*Elinvo (Electronics, Informatics, and Vocational Education)*
Vol 10, No. 2, 2025, pp. 181-190
ISSN 2580-6424 (printed) | ISSN 2477-2399 (online)

# A Text-Based Analysis of SDGs Goal 4 Representation in Informatics and Computer Science Syllabi

**Indah Rahma Ilmiana*[1], Muhammad Kunta Biddinika[2], Herman[3]**

[1, 2, 3] Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[1] 2407048012@webmail.uad.ac.id*; [2] muhammad.kunta@mti.uad.ac.id; [3] hermankaha@mti.uad.ac.id
* corresponding author

**Article Info**

***Article history:***
Received September 25, 2025
Revised December 1, 2025
Accepted December 3, 2025
Available Online December, 2025

***Keywords:***
SDGs Goal 4; Curriculum Mapping; Informatics Education; Text Mining; Machine Learning.

**Abstract**

Higher education curricula play an important role in supporting the achievement of Sustainable Development Goals (SDGs), particularly SDGs Goal 4: Quality Education, through the development of relevant and sustainable learning. The purpose of this study is to look at the connections between the SDGs Goal 4 targets and the computer science and informatics curricula at universities on the island of Java. 552 course syllabi from current APTIKOM member universities are examined in this study using a text-based document analysis design. The analysis focuses on three frequently accessible syllabus components: course name, course description, and learning outcomes. The Term Frequency-Inverse Document Frequency (TF-IDF) approach was used to represent the textual data after it had been pre-processed and mapped using SDGs keywords to create initial labels. Support Vector Machine (SVM) was used as the comparison model and K-Nearest Neighbors (KNN) as the baseline model for the classification procedure. The results showed a strong relationship between most curricula and Target 4.4, which stresses the development of technical and digital skills. In terms of performance, the KNN model achieved an accuracy of 83%, while the SVM model showed a higher accuracy of 86% with more consistent performance on high-dimensional data and imbalanced class distributions. In conclusion, the syllabus analysis methodology based on keyword mapping and machine learning can be used as a systematic exploratory framework for identifying the relationship between higher education curriculum and SDGs.



## INTRODUCTION

As the cornerstone of a nation's social and economic growth, education is a human right that is delivered through national and international education systems [1], [2]. The Sustainable Development Goals (SDGs), especially Goal 4, which highlights the provision of inclusive, egalitarian, and high-quality education for all, are the focus of international attention in the 2030 Agenda for Sustainable Development. In order to assist sustainable development, SDG 4 not only calls for the implementation of education policies but also promotes the reform of education governance, the use of inclusive pedagogical techniques, and the development of human resource capacity in education. [3].

Higher education institutions play a strategic role in achieving SDGs 4 through the implementation of the Tri Dharma Perguruan Tinggi, namely education, research, and community service [4]. In addition to contributions through research and community service, the curriculum and

ISSN 2580-6424 (printed) | ISSN 2477-2399 (online)

learning process are the most direct and systematic ways to integrate the principles of sustainable development into educational practices [5]. The curriculum serves as the primary medium for translating sustainability values such as digital skills development, equitable access to education, strengthening literacy and numeracy, and improving educator capacity into concrete learning outcomes [6]. Therefore, analyzing syllabi as an operational representation of the curriculum is a relevant approach to evaluating the extent to which higher education institutions support the SDGs agenda.

Study programs in the field of Informatics and Computer Science are strongly linked to SDGs 4, especially in the development of digital and technical skills that are the main needs of a technology-based society [7]. Graduates from this field are expected to not only master technical competencies, but also be able to act as agents of change in designing effective, ethical, and sustainable digital solutions [8]. However, the extent to which these sustainability principles are explicitly reflected in the syllabus of Informatics and Computer Science courses has not been widely studied, especially in the context of higher education in Indonesia [9].

This study applied such a method of text-based analytical using the SDGs Keyword Mapper by the European Commission on purpose to help assessing the way principles of sustainability are taken into curriculum texts [10]. The tool isused in various studies to help determining the content of learning to sustainable objectives of development. It serves a mapping of keywords linked to certain SDG targets. [11]. The keyword-based mapping approach facilitates data driven curriculum evaluation and allows for the systematic study of numerous academic texts.

