Predicting Sustainability Literacy in Engineering Students: A Random Forest Analysis

Dr. Sandra Jennina Sanchez, Florida International University

Ph.D. in Higher Education at Florida International University. Ph.D. in Business Competitiveness and Economic Development from University of Deusto in Spain. Master of Business Administration (MBA) from University of Dallas, Texas. B.S. in Business Administration from Externado University in Colombia. Bilingual in English, and Spanish. Expert in transforming educational models, with 12 years of experience in the education sector, managing and leading programs of high academic recognition and international positioning. 10 years of business experience in financial and administrative areas, leading organizational change management processes

Dr. MILTON JANUARIO RUEDA, Universidad Ean

Statistician, specialist in financial analysis and management, Master in Statistics, and Ph.D. in Statistics from the University of Konstanz, Germany. Research professor – Director of Doctoral Programs at Ean University, with over 30 years of experience in data analysis. Extensive teaching experience at various universities, speaker, and consultant for project development in public and private sector organizations. Specialist in advanced information analysis techniques, statistics and applications, data analysis and information methodologies, projects, engineering, economics, and business.

Julian Rodrigo Sosa-Molano, Florida International University

Julian is a graduate research assistant at Florida International University. He holds a BSc in Electronics Engineering from Pontificia Universidad Javeriana in Colombia and a MSc in Electrical and Computer Engineering from The University of Arizona. He is pursuing his PhD degree in Engineering and Computing Education at Florida International University. He has professional experience in Information Technology, Semiconductors, and Telecommunications in international companies such as Ecopetrol, Texas Instruments, and Ericsson. His research interests focus on inclusive STEM learning and teaching methodologies for students with physical disabilities.

PABLO CESAR OCAMPO, EAN University

©American Society for Engineering Education, 2026

WIP: FACTORS THAT INFLUENCE SUSTAINABILITY LITERACY IN ENGINEERING STUDENTS: A RANDOM FOREST ANALYSIS

Abstract

Sustainability literacy has become a critical component in advancing Education for Sustainable Development (ESD). Integrating ESD into the design of global engineering curricula involves embedding sustainability literacy as a core element. As a result, teaching and assessing sustainability knowledge in higher education institutions is vital to raising sustainability awareness and promoting ESD.

This study aims to identify the most influential predictors of sustainability literacy using a Random Forest regression model. The sample consisted of 105 undergraduate engineering students from a university in Bogotá, Colombia. Sustainability literacy was measured using the Sustainability Literacy Test (Sulitest), a validated international assessment instrument. The Random Forest Analysis included nine predictors mixing between four numerical variables such as Sustainable Humanity, Transition Towards Sustainability, Global–Local Human Interactions, and Role to Play and five categorical variables like academic program, age, years of working experience, semester, gender.

The model achieved strong predictive performance explaining 86.61% of variance in literacy scores. Variable importance metrics revealed that Sustainable Humanity (31.26%) and Transition Towards Sustainability (26.63%) were the most influential predictors, while demographic variables showed negligible impact. Consequently, this research demonstrated Sustainable Humanity was the strongest predictor of sustainability literacy, followed by Transition Towards and Global–Local Human, underscoring the importance of human awareness, transitional dynamics, and global-local perspectives in shaping sustainability knowledge.

Key words:

Education for sustainable development, sustainability literacy, sustainability awareness

Introduction

Higher Education Institutions (HEIs) play a vital role in creating sustainability awareness to support the development of a prosperous and equitable society [1]. Given the significant impact of Education for Sustainable Development (ESD) on enhancing Sustainability Literacy, prior research has emphasized the importance of incorporating sustainability into curricula [2, 3, 4, 5, 6].

In this sense, there is growing demand for sustainability education in engineering programs. For instance, the Accreditation Board for Engineering and Technology (ABET), as part of the student outcomes of an Engineering Program, included the accountability of global, cultural, social, environmental, and economic factors, as well as ethical responsibilities encompassing societal and environmental impact that reflect basic sustainability competencies in the design of academic engineering programs [7, 8,]. Educating engineers on sustainability issues within HEIs is essential, as they play a key role in developing sustainable products, services, and solutions worldwide [9, 10].

Although integrating ESD into the design of global engineering curricula involves embedding sustainability literacy as a core element, and HEIs have responsibility in teaching Sustainability Literacy (SL) [11, 12,], HEIs have faced persistent challenges in integrating sustainability into their curricula [13, 14]. These challenges are particularly evident among engineering faculty striving to enhance sustainability literacy within HEIs.

