treiblmaier (2022)
Practice on the sustainable development of talent cultivation mode in the context of big data	Digital media technology professional talent training mode	China	Chen (2020)
Sustainable knowledge management in academia and research organisations in the innovation context	Education for sustainable development	Worldwide	Gómez-Marín et al. (2022)
Towards the revolution and democratization of education: a framework to overcome challenges and explore opportunities through industry 4.0	Education 4.0 democratised access to quality education through technology	Brazil	Costa et al. (2022)
Which factors affect the performance of technology business incubators in China? An entrepreneurial ecosystem perspective	Performance of technology business incubators	China	Yuan et al. (2022)

equality in South Korea (Kwok & Treiblmaier, 2022). In addition, in Sri Lanka, Dorabawila et al. (2022) have examined the status of inclusive education, reporting that the number of students with disabilities in HEIs is low, owing to a poor policy framework. In the process of making inclusive education successful and impacting the industry experiencing an ongoing digital transformation, significant challenges are anticipated, including the education and training of HEI staff at the technological level (Guşe & Mangiuc, 2022).

Table 3 highlights how important it is to advance data science endeavours under the guise of digital transformation, demanding further additional efforts in this respect. The information collected also exposes the challenges involved in terms of better addressing SD in HEIs through data science. Effective use of data science can positively impact all actors and stakeholders around educational institutions, allowing them to advance the necessary changes to build an SD future in all areas of knowledge, and be able to address skill gaps, data management, and research innovation, guided by technology (Costa et al., 2022; González-Pérez & Ramírez-Montoya, 2022). Data science must be accounted for when considering the need for innovation management inside academia (Gómez-Marín et al., 2022) and reorienting models to innovation, instead of focusing on business models. In this way, it will be possible to maximise human, organisational, relational, and research and development (R&D) capitals to reshape academia in order to promote SD worldwide, resulting in new forms of knowledge production and responsible digital transformation advancing sustainable change (Dwivedi et al., 2022).

Figure 2 provides an overview of the interconnections between data science and SD. It was designed based on the discussion resulting from the bibliometric analysis and the case studies, summarising the information into a new framework, evidencing how researchers are exploring the contributions of data science for SD in HE. Figure 2 presents three main interconnected components: (1) HEIs' core practices, (2) the methods and techniques related to data science, and (3) desired outcome: achieving the SDGs.

Figure 2 illustrates how HEIs perceive the SD challenges and consider their internal systems and practices, that is, research, teaching, outreach, governance, and campus operations, towards contributing to the SDGs. All these practices are traditionally conducted in several ways in the context of SD. However, when combined with data science methods and techniques, for example, logistic regression, learning systems, predictive models, ML, big data, and so forth, HEIs could maximise their contributions. For example, data science methods can help researchers engaged in SD research to collect reliable data and adopt powerful data analysis strategies (Guo, Chen, et al., 2021; Shaffer et al., 2019). Teaching practices for SD also may be supported by data science (teaching and learning analytics), which could not only improve classroom-based teaching (Sampson, 2017) but also bring more learning options to develop sustainability literacy and contribute

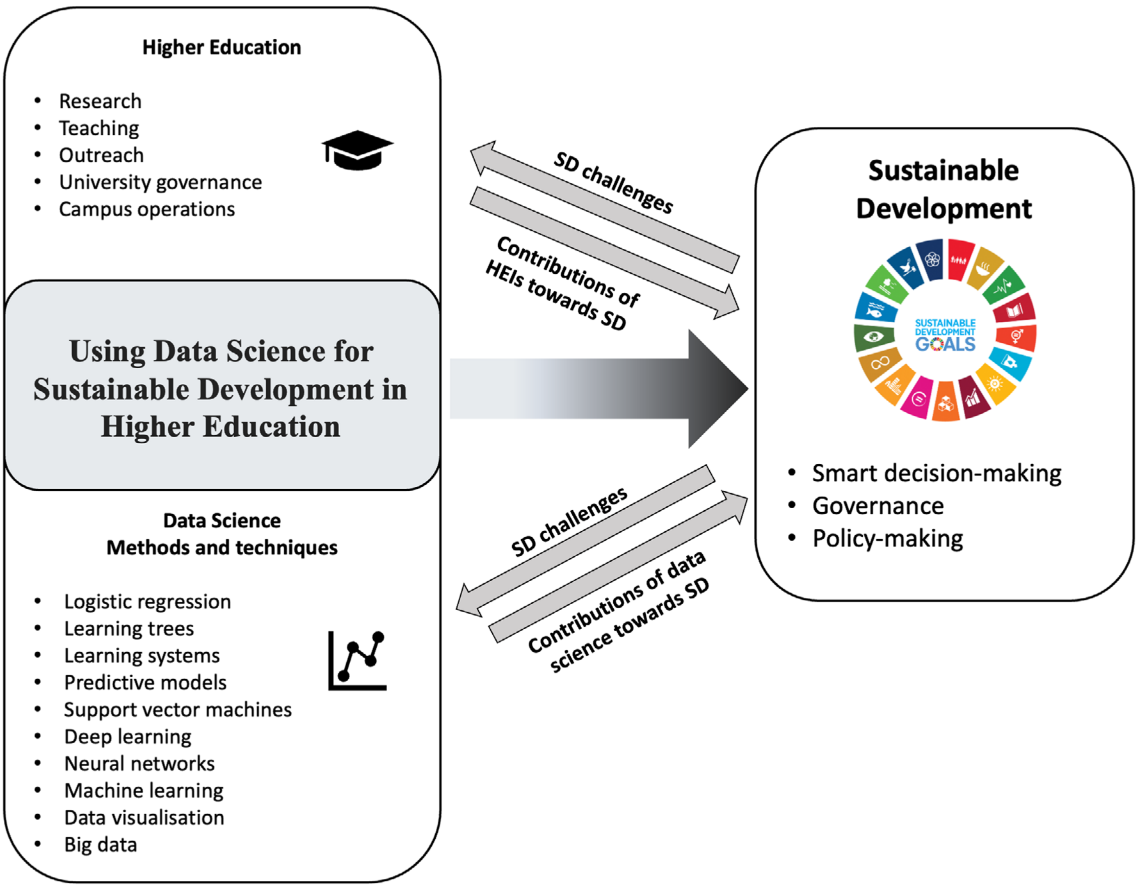

FIGURE 2 An overview of the interconnections between data science and SD. *Source*: The authors.

to the SDG 4 – quality education (Li, Fei, & Liu, 2021; Li, Yueli, et al., 2021). HEIs also can take advantage of data science since it can enhance university governance, help implement smart campus management by using data-based decision-making systems and ensure resource efficiency (Osuwa et al., 2019; Viñán-Ludeña et al., 2020).