This research also applied approach of a machine learning that becomes analytical tool to help figuring out patterns of correlation of the SDGs targets and also the syllabus. K-Nearest Neighbors (KNN) and also Support Vector Machine (SVM) are the two techniques of classification sppi. KNN serves as an interpretive classifier baseline that is appropriate for exploratory study according ti text document [12], [13]. Meanwhile, SVM is employed as a comparative technique to assess how similarity of sensitive classification outcomes are to the choice of algorithms that are more resilient when dealing with high-dimensional data [14]. Instead, than concentrating on creating a prediction system, this method seeks to improve comprehension of the patterns found.

The incorporation of SDGs into higher education courses in a variety of settings was the subject of a prior study. A study in Ireland showed that mapping thousands of modules using the TF-IDF approach enabled the identification of curriculum relevance to SDGs targets [15]. Another study on engineering study programs revealed that SDGs goals are reflected in various ways in learning outcomes [16]. In addition, various pedagogical and interdisciplinary approaches have been reported to contribute to strengthening digital and sustainability competencies in higher education [17],[18]. Nevertheless, the curricula of Indonesian computer science and informatics programs have not been particularly covered in these research, nor have they focused their analysis on mapping SDGs 4 targets at the course level using an exploratory machine learning-based document analysis approach.

Based on that research gap, this study aims to explore the extent to which selected targets in SDGs 4 are reflected in the syllabus of Informatics and Computer Science [19] study programs at universities on the island of Java. The analysis focuses on five SDGs 4 targets relevant to the role of higher education, namely Target 4.4 (development of digital and technical skills), Target 4.5 (equitable access to education), Target 4.6 (literacy and numeracy), Target 4.7 (development of educator capacity), and Target 4.a (provision of inclusive education facilities and infrastructure) [20]. The results of this study should give an indicative picture of the trends and any gaps in the curriculum's integration of the SDGs, which may be used as a basis for future curriculum development and assessment.

ISSN 2580-6424 (printed) | ISSN 2477-2399 (online)

## METHODS

Course syllabi from computer science and informatics study programs at Indonesian institutions on the island of Java make up the research object. The information of syllabus was taken from websites of official university of November 2024 and also January 2025. Ten universities in Java served their curricula. The high concentration of universities and also the accessibility of educational and also IT infrastructure which give facilitates to online access to publications of scholar served according to the study targeted selection. Only some institutions that were listed as active as members of APTIKOM were taken to be data,T he official APTIKOM list of membership thtat was available on the website of APTIKOM during the process of collecting data, was applied on purpose to help confirming status of institutional membership. Syllabi which came from related programs of study, were openly give access to everyone, and also featured course expanations of requirements for inclusion.

This research applied a structured, strategy of process-oriented for analysis of text-based document [21]. Data collection, text preprocessing, SDGs keyword mapping, feature extraction, machine learning-assisted classification, and outcome analysis were all steps in the research procedure. This method was exploratory and also sought to help determining the way of certain SDGs 4 targets and also content of syllabus referred to each other.

1. Data Collection
   Documents of syllabus were taken from the official universities websites enrolled as active members of APTIKOM to get collecting data. This analysis only employed publicly accessible online curricula from Computer Science and Informatics study programs. The analysis was restricted to three primary components such as course names, course descriptions, and learning outcomes, because different universities have different syllabus forms [22]. Because they weren't regularly included, other components including learning techniques and evaluation methodologies weren't examined [23], which could lead to structural bias.
2. Pre-processing
   Pre-processing was applied to the collected textual data in order to improve data quality and consistency [24]. Preprocessing procedures included removing punctuation and non-alphabetic characters, removing stopwords using a standard Indonesian stopword list created specifically for this study, and case normalization (turning all text to lowercase). Stemming and lemmatization did not maintain the integrity of discipline-specific vocabulary and technical expressions commonly used in computer science and informatics. Manual checks were performed on a portion of the data to make sure that SDG-related phrases were not unintentionally eliminated during pre-processing. Python was used for all pre-processing, feature extraction, and classification processes, with scikit-learn serving as the main machine learning package [25].
3. SDGs Keyword Mapping
   The SDGs Keyword Mapper, created by the European Commission, is then used to map the pre-processed material to SDGs keywords [10]. The goal of this mapping is to produce first labels that show how the syllabus relates to SDGs 4 targets, specifically Targets 4.4, 4.5, 4.6, 4.7, and 4.a. Word and phrase matching is the basis for the automatic mapping process [26]. In cases where a syllabus contained keywords from more than one SDGs target, the label was determined based on the frequency of occurrence of the keywords. The SDG keyword list was directly translated into Indonesian while preserving each keyword's original semantic meaning because most of the syllabus documents are written in Indonesian. The Discussion section specifically addresses the limits of this technique because it was exploratory in nature and was not meant to be a validated manual annotation. The SDGs Keyword Mapper used in

