The Journal of Open Source Software

seesus: a social, environmental, and economic sustainability classifier for Python

Meng Cai [1,2¶], **Yingjie Li** [3], **Dirk Colbry** [4], **Veronica F. Frans** [5,6,7], and **Yuqian Zhang** [5,8]

1 School of Planning, Design and Construction, Michigan State University, East Lansing, MI, 48824, United States **2** Department of Civil and Environmental Engineering, Technical University of Darmstadt, Darmstadt 64287, Germany **3** Natural Capital Project, Woods Institute for the Environment, Stanford University, Stanford, CA, 94305, United States **4** Department of Computational Mathematics, Science and Engineering, Michigan State University, East Lansing, MI 48824, United States **5** Center for Systems Integration and Sustainability, Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI 48823, United States **6** Ecology, Evolution, and Behavior Program, Michigan State University, East Lansing, MI 48824, United States **7** W.K. Kellogg Biological Station, Michigan State University, Hickory Corners, MI 49060, United States **8** Environmental Science and Policy Program, Michigan State University, East Lansing, MI 48823, United States ¶ Corresponding author

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Open Journals

Reviewers:

- @openjournals

Submitted: 01 January 1970
Published: unpublished

Summary

seesus is an open-source Python package that evaluates whether a textual expression aligns with the concept of sustainability as defined by the United Nations Sustainable Development Goals (SDGs). It labels a statement with the 17 SDGs as well as 169 specific targets and categorizes the statement into social, environmental, or economic sustainability. Compared to existing SDG classification software in Python, seesus is the only one that is tuned to a finer scale – the SDG target level – and offers transparency and controllability, allowing users to examine and customize the syntax of text classification according to their needs.

Statement of need

Sustainability is an important topic in contemporary discourse. However, the delineation and interpretation of this concept are often different across disciplines (Salas-Zapata et al., 2017), which hinders effective communication, causes inconsistencies in research and practice, and impedes measurable actions to achieve sustainability (Waseem & Kota, 2017; Yamada et al., 2022). With the increasing popularity of text-based assessments (Amini et al., 2018; Olsen & Fenhann, 2008; Singh et al., 2023), these issues have become more prominent, as the criteria vary for evaluating sustainability commitments and contributions.

seesus, based on the United Nations (UN) Sustainable Development Goals (SDGs), addresses the critical need in text analysis to capture the concept of sustainability with a rigorous and credible definition. The SDGs provide an international framework and a shared understanding of what it means to be sustainable, balancing the environmental, economic, and social dimensions of sustainability (UN, 2015). Automated text analysis to align and classify statements according to the SDGs can help identify the focal points for sustainable development strategies and facilitate data-driven decision-making processes in pursuit of the SDGs. seesus identifies expressions regarding the 17 SDGs and their associated 169 targets within text and labels whether the expressions pertain to social, environmental, or economic sustainability.

In an era of large language models, seesus chooses to use predefined regular expression patterns instead of machine learning for text classification, because this method is more

transparent, replicable, and controllable. Users of seesus can examine the matching logic and customize the syntax if necessary, so users can always understand and maintain control over the results. In addition, compared to other text classifiers based on the SDGs in Python, including `SDG-Classifier` (Rawat, 2022), `SDG Auto Labeller` (Glass, 2020), `UN-SDG-Classifier` (Lamichaney, 2021), `EUR-SDG-Mapper` (Jelicic et al., 2022), and `OSDG` (Pukelis et al., 2022), seesus is the only one that covers all the SDGs and is fine-tuned to the target level. In particular, we developed regular expression matching syntax for the 17 SDGs and the 169 SDG targets, including both direct and indirect matching. We manually tested, reviewed, and improved the accuracy of the matching syntax using randomly selected statements from corporate reports. We went through three rounds of adjustments to finalize the syntax.

seesus achieves an accuracy rate of 76%, as determined by alignment with manual coding. Human intercoder agreement on the same text stands at 83%. Considering the inherent ambiguity and complexity of language, as well as the interconnected nature of the SDGs, the accuracy of seesus is rather high. Other SDG text classifiers did not report accuracy evaluations. Detailed information on our accuracy evaluation and manual refinement can be found in `SDGdetector` (Li et al., 2023), our R package employing the same matching logic as seesus.