In sum, the findings and discussion provided relevant insight into the previously developed research questions (RQs). For the RQ1 'How is data science contributing to efforts to pursue sustainability in Higher Education Institutions (HEIs)?' the bibliometric analysis unveiled five research strands of how data science contributed to SD in HEIs, considering the use of data science techniques in teaching and research practices, application of data science for decision-making in building efficiency and campus operations in general, and integration of data science into the curricula. Additionally, data science tactics are being included in the SD curriculum in response to the COVID-19 pandemic. Furthermore, a theoretical framework synthesising the interconnections between data science and SD was created to illustrate the contributions of data science to sustainability in HEIs.

Finally, the findings related to the RQ2 'Which examples exist and illustrate how the topic is being handled by them?' were based on the case studies found in the literature and judiciously selected by experts. These could offer tangible examples of the application of data science in addressing sustainability challenges in the context of HEIs, evidencing particular elements such as core components of education 4.0 in the 21st Century, inclusive education, the impact of digitalisation-based approaches to academic education, and many other aspects that tackle the role of data science in addressing sustainability challenges in a direct or indirect way, fostering positive externalities and hindering the negative ones.

5 | CONCLUSIONS

This study aimed to foster knowledge of how data science can contribute to SD in HE. To address this goal, the approach used relied on two main methods: a bibliometric assessment conducted through the co-occurrence of terms technique and an expert-driven literature review resulting in 10 relevant case studies which exemplify issues in the literature and illustrate the outcomes of the bibliometric analysis.

The results from the bibliometrics indicate the presence of five thematic clusters, which can be considered key research strands in the extant literature. They are:

1. the leading data science methods and techniques used in SD research,
2. the relevance of data science in campus operations management and resource efficiency
3. how data science is being used in teaching and learning practices,
4. how AI, big data, and other data science methods are being deployed in engineering programmes, and
5. the relevance of data science for ESD.

Combined with the case studies, this analysis indicates a wide range of ways in which data science and SD in HE intersect. The implications of this analysis are threefold: the study highlights how data science can be used in HE, it offers a framework for enhancing SD in practice, and it supports further research into SD in HE. Importantly, the findings illustrate the contribution which data science could make to achieving the sustainable university ideal (see Sterling et al., 2013), in which sustainability is embedded within all the different aspects of the institution's activities, including teaching, research, estates management, outreach and governance (see Figure 2), as well as links with the wider community. Many HEIs have sustainability in their strategic aims or values, but far fewer have managed to implement effective means of supporting SD across each of these areas.

Through using data science, HEIs could enhance estates management by utilising more effective energy management systems, using data analytics to streamline or enhance their business activities and productivity and quantify the social and environmental impacts, both positive and negative, of their activities. In research, better use of big open data platforms, and enhanced use of remote sensing or mapping abilities based on data science in SD research, could enhance analysis of global achievements towards the SDGs. In support of this development, HEIs should also actively promote open data in their research and other activities. Enabling opportunities for data sharing will enhance the potential for achieving sustainability goals at a regional, national, or international level. In the educational context, data science can contribute to measuring and evaluating educational gains around SD using learning analytics, assist with mapping SDGs in teaching, and offer AI-assisted personalised learning opportunities using the affordances offered by new AI developments such as ChatGPT (openai.com/blog/chatgpt). Although there was less research identified in the cluster analysis related to education (see Figure 1), this is likely to change with the new generative AI taking education by storm and offering huge challenges to traditional curricula and pedagogies in HE (Cotton et al., 2023). In short, governance and decision-making across HEIs could be more firmly evidence-based using data science to bring together and collate or visualise all the myriad of data sources that are available.

The conversation on data science in a SD context may be advanced in the following ways:

1. By promoting data literacy and education, encouraging individuals, and institutions to learn the basics of data science and to develop an understanding of how data can be used to support their work.
2. By supporting the development of data-driven products and services: developing new products and services that may make use of data science to improve user experiences and create added value.
3. By fostering an open dialogue about challenges and opportunities in data science: discussing the new doors that data science opens, and how to overcome the challenges involved.
4. By supporting the development of other sustainability-oriented data science tools and platforms: invest in tools and platforms that may make it easier for sustainability experts to work with data, such as open-source software and cloud-based services.

This study has limitations which could also be seen as opportunities for future studies. The first one is that the bibliometric analysis used the most logical strings without an emphasis on subsidiary themes such as AI. Second, the research focused on data science and not on other elements related to information technology. Further, the emphasis was on SD issues, which limits the scope of the findings to this sector. Nonetheless, the lack of research in this area means this paper provides a positive addition to the literature in that it sheds light on the associations between data science, HE, and SD and offers insights into the potential implications of data science for SD in HE. Future research might usefully offer a more detailed analysis of the potential contribution of data science to specific areas such as teaching, research, campus operations, or living labs.