4. Feature Extraction
   It was done by applying method of the TF-IDF with a predefined SDGs vocabulary of keyword. This TF-IDF tended to be said that generate features from the all corpu so it also restricted representation feature to keywords of SDGs-related derived from the SDGs Keyword Mapper [27]. It tends to guarantee that the derived attributes usually reflect the syllabus relevance to certain SDG Goal 4 aims. No further normalization was aimed at this point due to the entire texts of syllabus been already lowercased during pre-processing. The value of relative of SDG-related terms in every course was reflected by a numerical vector with TF-IDF weights. Parameters like range in n-gram or minimum and maximum frequency of document and also feature size were taken to be limited by the list of keyword due to a fixed vocabulary of keyword was done. By applying on SDG-aligned patterns of content rather than helping to maximize general purpose text performance of classification, this keyword-guided TF-IDF reflection advances the exploratory goal of study [28].
5. Machine Learning Classification
   A 70:30 ratio was applied on purpose to help split the dataset into subsets of training and also testing. A stratified train-test split was applied on prupose to help maintaining proportions of class in both groups because of unbalanced distribution of SDGs Goal 4 target labels. The algorithm of KNN was done as a method of interpretive baseline for metric document similarity-according to the classification by doing the Euclidean distance. Model sensitivity was done by using a range of k values (3, 5, and 7). According to performance of classification, k=3 was determined to be the ideal configuration [29]. As a comparative approach and SVM model with a linear kernel was done on purpose that its suitability for high-dimensional TF-IDF representations of feature [30]. Since the main goal was analysis of exploratory that curriculum patterns of representation  ring how than optimization of accuracy in prediction and neither model taken class strategies in weighting or also resampling.
6. Handling Data Imbalance
   There is found any notable imbalance in the distribution of data of these 4 objectives of SDGs, particularly in Target 4.5 that consisted of less syllabi than the other ones. Following the keyword mapping and also categorization procedure completion from this condition was found. Methods of data balancing like class weighting and also the Synthetic Minority Over-sampling Technique (SMOTE) were not applied in this process of investigation [31]. This option was deliberate since the primary goal of study was to help investigating document-according to curricular patterns of representation rather than to help maximizing classification capabilities of models' prediction. The use of oversampling techniques may result in artificial data that is not representative of real-world teaching methods. Therefore, data imbalance is treated as an empirical characteristic of syllabus documents, and its impact on classification results is critically analyzed at the stage of interpreting and discussing the research results.
7. Analysis and Interpretation of Results
   The distribution of 4 targets of SDGs and also the imbalance in the data were gathered into account while doing analyzing the categorization outcomes [32]. The main goal of this analysis was to help pinpoint prevalent trends and also weaknesses in the way concepts of sustainability are incorporated into the science of computer and also informatics curricula. To help giving a fairer picture of the performance of model certainly in conditions with

imbalanced data so the outcomes of classification were assessed using an amount of measures, consisting of accuracy and precision and recall and also F1-score.

## RESULT AND DISCUSSION

### Data Collection and Text Preprocessing

A total of 552 courses from different univiersities in Java, Indonesia becomes such an active members of APTIKOM were acquired according to the method of data collection. Tose came from publicly available syllabuses of course for programs of computer science and also informatics, which is able to be stated on the official websites of them. The data of text was pre-processed to help increasing consistency and also prepare it for analysis before giving them labelling and also classification. This helps stopwording are determined and punctuation and also non-alphabetic characters are cleared, and letters are getting normalized. It was done on urpose to help reducing noise and also minimize differences in style of writing of syllabi. An example of the preprocessing outcomes is shown in Table 1.

Table 1 Example of Preprocessing Stages

| Text Before Preprocessing | Text After Preprocessing |
|---|---|
| Mata kuliah ini menjelaskan tentang sistem bilangan, menjelaskan fungsi Boolean … | mata kuliah menjelaskan tentang sistem bilangan menjelaskan fungsi boolean |
| Melalui mata kuliah Sistem Basis Data, mahasiswa dapat mempelajari pemodelan data … | melalui mata kuliah sistem basis data mahasiswa mempelajari pemodelan data |

