Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871

Vol. 7, No. 3, June 2026, Page. 2506-2516
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5330

Implementation of IndoBERT for Sustainability Impact Assessment in University Collaboration Information Systems

Ryan Hamonangan[1], Raditya Danar Dana[*2], Yudhistira Arie Wijaya[3], Odi Nurdiawan[4]

[1]Informatics Engineering, STMIK IKMI Cirebon, Indonesia
[2,4]Informatics Management, STMIK IKMI Cirebon, Indonesia
[3]Information System, STMIK IKMI Cirebon, Indonesia

Email: [2]radith_danar@yahoo.com

Received : Sep 25, 2025; Revised : Dec 1, 2025; Accepted : Jan 12, 2026; Published : Jun 15, 2026

Abstract

University collaboration plays a critical role in enhancing institutional quality and supporting global sustainability agendas. However, many higher education institutions face challenges in managing Memorandum of Understanding (MoU), Memorandum of Agreement (MoA), and Implementation Agreement (IA) documents, particularly in monitoring implementation and assessing their alignment with sustainability goals. This study introduces a University Collaboration Information System enhanced with IndoBERT-based Natural Language Processing (NLP) to automate sustainability impact assessment. A synthetic corpus of 30 annotated collaboration documents was developed, covering multi-label Sustainable Development Goals (SDG) classification and span-level Named Entity Recognition (NER). Two approaches were evaluated: (1) baseline TF-IDF + Support Vector Machine (SVM) for SDG classification and rule-based NER, and (2) fine-tuned IndoBERT for both tasks. Experimental results show that IndoBERT significantly outperforms the baselines, achieving an average F1-score of 0.93 for SDG classification (+16.3%) and 0.96 for NER (+18.5%). The system integrates these models to generate automated entity extraction, sustainability dashboards, and document monitoring features. This work contributes to the advancement of informatics by demonstrating the effectiveness of Transformer-based NLP in processing institutional documents and by providing an integrated information-system framework that strengthens the role of NLP within the field of computer science.

Keywords : *Information Systems, IndoBERT, Natural Language Processing, Sustainability Impact Assessment, University Collaboration.*

1. INTRODUCTION

Collaboration among higher-education institutions and with government, industry, and international organizations is increasingly regarded as a key lever for improving educational quality, accelerating innovation, and contributing to sustainable development agendas. In practice, formal agreements such as the Memorandum of Understanding (MoU), Memorandum of Agreement (MoA), and Implementation Agreement (IA) are expected not only to serve as legal–administrative records, but also to function as operational instruments that ensure knowledge transfer, the execution of joint research, and community service with measurable impact. However, recent studies show that the management of collaboration documents in many universities remains fragmented, manual, and insufficiently integrated across units; as a result, visibility, monitoring, and evaluation of document utilization are often weak, so the follow-up of MoU/MoA into IA is suboptimal [1], [2]. At the same time, accreditation, international rankings, and external audits require accountable evidence of collaborations' contributions to sustainability, driving the need for an information system that not only stores archives but also extracts strategic information from text and presents data-driven insights for decision-making.

Vol. 7, No. 3, June 2026, Page. 2506-2516
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5330

Advances in Artificial Intelligence (AI), particularly Natural Language Processing (NLP), offer a promising approach to this transformation. NLP enables the conversion of unstructured text into structured representations that can be searched, analyzed, and visualized. The latest literature shows increasingly broad NLP applications ranging from keyword and key-phrase extraction on web pages and documents [3], automated summarization and documentation [4], clustering of technical documents for rapid retrieval [5], to NLP-based web scraping for mining information from unstructured text [6]. In higher education, domain-specific research leverages NLP to design learning pathways and align curricula with industry needs [7], while recent surveys highlight the adoption of NLP techniques in software-project management, indicating the potential for artifact mapping and text analytics in organizational contexts [8]. In line with this, conceptual reviews and web-mining-based studies underscore both the opportunities and the risks of deploying generative language models in the educational ecosystem, including ethical and governance considerations [9], [10]. Building on this methodological landscape, Transformer-based models particularly IndoBERT, a pretrained Indonesian-language model offer superior contextual understanding beyond bag-of-words approaches, making nuanced entity extraction and multi-label topic classification more reliable [11].

