Idealis: Indonesia Journal Information System
Volume 9, Nomor 2, Juli 2026
ISSN 2684-7280 (*online*)
Halaman 259-267
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

Intelligent One-Gate System Based on Natural Language Processing for Enhancing Academic Information Services

Jaka Purnama[1*], Yayuk Ike Meilani[2]

[1,2]Information Systems, PalComTech Institute of Technology and Business, Palembang, Indonesia
Email: [1*]jaka_purnama@palcomtech.ac.id, [2]yayuk_ike@palcomtech.ac.id
(*: *corresponding author*)

Abstract - The rapid digitalization of higher education requires academic information services that are fast, integrated, and accessible. Fragmented service channels, high administrative workloads, and slow response times remain persistent operational challenges. This study develops an Intelligent Transformer-Based One-Gate System that centralizes academic services by integrating text classification, abstractive summarization, and chatbot modules. Using a Research and Development approach, the system was built from 1,000 academic documents and service queries collected from FAQs, academic regulations, guides, and digital service archives. The corpus was cleaned, tokenized, encoded, and divided into training, validation, and testing subsets. BERT was applied for document and query classification, while BART and PEGASUS were evaluated for abstractive summarization using ROUGE metrics. The chatbot was assessed through a Likert-scale user acceptance survey. The results of this research show that the BERT classifier achieved 88% accuracy and an F1-score of 0.875, PEGASUS outperformed BART with ROUGE-1 = 0.74, ROUGE-2 = 0.66, and ROUGE-L = 0.71, and the chatbot achieved an average user satisfaction score of 82%. The main contribution of this research is an integrated transformer-based one-gate architecture that combines document routing, academic document summarization, and conversational assistance in a single service platform, offering practical value for reducing fragmented academic information services and methodological value as a reference model for higher education NLP implementation.

Keywords: Academic Information Services, Chatbot, Natural Language Processing, One-Gate System, Transformer-based Models.

1. INTRODUCTION

Digital transformation in higher education has significantly changed how academic services are delivered. Traditional service models are often fragmented across multiple administrative units, resulting in inefficiencies, delays, and increased workload for staff. Students face difficulties in obtaining accurate and timely academic information, which negatively affects user satisfaction and institutional performance [1]. In the current digital era, the development of information technology has a very significant influence on various aspects of human life[2].

The concept of a one-gate system has been introduced to centralize service delivery by integrating multiple channels into a single access point. This approach enhances accessibility and reduces redundancy in administrative tasks, supporting the development of smart campuses [3]. When combined with Natural Language Processing (NLP), one-gate systems can be further optimized to understand and respond to user queries in natural language. NLP technologies, such as text classification, document summarization, and chatbots, have been widely applied in educational contexts to automate information retrieval and improve service efficiency [4],[5].

Recent advancements in transformer-based models such as BERT, GPT, and T5 have shown remarkable improvements in natural language understanding, particularly for classification, summarization, and question-answering tasks [6], [7]. These capabilities make transformer models suitable for academic service environments, where user requests often contain domain-specific terminology, ambiguous intents, and long policy documents.

These models enable systems to capture semantic meaning more effectively, which is crucial in handling complex academic queries. Studies in educational institutions have reported that NLP-based service systems can reduce administrative workload by up to 40% while significantly improving student satisfaction [8]. Artificial Intelligence (AI) is a constantly evolving field of research, with computer vision as one of its branches that studies the ability of computers to recognize and interpret observed objects[9].