Based on Table 1, the text after preprocessing becomes cleaner and more structured. This way also generates more logical content which help them easy to get review later by clearing such unimportant elements like capital letters and numbers and also punctuation. And also standardization guarantees which the model of classification may focus on significant patterns in the data also reducing noise from terms of irrelevant. Evonmore, preprocessing can be essential to help guaranteeing accuracy and also fairness in text analysis. For instance, words like "Database" and "database" would be considered as two distinct tokens in the absence of case normalization giving such in duplication and also decreased feature extraction accuracy. Practically speaking so the preprocessing step makes sure that the analysis is not hampered by some in writing styles among colleges. Beacause of differences in syllabus layouts and also one school may adopt lengthy so detailed sentences while another may employ succinct points of bullet. These structural variations are minimixed by text cleaning and also normalization and also generating such a standardized dataset. This help enhanceo cross-institutional validity comparisons while also enhancing machine learning performance.

### SDGs Target Distribution and Data Imbalance

This table displays the distribution of labels for the 552 syllabi according I the keyword mapping results.

Table 2 The Results of Distribution Labels

| SDGs Goals 4 Target | Number of Courses |
|---|---|
| Target 4.4 | 262 |
| Target 4.5 | 5 |
| Target 4.6 | 71 |
| Target 4.7 | 21 |
| Target 4.a | 117 |
| Passive | 76 |
| Total | 552 |

Targets 4.a and also 4.6 shows which are referred to foundational capabilities and also supporting environments in learning have representation in moderate in Table 2 distribution stating that some courses also cover more comprehensive educational factors than just technical skills. Targets 4.7 and 4.5, they mention showing that the determined syllabi place less emphasis in the elements. Furthermore, a significant percentage of courses are categorized as Passive, which means that neither their descriptions nor their studying objectives consisten in any clear SDGs Goal 4-referref keywords.

When all is said and done, this distribution confirms that there is a significant class difference across SDG Goal 4 targets. Such an imbalance is seen as a genuine aspect of the course texts rather than a methodological error in this study. This context is necessary to comprehend the observed differences between SDG target categories and evaluate the effectiveness of later machine learning classification conclusions.

**Feature Extraction and ML Classification Results**

At the moment the approach of Term TF-IDF is app;ied was to represent the pre-processed material of syllabus as numerical vectors. It was done to lessen the f general weight terms that are less instructive and also to highlight terms in the material of syllabus that have significant value of thematic. The SDGs-related keywords that made the biggest contributions to syllabus categorization may be found thanks to the representation of TF-IDF. KNN and SVM were then applied In order to classify the TF-IDF vectors. SVM was eapplied mployed as a more reliable comparison technique for high-dimensional data whereas KNN pereserved as a method baseline according to similarity of document.

After the feature extraction process using TF-IDF, the syllabus text vectors were used as input for the classification step using the K-Nearest Neighbors (KNN) method. The Euclidean distance measure was used to test various values of k in order to assess model performance. Following the train-test split outlined in the Methods section, all classification performance metrics are presented on the test set, which makes up 30% of the entire dataset (n = 166 syllabi). As shown in Table 3, the experimental findings show that the arrangement with k=3 produced the best overall performance.

Table 3 Classification Performance KNN

| | *Precision* | *Recall* | *f1-score* | *Support* |
|---|---|---|---|---|
| Target 4.4 | 0.87 | 0.87 | 0.87 | 79 |
| Target 4.5 | 1.00 | 0.50 | 0.67 | 2 |
| Target 4.6 | 0.79 | 0.71 | 0.75 | 21 |
| Target 4.7 | 0.50 | 0.83 | 0.62 | 6 |
| Target 4.a | 0.78 | 0.71 | 0.75 | 35 |
| Passive | 0.92 | 1.00 | 0.96 | 23 |
| *Accuracy* | | | 0.83 | 166 |
| *Macro avg* | 0.81 | 0.77 | 0.77 | 166 |
| *Weighted avg* | 0.84 | 0.83 | 0.83 | 166 |

Table 3 displays the classification results using the KNN technique with a value of k=3. It shows that the KNN model obtained an accuracy of 83% with a weighted average F1-score of 0.83. The best-performing KNN configuration was determined to be k=3 with the Euclidean distance metric since it gave the most balanced performance among the examined configurations. This demonstrates that the model performed fairly well in classifying most of the syllabi, particularly in classes with a significant amount of data. And, with precision, recall, F1-score values of 0.87, Target 4.4 exhibits the best performance, suggesting that the model can reliably identify courses that emphasize the development of technical and digital abilities. With F1-scores of 0.75, targets 4.6 and 4.a likewise exhibit comparatively

ISSN 2580-6424 (printed) | ISSN 2477-2399 (online)

strong performance, demonstrating the model's capacity to identify information pertaining to reading, numeracy, and supporting learning characteristics. Conversely, recall and F1-scores that are lower than those of the majority classes show that performance on Targets 4.5 and 4.7 is typically poorer. The KNN model is skewed toward classes with higher frequencies because of the extremely small amount of data in these classes. With an F1-score of 0.96, the Passive category performs quite well, demonstrating that the algorithm can correctly identify curricula that are not directly related to SDGs 4 aims.

