Sustainability in sports: a three-decade scientific review

Lucio Palazzo[1], Tullio Menini[1], Violetta Simonacci[2], Michele Gallo[1]
[1]Department of Social Sciences and Humanities, University of Naples L'Orientale
[2]Department of Political Sciences, University of Naples Federico II

Sustainability in sports: a three-decade scientific review

Lucio Palazzo, Tullio Menini, Violetta Simonacci, Michele Gallo

Abstract This study uses text mining to investigate the alignment between scientific research on sustainability in sports and the United Nations Sustainable Development Goals (SDGs). We analyzed 1499 research papers published between 1991 and 2024, extracted from Scopus using the keywords "sport" and "sustainability". Using text mining techniques and cosine similarity for topic matching, we categorized each paper according to its most relevant SDG. Our analysis reveals that the most frequently addressed SDGs in sports research are Gender Equality (SDG 5), Decent Work and Economic Growth (SDG 8), and Partnerships for the Goals (SDG 17). However, SDGs focusing on environmental sustainability and technological innovation show relatively lower representation. Some distinct disciplinary patterns show concentrated alignments with specific SDGs, while more multidisciplinary research areas display a more varied distribution across multiple SDGs. These findings provide insight into the evolving sustainability landscape in sports research and highlight areas where further investigation could expand our understanding of the sector's role in achieving global sustainability objectives.

Keywords: Sustainable Development Goals, SDG, sport research, social impact

1. Introduction

Sustainability in sports has become an increasingly significant area of research over the past few decades, driven by the growing recognition of sporting activities' environmental, economic, and social impacts [1, 2, 3]. As the global sports industry expands, the need for sustainable practices within this sector is more pressing than ever. The United Nations Sustainable Development Goals[1] (SDGs), which aim to address global challenges such as climate change, inequality, and economic development, are highly relevant to the sports world. Sports are recognized as potentially powerful tools for significantly achieving the SDGs. From reducing carbon footprints in large-scale events to promoting social inclusion and economic sustainability in local sports initiatives, the opportunities for positive impact are vast. Furthermore, as sustainability becomes an essential component of sports governance, organizations, and policies, the academic community has responded with a growing body of work exploring the intersections between sustainability and sports [4, 5, 6].

Despite the growing importance of these connections, the scientific research on the topic can make it difficult to establish clear patterns and alignments with specific SDGs. For this reason, this paper investigates the connections between sport-related scientific papers and SDGs by applying topic matching. We reviewed scientific research on sustainability in sports, starting from a sample of 1499 scientific articles from Scopus using the keywords "sport" and "sustainability", published between 1991 and 2024. We classified these papers into relevant SDGs using topic matching combined with cosine similarity. This work offers

[1]Source: `https://sdgs.un.org/goals`.

a preliminary view of the key sustainability themes emerging in sports research. The aim is to highlight the growing body of knowledge and its alignment with global sustainability priorities.

2. Methods

To classify research papers on sports and sustainability within the framework of SDGs, we employed a text-mining approach based on topic matching. The primary information in the research papers' titles, abstracts, and keywords was combined with the SDG descriptions. Data is preprocessed by converting text to lowercase, removing punctuation, stopwords, and special characters, and applying stemming to standardize word forms. This step helps reduce dimensionality and avoids the generation of outliers and biases in subsequent analysis steps. The cleaned text is then transformed into a Term Frequency-Inverse Document Frequency (TF-IDF) weighting matrix. The TF-IDF converts documents into vectors based on the relevancy of the word. Given a (cleaned) dataset of T terms and N documents, the TF-IDF is defined as:

$$\text{TF-IDF}(t,d) = TF(t,d) \log \frac{N}{DF(t)} \tag{1}$$

where

$$\begin{aligned} TF(t,d) =& \frac{f_{t,d}}{\sum_{t=1}^{T} f_{t,d}}, \\ DF(t) =& \sum_{d=1}^{N} I\left(f_{t,d} > 0\right), \end{aligned}$$

given $f_{t,d}$ as the raw count of term t in the d-th document; $TF(t,d)$ is the relative frequency of term t in document d and $DF(t)$ is the number of documents containing term t. The emphasis is placed on significant terms within a document, while standard terms across the corpus are penalized.

Employing a proper similarity measure to classify articles under the most relevant SDG is possible. One of the most widely used measures in the text mining framework is the cosine similarity [7], which measures the similarity between documents as a function of the cosine of the angle generated among their vectorized forms. Given two documents i and j, let $\mathbf{d}_i$ and $\mathbf{d}_j$ be their respective TF-IDF weighted vectors, then the cosine similarity among them is equal to:

$$\cos(\theta_{ij}) = \frac{\mathbf{d}_i \cdot \mathbf{d}_j}{||\mathbf{d}_i||\,||\mathbf{d}_j||} \tag{2}$$

with $||\cdot||$ the Euclidean norm and $\mathbf{d}_i \cdot \mathbf{d}_j$ is the dot product between two vectors.

