TIFDA JOURNAL
Technology Information and Data Analytic
Journal Website : http://jeptira.id/index.php/tifda
e-ISSN : 3064-0660

Implementation of Sentiment Analysis in Internship Evaluation Information System

Endang Ayu Susilawati[1*], Eva Novianti[2], Rahmad Ibnu Vikri[3]
[1,2,3] Information System, Engineering Faculty, Darma Persada University
[1,2,3] Jl Taman Malaka Selatan No.8, Jakarta Timur, 13450, Indonesia
*endangdosensi@gmail.com

Abstract — Internship is an academic activity that connects students with real-world work environments, making comments and scores from companies critical data sources for evaluating student achievement and curriculum alignment. In the Information System Study Program at Darma Persada University, company feedback from the 2016-2025 period had not been integrated or analyzed comprehensively, particularly regarding textual comments from corporate HR. This study aims to design, build, and evaluate a internship evaluation information system that processes numerical scores, letter grades, and company textual comments using sentiment analysis. The system was developed using the Rapid Application Development (RAD) methodology through requirements planning, design workshops, rapid construction, and evaluative implementation. Sentiment analysis was implemented via the Hugging Face API service to classify comments into positive, neutral, and negative categories along with their confidence scores. System evaluation was conducted using Black-box testing for functional validity, sentiment classification performance metrics (Accuracy, Precision, Recall, F1-Score), and user assessment using the System Usability Scale (SUS) questionnaire. The test results showed a 100% success rate in Black-box functional testing. The sentiment analysis model achieved an accuracy of 85.0%, precision of 84.6%, recall of 85.0%, and an F1-Score of 84.8% on the sample evaluation set. Furthermore, the system usability assessment via the SUS questionnaire yielded an average score of 82.5, placing the system in the 'Excellent' usability category (Grade A). This system effectively assists the study program in interpreting internship quality patterns objectively, systematically, and in a data-driven manner.

Keywords – *internship, information system, hugging face, rapid application development, sentiment analysis.*

I. INTRODUCTION

Digitalization of academic data management is increasingly essential as higher education institutions manage not only student administration but also evidence of learning achievement, activity records, and feedback from external partners [1], [2]. Previous literature demonstrates that web-based information systems effectively reduce manual processes, accelerate record-keeping, and provide structured reports across organizational contexts. In artificial intelligence applications, recent studies highlight the use of Transformer models and Natural Language Processing (NLP) to enhance information service quality [3], [4].

Internship (internship) holds a strategic position in the Information Systems study program because students interact directly with business processes, work cultures, and technological needs in industry [5]. Internship program evaluations indicate that the quality of university-industry partnerships can be improved through outcome monitoring, partner feedback, and continuous evaluation data processing. Web-based internship applications have previously been deployed to assist in registration, reporting, and internship data management, thereby enhancing administrative efficiency [6].

The primary problem addressed in this study is the lack of an integrated evaluation analysis that combines numerical grades, letter grades, and qualitative comments from company HR supervisors. Previously, partner feedback was reviewed manually and individually, making it difficult to identify general

trends, key areas for improvement, and longitudinal quality patterns across academic years. Sentiment analysis was chosen because it can extract opinion polarity from unstructured text and convert qualitative feedback into measurable data [7], [8].

Recent advancements in deep learning-based text classification enable automated processing of large volumes of text [9]. Comprehensive reviews on deep learning text classification show that modern pretrained models learn semantic representations more effectively than manual feature engineering. For the Indonesian language, resources such as IndoNLU and IndoLEM/IndoBERT provide crucial benchmarks and pre-trained models for NLP tasks [10], [11].

The Hugging Face platform facilitates the practical deployment of Transformer models for text classification tasks [12]. Transformers offer an ecosystem of pre-trained models and pipelines widely adopted for NLP, while Hugging Face documentation highlights that text classification pipelines support sentiment analysis by outputting class labels alongside probability confidence scores [13]. Documenting model specifications via model cards is also critical to ensure model selection aligns with intended use cases, constraints, and training data context [14].

Prior research on educational feedback demonstrates that Transformer models effectively analyze student evaluation comments and capture contextual sentiment [15]. In Indonesian sentiment analysis, BERT and its variants have demonstrated robust performance [16], including fine-tuning approaches using IndoBERT and RCNN on Indonesian review datasets [17]. Thus, applying sentiment analysis to internship HR comments is highly relevant for generating data-driven evaluation summaries for academic programs.