Given the interdisciplinary nature of the sustainability concept, the usage of this package is not confined to a specific scientific context. It has a wide application in research based on text analysis across various domains. For example, sustainability scientists can use seesus to label academic publications to quantify which dimension of sustainability receives the most attention. Policy analysts can utilize seesus to conduct large-scale scans of planning documents to assess efforts toward urban sustainability and track the changes over time. Scholars in business research engaged in environmental, social, and governance reporting can employ seesus to evaluate the alignment of corporate messaging with the SDGs. In K12 education, teachers and students can use this tool to delve into community sustainability studies. Individuals who are actively engaged in civic participation may leverage this tool to examine local sustainability plans and efforts. In addition, seesus can be used in combination with translation software to support text analysis in languages other than English.

It is worth noting the limitations of seesus. Because of regular expressions' limited logic capability and lack of context awareness, seesus is not able to capture negative connotations. In other words, it identifies if an expression is related to the SDGs and their targets, but it cannot distinguish whether the expression is about achieving the SDGs or failing to do so. Following the best practices of open-source software, we welcome and encourage users to contribute to improving seesus. We also recommend users cross-validate their results by different text analysis tools and manual checking.

Functionality

seesus currently has four main functions: (1) evaluating whether a statement aligns with the concept of sustainability; (2) identifying SDGs and associated targets in a statement; (3) classifying a statement into social, environmental, and economic sustainability; (4) customizing match syntax.

Acknowledgements

MC was supported by the Michigan State University Dissertation Completion Fellowship. VFF was supported by the National Science Foundation Graduate Research Fellowship Program (Fellow ID: 2018253044) and the Michigan State University Enrichment Fellowship. This project was partly funded by the European Union (ERC, scAInce, 101087218). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those

of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

References

Amini, M., Bienstock, C. C., & Narcum, J. A. (2018). Status of corporate sustainability: A content analysis of fortune 500 companies. *Business Strategy and the Environment*, *27*(8), 1450–1461. https://doi.org/10.1002/bse.2195

Glass, A. (2020). *SDG auto labeller* [Python]. https://github.com/adsglass/SDG-auto-labeller

Jelicic, N., Vorst, T. van der, Ranjbar, B., & Mijnhardt, W. (2022). *EUR-SDG-mapper* [Python]. https://github.com/dialogicnl/eur-sdg

Lamichaney, S. (2021). *UN-SDG-classifier*. https://github.com/SURYA-LAMICHANEY/UN-SDG-Classifier

Li, Y., Frans, V. F., Song, Y., Cai, M., Zhang, Y., & Liu, J. (2023). SDGdetector: An r-based text mining tool for quantifying efforts toward sustainable development goals. *Journal of Open Source Software*, *8*(84), 5124. https://doi.org/10.21105/joss.05124

Olsen, K. H., & Fenhann, J. (2008). Sustainable development benefits of clean development mechanism projects: A new methodology for sustainability assessment based on text analysis of the project design documents submitted for validation. *Energy Policy*, *36*(8), 2819–2830. https://doi.org/10.1016/j.enpol.2008.02.039

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:2211.11252*. https://doi.org/10.48550/arXiv.2211.11252

Rawat, J. (2022). *SDG-classifier* [Jupyter Notebook]. https://github.com/JiteshRawat/SDG-Classifier

Salas-Zapata, W. A., Ríos-Osorio, L. A., & Cardona-Arias, J. A. (2017). Methodological characteristics of sustainability science: A systematic review. *Environment, Development and Sustainability*, *19*(4), 1127–1140. https://doi.org/10.1007/s10668-016-9801-z

Singh, A. B., Meena, H. K., Khandelwal, C., & Dangayach, G. S. (2023). Sustainability assessment of higher education institutions: A systematic literature review. *Engineering Proceedings*, *37*(11), 23. https://doi.org/10.3390/ECP2023-14728

UN. (2015). Transforming our world: The 2030 agenda for sustainable development. *Resolution Adopted by the General Assembly*. https://doi.org/10.5040/9781509934058.0025

Waseem, N., & Kota, S. (2017). Sustainability definitions—an analysis. In A. Chakrabarti & D. Chakrabarti (Eds.), *Research into design for communities, volume 2* (pp. 361–371). Springer. https://doi.org/10.1007/978-981-10-3521-0_31

Yamada, S., Kanoi, L., Koh, V., Lim, A., & Dove, M. R. (2022). Sustainability as a moral discourse: Its shifting meanings, exclusions, and anxieties. *Sustainability*, *14*(55), 3095. https://doi.org/10.3390/su14053095