It could also explore how data science may help HEIs to have a greater impact in their communities. The data from this study has a significant meaning to the HE community. First, it offers an overview of the complexity of data science through the lenses of SD. Second, it provides insights into the factors that contribute to the use of data science in supporting SD initiatives. Third, it reiterates the various roles data science can play in support of teaching and research, and learning opportunities. The findings also can be used to inform educational policy and practice to provide students with the best possible opportunities for success using the current technologies.

Finally, further research with in-depth explorations of the opinions of students and staff, through interviews, for example, could unveil other relevant aspects which could contribute to furthering the use of data science in an SD context.

ACKNOWLEDGEMENTS

This paper is part of the "100 papers to accelerate the implementation of the UN Sustainable Development Goals" initiative promoted by the Research and Transfer Centre Sustainability & Climate Change Management (FTZ-NK).

The authors are grateful for the full support received from the Inter-University Sustainable Development Research Programme (IUSDRP), the Hamburg University of Applied Sciences (HAW-Hamburg) and the University of Sao Paulo (USP) during the development of this study.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

ORCID

Walter Leal Filho https://orcid.org/0000-0002-1241-5225
João Henrique Paulino Pires Eustachio https://orcid.org/0000-0002-6782-3904
Andreea Corina Nita (Danila) https://orcid.org/0000-0001-6122-3134
Maria Alzira Pimenta Dinis https://orcid.org/0000-0002-2198-6740
Amanda Lange Salvia https://orcid.org/0000-0002-4549-7685
Debby R. E. Cotton https://orcid.org/0000-0001-7675-8211
Kamila Frizzo https://orcid.org/0000-0002-0858-7614
Laís Viera Trevisan https://orcid.org/0000-0003-3673-6573
Thais Dibbern https://orcid.org/0000-0003-4826-4614

REFERENCES

Abdella, G. M., Kucukvar, M., Onat, N. C., Al-Yafay, H. M., & Bulak, M. E. (2020). Sustainability assessment and modeling based on supervised machine learning techniques: The case for food consumption. *Journal of Cleaner Production*, *251*, 119661. https://doi.org/10.1016/j.jclepro.2019.119661

Agarwal, R., & Dhar, V. (2014). Big data, data science, and analytics: The opportunity and challenge for IS research. *Information Systems Research*, *25*(3), 443–448. https://doi.org/10.1287/isre.2014.0546

Akter, S., & Wamba, S. F. (2019). Big data and disaster management: A systematic review and agenda for future research. *Annals of Operations Research*, *283*(1), 939–959. https://doi.org/10.1007/s10479-017-2584-2

Allen, J. L., McMullin, R. T., Tripp, E. A., & Lendemer, J. C. (2019). Lichen conservation in North America: A review of current practices and research in Canada and the United States. *Biodiversity and Conservation*, *28*(12), 3103–3138. https://doi.org/10.1007/s10531-019-01827-3

Al-masni, M. A., Al-antari, M. A., Park, J.-M., Gi, G., Kim, T.-Y., Rivera, P., Valarezo, E., Choi, M.-T., Han, S.-M., & Kim, T.-S. (2018). Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system. *Computer Methods and Programs in Biomedicine*, *157*, 85–94. https://doi.org/10.1016/j.cmpb.2018.01.017

Alsharkawi, A., Al-Fetyani, M., Dawas, M., Saadeh, H., & Alyaman, M. (2021). Poverty classification using machine learning: The case of Jordan. *Sustainability*, *13*(3), 1412. https://doi.org/10.3390/su13031412

Arango-Uribe, M. L., Barrera-Causil, C. J., Pallares, V., Rojas, J. M., Mercado Díaz, L. R., Marrone, R., & Marmolejo-Ramos, F. (2022). Statistical modelling of the impact of online courses in higher education on sustainable development. *International Journal of Sustainability in Higher Education*, *24*, 404–425. https://doi.org/10.1108/IJSHE-12-2021-0495

Baesens, B. (2014). *Analytics in a big data world: The essential guide to data science and its applications*. John Wiley & Sons.

Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, *83*, 405–417. https://doi.org/10.1016/j.eswa.2017.04.006

Berman, F., Rutenbar, R., Hailpern, B., Christensen, H., Davidson, S., Estrin, D., Franklin, M., Martonosi, M., Raghavan, P., Stodden, V., & Szalay, A. S. (2018). Realizing the potential of data science. *Communications of the ACM*, *61*(4), 67–72. https://doi.org/10.1145/3188721

Bibri, S. E. (2018). The IoT for smart sustainable cities of the future: An analytical framework for sensor-based big data applications for environmental sustainability. *Sustainable Cities and Society*, *38*, 230–253. https://doi.org/10.1016/j.scs.2017.12.034

Bonini, P. (2020). When tomorrow comes: Technology and the future of sustainability learning in higher education. *Environment: Science and Policy for Sustainable Development*, *62*(4), 39–48. https://doi.org/10.1080/00139157.2020.1764300

Bozkurt, A. (2022). Resilience, adaptability, and sustainability of higher education: A systematic mapping study on the impact of the coronavirus (Covid-19) pandemic and the transition to the new normal. *Journal of Learning for Development*, *9*(1), 1–16. https://doi.org/10.56059/jl4d.v9i1.590

Braunschweig, K., Eberius, J., Thiele, M., & Lehner, W. (2012). The state of open data. *Limits of Current Open Data Platforms*, *1*, 72–72.

Budaya, I. G. B. A., Martha, G. I. R., Agustino, D. P., Wijaya, I. M. P. P., & Harsemadi, I. G. (2021). Prediction model for the tenant's potential failure from business incubation process during COVID-19 period using supervised learning. In *2021 3rd international conference on cybernetics and intelligent system (ICORIS)* (pp. 1–5). IEEE. https://doi.org/10.1109/ICORIS52787.2021.9649505

Cai, Y. P., Huang, G. H., Lin, Q. G., Nie, X. H., & Tan, Q. (2009). An optimization-model-based interactive decision support system for regional energy management systems planning under uncertainty. *Expert Systems with Applications*, *36*(2), 3470–3482. https://doi.org/10.1016/j.eswa.2008.02.036

Cao, L. (2017). Data science: challenges and directions. *Communications of the ACM*, *60*(8), 59–68.