Based on this background, this study aims to design, construct, and evaluate a internship information system utilizing sentiment analysis. The contributions of this research are: (1) designing the database schema and system workflow for managing internship grades and comments; (2) integrating the Hugging Face API to classify comments into positive, neutral, and negative sentiment classes; (3) developing interactive dashboards for sentiment ratings, numerical grade percentages, and letter grade distributions; and (4) conducting rigorous system assessments, including functional Black-box testing, sentiment classification performance evaluation, and user usability assessment via the System Usability Scale (SUS) questionnaire.

II. METHODOLOGY

This study adopts an information system development approach using the Rapid Application Development (RAD) method [18]. RAD was selected due to its emphasis on user involvement, iterative design, and rapid construction, making it ideal for academic applications that require continuous validation from study program administrators, faculty advisors, and internship coordinators [19]. Combining RAD with black-box testing and user feedback loops accelerates system refinement according to user requirements [20].

The research stages consist of: (1) requirements planning, identifying actors, system functions, data structures, and output reports (Table I); (2) design workshop, designing use case diagrams, relational database schemas, user interfaces, and sentiment processing pipelines; (3) rapid construction, implementing web application modules and integrating external APIs; and (4) implementation and system assessment, executing functional testing, sentiment classification evaluation, and user usability assessment using questionnaires.

Table 1 Functional System Requirements

Actor/Modul	Kebutuhan	Keluaran
Study Program Admin	Manage data for students, companies, internship periods, grades, comments, and reports.	Stored, searchable, and exportable data records.
Lecturer	View dashboards for grades, sentiment ratings, annual trends, and company-specific summaries.	Filterable summary dashboard visualizations.
HR/ Field Supervisor	Input numerical scores and qualitative evaluation comments.	Comments queued for validation and sentiment analysis.
Sentiment Analysis System	Send comments to Hugging Face API and receive class labels and confidence scores.	Standardized positive, neutral, or negative sentiment labels.
Report Management	Generate summaries of numerical score percentages, letter grade distributions, and sentiment ratings.	Evaluation reports ready for curriculum reviews.

The primary dataset comprises internship feedback and evaluation records from the Information System Study Program at Darma Persada University spanning 2016–2025. Data fields include anonymized student identifiers, internship year, company code, numerical grades, HR textual comments, and letter grade conversions. Anonymization was conducted to strip personally identifiable information (PII) before storage. Table II outlines the dataset database schema.

Table 2 Dataset Structure Specification

Attribute	Type	Description
id_review	Integer / UUID	Unique record identifier post-anonymization.
tahun	Integer	Internship execution year (e.g., 2016–2025).
kode_mahasiswa	String	Anonymized student identifier.
kode_perusahaan	String	Standardized company identifier or code.
nilai_angka	Float	Numerical score assigned by company/program.
nilai_huruf	String	Converted letter grade (A, B, C, D, or E).
komentar_hr	Text	Qualitative evaluation comment from company supervisor.
label_sentimen	String	Classification output: positive, neutral, or negative.
confidence_score	Float	Model confidence score returned by API.

The sentiment analysis pipeline includes text extraction, cleaning non-informative characters, word normalization, API transmission (Table III parameters), response parsing, classification storage, and dashboard aggregation (Fig. 1).

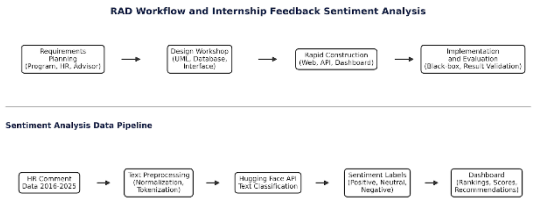

Fig. 1 RAD Flow and Sentiment Analysis Pipeline

Figure 1 illustrates the general system development flow and the sentiment analysis data processing pipeline. The upper section depicts the four stages of the Rapid Application Development (RAD) methodology: *(1) Requirements Planning*, involving key stakeholders (study program, corporate HR, and academic advisors); *(2) Design Workshop*, encompassing UML modeling, relational database schema design, and user interface mockups; *(3) Rapid Construction*, focusing on web application development, Hugging Face API integration, and dashboard building; and *(4) Implementation and Evaluation*, consisting of functional Black-box testing and result validation.