The Support Vector Machine (SVM) algorithm was used for the subsequent classification step as a comparison technique to KNN. Because SVM is more resilient when dealing with high-dimensional data and possible class imbalance in text data, it was selected. The same metrics such as precision, recall, F1-score, and accuracy were used to assess SVM's performance so that the outcomes could be directly compared to KNN's performance. Table 4 summarizes the SVM classification findings.

Table 4 Classification Performance SVM

| | *Precision* | *Recall* | *f1-score* | *Support* |
|---|---|---|---|---|
| Target 4.4 | 0.89 | 0.91 | 0.90 | 79 |
| Target 4.5 | 1.00 | 0.50 | 0.67 | 2 |
| Target 4.6 | 0.93 | 0.67 | 0.78 | 21 |
| Target 4.7 | 0.60 | 1.00 | 0.75 | 6 |
| Target 4.a | 0.74 | 0.74 | 0.74 | 35 |
| Passive | 0.96 | 1.00 | 0.98 | 23 |
| *Accuracy* | | | 0.86 | 166 |
| *Macro avg* | 0.85 | 0.80 | 0.80 | 166 |
| *Weighted avg* | 0.86 | 0.86 | 0.85 | 166 |

The test results in Table 4 show that SVM achieved an accuracy of 86%, higher than KNN. The weighted average F1-score obtained was 0.85, indicating a more stable overall classification performance. Target 4.4 again showed the most consistent performance with an F1-score value of 0.90, confirming the dominance of technical and digital skills development content in the Informatics and Computer Science syllabus. Targets 4.6 and 4.7 also experienced an increase in performance compared to KNN, particularly in terms of recall, which shows that SVM is better able to recognize minority classes than KNN. The Passive category said that the highest performance that resulted in F1-score of 0.98 showing that SVM is becoming effective in distinguishing syllabi which are not directly related to the analyzed SDGs targets. However, Target 4.5 still shows performance limitationsn andalso again demonstrated the impact of data distribution imbalance on classification results.

**Discussion**

The classification outcomes by applying TF-IDF-based approach and also machine learning show pattern in a consistent way relating to the the orientation of the Informatics and Computer Science study curriculum of program on Java Island by reaching SDGs Goal 4 (Quality Education). Overall, the KNN and SVM models could help classify syllabi with a comparatively high degree of accuracy (83% and also 86%, respectively) sttaing that the text-based epresentation feature adopted was highly successful in ceshting the syllabus's thematic qualities.

The dominance of Target 4.4 was becoming such a most important and also crucial discovery. Moreover to result such a the largest data so this target gives such a the most reliable classification process with good kind of precision and also recall and values F1-score on both algorithms. This finding bolsters the argument that computer science and also informatics curricula inherently prioritize the of technical and digital skills development which are important to the discipline and also supported by demands of higher of education's digital. The categorization results show that, in addition to Target 4.4, Target 4.6 and category 4.a are nonetheless substantially represented, although doing worse than the

majority class. It indicates that some courses are not only led towards technical skills, but also contribute to strengthening basic competencies that support lifelong learning. However the F1-score value which is small than Target 4.4 indicates that the united literacy and also numeracy factors in the syllabus is still implicit and also not always explicitly demonstrated through semantically strong terms.

Conversely, Target 4.5 and Target 4.7 show relatively low classification performance, especially in terms of recall. This condition reflects two main things. First, empirically, the number of syllabi that explicitly emphasize aspects of inclusivity, equality, or diversity is still very limited in the context of Informatics and Computer Science. Second, from a methodological perspective, the imbalance in data distribution between classes causes the model to be biased towards the majority class. This is a common limitation in curriculum document-based text classification and has been anticipated as a characteristic of the data, not merely a weakness of the model. The Passive category showed very high performance in both algorithms, particularly in SVM with an F1-score of 0.98. This finding indicates that syllabi that have no explicit connection to SDGs 4 targets can be consistently recognized. Substantively, this indicates that there are still courses that are purely oriented towards technical or theoretical content without explicitly linking them to the principles of sustainability or the educational dimensions emphasized in the SDGs. This finding is important because it shows that there is room to enrich the narrative of sustainability in the formulation of the syllabus, without having to fundamentally change the scientific substance of the course.