Cao, Y., Song, X., & Wang, T. (2015). Development of an energy-aware intelligent facility management system for campus facilities. *Procedia Engineering*, *118*, 449–456. https://doi.org/10.1016/j.proeng.2015.08.446

Caruso, P. F., Angelotti, G., Greco, M., Guzzetta, G., Cereda, D., Merler, S., & Cecconi, M. (2022). Early prediction of SARS-CoV-2 reproductive number from environmental, atmospheric and mobility data: A supervised machine learning approach. *International Journal of Medical Informatics*, *162*, 104755. https://doi.org/10.1016/j.ijmedinf.2022.104755

Chen, L. (2020). Practice on the sustainable development of talent cultivation mode in the context of big data. In *Cyber security intelligence and analytics* (pp. 682–691). Springer International Publishing. https://doi.org/10.1007/978-3-030-15235-2_94

Costa, A. C. F., de Mello Santos, V. H., & de Oliveira, O. J. (2022). Towards the revolution and democratization of education: A framework to overcome challenges and explore opportunities through industry 4.0. *Informatics in Education*, *21*(1), 1–32.

Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 1–12. https://doi.org/10.1080/14703297.2023.2190148

de Moraes, G. H. S. M., Pelegrini, G. C., de Marchi, L. P., Pinheiro, G. T., & Cappellozza, A. (2022). Antecedents of big data analytics adoption: An analysis with future managers in a developing country. *The Bottom Line*, *35*, 73–89. https://doi.org/10.1108/bl-06-2021-0068

Dhar, V. (2013). Data science and prediction. *Communications of the ACM*, *56*(12), 64–73. https://doi.org/10.1145/2500499

Dhyani, S., Bartlett, D., Kadaverugu, R., Dasgupta, R., Pujari, P., & Verma, P. (2020). Integrated climate sensitive restoration framework for transformative changes to sustainable land restoration. *Restoration Ecology*, *28*(5), 1026–1031. https://doi.org/10.1111/rec.13230

Dorabawila, S., Yatigammana, S., & Abhayaratne, A. (2022). Inclusive education and sustainable development. In *The Wiley handbook of sustainability in higher education learning and teaching* (pp. 377–395). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119852858.ch19

Dubey, R., Gunasekaran, A., Childe, S. J., Papadopoulos, T., Luo, Z., Wamba, S. F., & Roubaud, D. (2019). Can big data and predictive analytics improve social and environmental sustainability? *Technological Forecasting and Social Change*, *144*, 534–545. https://doi.org/10.1016/j.techfore.2017.06.020

Dwivedi, Y. K., Hughes, L., Kar, A. K., Baabdullah, A. M., Grover, P., Abbas, R., Andreini, D., Abumoghli, I., Barlette, Y., Bunker, D., Chandra Kruse, L., Constantiou, I., Davison, R. M., De', R., Dubey, R., Fenby-Taylor, H., Gupta, B., He, W., Kodama, M., … Wade, M. (2022). Climate change and COP26: Are digital technologies and information management part of the problem or the solution? An editorial reflection and call to action. *International Journal of Information Management*, *63*, 102456. https://doi.org/10.1016/j.ijinfomgt.2021.102456

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639), 115–118. https://doi.org/10.1038/nature21056

European Commission (EC). (2019). *Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions. The European Green Deal (COM(2019) 640 final)*. European Comission. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52019DC0640&from=EN

Eustachio, J. H. P. P., Caldana, A. C. F., Liboni, L. B., & Martinelli, D. P. (2019). Systemic indicator of sustainable development: Proposal and application of a framework. *Journal of Cleaner Production*, *241*, 118383. https://doi.org/10.1016/j.jclepro.2019.118383

Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM*, *56*(4), 82–89. https://doi.org/10.1145/2436256.2436274

Gholami, H., Bachok, M. F., Saman, M. Z. M., Streimikiene, D., Sharif, S., & Zakuan, N. (2020). An ISM approach for the barrier analysis in implementing green campus operations: Towards higher education sustainability. *Sustainability*, *12*(1), 363. https://doi.org/10.3390/su12010363

Giuliani, G., Mazzetti, P., Santoro, M., Nativi, S., Bemmelen, J., Colangeli, G., & Lehmann, A. (2020). Knowledge generation using satellite earth observations to support sustainable development goals (SDG): A use case on land degradation. *International Journal of Applied Earth Observation and Geoinformation*, *88*, 102068. https://doi.org/10.1016/j.jag.2020.102068

Gómez-Marín, N., Cara-Jiménez, J., Bernardo-Sánchez, A., Álvarez-de-Prado, L., & Ortega-Fernández, F. (2022). Sustainable knowledge management in academia and research organizations in the innovation context. *The International Journal of Management Education*, *20*(1), 100601. https://doi.org/10.1016/j.ijme.2022.100601

González-Pérez, L. I., & Ramírez-Montoya, M. S. (2022). Components of education 4.0 in 21st century skills frameworks: Systematic review. *Sustainability*, *14*(3), 1–31. https://doi.org/10.3390/su14031493

Goralski, M. A., & Tan, T. K. (2020). Artificial intelligence and sustainable development. *The International Journal of Management Education*, *18*(1), 100330. https://doi.org/10.1016/j.ijme.2019.100330

Grimaldi, D., & Carrasco-Farré, C. (2021). *Implementing data-driven strategies in smart cities: A roadmap for urban transformation*. Elsevier.