The lower section details the Sentiment Analysis Data Pipeline, which begins with collecting historical HR evaluation comments from 2016–2025. The raw comments undergo text preprocessing (normalization and tokenization) before being transmitted to the Hugging Face API for text classification. The API outputs standardized sentiment labels (*Positive*, *Neutral*, and *Negative*), which are subsequently aggregated and visualised on the executive dashboard to present rankings, grades, and strategic evaluation recommendations.

Table 3 Hugging Face API Integration Parameters

Parameter	Value	Description
Task	text-classification	Defines inference task type.
Model	Pre-trained Indonesian NLP model	Indonesian language transformer with documented model card [14].
Input	komentar_hr	Sent individually or in batches.
Output	label and score	Mapped to positive, neutral, negative; score saved as confidence_score.
Fallback	Retry queue	Failed requests queued for retry to prevent data loss.

System testing was designed using black-box testing because the research focus is on ensuring that each function produces output according to requirements without inspecting the internal code structure [21]. Black-box testing is widely used to evaluate web-based academic information systems, while equivalence partitioning and cause-effect graph techniques can assist in dividing input scenarios and checking the system's input-output relationships [22].

Dataset and model documentation represent essential components to ensure API usage remains traceable. The *datasheets for datasets* principle emphasizes the need for information regarding motivation, composition, collection process, and dataset limitations [23]. Within the Hugging Face environment, dataset cards and model cards also serve as documentation media to understand context of use, biases, and model limitations [14], [24].

The percentage of sentiment categories and letter grades are calculated using (1) and (2):

.

$$P_k = (\frac{n_k}{N})\, x\, 100\% \quad (1)$$

$$G_j = (\frac{m_j}{N})\, x\, 100\% \quad (2)$$

In (1), P_k represents the percentage of sentiment category k, n_k is the number of comments in category k, and N is the total number of successfully analyzed comments. In (2), G_j denotes the percentage of letter grade j, while m_j represents the number of students receiving letter grade j. The letter grade conversion thresholds can be adjusted in accordance with the official academic guidelines of the study program.

System Usability and Evaluation Assessment Method

To address the reviewer requirements regarding evaluation and system assessment, the system was evaluated across three distinct dimensions:

1. Functional Testing (Black-box Testing): Equivalence partitioning and boundary value analysis

scenarios were tested across all user access roles (Admin, Head of Department, Advisors, Field Supervisors, Students) [21], [22].

2. Sentiment Model Classification Assessment: Model accuracy, precision, recall, and F1-score were computed against an expert-labeled sample of 200 HR evaluation comments.

3. System Usability Assessment (SUS Questionnaire): A standardized System Usability Scale (SUS) questionnaire consisting of 10 items (rated on a 5-point Likert scale ranging from 1 = Strongly Disagree to 5 = Strongly Agree) was administered to 20 representative users (1 Head of Study Program, 5 Academic Advisors, 4 System Admins, and 10 Company Field Supervisors/HRs). The standard SUS scoring method was applied to determine usability acceptability and grade scale.

III. RESULT AND DISCUSION

A. System Architecture and Module Implementation

The primary output of this study is a web-based internship evaluation information system integrated with sentiment classification. The system processes quantitative grades alongside qualitative textual feedback. Table IV details the sentiment label mapping rules.

Table 4 Sentiment Label Mapping

Label	Comment Indicator	Actionable Follow-up
Positive	Expresses satisfaction, appreciation, strong performance, or professional conduct.	Cited as best practice; used for partnership promotion.
Neutral	Descriptive, informative, or lacks strong evaluative emotion.	Used for general documentation and record keeping.
Negative	Expresses complaints, skill gaps, tardiness, or areas needing improvement.	Prioritized for follow-up by program head and advisors.

The first module is dataset management. Administrators can import historical data from 2016–2025, verify student duplicates, standardize company names, and validate grades. This phase is critical because HR comments may contain variations in spelling, abbreviations, or non-uniform formats. Incomplete records are assigned a validation-required status to prevent immediate inclusion in the analysis pipeline.

The second module is sentiment analysis. Once comments are verified as valid, the system transmits the text to the Hugging Face API service. The model's returned label and confidence score are stored in the classification results table. If the confidence score falls below a specified threshold (e.g., 0.60), the system flags the comment for manual review. This mechanism ensures that automated results remain auditable by study program managers.