SVM often performs better than KNN when comparing algorithms, especially when it comes to macro average and weighted average F1-score. This indicates that SVM is better able to handle high dimensional data and unbalanced class distributions. However, KNN remains relevant as an interpretive baseline method and is suitable for exploratory analysis based on document similarity. The simultaneous use of both algorithms strengthens the reliability of result interpretation and reduces dependence on a single classification approach. Overall, the results of this study show that a combined approach of SDGs keyword mapping, TF-IDF feature extraction, and machine learning classification can be used effectively to evaluate the relevance of the syllabus to SDGs Goal 4 targets. These findings are not intended as a normative evaluation of curriculum quality, but rather as empirical indications of areas that have been well integrated and aspects that still have the potential to be strengthened. Thus, this study provides an analytical basis that can be used by universities as material for reflection in developing a more balanced and sustainability-oriented curriculum.

## CONCLUSION

This study examined the representation of selected SDGs Goal 4 targets within syllabus documents of Informatics and Computer Science study programs at universities on the island of Java using a text-based document analysis and machine learning approach. Through the integration of SDG-oriented keyword mapping, TF-IDF feature representation, and classification using KNN and SVM, this research enables a systematic and reproducible analysis of the presence of sustainability-related educational themes in curriculum documentation. According to the findings, Target 4.4, which emphasizes the growth of technical and digital abilities, is the most widely represented dimension and reflects the fundamental focus of computer science and informatics education. Targets 4.5 and 4.7 continue to have very little representation throughout the examined syllabi, but Targets 4.6 and 4.a show moderate participation. These trends imply that technical competencies are now the focus of sustainable integration in the curriculum, with less explicit attention paid to inclusion and instructor capacity building.

When it comes to processing high-dimensional text data and unbalanced class distributions, SVM outperforms KNN in terms of methodological stability. Nevertheless, optimizing predictive performance is not the goal of this study. Rather, it presents machine learning as an analytical exploratory tool for finding curriculum representation patterns. Consequently, rather than being a

constraint that needs to be fixed, data imbalance is viewed as an empirical feature of the syllabus corpus. To sum up, this study provides an exploratory analytical methodology for evaluating how well higher education curriculum align with SDGs Goal 4 at the course level. The results are meant to serve as an empirical foundation for reflection and future curriculum improvement toward better balanced and sustainable education in informatics and computer science programs, rather than as a normative assessment of curriculum quality.

## ETHICS AND DATA AVAILABILITY STATEMENT

This research relied exclusively on publicly available syllabus documents obtained from official university websites. As no human subjects, private information, or sensitive institutional data were involved, institutional review board (IRB) approval was not required. The analysis reports aggregated findings only and does not disclose the identities of individual institutions. The entire step of procedures were undertaken solely for purposes of academic research. The data of syllabus came from publically available. The dataset gathering process is not taken publicly available bevause of institutional concerns. sources however upon reasonable request so the respective author is able to give providing the SDGs keyword list and also comprehensive methodology details.

## REFERENCES

[1] I. D. G. Palguna and B. K. Dwi Atmaja, "Konsepsi Pendidikan Sebagai Hak Konstitusional," *Jurnal Hukum Ius Quia Iustum*, vol. 30, no. 2, pp. 350–370, May 2023, doi: 10.20885/iustum.vol30.iss2.art6.

[2] F. Tibbitts, "Revitalizing the mission of higher education through a human rights-based approach," *Prospects (Paris)*, vol. 54, no. 2, pp. 401–409, Jun. 2024, doi: 10.1007/s11125-023-09654-9.

[3] D. B. Edwards Jr, M. N. Asadullah, and A. Webb, "Critical perspectives at the mid-point of Sustainable Development Goal 4: Quality education for all—progress, persistent gaps, problematic paradigms, and the path to 2030," *Int J Educ Dev*, vol. 107, p. 103031, May 2024, doi: 10.1016/j.ijedudev.2024.103031.