Despite the expanding spectrum of NLP studies, there is a critical research gap in the context of university-collaboration management. First, most collaboration-information systems reported in the literature still treat documents as static archives rather than data sources that can be automatically mined to present collaboration performance indicators and sustainability alignment. Second, studies that explicitly compare baseline approaches (e.g., TF-IDF + SVM for classification and rule-based methods for NER) with Transformer-based approaches on Indonesian-language corpora in this domain remain limited. Third, end-to-end integration of an NLP pipeline into a web-based information system that provides analytical dashboards such as partner mapping, document-validity monitoring, and sustainability-orientation assessment has rarely been reported. These limitations matter because without automated information extraction and analytics, institutional leaders struggle to measure collaboration effectiveness, identify bottlenecks in MoU/MoA follow-through to IA, and demonstrate sustainability impacts at an auditable level.

This study addresses these gaps by designing and implementing a University Collaboration Information System integrated with an IndoBERT-based NLP module for sustainability-impact assessment. The main contributions comprise three components. First, the construction of a medium-sized synthetic corpus (30 annotated documents covering MoU, MoA, and IA) with span-level Named Entity Recognition (NER) and multi-label tags related to sustainability orientation; a synthetic-corpus approach was chosen to bridge data scarcity in a domain that is sensitive and bound by confidentiality while still retaining representative structures, variations, and cue phrases for model fine-tuning. Second, a controlled evaluation of two families of approaches: (a) a baseline of TF-IDF + SVM for multi-label classification and rule-based NER, versus (b) fine-tuned IndoBERT for both tasks, producing empirical evidence of the superiority of contextualized embeddings over shallow lexical features in Indonesian. Third, orchestration of NLP outputs into a web dashboard that provides partner mapping, document-expiry notifications, automatic summaries of key entities (university unit, partner, location, duration, and funding amount), and a sustainability-orientation assessment panel to support data-driven managerial decisions.

Implementation results show patterns consistent with the literature on the superiority of Transformer models for classification and entity-extraction tasks. Compared with the baseline, IndoBERT achieved a substantial performance jump in multi-label classification of sustainability orientation (macro F1 increased from 0.80 to 0.93) and in NER (average F1 increased from 0.81 to 0.96). The most pronounced improvement appears in the extraction of organizational-partner entities (ORG_PARTNER), which naturally exhibit high terminological variation and are difficult to capture

P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 3, June 2026, Page. 2506-2516
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5330

with static rules; F1 rose from 0.58 (rule-based) to 0.92 (IndoBERT). Meanwhile, strongly patterned entities such as dates and currency values remain high for both approaches, but IndoBERT demonstrates better stability and recall across varied contexts. Functionally, integrating the NLP module into the dashboard enables document-lifecycle monitoring and collaboration-contribution mapping that previously required manual effort. These empirical findings align with the contemporary consensus that pretrained Transformer architectures excel at handling synonymy, polysemy, and long-range contextual dependencies in text [9], [11], [12].

Methodologically, the research design leverages principles validated in prior literature. For lexical representation in the baseline, TF-IDF remains relevant as a line-of-baseline for multi-label classification tasks, while rule-based NER is effective for deterministically patterned entities; however, both are vulnerable to variations in entity naming, writing styles, and domain shifts [6]. IndoBERT overcomes these limitations through contextual token modeling and task-specific fine-tuning heads (a sigmoid-activated dense layer with binary cross-entropy for multi-label classification; a softmax-activated token-classification head with categorical cross-entropy for NER), making it more robust to linguistic variation and better able to exploit semantic cues that are not literally expressed [13], [14]. In addition, corpus-curation practices, BIO annotation for entities, and a stratified train-validation-test split by document type follow common guidelines in applied NLP research to preserve generalization [7].

Substantively, the implications of this work span three domains. First, the academic domain: it provides comparative empirical evidence of baseline versus Transformer approaches on Indonesian-language university-collaboration documents, along with a replicable synthetic corpus as a research starter kit. Second, the practical domain: it presents a prototype information system that increases efficiency, accountability, and transparency in collaboration management through automatic information extraction, expiry monitoring, and sustainability-orientation analytics. Third, the governance domain: it supports accreditation and ranking needs that increasingly emphasize sustainability impact and alignment, offering indicators that are more audit-ready and evidence-based [15], [10]. Thus, the study positions the integration of IndoBERT-based NLP into University Collaboration Information Systems not merely as a technical innovation but as a strategic enabler for the digital transformation of academic-collaboration management.