Accordingly, this study responds to the urgent call to advance sustainability literacy in engineering education by supporting faculty in identifying critical factors that strengthen pedagogical practices and promote sustainability awareness among future engineers. As a result, this research aims to determine the most influential predictors of sustainability literacy among engineering students.

Research design

To determine the most influential predictors of sustainability literacy among engineering students, this cross-sectional study utilized the Sustainability Literacy Test (Sulitest), a globally recognized assessment tool developed to assess sustainability literacy [15]. Prior international research has demonstrated the psychometric validity of Sulitest, reporting a Cronbach's alpha reliability of 0.79 and a standardized Cronbach's alpha of 0.8 [16], as well as a Cronbach's alpha of 0.792 in another study [17].

The sample consisted of 105 undergraduate students (N = 105) engineering related programs in a university in Bogota, Colombia. The mean age was 23.0 years (SD = 4.27), with an age range from 16 to 35 years, indicating a predominantly young sample with moderate variability in age. Regarding academic progression, participants were enrolled on average in approximately the fifth semester (Mean = 4.92; SD = 2.56), suggesting variability in academic standing within the group. In terms of work experience, participants reported an average of 2.48 years (SD = 3.48), reflecting substantial dispersion and indicating that while some individuals had considerable work experience, others had little or none. Finally, the mean score on the measured variable was 57.0 (SD = 16.8), demonstrating moderate variability in performance across participants.

To analyze data, this research applied a Random Forest Regression. A Random Forest model was employed to analyze the extent to which critical sustainability factors such as *Sustainable Humanity*, *Transition Towards Sustainability*, *Global–Local Human Interactions*, and *Role to Play* can predict overall sustainability literacy levels. Random Forest regression was selected due to its robustness in handling mixed data types, modeling non-linear relationships, and ranking predictor importance without restrictive distributional assumptions [18]. The dataset initially contained 13 features; after preprocessing, nine predictors were retained. Data were split into training (80%) and testing (20%) sets. The model was configured with 500 trees (ntree = 500) and mtry = 3. Variable importance was evaluated using Percentage Increase in Mean Squared Error (%IncMSE) and Increase in Node Purity.

The instrument

The Sulitest consists of a standardized online questionnaire of 30 multiple-choice questions that measures the following factors [15].: *Sustainable Humanity*, *Transition Towards Sustainability*, *Global–Local Human Interactions*, and *Role to Play. Those dimensions* can predict overall sustainability literacy levels. Sustainable Humanity assesses understanding of ecosystems, humanity's relationship with nature, and the ecological and social aspects of sustainability. Global and Local Human-Constructed Systems measure knowledge of social, economic, and governance structures, including education, culture, production, consumption, and resource management. Transitions Towards Sustainability evaluates awareness of systemic change processes and familiarity with key sustainability initiatives and concepts such as the Global Compact, GRI, and ecological footprint. Roles to Play dimension measures awareness of how individuals and groups can drive systemic change.

Results

The sociodemographic analysis indicates that sustainability literacy among engineering students is more strongly associated with academic and attitudinal factors than with traditional demographic characteristics. A consistent upward trend in scores was observed as academic level increased, with the highest mean score reported among students in their 7th semester or beyond (62.7 ± 16.9).

Students without work experience obtained higher scores (59.2 ± 16.4) than those with work experience. Women demonstrated higher performance (59.1 ± 16.8) despite representing a minority within the sample. Additionally, students enrolled in the morning program outperformed those in the evening program. Regarding age, participants aged 22–24 years achieved the highest mean score (62.7 ± 13.6), while the youngest group (18–19 years) showed the lowest performance.