This research differs from previous studies in three main aspects. First, earlier work generally focused on a single NLP function, such as chatbot-based student consultation [10] or transformer-based long document classification [11]. Second, those studies did not integrate document routing, automatic summarization, and conversational response generation into one operational service gateway. Third, the proposed system is designed for academic information services, where administrative users need a unified platform rather than several separate applications. Therefore, the research gap addressed in this study is the absence of an integrated NLP-based one-gate framework that simultaneously supports document classification, summarization, and real-time academic inquiry handling. Based on this gap, the contribution of this study is the development and evaluation of an Intelligent One-Gate System that combines text classification, auto-summarization, and chatbot interaction in a single academic service ecosystem. The system employs transformer-based language models, namely BERT, BART, and PEGASUS, to support automatic document routing, concise policy summarization, and timely

Idealis: Indonesia Journal Information System
Volume 9, Nomor 2, Juli 2026
ISSN 2684-7280 (*online*)
Halaman 259-267
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

responses to routine academic questions. Practically, the system is expected to reduce repetitive administrative work, accelerate information delivery, and improve the accessibility of academic services for students and staff. The explicit benefit of the proposed system is that students receive faster and more consistent academic information, while administrative staff can reduce repetitive manual responses and focus on more complex service cases.

Therefore, this study introduces an Intelligent One-Gate System powered by NLP to address the challenges of fragmented academic information services. The system integrates three core modules: text classification to organize documents, auto-summarization to condense lengthy content, and a chatbot to provide real-time responses. The aim is to improve service accuracy, efficiency, and user experience while supporting the broader digital transformation agenda in higher education.

2. RESEARCH METHOD

This study adopts a Research and Development (R&D) approach to construct and evaluate the Intelligent One-Gate System. As depicted in the system architecture, the methodology is operationalized through a continuous pipeline structured around an Input-Process-Output paradigm, encompassing data collection, preprocessing, feature engineering, machine learning modeling, and system integration. The workflow is illustrated in Figure 1.

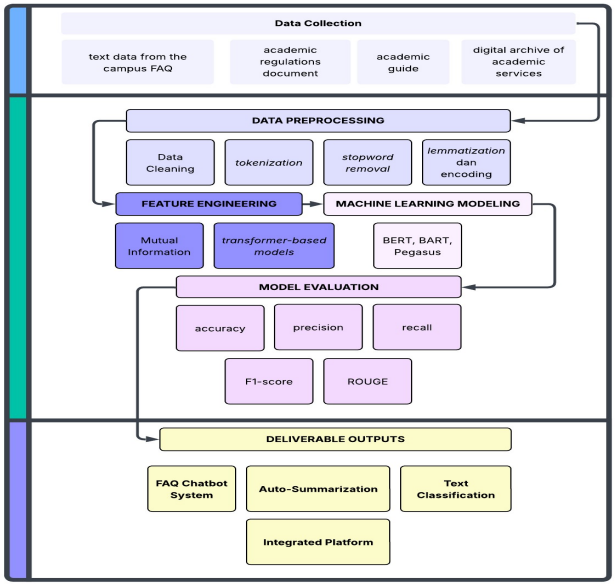

Fig 1. Research and Development (R&D) Methods

Figure 1 illustrates the end-to-end architecture of the proposed NLP-based Intelligent One-Gate system. The process starts with data collection from multiple academic sources, including FAQs, regulations, guidelines, and digital service archives, forming a heterogeneous corpus. The data then undergoes preprocessing involving cleaning, tokenization, stopword removal, lemmatization, and encoding to produce structured inputs for modeling. Feature engineering combines mutual information and transformer-based representations to capture semantic relationships. The modeling stage employs BERT for classification and BART–PEGASUS for abstractive summarization. Outputs are evaluated using accuracy, precision, recall, F1-score, and ROUGE metrics. The final outputs are integrated into a chatbot, auto-summarization, and text classification system within a unified academic service platform, enabling automated classification, summarization, and interaction..