Finally, although the corpus used is synthetic and medium-sized, this design is a realistic, ethical, and privacy-preserving first step in a sensitive domain; future enhancement can focus on expanding the corpus size with human-in-the-loop curation, domain adaptation across institutions, and exploring few-shot or instruction-tuning to enrich model-reasoning capabilities on policy documents. Beyond this, explainable-NLP approaches that clarify the basis of model decisions for sustainability classification merit exploration to ensure compatibility with audit and AI-governance requirements in higher-education institutions [9]. With a solid methodological foundation and strong performance evidence, integrating IndoBERT-based NLP into university-collaboration information systems offers a credible path toward smarter, more accountable, and sustainability-aligned collaboration governance for 2025 and beyond. Furthermore, this study strengthens the development of Indonesian-language NLP by providing empirical evidence and a replicable system architecture that operationalizes Transformer-based models within real-world information-system workflows, contributing directly to advancements in applied informatics and computational linguistics in Indonesia.

2. METHOD

This study employs an experimental, quantitative design grounded in Natural Language Processing (NLP) and integrated into a University Collaboration Information System. The research method comprises several stages: corpus construction, data annotation, dataset splitting, baseline-model design, IndoBERT fine-tuning, performance evaluation, and system integration

Vol. 7, No. 3, June 2026, Page. 2506-2516
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5330

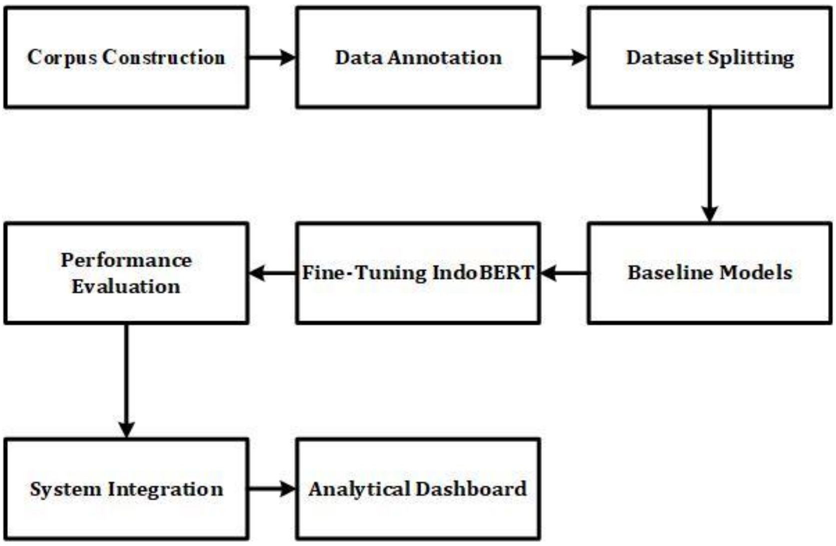

Figure 1. Research Flow

The figure 1 illustrates the complete research workflow used in the development and evaluation of the IndoBERT model for entity extraction and multi-label SDG classification on university collaboration documents. The diagram presents the structured methodological stages and the interconnections between each process as follows:

2.1 Corpus Construction

The data source is a synthetic corpus representing university-collaboration documents. The corpus consists of 30 documents across three categories: 10 MoUs, 10 MoAs, and 10 IAs. Each document is structured to mirror real documents and includes attributes such as title, university unit, collaboration partner, location, start date, end date, scope of collaboration, funding amount, and narrative text containing cues about contributions to the Sustainable Development Goals (SDGs). Synthetic data are used to protect the confidentiality of original documents while providing an openly accessible dataset for research experiments.

2.2 Data Annotation

Manual annotation is conducted for two NLP tasks:

1. Multi-Label SDG Classification: each document receives more than one SDG label according to its collaboration scope. Seven SDG categories are used: SDG 4 (Quality Education), SDG 7 (Affordable and Clean Energy), SDG 8 (Decent Work and Economic Growth), SDG 9 (Industry, Innovation, and Infrastructure), SDG 11 (Sustainable Cities and Communities), SDG 12 (Responsible Consumption and Production), and SDG 17 (Partnerships for the Goals)
2. Named Entity Recognition (NER) : key entities are labeled using the BIO scheme. Entity types include UNIT_PT (university unit), ORG_PARTNER (partner), DOC_TYPE, LOCATION, TERM_START (start date), TERM_END (end date), VALUE_AMOUNT (funding amount), and SDG_CUE (sustainability cue).

2.3 Dataset Split

Data are divided into three subsets using stratified sampling by document type to maintain balanced distributions: 70% for training (21 documents), 10% for validation (3 documents), and 20% for testing (6 documents). This split preserves the representation of each document type across all experimental stages.