This research employed a random forest regression approach to predict the *Score* variable using a dataset of 105 observations with 13 initial features. Following preprocessing—which included converting categorical variables to factors, the dataset was reduced to 9 predictors. The data were split into training (80%) and testing (20%) subsets. The model was configured with 500 decision trees (*ntree* = 500) and 3 variables considered per split (*mtry* = 3), utilizing bootstrap aggregating (bagging) to enhance predictive accuracy and prevent overfitting. Variable importance was evaluated using both the mean decrease in accuracy (%IncMSE) and node purity, providing insights into the contribution of each feature to the model's predictive performance. Figure 1 illustrates the process:

FIGURE 1
Random Forest Analysis

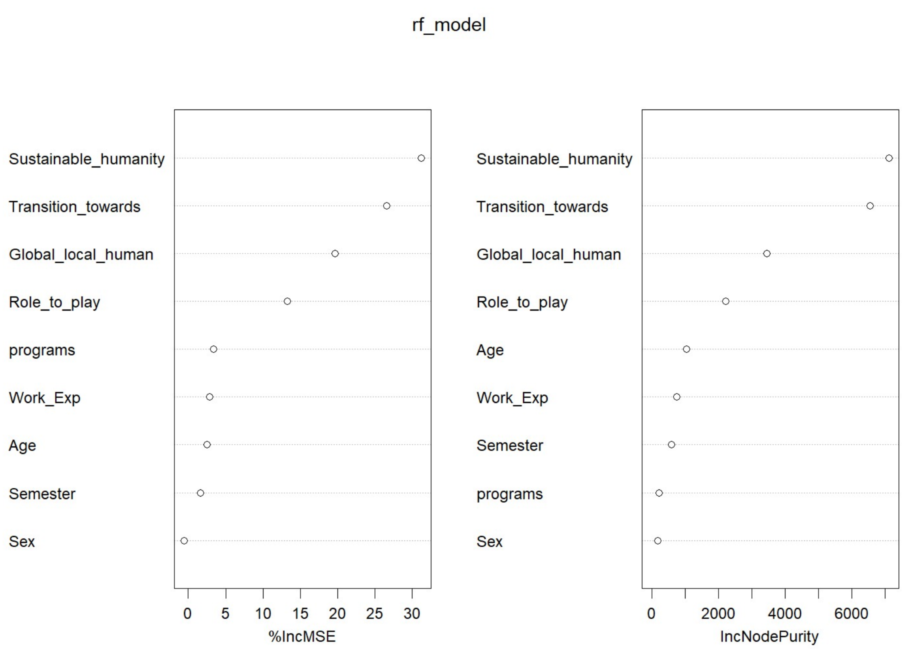
The random forest model demonstrated excellent predictive performance, explaining 86.61% of the variance in the *Score* variable, with a test R^2 of 0.878. Prediction errors were relatively low Root Mean Square Error (RMSE = 7.94) and Mean Absolute Error (MAE = 5.95), indicating high accuracy. Variable importance analysis identified *Sustainable humanity* (%IncMSE = 31.26%) as the strongest predictor, followed by *Transition towards* (26.63%), *Global local human* (19.64%), and *Role to play* (13.32%). In contrast, demographic variables such as Sex, Age, and Semester had minimal predictive power, with Sex even slightly reducing model accuracy when included. The model exhibited no signs of overfitting, as performance was consistent across training and testing sets.

The results of the random forest analysis highlight the key factors influencing sustainability literacy. *Sustainable humanity* emerged as the most critical predictor, followed by *Transition towards* and *Global local human*, indicating that dimensions related to human sustainability awareness, progression toward sustainable practices, and global-local interconnections are central to explaining test performance. Conversely, demographic factors such as Sex (%IncMSE = -0.58%), Semester (1.62%), and Age (2.54%) contributed negligibly, or even negatively, to model predictions, suggesting that attitudinal and cognitive factors play a far more significant role than sociodemographic characteristics in determining sustainability literacy.

Furthermore, Figure 2 shows the consistency between the %IncMSE (predictive accuracy) and IncNodePurity (node purity) metrics that reinforce the robustness of these variable hierarchies, emphasizing that sustainability education should prioritize the development of targeted competencies over general population characteristics.