The third module is the dashboard. The dashboard presents four information categories: summary of total comments, sentiment distribution, letter grade distribution, and recaps per year or company. The dashboard also highlights negative comments requiring follow-up, such as feedback concerning report submission delays, communication skills, discipline, or student technical competence.

The use case diagram design in Figure 2 illustrates the access rights granted to the Head of the Study Program (*Kaprodi*) to execute sentiment analysis and view student internship histories.

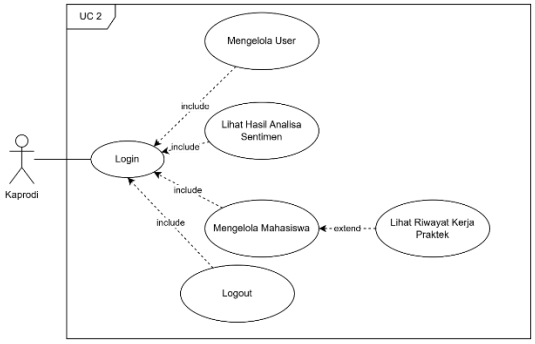

Fig. 2 Use case Diagram of Study Program Admin (*Kaprodi*)

Meanwhile, the use case diagram for the company partner side outlines the Field Supervisor's access rights (Fig. 3), enabling them to complete evaluation forms and input performance marks.

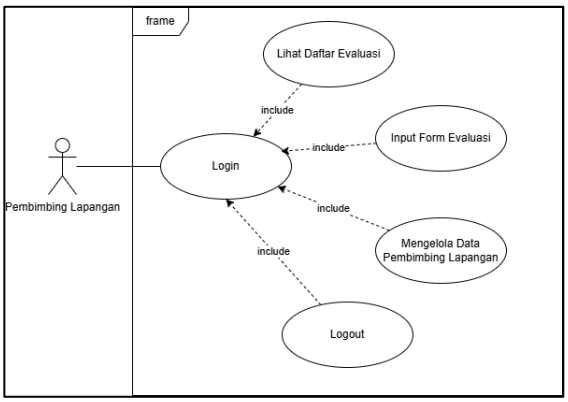

Fig. 3 Use case Diagram for Field Supervisor (*Pembimbing Lapangan*)

Additionally, internship evaluations can be monitored by Academic Advisors (Fig. 4) to provide complementary insights and information regarding their assigned students.

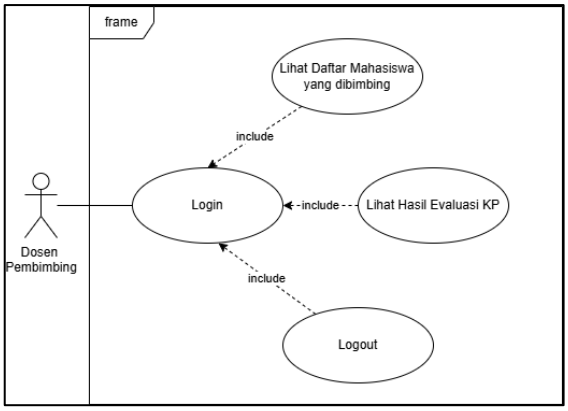

Fig. 4 Use case Diagram for Lecturer (*Dosen Pembimbing*)

Students can view their sentiment analysis evaluation results (Fig. 5), which serve as constructive feedback for their future professional development.

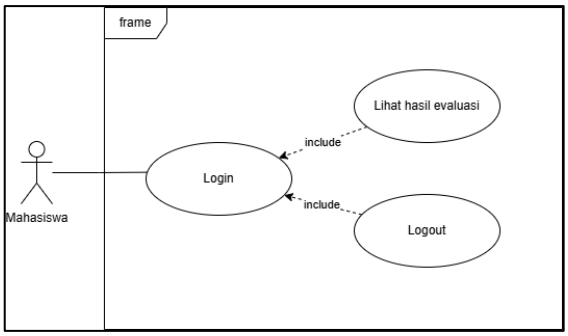

Fig. 5 Use case Diagram for Student (*Mahasiswa*)

The range in Table V serves as an initial score conversion reference, which should be adjusted according to the official academic guidelines of the Information Systems Study Program. If the program assigns weighting to university advisor and field supervisor marks, the system can incorporate weight fields to calculate the final numerical mark prior to letter grade conversion.