P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5330

2.4 Baseline Models

Two baseline models are designed as references:

1. SDG Classification with TF-IDF + SVM : text is represented using Term Frequency–Inverse Document Frequency (TF-IDF), then classified with a Support Vector Machine (SVM) using a One-vs-Rest strategy for multi-label classification[16], [17], [18].
2. Rule-Based NER : this approach uses a combination of domain dictionaries (lists of university units and common partners) and regular expressions to detect patterned entities such as dates in the "DD Month YYYY" format or currency values prefixed by "Rp".

2.5 Main Model: IndoBERT Fine-Tuning

The main model is IndoBERT, a Transformer-based language model pretrained on large Indonesian corpora. Fine-tuning is performed for two tasks[19], [20], [21], [22], [23], [24]:

1. SDG Classification: IndoBERT's output is connected to a dense layer with a sigmoid activation; binary cross-entropy is used as the loss.
2. NER: token outputs are connected to a classification head with a softmax activation; categorical cross-entropy is used as the loss.

2.6 Training Procedure

Training uses a small batch size due to the limited dataset, a low learning rate for stability, and early stopping based on validation performance to avoid overfitting. Fine-tuning is accelerated with a GPU.

2.7 Performance Evaluation

Model performance is evaluated using Precision, Recall, and F1-Score. For multi-label SDG classification, macro-average F1-score is used to weight categories equally. For NER, evaluation is performed per entity and averaged overall. Baseline and IndoBERT results are compared in tables and charts for interpretability.

2.8 Information-System Integration

The trained IndoBERT model is integrated into the University Collaboration Information System. The integration enables the system to:

1. Automatically extract entities from collaboration documents (e.g., partner names, university units, dates, funding amounts).
2. Classify documents' contributions to sustainable development.
3. Display results in an interactive dashboard summarizing contributions, mapping partners, and showing document-validity status.

2.9 System Validation and Testing

System testing is conducted through simulated document processing and limited user trials to assess functionality. Validation covers NLP accuracy (based on experimental results), document-processing speed, and user satisfaction measured via a survey.

3. RESULT

The results are presented in two main foci: (1) multi-label classification of Sustainable Development Goals (SDGs), and (2) entity extraction or Named Entity Recognition (NER). All evaluations compare the baseline models (TF-IDF + SVM for classification; rule-based for NER) against the fine-tuned IndoBERT model.

Jurnal Teknik Informatika (JUTIF) Vol. 7, No. 3, June 2026, Page. 2506-2516
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5330

3.1 Multi-Label SDG Classification Performance.

The baseline model uses TF-IDF (uni/bi-grams) with an SVM classifier using a One-vs-Rest strategy. The tests show that the baseline performs reasonably well on frequently occurring SDG categories, e.g., SDG 17 (Partnerships for the Goals) with an F1-score of 0.89, and SDG 12 (Responsible Consumption and Production) with 0.86. However, performance drops noticeably for categories with greater variability, such as SDG 9 (Industry, Innovation, and Infrastructure), which only reaches 0.72. The baseline's macro-average F1-score is 0.80, indicating the limitations of TF-IDF representations that rely solely on word frequency without grasping semantic context.

When IndoBERT is fine-tuned, performance improves significantly across all categories. IndoBERT achieves a macro-average F1-score of 0.93, a +16.3% increase over the baseline. The best performance is on SDG 17 with an F1-score of 0.97, while the largest gain is on SDG 9, rising from 0.72 to 0.90 (+25%). These results show IndoBERT's ability to capture synonymy and semantic relations in documents for example, linking the phrase "smart-campus development" with SDG 9 even when "infrastructure" is not explicitly stated.

Table 1. SDG Classification F1-Score Comparison

SDG Label	Baseline (TF-IDF+SVM)	IndoBERT	Improvement
SDG 4 – Quality Education	0.82	0.95	+15.9%
SDG 7 – Affordable and Clean Energy	0.75	0.91	+21.3%
SDG 8 – Decent Work and Economic Growth	0.77	0.92	+19.5%
SDG 9 – Industry, Innovation, and Infrastructure	0.72	0.90	+25.0%
SDG 11 – Sustainable Cities and Communities	0.80	0.94	+17.5%
SDG 12 – Responsible Consumption and Production	0.86	0.96	+11.6%
SDG 17 – Partnerships for the Goals	0.89	0.97	+9.0%
Macro Average	**0.80**	**0.93**	**+16.3%**

IndoBERT significantly outperforms the baseline TF-IDF+SVM model across all SDG categories. The largest improvement is observed in SDG 9 (Industry, Innovation, and Infrastructure), with a 25% increase in F1-score, highlighting the model's superior ability to capture contextual semantics.