FIGURE 2

Factors influencing sustainability literacy

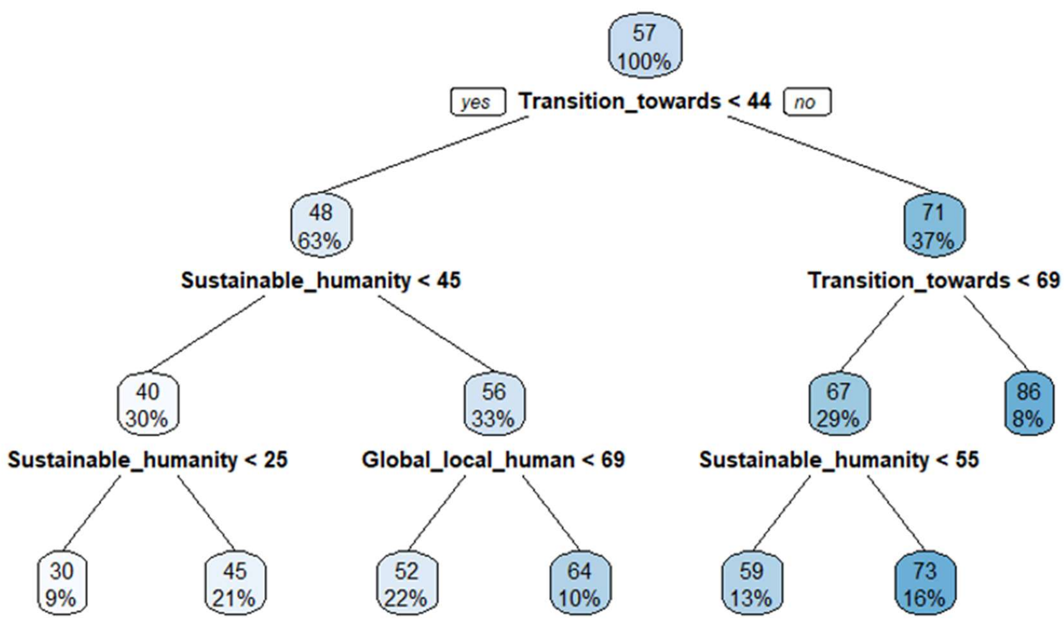


The correlation matrix that confirms strong positive relationships between the sustainability literacy score (Score) and key dimensions. Sustainable Humanity exhibits the strongest correlation (r = 0.845), followed by Transition towards (r = 0.741) and Global local human (r = 0.641), suggesting that awareness of humanity's role in sustainability, understanding of sustainability transitions, and global-local perspectives are the most influential factors in achieving higher literacy scores. Role to play shows a moderate correlation (r = 0.588), underscoring the importance of individual and collective agency. Inter-dimensional correlations, such as the 0.486 observed between Sustainable Humanity and Transition towards, indicate thematic overlap while remaining below thresholds of concern for multicollinearity.

In addition, figure 4 is the representation of a single decision tree that serves as an interpretable benchmark for comparison with the ensemble structure of the Random Forest model. Although this individual tree lacks the predictive robustness and stability of the ensemble, due to its greater susceptibility to overfitting and variability, it plays a crucial pedagogical role by clearly visualizing nodal splitting rules and variable hierarchies in predicting sustainability literacy. The decision tree highlights not only the superiority of ensemble methods in capturing complex, non-linear interactions within the data but also the inherent trade-off between the interpretability of simple models and the predictive accuracy of advanced machine learning architectures.

FIGURE 3

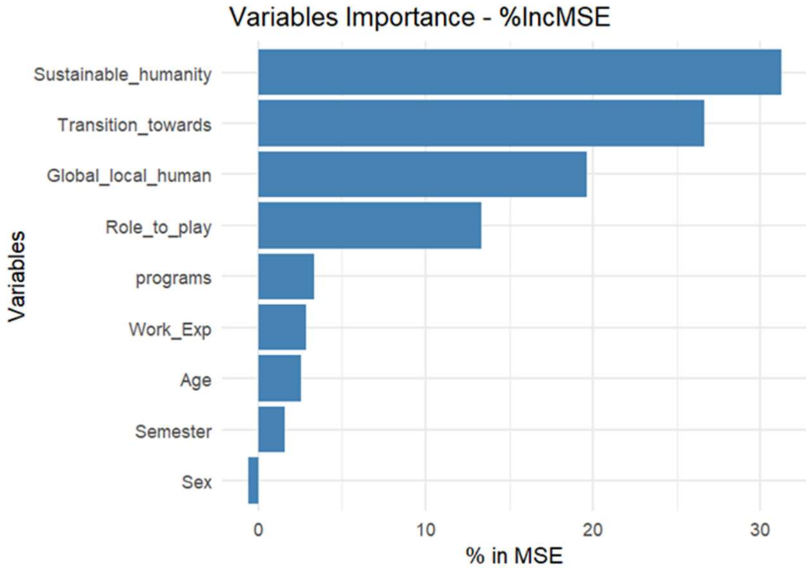
Decision tree



Analysis of Variable Importance Results

The variable importance metrics derived from the Random Forest model provide key insights into the primary drivers of sustainability literacy scores:

FIGURE 4
Variable Importance



The primary predictor is Sustainable humanity (%IncMSE = 31.26; IncNodePurity = 7126.3). It emerges as the most influential variable, indicating that awareness of humanity's role in sustainability is the strongest predictor of literacy scores. In addition, transition towards (26.63% IncMSE) and Global local human (19.64% IncMSE) are secondary pillars, reflecting the importance of transitional dynamics and glocal (global-local) perspectives in sustainability knowledge.