Guo, H., Chen, F., Sun, Z., Liu, J., & Liang, D. (2021). Big earth data: A practice of sustainability science to achieve the sustainable development goals. *Science Bulletin*, *66*(11), 1050–1053. https://doi.org/10.1016/j.scib.2021.01.012

Guo, H., Liang, D., Chen, F., & Shirazi, Z. (2021). Innovative approaches to the sustainable development goals using Big Earth Data. *Big Earth Data*, *5*(3), 263–276. https://doi.org/10.1080/20964471.2021.1939989

Guo, H., Nativi, S., Liang, D., Craglia, M., Wang, L., Schade, S., Corban, C., He, G., Pesaresi, M., Li, J., Shirazi, Z., Liu, J., & Annoni, A. (2020). Big Earth Data science: An information framework for a sustainable planet. *International Journal of Digital Earth*, *13*(7), 743–767. https://doi.org/10.1080/17538947.2020.1743785

Guo, H., Liang, D., Sun, Z., Chen, F., Wang, X., Li, J., Zhu, L., Bian, J., Wei, Y., Huang, L., Chen, Y., Peng, D., Li, X., Lu, S., Liu, J., & Shirazi, Z. (2022). Measuring and evaluating SDG indicators with Big Earth Data. *Science Bulletin*, *67*(17), 1792–1801. https://doi.org/10.1016/j.scib.2022.07.015

Gürdür Broo, D., Kaynak, O., & Sait, S. M. (2022). Rethinking engineering education at the age of industry 5.0. *Journal of Industrial Information Integration*, *25*, 100311. https://doi.org/10.1016/j.jii.2021.100311

Gușe, G. R., & Mangiuc, M. D. (2022). Digital transformation in Romanian accounting practice and education: Impact and perspectives. *Amfiteatru Economic*, *24*(59), 252–267. https://doi.org/10.24818/EA/2022/59/252

Haenssle, H. A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., Kalloo, A., Hassen, A. B. H., Thomas, L., Enk, A., Uhlmann, L., Alt, C., Arenbergerova, M., Bakos, R., Baltzer, A., Bertlich, I., Blum, A., Bokor-Billmann, T., Bowling, J., … Zalaudek, I. (2018). Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to

58 dermatologists. *Annals of Oncology*, *29*(8), 1836–1842. https://doi.org/10.1093/annonc/mdy166

Hamrani, A., Akbarzadeh, A., & Madramootoo, C. A. (2020). Machine learning for predicting greenhouse gas emissions from agricultural soils. *Science of the Total Environment*, *741*, 140338. https://doi.org/10.1016/j.scitotenv.2020.140338

Hazen, B. T., Skipper, J. B., Ezell, J. D., & Boone, C. A. (2016). Big data and predictive analytics for supply chain sustainability: A theory-driven research agenda. *Computers & Industrial Engineering*, *101*, 592–598. https://doi.org/10.1016/j.cie.2016.06.030

Helbig, R., von Höveling, S., Solsbach, A., & Marx Gómez, J. (2021). Strategic analysis of providing corporate sustainability open data. *Intelligent Systems in Accounting, Finance and Management*, *28*(3), 195–214. https://doi.org/10.1002/isaf.1501

Hofman, W., & Rajagopal, M. (2014). A technical framework for data sharing. *Journal of Theoretical and Applied Electronic Commerce Research*, *9*(3), 45–58. https://doi.org/10.4067/S0718-18762014000300005

Huang, Y.-S., Hung, J.-S., & Ho, J.-W. (2017). A study on information sharing for supply chains with multiple suppliers. *Computers & Industrial Engineering*, *104*, 114–123. https://doi.org/10.1016/j.cie.2016.12.014

Idreos, S., Papaemmanouil, O., & Chaudhuri, S. (2015). Overview of data exploration techniques. Proceedings of the 2015 ACM SIGMOD international conference on management of data, 277–281. https://doi.org/10.1145/2723372.2731084

Izzo, D., Märtens, M., & Pan, B. (2019). A survey on artificial intelligence trends in spacecraft guidance dynamics and control. *Astrodynamics*, *3*(4), 287–299. https://doi.org/10.1007/s42064-018-0053-6

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, *353*(6301), 790–794. https://doi.org/10.1126/science.aaf7894

Kamran, F., Shahzad, M., & Shafait, F. (2018). Automated military vehicle detection from low-altitude aerial images. 2018 Digital Image Computing: Techniques and Applications (DICTA), 1–8. https://doi.org/10.1109/DICTA.2018.8615865

Kapoor, K., Weerakkody, V., & Sivarajah, U. (2015). Open data platforms and their usability: Proposing a framework for evaluating citizen intentions. In M. Janssen, M. Mäntymäki, J. Hidders, B. Klievink, W. Lamersdorf, B. van Loenen, & A. Zuiderwijk (Eds.), *Open and big data management and innovation* (pp. 261–271). Springer International Publishing. https://doi.org/10.1007/978-3-319-25013-7_21

Kwok, A. O. J., & Treiblmaier, H. (2022). No one left behind in education: Blockchain-based transformation and its potential for social inclusion. *Asia Pacific Education Review*, *23*, 445–455. https://doi.org/10.1007/s12564-021-09735-4

Leal Filho, W., Abubakar, I. R., Mifsud, M. C., Eustachio, J. H. P. P., Albrecht, C. F., Dinis, M. A. P., Borsari, B., Sharifi, A., Levesque, V. R., Ribeiro, P. C. C., LeVasseur, T. J., Pace, P., Trevisan, L. V., & Dibbern, T. A. (2023). Governance in the implementation of the UN sustainable development goals in higher education: Global trends. *Environment, Development and Sustainability*, 1–24. https://doi.org/10.1007/s10668-023-03278-x

Leal Filho, W., Eustachio, J. H. P. P., Dinis, M. A. P., Sharifi, A., Venkatesan, M., Donkor, F. K., Doni, F., Abubakar, I. R., Cichos, K., & Vargas-Hernández, J. (2022). Transient poverty in a sustainable development context. *International Journal of Sustainable Development & World Ecology*, *29*(5), 415–428. https://doi.org/10.1080/13504509.2022.2029612