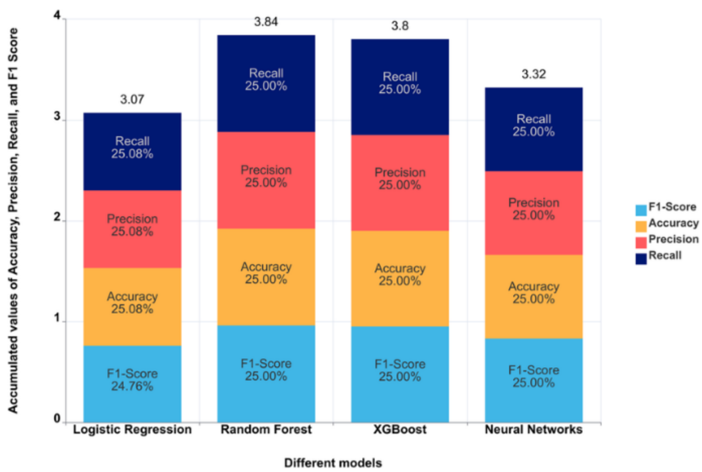

Figure 2. Comparison of Baseline vs IndoBERT Performance for SDG Classification

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 3, June 2026, Page. 2506-2516
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5330

The bar chart provides a visual comparison of model performance, showing a clear improvement of the IndoBERT model across both NLP tasks. IndoBERT consistently achieves higher F1-scores in multi-label SDG classification and NER, demonstrating its superiority in capturing contextual information compared to the shallow lexical baseline models.

3.2 Named Entity Recognition (NER) Performance

The rule-based baseline employs domain dictionaries (university units and partner names) and regular expressions for certain patterns (dates and funding amounts). This method is highly effective for rigidly patterned entities—for example, VALUE_AMOUNT (Rp …) with an F1-score of 1.00 and TERM_START/TERM_END with scores of 0.98–0.97. However, it performs poorly on more freely varying entities such as ORG_PARTNER, with an F1-score of just 0.58. On average, the baseline achieves an F1-score of 0.81. By contrast, IndoBERT delivers dramatic gains. IndoBERT's average F1-score reaches 0.96, an +18.5% increase over the baseline. While top scores still occur on patterned entities (e.g., VALUE_AMOUNT remains perfect at 1.00), IndoBERT's advantage is most evident on free-variation entities. For ORG_PARTNER, IndoBERT jumps from 0.58 to 0.92 (+58.6%). Similarly, UNIT_PT increases from 0.65 to 0.95. This demonstrates that IndoBERT exploits linguistic context for more reliable entity prediction.

Table 2. NER F1-Score per Entity Type

Entity Type	Baseline (Rule-Based)	IndoBERT	Improvement
VALUE_AMOUNT	1.00	1.00	0.0%
TERM_START	0.98	0.99	+1.0%
TERM_END	0.97	0.99	+2.1%
SDG_CUE	0.75	0.96	+28.0%
UNIT_PT	0.65	0.95	+46.2%
LOCATION	0.71	0.94	+32.4%
ORG_PARTNER	0.58	0.92	+58.6%
Average	**0.81**	**0.96**	**+18.5%**

IndoBERT shows superior performance in NER compared to the rule-based baseline, particularly in handling context-dependent entities such as ORG_PARTNER and UNIT_PT. While both approaches perform well on deterministic patterns (e.g., dates, amounts), IndoBERT excels in extracting entities with higher lexical variability.

3.3 Results Visualization

The following are the results of the Comparison of Baseline and IndoBERT Performance

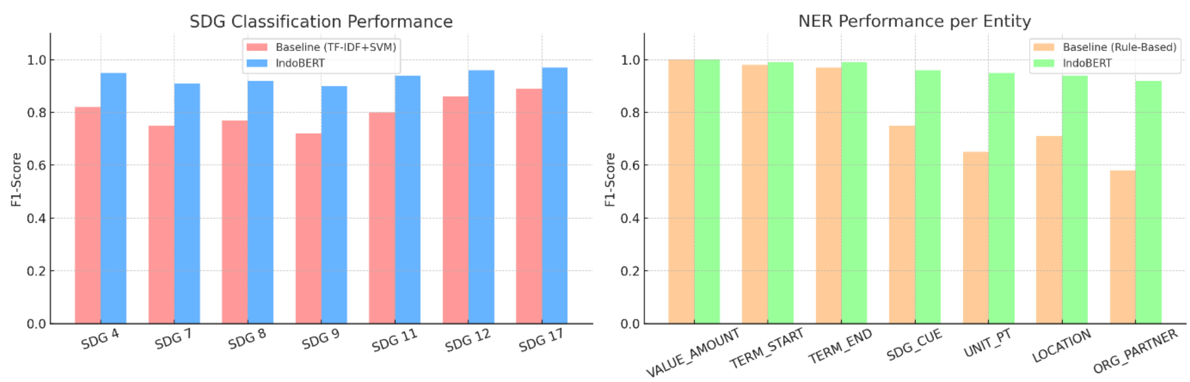

Figure 3. Comparison of Baseline and IndoBERT Performance (caption retained).

