

## THE SUSTAINABLE DEVELOPMENT GOALS IN HIGHER EDUCATION: A MACHINE-LEARNING APPROACH

António Lopes<sup>1</sup>, Catarina Roseta-Palma<sup>2\*</sup> e Ana Simaens<sup>2</sup>

1: Instituto de Telecomunicações, Instituto Universitário de Lisboa (ISCTE-IUL)  
Av. Forças Armadas, 1649-026 Lisboa  
e-mail: antonio.luis@iscte-iul.pt

2: Business Research Unit (BRU-IUL), Instituto Universitário de Lisboa (ISCTE-IUL)  
Av. Forças Armadas, 1649-026 Lisboa  
e-mail: catarina.roseta@iscte-iul.pt, ana.simaens@iscte-iul.pt

**Palavras chave:** sustainable development goals (SDG), machine learning, text classification, higher education institutions, scientific publications

### Resumo

The Sustainable Development Goals (SDGs) were defined by the United Nations in 2015 to provide goalposts for humanity's ambition to move towards a better planet, a prosperous economy, and an inclusive society by 2030. We stand roughly at the midpoint of the SDG implementation period, a good moment to take stock of developments. Higher education institutions (HEIs) play a fundamental role in the creation of knowledge and its dissemination, so they are crucial levers to ensure that the SDGs reach a wider audience. Accordingly, various studies have summarized the contributions of HEIs to the various SDGs in terms of their strategy (Leal Filho et al., 2023), education, sustainability reporting in the sector and, especially, research (Agnew et al., 2020). For the latter, Elsevier has now produced two reports regarding the published scientific contributions that relate to sustainability (six areas in 2015 and the 17 SDGs in 2020), using the whole Scopus database in their analysis. The 2015 report focuses on the status of sustainability science as a research field, pointing out that the number of relevant publications considered to have sustainability relevance grew from 56,390 in 2009 to 75,602 in 2013. The 2020 report, on the other hand, uses much more detailed bibliometric information to note that in the five-year period between 2015 and 2019 "SDG-related publications have reached a staggering 4.1 million articles", with a dominance of SDG3 (Good Health), at 3 million articles, followed by SDG7 (Clean Energy), at 383,000 articles. Nonetheless, the report emphasizes the importance of "closing the gap between science, policy and society" (Agnew et al., 2020, p. 5), as many of the 2030 targets are not achievable based on current trends. On the other hand, Armitage et al. (2020) show that different bibliometric queries engender large differences in the way scholarly publications are classified on the SDGs, challenging the use of existing tools and rankings.

Many individual HEIs have committed to Agenda 2030 and wish to assess their own contributions to the SDGs, yet not all have the resources to apply such methodologies themselves: many lack an adequate database of publications with SDG relevance, since manual assessments are time consuming and might not reflect widely accepted categories. Machine-learning can be a valuable tool in this task of automatically classifying scientific publications as to their contribution to the SDGs (Angin et al., 2022; Morales-Hernández et al., 2022), allowing HEIs to monitor their own contributions and appraise their impact as well as

improving communication and increasing the engagement of the academic community. Automatic text-classification tools to assess the SDGs in wider contexts have also appeared recently, such as SDG-meter and OSDG 2.0 (Guisiano et al. 2022; Pukelis et al 2022, respectively).

This paper takes stock of the methodologies that have been applied to the assessment of research outputs as they relate to the SDGs, including machine-learning-based approaches that help understand these contributions. We also present our machine-learning-based approach as applied to ISCTE’s scientific publications and compare these to the actual choices of the authors when asked to self-report on the suitable SDGs. Overall, the paper contributes to the literature on the assessment of the contributions of HEIs to the SDGs and provides a practical example of how machine learning can be used to automate the process of classifying scientific publications based on their relevance to the SDGs.

### Acknowledgements

This work was supported by Fundação para a Ciência e a Tecnologia, grant UIDB/00315/2020.

### References

- Agnew, K., Francescon, D., Martin, R., Rhannam, M., Schemm, Y., Balisciano, M., Bayazit, K., Bos, C., Erkal, E., & Falk-Krzesinski, H. J. (2020). “The Power of Data to Advance the SDGs: Mapping research for the Sustainable Development Goals.” [https://www.elsevier.com/\\_data/assets/pdf\\_file/0004/1058179/Elsevier-SDG-Report-2020.pdf](https://www.elsevier.com/_data/assets/pdf_file/0004/1058179/Elsevier-SDG-Report-2020.pdf)
- Angin, M., Taşdemir, B., Yılmaz, C. A., Demiralp, G., Atay, M., Angin, P., & Dikmener, G. (2022). A RoBERTa Approach for Automated Processing of Sustainability Reports. In Sustainability (Vol. 14, Issue 23). <https://doi.org/10.3390/su142316139>
- Armitage, C. S., Lorenz, M., & Mikki, S. (2020). Mapping scholarly publications related to the Sustainable Development Goals: Do independent bibliometric approaches get the same results?. Quantitative Science Studies, 1(3), 1092-1108. [https://doi.org/10.1162/qss\\_a\\_00071](https://doi.org/10.1162/qss_a_00071)
- Guisiano, J.E., Chiky, R., De Mello, J. (2022). SDG-Meter: A Deep Learning Based Tool for Automatic Text Classification of the Sustainable Development Goals. In: Nguyen, N.T., Tran, T.K., Tukayev, U., Hong, TP., Trawiński, B., Szczerbicki, E. (eds) Intelligent Information and Database Systems. ACIIDS 2022. Lecture Notes in Computer Science, vol 13757. Springer, Cham. [https://doi.org/10.1007/978-3-031-21743-2\\_21](https://doi.org/10.1007/978-3-031-21743-2_21)
- Leal Filho, W., Simaens, A., Paço, A., Hernandez-Diaz, P. M., Vasconcelos, C. R. P., Fritzen, B., & Mac-Lean, C. (2023). Integrating the Sustainable Development Goals into the strategy of higher education institutions. International Journal of Sustainable Development & World Ecology, 1–12. <https://doi.org/10.1080/13504509.2023.2167884>
- Morales-Hernández, R. C., Jagüey, J. G., & Becerra-Alonso, D. (2022). A Comparison of Multi-Label Text Classification Models in Research Articles Labeled With Sustainable Development Goals. IEEE Access, 10, 123534–123548. <https://doi.org/10.1109/ACCESS.2022.3223094>
- Pukelis, L., Bautista-Puig, N., Statulevičiūtė, G., Stančiauskas, V., Dikmener, G., & Akylbekova, D. (2022). OSDG 2.0: a multilingual tool for classifying text data by UN Sustainable Development Goals (SDGs). arXiv preprint <https://doi.org/10.48550/arXiv.2211.11252>