Leal Filho, W., Levesque, V., Sivapalan, S., Salvia, A. L., Fritzen, B., Deckert, R., Kozlova, V., LeVasseur, T. J., Emblen-Perry, K., Azeiteiro, U. M., Paço, A., Borsari, B., & Shiel, C. (2022). Social values and sustainable development: Community experiences. *Environmental Sciences Europe*, *34*(1), 67. https://doi.org/10.1186/s12302-022-00641-z

Leal Filho, W., Trevisan, L. V., Rampasso, I. S., Anholon, R., Dinis, M. A. P., Brandli, L. L., Sierra, J., Salvia, A. L., Pretorius, R., Nicolau, M., Eustachio, J. P. P., & Mazutti, J. (2023). When the alarm bells ring: Why the UN sustainable development goals may not be achieved by 2030. *Journal of Cleaner Production*, *407*, 137108. https://doi.org/10.1016/j.jclepro.2023.137108

Leal Filho, W., Vasconcelos, C. R. P., Dinis, M. A. P., & Trevisan, L. V. (2022). Commentary – empty promises: Why declarations and international cooperation on sustainable development often fail to deliver. *International Journal of Sustainable Development & World Ecology*, *29*(8), 850–857. https://doi.org/10.1080/13504509.2022.2107108

Leal Filho, W., Yang, P., Eustachio, J. H. P. P., Azul, A. M., Gellers, J. C., Gielczyk, A., Dinis, M. A. P., & Kozlova, V. (2022). Deploying digitalisation and artificial intelligence in sustainable development research. *Environment, Development and Sustainability*, *25*, 4957–4988. https://doi.org/10.1007/s10668-022-02252-3

Li, B., Fei, Y., & Liu, H. (2021). An artificial intelligence based model for evaluation of college students' ability and characteristics through teaching evaluation. *Journal of Intelligent & Fuzzy Systems*, *40*(2), 3397–3407. https://doi.org/10.3233/JIFS-189378

Li, M., Yueli, D., Yunfei, Z., & Yingdong, X. (2021). Research on professional talent training mode on data science and big data technology in local application-oriented universities. 2021 International Conference on Big Data Engineering and Education (BDEE), 56–59. https://doi.org/10.1109/BDEE52938.2021.00016

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, *18*(8), 2674. https://doi.org/10.3390/s18082674

Liberatore, F., Quijano-Sanchez, L., & Camacho-Collados, M. (2019). Applications of data science in policing: European law enforcement research bulletin, 4 SCE, 89–96.

Lindsey, R., Daluiski, A., Chopra, S., Lachapelle, A., Mozer, M., Sicular, S., Hanel, D., Gardner, M., Gupta, A., Hotchkiss, R., & Potter, H. (2018). Deep neural network improves fracture detection by clinicians. *Proceedings of the National Academy of Sciences*, *115*(45), 11591–11596. https://doi.org/10.1073/pnas.1806905115

Linkov, I., Trump, B. D., Poinsatte-Jones, K., & Florin, M.-V. (2018). Governance strategies for a sustainable digital world. *Sustainability*, *10*(2), 440. https://doi.org/10.3390/su10020440

Liu, L., Liu, P., Luo, J., & Zhao, S. (2021). Big data-based dynamic decision-making algorithm for power enterprise operation risk management. 2021 international conference on big data and intelligent decision making (BDIDM), pp. 70–74. IEEE. https://doi.org/10.1109/BDIDM53834.2021.00021

Machicao, J., Specht, A., Vellenich, D., Meneguzzi, L., David, R., Stall, S., Ferraz, K., Mabile, L., O'Brien, M., & Corrêa, P. (2022). A deep-learning method for the prediction of socio-economic indicators from street-view imagery using a case study from Brazil. *Data Science Journal*, *21*. https://doi.org/10.5334/dsj-2022-006

Majji, K. C., & Baskaran, K. (2021). Artificial intelligence analytics—Virtual assistant in UAE automotive industry. In *Inventive systems and control: Proceedings of ICISC 2021* (pp. 309–322). Springer Singapore. https://doi.org/10.1007/978-981-16-1395-1_24

Malhotra, C., Anand, R., & Singh, S. (2018). Applying big data analytics in governance to achieve sustainable development goals (SDGs) in India. In U. M. Munshi & N. Verma (Eds.), *Data science landscape: Towards research standards and protocols* (pp. 273–291). Springer. https://doi.org/10.1007/978-981-10-7515-5_19

Mancisidor, R. A., Kampffmeyer, M., Aas, K., & Jenssen, R. (2020). Deep generative models for reject inference in credit scoring. *Knowledge-Based Systems*, *196*, 105758. https://doi.org/10.1016/j.knosys.2020.105758

Marjanovic, O., & Cecez-Kecmanovic, D. (2020). Open government data platforms – A complex adaptive sociomaterial systems perspective.

Information and Organization, *30*(4), 100323. https://doi.org/10.1016/j.infoandorg.2020.100323

Mian, S. H., Salah, B., Ameen, W., Moiduddin, K., & Alkhalefah, H. (2020). Adapting universities for sustainability education in industry 4.0: Channel of challenges and opportunities. *Sustainability*, *12*(15), 6100. https://doi.org/10.3390/su12156100

Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data*, *2*(1), 1–21. https://doi.org/10.1186/s40537-014-0007-7

OECD. (2019). *Enhancing access to and sharing of data reconciling risks and benefits for data re-use across societies*. OECD Publishing.