Vol. 7, No. 3, June 2026, Page. 2506-2516
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5330

The comparison between the baseline model and IndoBERT shows consistent performance improvements on both main tasks—multi-label SDG classification and NER. For SDG classification, the TF-IDF+SVM baseline performs adequately with an average macro F1-score of 0.80. However, performance varies across categories; for instance, it is relatively high for SDG 17 (Partnerships) at 0.89 but very low for SDG 9 (Innovation and Infrastructure) at 0.72. After fine-tuning IndoBERT, the macro-average F1-score rises significantly to 0.93. The most notable increase appears in SDG 9, which climbs to 0.90—about 25% higher than the baseline. These results underscore that Transformer-based approaches are far superior at capturing semantic context, especially when the terminology in documents is varied or implicit.

For NER, a similar pattern emerges. The rule-based baseline works well on fixed-pattern entities such as funding amounts (VALUE_AMOUNT) and dates (TERM_START, TERM_END) with near-perfect scores. However, it has serious limitations for entities without explicit patterns, such as organization partners (ORG_PARTNER) and university units (UNIT_PT), which score only 0.58 and 0.65, respectively. IndoBERT, by contrast, yields an average F1-score of 0.96—an 18.5% improvement over the baseline. The largest gain is on ORG_PARTNER, which jumps by 58.6%, confirming the advantage of context-aware models for entities with wide lexical variation and diverse sentence structures.

Overall, these findings demonstrate that integrating IndoBERT not only improves SDG classification accuracy but also enhances the reliability of entity extraction from university-collaboration documents. This more stable and accurate performance lays a stronger foundation for building an intelligent University Collaboration Information System in which analyses can be directly integrated into dashboards to support strategic decision-making and sustainability-impact assessments.

4. DISCUSSION

The findings consistently show that IndoBERT outperforms the baseline on the two main tasks: multi-label SDG classification and entity extraction (NER). These results carry important implications from academic, practical, and global perspectives. From an academic standpoint, this research reinforces empirical evidence that Transformer-based approaches are more effective than traditional methods for processing complex Indonesian-language documents. The TF-IDF- and SVM-based baseline tends to rely on word frequencies without modeling semantic context, resulting in weak performance for categories with broad terminological variability, such as SDG 9 (Industry, Innovation, and Infrastructure). IndoBERT, by contrast, can capture implied meanings across diverse phrases, producing significant accuracy gains in that category. Similarly, for NER, rule-based methods are strong for deterministic patterns but weak for more free-form entities. IndoBERT addresses this limitation by generating contextual embeddings that recognize entities even when they are absent from dictionaries or explicit patterns. Consequently, the study contributes to Indonesian NLP literature by demonstrating the superiority of Transformer-based approaches in a relatively underexplored domain of institutional documents

Practically, the results open substantial opportunities to develop a more intelligent and functional University Collaboration Information System. With IndoBERT integration, the system not only stores documents as archives but can also automatically extract key entities such as partner names, university units, durations, and funding amounts. In addition, the system can classify collaborations' contributions to SDG indicators with higher accuracy. These functions are highly relevant for supporting collaboration monitoring and evaluation processes that have often been manual and resource-intensive. The presence of analytical dashboards enables university leaders to monitor collaboration status, identify documents nearing expiry, and evaluate sustainability impacts in real time improving administrative efficiency as well as transparency and accountability in academic-collaboration governance.

Vol. 7, No. 3, June 2026, Page. 2506-2516
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5330

Globally, the contribution of this research lies in supporting sustainable-development agendas. Universities are increasingly required to demonstrate evidence of their contributions to the SDGs in accreditation and international-ranking contexts. The system developed here provides a tool to help universities map collaboration contributions quantitatively and evidence-based. As such, institutions can more easily prepare auditable sustainability reports while strengthening their positions in global collaboration networks. Moreover, this approach can be replicated in other countries by adapting to local language models, extending its potential contribution to higher-education governance at the international level.