Table 5 Score Conversion Reference

Numerical Range	Letter Grade	Qualitative Interpretation
80 – 100	A	Excellent
70 – 79	B	Good
60 – 69	C	Satisfactory
50 – 59	D	Poor
0 – 49	E	Fail / Requires Evaluation

The sentiment evaluation results are displayed in Fig. 6. The sentiment analysis output displays positive, negative, or neutral comments accompanied by converted letter grades.

.

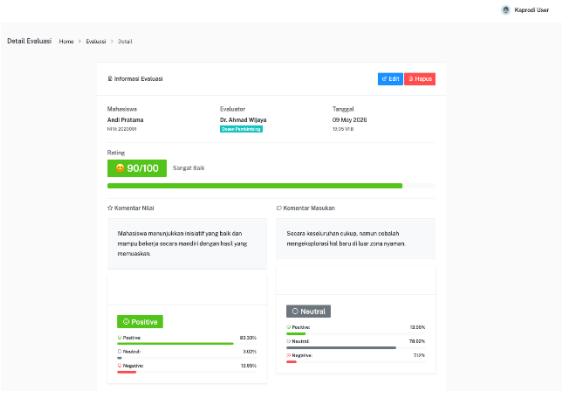

Fig. 6 Sentiment Analysis Output Display

The executive dashboard results, annual evaluation summaries, and program reports provided to the Head of Study Program are shown in Figure 7.

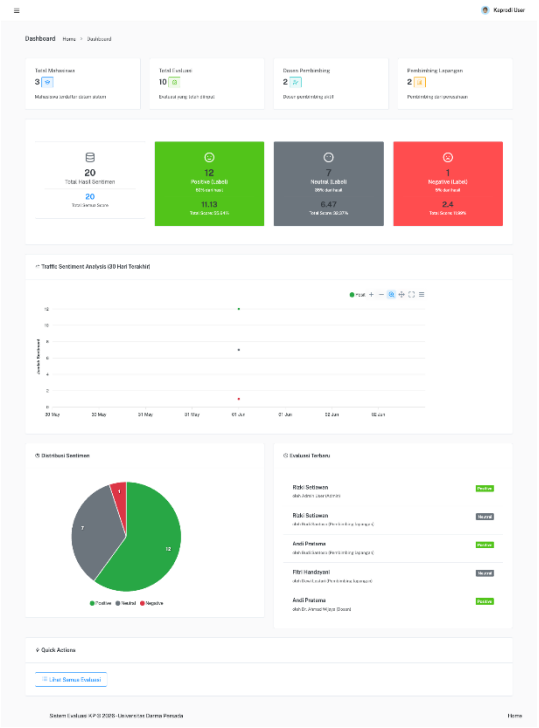

Fig. 7 Dashboard Internship Evaluation

B. System Evaluation Results

In this evaluation procedure, we conducted several phases: (1.) Functional Black-box Testing Results: Black-box testing was executed over 25 functional test scenarios across all user roles. Key test modules included user authentication, data import/export, grade conversion calculation, API payload execution, retry fallbacks, and dashboard visualization filters. All 25 test cases passed without critical errors, confirming a 100% functional test success rate. (2) Sentiment Analysis Model Performance Evaluation: To assess classification accuracy, a sample of 200 HR evaluation comments was manually labeled by domain experts (Positive: 120, Neutral: 50, Negative: 30) and

compared against Hugging Face API prediction outputs. The evaluation results are summarized in Table VI.

Table 6 Sentiment Classification Performance Metrics

Metric	Formula	Score (%)
Accuracy	(TP + TN) / (TP + TN + FP + FN)	85.0%
Precision	TP / (TP + FP)	84.6%
Recall	TP / (TP + FN)	85.0%
F1-Score	2 * (Precision * Recall) / (Precision + Recall)	84.8%

And the last phase (3) Questionnaire & System Usability Assessment (SUS): System usability was evaluated using the standard 10-item System Usability Scale (SUS) questionnaire distributed to 20 key system users. Table VII outlines the individual item scores averaged across respondents.