Okoroafor, E. R., Offor, C. P., & Prince, E. I. (2022). Mapping relevant petroleum engineering skillsets for the transition to renewable energy and sustainable energy. SPE Nigeria annual international conference and exhibition. https://doi.org/10.2118/212040-MS

Open Data Institute. (2016). *What is open data?* https://www.theodi.org/article/what-is-open-data/

Osuwa, A. A., Katende, J. O., & Osuwa, A. A. (2019). Perception of smart campus big data analytics in university. 2019 third world conference on smart trends in systems security and sustainablity (WorldS4), 48–58. https://doi.org/10.1109/WorldS4.2019.8903933

Ozturk, T., Talo, M., Yildirim, E. A., Baloglu, U. B., Yildirim, O., & Rajendra Acharya, U. (2020). Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Computers in Biology and Medicine*, *121*, 103792. https://doi.org/10.1016/j.compbiomed.2020.103792

Paulauskaite-Taraseviciene, A., Lagzdinyte-Budnike, I., Gaiziuniene, L., Sukacke, V., & Daniuseviciute-Brazaite, L. (2022). Assessing education for sustainable development in engineering study programs: A case of AI ecosystem creation. *Sustainability*, *14*(3), 1702. https://doi.org/10.3390/su14031702

Pedamkar, P. (2022). Introduction to data science | Main components and characteristics. EDUCBA. https://www.educba.com/introduction-to-data-science/

Pennington, D., Ebert-Uphoff, I., Freed, N., Martin, J., & Pierce, S. A. (2020). Bridging sustainability science, earth science, and data science through interdisciplinary education. *Sustainability Science*, *15*(2), 647–661. https://doi.org/10.1007/s11625-019-00735-3

Pitts, J., Gopal, S., Ma, Y., Koch, M., Boumans, R. M., & Kaufman, L. (2020). Leveraging big data and analytics to improve food, energy, and water system sustainability. *Frontiers in Big Data*, *3*, 1–16. https://www.frontiersin.org/article/10.3389/fdata.2020.00013

Poudyal, S., Nagahi, M., Nagahisarchoghaei, M., & Ghanbari, G. (2020). Machine learning techniques for determining students' academic performance: A sustainable development case for engineering education. 2020 international conference on decision aid sciences and application (DASA), 920–924. https://doi.org/10.1109/DASA51403.2020.9317178

Pumsirirat, A., & Yan, L. (2018). Credit card fraud detection using deep learning based on auto-encoder and restricted Boltzmann machine. *International Journal of Advanced Computer Science and Applications*, *9*(1), 18–25. https://doi.org/10.14569/IJACSA.2018.090103

Rahul, M. K., Uppunda, P. L., Vinayaka, R. S., Sumukh, B., & Gururaj, C. (2022). Simulation of self-driving cars using deep learning. *Fundamentals and Methods of Machine and Deep Learning: Algorithms, Tools and Applications*, 379–396.

Revathy, G., Alghamdi, S. A., Alahmari, S. M., Yonbawi, S. R., Kumar, A., & Haq, M. A. (2022). Sentiment analysis using machine learning: Progress in the machine intelligence for data science. *Sustainable Energy Technologies and Assessments*, *53*, 102557. https://doi.org/10.1016/j.seta.2022.102557

Sampson, D. (2017). Teaching and learning analytics to support teacher inquiry. 2017 IEEE global engineering education conference (EDUCON), 1881–1882. https://doi.org/10.1109/EDUCON.2017.7943109

Sarker, I. H. (2022). Smart City Data Science: Towards data-driven smart cities with open research issues. *Internet of Things*, *19*, 100528. https://doi.org/10.1016/j.iot.2022.100528

Scopus. (2022). *Content: How Scopus works*. Elsevier Solutions https://www.elsevier.com/solutions/scopus/how-scopus-works/content?dgcid=RN_AGCM_Sourced_300005030

Seles, B. M. R. P., de Sousa, L., Jabbour, A. B., Jabbour, C. J. C., de Camargo Fiorini, P., Mohd-Yusoff, Y., & Tavares Thomé, A. M. (2018). Business opportunities and challenges as the two sides of the climate change: Corporate responses and potential implications for big data management towards a low carbon society. *Journal of Cleaner Production*, *189*, 763–774. https://doi.org/10.1016/j.jclepro.2018.04.113

Senay, G. B., Schauer, M., Friedrichs, M., Velpuri, N. M., & Singh, R. K. (2017). Satellite-based water use dynamics using historical Landsat data (1984–2014) in the southwestern United States. *Remote Sensing of Environment*, *202*, 98–112. https://doi.org/10.1016/j.rse.2017.05.005

Seyedzadeh, S., Rahimian, F. P., Glesk, I., & Roper, M. (2018). Machine learning for estimation of building energy consumption and performance: A review. *Visualization in Engineering*, *6*(1), 5. https://doi.org/10.1186/s40327-018-0064-7

Shaffer, J. G., Mather, F. J., Wele, M., Li, J., Tangara, C. O., Kassogue, Y., Srivastav, S. K., Thiero, O., Diakite, M., Sangare, M., Dabitao, D., Toure, M., Djimde, A. A., Traore, S., Diakite, B., Coulibaly, M. B., Liu, Y., Lacey, M., Lefante, J. J., … Doumbia, S. O. (2019). Expanding research capacity in sub-Saharan Africa through informatics, bioinformatics, and data science training programs in Mali. *Frontiers in Genetics*, *10*, 331. https://doi.org/10.3389/fgene.2019.00331

Shaik, T., Tao, X., Dann, C., Xie, H., Li, Y., & Galligan, L. (2022). Sentiment analysis and opinion mining on educational data: A survey. *Natural Language Processing Journal*, *2*, 100003.

Sharma, A., Chaudhary, S., Malhotra, J., Saadi, M., Al Otaibi, S., Nebhen, J., & Wuttisittikulkij, L. (2021). A cost-effective photonic radar under adverse weather conditions for autonomous vehicles by incorporating a frequency-modulated direct detection scheme. *Frontiers in Physics*, 467.