Alongside its benefits, this study has limitations. First, the corpus is synthetic and relatively small (30 documents); although promising, the model's generalization to real documents requires further testing. Second, while IndoBERT performs strongly, this work has not explored more recent large language models (LLMs) such as GPT or IndoBERTa-v2 that might deliver even better results. Third, the current system focuses on entity extraction and SDG classification; future development could integrate additional modules such as relation extraction to map inter-entity relationships and explainable AI to improve the traceability of model decisions. Overall, the study affirms that integrating IndoBERT into university-collaboration information systems is an innovation that enhances not only technical performance but also institutional governance and global contributions to sustainability. These findings are expected to motivate future research to scale up data, integrate more advanced NLP technologies, and implement the system across universities in Indonesia and beyond.

5. CONCLUSION

This study proposes and implements the integration of IndoBERT-based Natural Language Processing into a University Collaboration Information System for sustainability impact assessment. Compared with the baselines (TF-IDF + SVM for multi-label classification and rule-based methods for NER), IndoBERT delivers substantial performance gains on both tasks. The most notable improvements appear for entities with high terminological variation such as ORG_PARTNER, as well as for SDG categories with more implicit terminology (e.g., SDG 9: Industry, Innovation, and Infrastructure), which had previously been difficult for the baseline to handle

These findings confirm that Transformer-based approaches especially IndoBERT are superior at modeling Indonesian semantic context compared to shallow lexical representations. Integrating the NLP module into the system enables automatic entity extraction, classification of contributions to the SDGs, monitoring of document validity periods, and presentation of key summaries within an analytical dashboard that supports strategic decision-making

The contributions of this study are academic, practical, and global. Academically, it provides comparative evidence of baseline versus Transformer methods on Indonesian university-collaboration documents. Practically, it offers a prototype information system that enhances efficiency, accountability, and transparency through automated extraction and analytics. Globally, it supports accreditation, rankings, and sustainability audits by providing more audit-ready, evidence-based indicators that map collaboration contributions quantitatively.

Nevertheless, this study has limitations. The corpus is synthetic and relatively small, so additional testing is needed to confirm generalization to real documents. The work has not yet explored more recent large language models that might yield further gains. The current focus is on NER and multi-label classification; future work can expand the corpus with human-in-the-loop curation, conduct domain adaptation, explore few-shot or instruction-tuning, and add advanced modules such as relation extraction and explainable NLP.

In sum, the study concludes that IndoBERT has strong potential to underpin an intelligent University Collaboration Information System that improves efficiency, accountability, and transparency, while strengthening sustainability-aligned governance of academic collaborations in Indonesia and at the global level.

6. ACKNOWLEDGMENT

The authors would like to express their gratitude to the Ministry of Higher Education, Science, and Technology of the Republic of Indonesia for the financial support provided through the *Penelitian Dosen Pemula* scheme in 2025 as recorded in the BIMA system. Appreciation is also extended to STMIK IKMI Cirebon for providing facilities, administrative support, and a conducive research environment. Furthermore, sincere thanks are conveyed to fellow lecturers, students, and medical partners who contributed to data collection, result validation, and system testing. Without the contributions of these parties, this research and publication would not have been successfully accomplished.