The moderate contributors are role to play (13.32% IncMSE) as it underscores the relevance of individual/collective agency in sustainability outcomes as well as work experience (2.86% IncMSE) and age (2.54% IncMSE) that show minimal but non-negligible effects, suggesting limited practical significance.

On the other hand, negligible or negative impact are semester (1.62% IncMSE), programs (3.38% IncMSE), and gender (-0.58% IncMSE) are those variables are statistically irrelevant. The negative %IncMSE for gender implies that including this variable degrades model accuracy, confirming its irrelevance to sustainability literacy.

Consequently, the alignment between %IncMSE (predictive accuracy loss) and IncNodePurity (node splitting efficacy) validates the robustness of these rankings. For example, Sustainable humanity dominates both metrics (31.26% IncMSE; 7126.3 NodePurity), confirming its primacy.

Implications for Engineering Faculty

The findings directly inform faculty seeking to enhance sustainability literacy. First, instructional design should prioritize strengthening students' understanding of humanity–nature interdependence and systems-level sustainability transitions. Second, curricular activities should incorporate case-based learning that links global frameworks (e.g., Global

Compact, GRI) with local engineering practice. Third, assessment strategies should focus on competency development rather than demographic segmentation. These results provide empirical justification for human-centric and transition-oriented pedagogical approaches in engineering education.

Conclusions

This study demonstrates that Random Forest modeling provides a rigorous analytical framework for identifying key determinants of sustainability literacy. Results confirm that sustainability-related cognitive and attitudinal dimensions, particularly Sustainable Humanity and Transition Towards Sustainability—are substantially more influential than demographic characteristics. By translating machine learning outputs into pedagogical priorities, this research bridges data analytics and engineering education practice, offering actionable guidance for strengthening sustainability literacy in higher education.

These findings confirm the multidimensional nature of sustainability literacy and highlight that educational strategies should prioritize human-centric and transitional dimensions to effectively enhance sustainability competencies. In addition, this research suggest that sustainability literacy is more strongly influenced by academic progression and continued exposure to sustainability-related content than by demographic characteristics or prior work experience. The results support the importance of sustained curricular integration of sustainability throughout the degree program.

Educational interventions should be reinforced during the early years of study and maintained longitudinally across the curriculum. Attention should also be given to potential differences between academic schedules to ensure equitable learning opportunities.

Future research should consider larger and more diverse datasets to validate these findings across different populations. Additionally, examining interaction effects among top predictors may reveal more complex relationships underlying sustainability literacy. Further model optimization, through hyperparameter tuning and comparison with other ensemble methods such as Gradient Boosting, could enhance predictive accuracy. The minimal influence of demographic variables indicates that educational assessment models should continue to focus primarily on behavioral and attitudinal factors rather than sociodemographic characteristics.

Limitation

This study is limited by its relatively small sample size (n = 105) and the single-institution context, which may affect generalizability. Although the Sulitest has established validity in prior studies, future research should further examine contextual reliability within Latin American engineering programs. Expanding the dataset and conducting cross-institutional comparisons would strengthen external validity.