Soares, N., Dias Pereira, L., Ferreira, J., Conceição, P., & Pereira da Silva, P. (2015). Energy efficiency of higher education buildings: A case study. *International Journal of Sustainability in Higher Education*, *16*(5), 669–691. https://doi.org/10.1108/IJSHE-11-2013-0147

Song, Y., Liu, T., Li, Y., & Liang, D. (2017). Region division of China's carbon market based on the provincial/municipal carbon intensity. *Journal of Cleaner Production*, *164*, 1312–1323. https://doi.org/10.1016/j.jclepro.2017.07.026

Sterling, S., Maxey, L., & Luna, H. (Eds.). (2013). *The sustainable university: Progress and prospects*. Routledge.

Sublime, J., & Kalinicheva, E. (2019). Automatic post-disaster damage mapping using deep-learning techniques for change detection: Case study of the Tohoku tsunami. *Remote Sensing*, *11*(9), 1123. https://doi.org/10.3390/rs11091123

Taylor, J., Graham, M., Louw, A., Lepheana, A., Madikizela, B., Dickens, C., Chapman, D., & Warner, S. (2021). Social change innovations, citizen science, miniSASS and the SDGs. *Water Policy*, *24*, 708–717. https://doi.org/10.2166/wp.2021.264

Tutsch, D., Callou, G., Maciel, P., & Ferreira, J. (2020). An algorithm to optimise the energy distribution of data centre electrical infrastructures. *International Journal of Grid and Utility Computing*, *11*, 419. https://doi.org/10.1504/IJGUC.2020.10028894

United Nations (UN). (1987). Report of the world commission on environment and development: "Our common future" [Brundtland report].

United Nations (UN). (2015). The 17 goals. https://sdgs.un.org/goals

Upasana. (2018). Data science applications. Top 10 use cases of data science. *Edureka*. https://www.edureka.co/blog/data-science-applications/

van Eck, N. J., & Waltman, L. (2011). Text mining and visualization using VOSviewer (arXiv:1109.2058). arXiv. https://doi.org/10.48550/arXiv.1109.2058

van Eck, N. J., & Waltman, L. (2017). Citation-based clustering of publications using CitNetExplorer and VOSviewer. *Scientometrics*, *111*(2), 1053–1070. https://doi.org/10.1007/s11192-017-2300-7

Van Lissa, C. J., Stroebe, W., van Dellen, M. R., Leander, N. P., Agostini, M., Draws, T., Grygoryshyn, A., Gützgow, B., Kreienkamp, J., Vetter, C. S., Abakoumkin, G., Abdul Khaiyom, J. H., Ahmedi, V., Akkas, H., Almenara, C. A., Atta, M., Bagci, S. C., Basel, S., Kida, E. B., … Bélanger, J. J. (2022). Using machine learning to identify important predictors of COVID-19 infection prevention behaviors during the early phase of the pandemic. *Patterns*, *3*(4), 100482. https://doi.org/10.1016/j.patter.2022.100482

Vance, E. A., & Love, K. (2021). Building statistics and data science capacity for development. *Chance*, *34*(3), 38–46. https://doi.org/10.1080/09332480.2021.1979810

Vassakis, K., Petrakis, E., & Kopanakis, I. (2018). Big data analytics: Applications, prospects and challenges. In G. Skourletopoulos, G. Mastorakis, C. X. Mavromoustakis, C. Dobre, & E. Pallis (Eds.), *Mobile big data: A roadmap from models to technologies* (pp. 3–20). Springer International Publishing. https://doi.org/10.1007/978-3-319-67925-9_1

Vega-Villalobos, A., Almanza-Ortega, N. N., Torres-Poveda, K., Pérez-Ortega, J., & Barahona, I. (2022). Correlation between mobility in mass transport and mortality due to COVID-19: A comparison of Mexico City, New York, and Madrid from a data science perspective. *PLoS One*, *17*(3), e0264713. https://doi.org/10.1371/journal.pone.0264713

Vicario, G., & Coleman, S. (2020). A review of data science in business and industry and a future view. *Applied Stochastic Models in Business and Industry*, *36*(1), 6–18. https://doi.org/10.1002/asmb.2488

Viñán-Ludeña, M. S., Jacome-Galarza, L. R., Montoya, L. R., Leon, A. V., & Ramírez, C. C. (2020). Smart university: An architecture proposal for information management using open data for research projects. In Á. Rocha, C. Ferrás, C. E. Montenegro Marin, & V. H. Medina García (Eds.), *Information technology and systems* (pp. 172–178). Springer International Publishing. https://doi.org/10.1007/978-3-030-40690-5_17

VOSviewer. (2022). VOSviewer—Visualizing scientific landscapes. VOSviewer. https://www.vosviewer.com/

Xiang, X., Li, Q., Khan, S., & Khalaf, O. I. (2021). Urban water resource management for sustainable environment planning using artificial intelligence techniques. *Environmental Impact Assessment Review*, *86*, 106515. https://doi.org/10.1016/j.eiar.2020.106515

Xu, Z., Zhai, S., & Phuong, N. X. (2019). Research on green transition development of energy enterprises taking mining industry as an example. *Nature Environment and Pollution Technology*, *18*(5), 1521–1526.

Yuan, X., Hao, H., Guan, C., & Pentland, A. (2022). Which factors affect the performance of technology business incubators in China? An entrepreneurial ecosystem perspective. *PLoS One*, *17*(1), e0261922. https://doi.org/10.1371/journal.pone.0261922

Zheng, Y., Ren, D., Guo, Z., Hu, Z., & Wen, Q. (2019). Research on integrated resource strategic planning based on complex uncertainty simulation with case study of China. *Energy*, *180*, 772–786. https://doi.org/10.1016/j.energy.2019.05.120

How to cite this article: Leal Filho, W., Eustachio, J. H. P. P., Nita (Danila), A. C., Dinis, M. A. P., Salvia, A. L., Cotton, D. R. E., Frizzo, K., Trevisan, L. V., & Dibbern, T. (2023). Using data science for sustainable development in higher education. *Sustainable Development*, 1–14. https://doi.org/10.1002/sd.2638