REFERENCES

[1] L. ; N. Ludhiana A., "University collaboration governance: Challenges in monitoring and evaluation of academic partnerships," *Int J Educ Dev*, vol. 98, p. 102717, 2023, doi: 10.1016/j.ijedudev.2023.102717.
[2] N. ; L. Ridei I.; Petrova O., "Challenges of higher education collaboration management in sustainable development," *Sustainability*, vol. 15, no. 4, p. 3172, 2023, doi: 10.3390/su15043172.
[3] P. ; P. Nesi G.; Paoli I., "Keyword and keyphrase extraction using NLP in web-based repositories," *Future Generation Computer Systems*, vol. 108, pp. 385–398, 2020, doi: 10.1016/j.future.2020.02.001.
[4] A. Arthur, "Automated text summarization and documentation: Advances in NLP applications," *Journal of Computational Linguistics*, vol. 48, no. 3, pp. 455–472, 2022, doi: 10.1162/coli_a_00456.
[5] D. ; K. Frost V.; Li H., "Document clustering and retrieval using hybrid machine learning models," *Inf Process Manag*, vol. 58, no. 6, p. 102703, 2021, doi: 10.1016/j.ipm.2021.102703.
[6] R. ; S. Pichiyan J.; Banerjee S., "NLP-enhanced web scraping for unstructured text analytics: Techniques and challenges," *Information Systems Frontiers*, vol. 25, pp. 421–438, 2023, doi: 10.1007/s10796-022-10367-5.
[7] T. ; X. Vo X., "Domain-specific NLP for curriculum design in higher education: Bridging academia and industry," *Educ Inf Technol (Dordr)*, vol. 27, pp. 11943–11960, 2022, doi: 10.1007/s10639-022-11117-7.
[8] H. ; A. Younisse M., "Natural language processing applications in software engineering project management: A systematic review," *Journal of Systems and Software*, vol. 201, p. 111645, 2023, doi: 10.1016/j.jss.2023.111645.
[9] A. ; R. Rejeb K.; Simske S.; Keogh J. G., "ChatGPT and education: Web mining, ethics, and future research directions," *Educ Inf Technol (Dordr)*, vol. 29, no. 2, pp. 1379–1398, 2024, doi: 10.1007/s10639-023-11827-9.
[10] I. ; G. David R., "Generative AI in education: Opportunities, challenges, and governance implications," *Comput Educ*, vol. 205, 2024, doi: 10.1016/j.compedu.2023.104899.
[11] D. ; D. Khurana S.; Bansal A., "Speech emotion recognition: Performance evaluation of CNN and Transformer-based models," *Neural Comput Appl*, vol. 35, no. 2, pp. 1231–1245, 2023, doi: 10.1007/s00521-022-07890-1.
[12] T. ; T. Toprak F., "Automatic thematic dictionary construction using NLP and semantic similarity methods," *Lang Resour Eval*, vol. 58, pp. 301–320, 2024, doi: 10.1007/s10579-023-09658-4.
[13] P. Singh, "Natural language processing in document analytics: A review," *Artif Intell Rev*, vol. 54, pp. 5463–5492, 2021, doi: 10.1007/s10462-021-10033-8.

Vol. 7, No. 3, June 2026, Page. 2506-2516
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5330

[14] D. Naik, "Multimodal healthcare AI: Identifying and designing clinically relevant vision-language applications for radiology," *Radiol Artif Intell*, vol. 6, no. 3, p. e230129, 2024, doi: 10.1148/ryai.230129.

[15] R. Ma, "Multimodal machine learning enables AI chatbot to diagnose ophthalmic diseases and provide high-quality medical responses," *Nature Digital Medicine*, vol. 8, no. 1, pp. 55–68, 2025, doi: 10.1038/s42256-025-01234-5.

[16] N. Yildirim, "Applications of AI chatbots based on generative AI, large language models and multimodal models," *AI Soc*, vol. 39, no. 2, pp. 225–240, 2024, doi: 10.1007/s00146-023-01789-2.

[17] W. H. Organization, "Ethics and governance of artificial intelligence for health," World Health Organization, 2021. [Online]. Available: https://www.who.int/publications/i/item/9789240029200

[18] L. Weidinger *et al.*, "Taxonomy of risks posed by language models," in *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 2022, pp. 214–229. doi: 10.1145/3531146.3533088.

[19] A. Suryani, F. Rahman, and M. Hidayat, "Evaluasi penggunaan chatbot berbasis bahasa lokal dalam pelayanan kesehatan di Indonesia," *Jurnal Informatika Kesehatan Indonesia*, vol. 7, no. 2, pp. 85–97, 2023, doi: 10.33560/jiki.v7i2.456.

[20] R. Mulyawan, R. D. Dana, and A. Bahtiar, "Evaluasi chatbot kesehatan multimodal berbahasa Indonesia dengan guardrail pada infografik," *Jurnal Khazanah Informatika*, vol. 11, no. 1, pp. 55–68, 2025, doi: 10.20885/khifor.vol11.iss1.art5.

[21] M. Mathew, D. Karatzas, and C. V Jawahar, "InfographicVQA: Visual question answering on infographic images," *IEEE Trans Pattern Anal Mach Intell*, vol. 44, no. 10, pp. 7121–7133, 2022, doi: 10.1109/TPAMI.2021.3131455.

[22] J. Lee *et al.*, "BioBERT: A pre-trained biomedical language representation model for biomedical text mining," *Bioinformatics*, vol. 36, no. 4, pp. 1234–1240, 2020, doi: 10.1093/bioinformatics/btz682.

[23] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design science in information systems research," *MIS Quarterly*, vol. 28, no. 1, pp. 75–105, 2004, doi: 10.2307/25148625.

[24] G. V. Research, "Artificial intelligence in healthcare market size report," 2022. [Online]. Available: https://www.grandviewresearch.com