[4] H. Ali, N. Zainatul, M. Kurniyati, I. Eka, and M. Riskiyah, *Tridarma Perguruan Tinggi (Teori dan Praktik)*, 1st ed. IAIN Madura Press, 2024. [Online]. Available: https://press.iainmadura.ac.id/

[5] W. Sribanasarn, A. Phumiphan, S. Kaewplang, M. Khotdee, O. Sivanpheng, and A. Kangrang, "Sustainability in Action: Analyzing Mahasarakham University's Integration of SDGs in Education, Research, and Operations," *Sustainability*, vol. 17, no. 14, p. 6378, Jul. 2025, doi: 10.3390/su17146378.

[6] E. Gregori-Giralt, C. Benítez-Robles, and J.-L. Menéndez-Varela, "Implementing the Sustainable Development Goals in the Curricula of University Degrees: Initial Steps," *Sustainability*, vol. 17, no. 14, p. 6355, Jul. 2025, doi: 10.3390/su17146355.

[7] M. E. Angelaki, F. Bersimis, T. Karvounidis, and C. Douligeris, "Towards more sustainable higher education institutions: Implementing the sustainable development goals and embedding sustainability into the information and computer technology curricula," *Educ Inf Technol (Dordr)*, vol. 29, no. 4, pp. 5079–5113, Mar. 2024, doi: 10.1007/s10639-023-12025-8.

[8] I. O. Pappas, P. Mikalef, Y. K. Dwivedi, L. Jaccheri, and J. Krogstie, "Responsible Digital Transformation for a Sustainable Society," *Information Systems Frontiers*, vol. 25, no. 3, pp. 945–953, Jun. 2023, doi: 10.1007/s10796-023-10406-5.

[9] M. Hamadi, U. Imtinan, and F. Namisango, "Sustainability education in information systems' curricula: A conceptual research framework," *Educ Inf Technol (Dordr)*, Jan. 2024, doi: 10.1007/s10639-023-12409-w.

[10] T. Adams, S. M. Jameel, and J. Goggins, "Education for Sustainable Development: Mapping the SDGs to University Curricula," *Sustainability*, vol. 15, no. 10, p. 8340, May 2023, doi: 10.3390/su15108340.

[11] M.-G. Barberà-Mariné, L. Fabregat-Aibar, V. Ferreira, and A. Terceño, "One Step Away from 2030: An Assessment of the Progress of Sustainable Development Goals (SDGs) in the European Union," *Eur J Dev Res*, vol. 36, no. 6, pp. 1372–1397, Dec. 2024, doi: 10.1057/s41287-024-00641-0.

[12] E. Ermin, R. Soekarta, L. Jupriadi Fakhri, M. Jundullah, and D. Retno Irianto, "ANALYSIS OF ONLINE TRANSPORTATION CUSTOMER SATISFACTION WITH THE APPLICATION OF TEXT CLASSIFICATION USING KNN METHOD," *Antivirus : Jurnal Ilmiah Teknik Informatika*, vol. 18, no. 2, pp. 244–252, Nov. 2024, doi: 10.35457/antivirus.v18i2.3994.

[13] Q. A'yuniyah and M. Reza, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Di Sma Negeri 15 Pekanbaru," *Indonesian Journal of Informatic Research and Software Engineering (IJIRSE)*, vol. 3, no. 1, pp. 39–45, Mar. 2023, doi: 10.57152/ijirse.v3i1.484.

[14] B. Wang, L. Zhou, Y. Gu, and H. Zou, "Density-Convoluted Support Vector Machines for High-Dimensional Classification," *IEEE Trans Inf Theory*, vol. 69, no. 4, pp. 2523–2536, Apr. 2023, doi: 10.1109/TIT.2022.3222767.

[15] P. Lemarchand, C. MacMahon, M. McKeever, and P. Owende, "An evaluation of a computational technique for measuring the embeddedness of sustainability in the curriculum aligned to AASHE-STARS and the United Nations Sustainable Development Goals," *Frontiers in Sustainability*, vol. 4, Mar. 2023, doi: 10.3389/frsus.2023.997509.

[16] T. Adams, S. Kumar, J. Goggins, and R. Manton, "Embedment of UN Sustainable Development Goals (SDG) within Engineering Degree Programmes," in *Civil Engineering Research in Ireland 2020*, NUI Galway School of Engineering., 2020.

[17] D. Schina, V. Esteve-González, M. Usart, J. L. Lázaro-Cantabrana, and M. Gisbert, "The integration of sustainable development goals in educational robotics: A teacher education experience," *Sustainability (Switzerland)*, vol. 12, no. 23, pp. 1–15, Dec. 2020, doi: 10.3390/su122310085.