3. Sustainability in sports research: key findings

The data were collected from Scopus using the keywords "sport" and "sustainability" for papers published between 1991 and 2024. This time frame allowed for a comprehensive review of the evolution of research at the intersection of these two topics over more than

three decades. After the processing pipeline, the TF-IDF matrix contains 1499 Documents. Each document's SDG with the highest similarity score is assigned to classify them, subject to a predefined threshold to filter weak associations. This threshold was set equal to the fifth percentile of the distribution of the maximum similarity scores. Table 1 represents a summary of the classified topics. The class "NO SDG" (75 papers) defines the class of unclassified articles, e.g., having similarity below the threshold for all the considered SDGs. Among all the SDGs, Gender equality (SDG 5) shows the highest frequency, followed by SDGs 8 and 17 corresponding to *decent work and economic growth* and *partnership for the goals*, respectively.

Table 1: Classification frequencies and description of the classes.

	Description	Freq.
NO SDG	Weak similarity with SDGs	75
SDG 1	No poverty	27
SDG 2	Zero hunger	61
SDG 3	Good health and well-being	66
SDG 4	Quality education	73
SDG 5	Gender equality	171
SDG 6	Clean water and sanitation	69
SDG 7	Affordable and clean energy	55
SDG 8	Decent work and economic growth	137
SDG 9	Industry, innovation and infrastructure	90
SDG 10	Reduced inequalities	56
SDG 11	Sustainable cities and communities	109
SDG 12	Responsible consumption and production	107
SDG 13	Climate action	71
SDG 14	Life below water	53
SDG 15	Life on land	64
SDG 16	Peace, justice, and strong institutions	86
SDG 17	Partnerships for the goals	129

The heatmap 1 illustrates the alignment between research fields and SDGs. Tiles color scale ranges from light to dark blue; darker shades indicate stronger associations between academic disciplines (x-axis) and SDGs (y-axis). White tiles highlight the absence of classification, emphasizing research areas without significant alignment with a particular SDG.

The SDGs related to topics unrelated to sports are absent. An example is given by *no poverty* and *zero hunger* (SGD 1 and 2), or goals concerning the industry, innovation, infrastructure and smart cities (SDGs 9, 10 and 11). Also, SDGs related to environmental sustainability, such as SDG 7 (Affordable and clean energy), SDG 12 (Responsible consumption and production), and SDG 13 (Climate action), although necessary, have comparatively