References

[1] J. Winter and D. Cotton, “Making the hidden curriculum visible: Sustainability literacy in higher education,” Environmental Education Research, vol. 18, no. 6, pp. 783–796, Dec. 2012, doi: 10.1080/13504622.2012.670207.
[2] K. Ceulemans, M. De Prins, V. Cappuyns, and W. De Coninck, “Integration of sustainable development in higher education's curricula of applied economics: Large-scale assessments, integration strategies and barriers,” Journal of Management & Organization, vol. 17, no. 5, pp. 621–640, 2011.
[3] R. Lozano, M. Y. Merrill, K. Sammalisto, K. Ceulemans, and F. J. Lozano, “Connecting competences and pedagogical approaches for sustainable development in higher education: A literature review and framework proposal,” Sustainability, vol. 9, no. 10, Art. no. 1889, 2017, doi: 10.3390/su9101889.
[4] C. J. Desha, K. Hargroves, and M. H. Smith, “Addressing the time lag dilemma in curriculum renewal towards engineering education for sustainable development,” International Journal of Sustainability in Higher Education, vol. 10, no. 2, pp. 184–199, 2009, doi: 10.1108/14676370910949356.
[5] A. Wiek, L. Withycombe, and C. L. Redman, “Key competencies in sustainability: A reference framework for academic program development,” Sustainability Science, vol. 6, no. 2, pp. 203–218, 2011, doi: 10.1007/s11625-011-0132-6.
[6] L. Zizka and P. Varga, “Teaching sustainability in higher education institutions: Assessing hospitality students’ sustainability literacy,” Journal of Hospitality & Tourism Education, vol. 33, no. 4, pp. 242–257, 2021, doi: 10.1080/10963758.2020.1726771.
[7] Accreditation Board for Engineering and Technology (ABET), Criteria for accrediting engineering programs, 2026–2027. Baltimore, MD, USA: ABET, 2024. [Online]. Available: https://www.abet.org/accreditation/accreditation-criteria/criteria-for-accrediting-engineering-programs-2026-2027/
[8] ABET, Criteria for accrediting engineering programs, 2018–2019. Baltimore, MD, USA: ABET, 2017. [Online]. Available: www.abet.org/wp-content/uploads/2018/02/E001-18-19-EAC-Criteria-11-29-17.pdf
[9] A. Azapagic, S. Perdan, and D. Shallcross, “How much do engineering students know about sustainable development? The findings of an international survey and possible implications for the engineering curriculum,” European Journal of Engineering Education, vol. 30, no. 1, pp. 1–19, Mar. 2005, doi: 10.1080/03043790512331313804.
[10] F. Sanchez-Carracedo, F. Sabate, and K. Gibert, “A methodology to assess the sustainability competencies in engineering undergraduate programs,” International Journal of Engineering Education, vol. 37, no. 5, pp. 1231–1243, 2021.
[11] P. H. Koehn and J. I. Uitto, “Evaluating sustainability education: Lessons from international development experience,” Higher Education, vol. 67, no. 5, pp. 621–635, 2014, doi: 10.1007/s10734-013-9669-x.
[12] G. Cebrián, D. Pascual, and Á. Moraleda, “Perception of sustainability competencies amongst Spanish pre-service secondary school teachers,” International Journal of Sustainability in Higher Education, vol. 20, no. 7, pp. 1171–1190, 2019, doi: 10.1108/IJSHE-10-2018-0168.
[13] M. Vallée, “How and why US universities fail to impart environmental literacy to all students,” International Journal of Sustainability in Higher Education, vol. 25, no. 9, pp. 60–77, 2023, doi: 10.1108/IJSHE-07-2022-0241.

[14] M. Weiss, M. Barth, and H. von Wehrden, “The patterns of curriculum change processes that embed sustainability in higher education institutions,” Sustainability Science, vol. 16, no. 5, pp. 1687–1702, 2021, doi: 10.1007/s11625-021-00984-1.
[15] A. Décamps, G. Barbat, J. C. Carteron, V. Hands, and C. Parkes, “Sulitest: A collaborative initiative to support and assess sustainability literacy in higher education,” The International Journal of Management Education, vol. 15, no. 2, pp. 138–152, Jul. 2017, doi: 10.1016/j.ijme.2017.02.006.
[16] S. J. Sanchez, D. Ballen, M. J. Rueda, and D. L. Robertson, “Associating sustainability literacy with educational level of industrial engineering students,” in Proc. 2024 ASEE Annual Conf. & Exposition, Jun. 2024. [Online]. Available: https://peer.asee.org/associating-sustainability-literacy-with-educational-level-of-industrial-engineering-students
[17] L. O. Cezarino, E. C. Abdala, M. A. Soares, and V. D. C. Fernandes, “Students' knowledge of sustainability issues in higher education,” Latin American Journal of Management for Sustainable Development, vol. 4, no. 1, pp. 24–40, 2018, doi: 10.1504/LAJMSD.2018.091318.
[18] L. Breiman, Consistency for a simple model of random forests, Tech. Rep. 670, University of California, Berkeley, CA, USA, 2004.