[18] J. M. D. S. Wijayarathne, G. M. Hassan, and M. J. Holmes, "Clean energy, clean water, and quality education: Prospects of achieving Sustainable Development Goals ( SDGs ) in Sri Lanka," *Nat Resour Forum*, vol. 47, no. 4, pp. 610–631, Nov. 2023, doi: 10.1111/1477-8947.12287.

[19] N. Alavala, N. Roy, M. McDaniel, M. M. Roozbahani, R. Borela, and P. Babolhavaeji, "Beyond Buzzwords: Making Sustainability a Pillar of the Computing Curriculum," in *Proceedings of the 30th ACM Conference on Innovation and Technology in Computer Science Education V. 1*, New York, NY, USA: ACM, Jun. 2025, pp. 389–395. doi: 10.1145/3724363.3729034.

[20] L. Alcántara-Rubio, R. Valderrama-Hernández, C. Solís-Espallargas, and J. Ruiz-Morales, "The implementation of the SDGs in universities: a systematic review," *Environ Educ Res*, vol. 28, no. 11, pp. 1585–1615, Nov. 2022, doi: 10.1080/13504622.2022.2063798.

[21] H. Morgan, "Conducting a Qualitative Document Analysis," *The Qualitative Report*, 2022, doi: 10.46743/2160-3715/2022.5044.

[22] R. Sodhi and J. Choudhary, "Text Analyzing Tool for Simplifying the Syllabus Creation Process," *International Journal of Information and Education Technology*, vol. 13, no. 2, pp. 409–416, 2023, doi: 10.18178/ijiet.2023.13.2.1820.

[23] N. Aulia, Sarinah, and Juanda, "Analisis Kurikulum Merdeka dan Kurikulum 2013," *Jurnal Literasi dan Pembelajaran Indonesia*, vol. 3, no. 1, pp. 14–20, 2023.

[24] C. P. Chai, "Comparison of text preprocessing methods," *Nat Lang Eng*, vol. 29, no. 3, pp. 509–553, May 2023, doi: 10.1017/S1351324922000213.

[25] S. Brajesh Kumar, V. Dinesh Kumar, and P. Prateek, "A Novel Framework for Text Preprocessing using NLP Approaches and Classification using Random Forest Grid Search Technique for Sentiment Analysis," *Econ Comput Econ Cybern Stud Res*, vol. 59, no. 2/2025, pp. 89–105, Jun. 2025, doi: 10.24818/18423264/59.2.25.06.

[26] F. Poza-Vilches *et al.*, "Greening of the syllabus in faculties of education sciences through sustainable development goals: the case of public Andalusian universities (Spain)," *International Journal of Sustainability in Higher Education*, vol. 23, no. 5, pp. 1019–1044, May 2022, doi: 10.1108/IJSHE-02-2021-0046.

[27] L. Yang, G. Wang, and H. Wang, "Reimagining Literary Analysis: Utilizing Artificial Intelligence to Classify Modernist French Poetry," *Information*, vol. 15, no. 2, p. 70, Jan. 2024, doi: 10.3390/info15020070.

[28] L.-C. Chen, "An extended TF-IDF method for improving keyword extraction in traditional corpus-based research: An example of a climate change corpus," *Data Knowl Eng*, vol. 153, p. 102322, Sep. 2024, doi: 10.1016/j.datak.2024.102322.

[29] D. Vijayanand, "From Data to Impact: Machine Learning Models for Sustainable Development Region Classification- Accelerating the Achievement of Sustainable Development Goals through Predictive Analytics," *International Journal For Multidisciplinary Research*, vol. 5, no. 5, Oct. 2023, doi: 10.36948/ijfmr.2023.v05i05.7180.

[30] L. Sindayigaya and A. Dey, "Machine Learning Algorithms: A Review," *International Journal of Science and Research (IJSR)*, vol. 11, no. 8, pp. 1127–1133, Aug. 2022, doi: 10.21275/SR22815163219.

[31] P. Gnip, L. Vokorokos, and P. Drotár, "Selective oversampling approach for strongly imbalanced data," *PeerJ Comput Sci*, vol. 7, p. e604, Jun. 2021, doi: 10.7717/peerj-cs.604.

[32] Berliana Sugiarti Putri, Lya Hulliyyatus Suadaa, and Efri Diah Utami, "A Multilevel and Hierarchical Approach for Multilabel Classification Model in SDGs Research," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 14, no. 1, pp. 52–61, Feb. 2025, doi: 10.22146/jnteti.v14i1.16265.