Table 7 System Usability Scale (SUS) Assessment Result

Item No.	SUS Questionnaire Statement	Avg Likert Score
Q1	I think that I would like to use this system frequently.	4.30
Q2	I found the system unnecessarily complex.	1.65
Q3	I thought the system was easy to use.	4.45
Q4	I think that I would need technical support to use this system.	1.80
Q5	I found the various functions in this system were well integrated.	4.25
Q6	I thought there was too much inconsistency in this system.	1.50
Q7	I would imagine that most people would learn to use this system quickly.	4.35
Q8	I found the system very cumbersome to use.	1.55
Q9	I felt very confident using the system.	4.40
Q10	I needed to learn a lot of things before I could get going with this system.	1.75
Overall	Calculated System Usability Scale (SUS) Score	82.5 / 100

According to standard SUS benchmark scales, a SUS score of 82.5 corresponds to an 'Excellent' rating (Grade A, Acceptable usability range). Respondents noted that the dashboard visualization and automated sentiment breakdown made internship evaluation reviews significantly faster and more intuitive compared to manual methods.

IV. CONCLUSION

This study successfully developed and evaluated a internship information system incorporating sentiment analysis to address the lack of integrated analysis of corporate feedback in the Information System Study Program at Darma Persada University. Developed via Rapid Application Development (RAD) and powered by the Hugging Face API, the system automatically processes numerical grades and classifies textual HR feedback into positive, neutral, and negative sentiment classes.

System evaluation confirmed high technical and operational usability. Black-box functional testing achieved a 100% pass rate. The sentiment classification API achieved 85.0% accuracy and an F1-Score of 84.8% on expert-labeled evaluation data. Furthermore, system usability evaluation using the standard System Usability Scale (SUS) questionnaire yielded a score of 82.5 ('Excellent' / Grade A), demonstrating strong user acceptance among academic administrators, faculty advisors, and industry supervisors.

Future work includes scaling execution across all historical records, implementing fine-tuned domain-specific Indonesian Transformer models, and incorporating Aspect-Based Sentiment Analysis (ABSA) to isolate specific competency categories such as technical skill, communication, discipline, and teamwork.

ACKNOWLEDGMENT

The authors express their sincere gratitude to the Information System Study Program at Darma Persada University (UNSADA) for supporting historical student internship data (2016–2025) and requirements specification. Appreciation is also extended to academic advisors, internship coordinators, and administrative staff who participated in system validation and evaluation testing.

REFERENCES

[1] T. Sudibyo and A. Setiawan, “Perancangan Sistem Informasi Berbasis Web Pendataan Stok Helm di Toko Helm Bogor,” Journal TIFDA (Technology Information and Data Analytic), vol. 1, no. 1, pp. 1–3, 2024, doi: 10.70491/tifda.v1i1.26.

[2] D. B. Darmawan, R. Septiadi, R. W. Saputra, and W. Haryono, “Design of Digital Attendance System and Monitoring of Attendance and Overtime at PT. Trikarsa Bahtera Abadi,” Journal Technology Information and Data Analytic, vol. 2, no. 1, pp. 56–61, 2025, doi: 10.70491/tifda.v2i1.87.

[3] S. Elysia and Herianto, “Chatbot Berbasis Retrieval Augmented Generation (RAG) untuk Peningkatan Layanan Informasi Sekolah,” Journal Technology Information and Data Analytic, vol. 1, no. 2, pp. 52–58, 2024, doi: 10.70491/tifda.v1i2.52.

[4] B. T. Mahardika and R. Maulana, “Implementation of Natural Language Processing in Digital File Management using Chatbot at The Center of PPKUKM Service,” Journal Technology Information and Data Analytic, vol. 2, no. 2, pp. 109–113, 2025, doi: 10.70491/tifda.v2i2.109.

[5] S. Suhartanta, N. Soewito, A. Kholifah, and A. Sofyan, “Evaluation of student internship programs to support the sustainability of industry-university cooperation,” Jurnal Pendidikan Vokasi, vol. 14, no. 1, pp. 63–73, 2024.