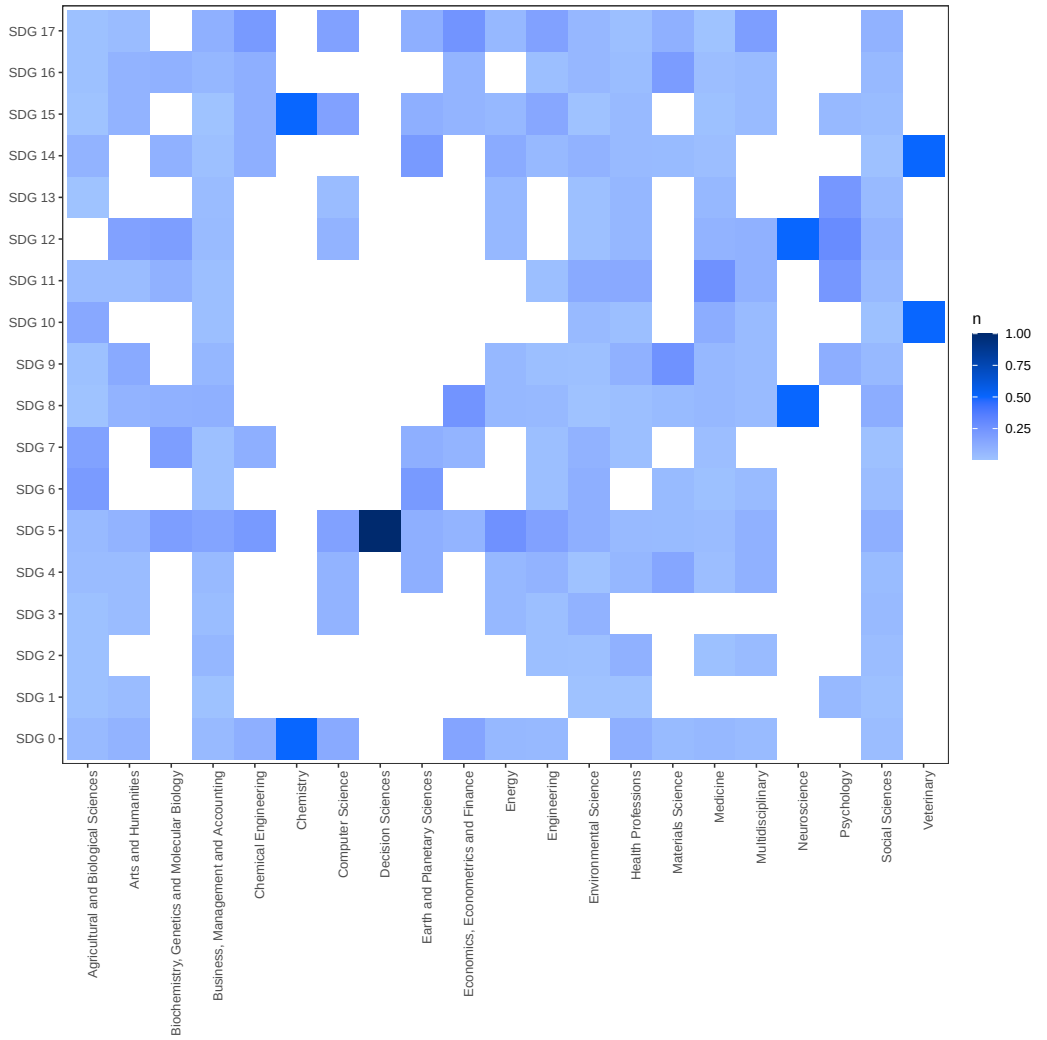

Figure 1: Conditional distribution of the research fields among the SDGs. SDG 0 class represents the group of articles that do not firmly belong to any SDG.

fewer documents classified under them; this result suggests that while environmental sustainability in sports is a growing area of interest, it may still be underrepresented relative to the more socially and economically focused SDGs.

Some research fields appear to be highly concentrated in a few SDGs, in particular, Decision Sciences (SDG 5), Chemistry (SDG 15), Neurosciences (SDG 8 and 12) and Veterinary (10 and 14). In contrast, some scientific areas, such as the social sciences, agriculture and multidisciplinary research, show a heterogeneous distribution of the SDGs, suggesting that the research is in development and sensitive to a cross-disciplinary approach to sports sustainability research.

In summary, the classification results reveal the increasing academic interest in sports sustainability across various social, economic, and environmental dimensions. However, certain SDGs, particularly those focused on environmental sustainability, innovation and smart cities, appear to be underrepresented, suggesting areas for future research. Specific disci-

plines, such as Environmental Science, Energy, and Earth and Planetary Sciences, exhibit strong links with SDG 7 (Affordable and Clean Energy), SDG 13 (Climate Action), and SDG 15 (Life on Land). This result reflects the well-established research focus on sustainability in energy, climate, and ecological sciences. Additionally, the heatmap indicates that while some disciplines are well-aligned with specific SDGs, there is still room for a more integrated and cross-disciplinary approach to sports sustainability research. As a next step, deeper investigations could provide further granularity and uncover additional relationships between sustainability in sports and the SDGs, highlighting potential research gaps and opportunities for future exploration.

References

[1] B. McCullough, Michael Pfahl, and S. Nguyen. The green waves of environmental sustainability in sport. *Sport in Society*, 19:1040–1065, 2016.

[2] Ine Hugaerts, J. Scheerder, Kobe Helsen, J. Corthouts, Erik Thibaut, and Thomas Könecke. Sustainability in Participatory Sports Events: The Development of a Research Instrument and Empirical Insights. *Sustainability*, 13:6034, 2021.

[3] Christopher Hautbois and Michel Desbordes. Sustainability in Sport: Sport, Part of the Problem... and of the Solution. *Sustainability*, 2023.

[4] E. Chatzigianni and C. Mallen. Exploring Congruence in Global Sport Governance between Environmental Policy and Practice. *Sustainability*, 2023.

[5] Patrizia Tettamanzi, Francesco Grazioli, and Michael Murgolo. The quest for sustainable sports management: evidence from the football industry. *Sport, Business and Management: An International Journal*, 2023.

[6] M. Varmus, Pavol Boško, Roman Adámik, and Ivan Greguška. Sports management in the context of sustainability. *Entrepreneurship and Sustainability Issues*, 2024.

[7] Alfirna Rizqi Lahitani, Adhistya Erna Permanasari, and Noor Akhmad Setiawan. Cosine similarity to determine similarity measure: Study case in online essay assessment. In *2016 4th International conference on cyber and IT service management*, 1–6, 2016.