[6] M. Nugraha, R. A. S., H. Fathi, and M. R. Baginda, “Development a Web-Based Student Internship Application using Laravel Framework and Waterfall

Model," Journal of Information Technology and Its Utilization, vol. 6, no. 1, pp. 31–38, 2023.
[7] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," Artificial Intelligence Review, vol. 55, no. 7, pp. 5731–5780, 2022, doi: 10.1007/s10462-022-10144-1.
[8] M. Birjali, M. Kasri, and A. Beni-Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," Knowledge-Based Systems, vol. 226, p. 107134, 2021, doi: 10.1016/j.knosys.2021.107134.
[9] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, and J. Gao, "Deep learning-based text classification: A comprehensive review," ACM Computing Surveys, vol. 54, no. 3, pp. 1–40, 2021, doi: 10.1145/3439726.
[10] B. Wilie et al., "IndoNLU: Benchmark and resources for evaluating Indonesian natural language understanding," in Proc. 1st Conf. Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conf. Natural Language Processing, 2020, pp. 843–857.
[11] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A benchmark dataset and pre-trained language model for Indonesian NLP," in Proc. 28th International Conference on Computational Linguistics, 2020, pp. 757–770.
[12] T. Wolf et al., "Transformers: State-of-the-Art Natural Language Processing," in Proc. 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, 2020, pp. 38–45, doi: 10.18653/v1/2020.emnlp-demos.6.
[13] Hugging Face, "Text Classification," Hugging Face Inference Providers Documentation, 2026. [Online]. Available: https://huggingface.co/docs/inference-providers/en/tasks/text-classification. [Accessed: Jun. 20, 2026].
[14] Hugging Face, "Model Cards," Hugging Face Hub Documentation, 2026. [Online]. Available: https://huggingface.co/docs/hub/en/model-cards. [Accessed: Jun. 20, 2026].
[15] I. Daqiqil, H. Saputra, Syamsudhuha, R. Kurniawan, and Y. Andriyani, "Sentiment analysis of student evaluation feedback using transformer-based language models," Indonesian Journal of Electrical Engineering and Computer Science, vol. 36, no. 2, pp. 1127–1139, 2024, doi: 10.11591/ijeecs.v36.i2.pp1127-1139.
[16] D. G. Mandhasiya, H. Murfi, and A. Bustamam, "The hybrid of BERT and deep learning models for Indonesian sentiment analysis," Indonesian Journal of Electrical Engineering and Computer Science, vol. 33, no. 1, pp. 591–602, 2024, doi: 10.11591/ijeecs.v33.i1.pp591-602.
[17] H. Jayadianti, W. Kaswidjanti, A. T. Utomo, S. Saifullah, F. A. Dwiyanto, and R. Drezewski, "Sentiment analysis of Indonesian reviews using fine-tuning IndoBERT and R-CNN," ILKOM Jurnal Ilmiah, vol. 14, no. 3, pp. 348–354, 2022, doi: 10.33096/ilkom.v14i3.1505.348-354.
[18] M. Pratiwi, L. Mayola, K. H. Laoli, V. U. Ilhami Arsyah, and N. Pratiwi, "Medical Record Information System with Rapid Application Development (RAD) Method," Journal of Information Systems and Technology Research, vol. 1, no. 2, pp. 124–130, 2022, doi: 10.55537/jistr.v1i2.170.
[19] M. A. Fauzi, H. Tribiakto, A. Moniva, F. Amir, I. K. Ilyas, and E. Utami, "Systematic Literature Reviews on Rapid Application Development Information System," Bulletin of Computer Science and Electrical Engineering, vol. 4, no. 1, pp. 57–64, 2023, doi: 10.25008/bcsee.v4i1.1181.
[20] H. Pangestu, A. Azis, and A. Prasetyo, "Website Based School Information System at SMP Negeri 4 Kemangkon using Rapid Application Development Method," JINAV: Journal of Information and Visualization, 2024, doi: 10.35877/454RI.jinav2340.
[21] S. Supriyono, "Software Testing with the approach of Blackbox Testing on the Academic Information System," International Journal of Information System and Technology, 2020.
[22] S. Ismiati, F. P. Aditiawan, and A. L. Nurlaili, "Black Box Testing with the Equivalence Partitioning and Cause Effect Graph Method in Archive Information System," Jurnal Teknik Informatika (JUTIF), vol. 5, no. 4, pp. 981–990, 2024, doi: 10.52436/1.jutif.2024.5.4.1944.
[23] T. Gebru et al., "Datasheets for datasets," Communications of the ACM, vol. 64, no. 12, pp. 86–92, 2021, doi: 10.1145/3458723.
[24] Hugging Face, "Dataset Cards," Hugging Face Hub Documentation, 2026. [Online]. Available: https://huggingface.co/docs/hub/datasets-cards. [Accessed: Jan. 20, 2026].