2.1. Data Collection and Corpus Construction.

This research relies on a comprehensive, domain-specific corpus constructed from heterogeneous textual sources within the university's digital ecosystem. The primary objective of this phase was to capture a wide spectrum of real-world academic interactions alongside formal institutional guidelines. A robust, domain-specific dataset is critical for training context-aware NLP models, particularly in higher education environments where administrative terminology is highly specialized [12]. The data was systematically acquired from four primary institutional repositories:

Idealis: Indonesia Journal Information System
Volume 9, Nomor 2, Juli 2026
ISSN 2684-7280 (*online*)
Halaman 259-267
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

a. Campus FAQ Text Data: This comprises historical question-and-answer pairs extracted from the university's official student portal. Examples of this data include routine procedural queries (e.g., "What is the procedure for proposing a leave of absence?", "How do I reset my academic portal password?") and their corresponding official answers.
b. Academic Regulations Documents: Formal institutional decrees and policy documents were collected to provide the models with authoritative knowledge. These documents typically involve complex, formal Indonesian/English phrasing, detailing policies such as graduation prerequisites, academic sanctions, and credit transfer rubrics.
c. Academic Guides: This source includes annual student handbooks and faculty guidebooks containing structured information regarding curriculum mapping, syllabus descriptions, and academic timelines.
d. Digital Archives of Academic Services: To capture authentic user intent and informal phrasing, anonymized textual logs were retrieved from the university's administrative ticketing system. This includes direct correspondences and emails from students seeking assistance.

To construct a unified dataset from these heterogeneous sources, a systematic extraction pipeline was employed. Unstructured formats, such as PDF regulation documents and web-based HTML FAQs, were parsed and converted into raw textual formats using optical character recognition (OCR) and web scraping techniques. From this raw repository, a representative sample of 1,000 distinct textual entries was curated. To facilitate the supervised learning aspect of the Text Classification module, these entries were manually annotated into five predefined categories: academic regulations, student administration, finance, facilities, and general information. The dataset was divided into 70% training data, 15% validation data, and 15% testing data, so that model development, hyperparameter tuning, and final evaluation used clearly separated subsets. For reproducibility, all experiments were conducted using a fixed random seed and the same train-validation-test split across classification and summarization tasks. Model performance was reported using accuracy, precision, recall, F1-score, and ROUGE metrics according to the task evaluated.

2.2. Data Preprocessing Pipeline

The raw textual corpus generated in the data collection phase contained significant syntactic noise, formatting inconsistencies, and extraneous characters inherent to web-scraped data and parsed PDFs. To transform this unstructured text into a machine-readable format, a rigorous sequential preprocessing pipeline was implemented. This stage is critical because modern Natural Language Processing approaches rely heavily on high-quality, normalized input sequences to maximize semantic understanding and downstream model performance [13]. The preprocessing pipeline was executed through the following sequential operations:

a. Data Cleaning (Lexical Normalization): The initial step focused on removing non-informative elements from the dataset. This included the programmatic stripping of residual HTML tags, URLs, special characters, and punctuation marks. Additionally, all text was converted to lowercase (case folding) to ensure that capitalization discrepancies (e.g., "Academic" vs. "academic") did not result in redundant feature representations.
b. Context-Aware Tokenization: Following normalization, the text strings were segmented into smaller computational units, or tokens. Given that the subsequent modeling phase utilizes transformer architectures (BERT, BART, PEGASUS), a subword tokenization approach (such as WordPiece or Byte-Pair Encoding) was applied rather than traditional whitespace tokenization. This method effectively handles out-of-vocabulary (OOV) academic terms and complex administrative jargon by breaking down unknown words into recognizable subword morphological units.
c. Stopword Removal: To further reduce data dimensionality and eliminate semantic noise, highly frequent but contextually meaningless words (e.g., conjunctions and prepositions like 'and', 'the', 'or') were systematically filtered out using predefined lexical libraries.
d. Lemmatization: The remaining tokens were subjected to lemmatization, which algorithmically maps variant forms of a word back to their base dictionary form or lemma (for instance, mapping "regulations" and "regulating" to "regulate"). Unlike basic stemming, lemmatization utilizes morphological analysis and vocabulary dictionaries, ensuring that the semantic core of the academic terminology is preserved.
e. Encoding and Vectorization: In the final preprocessing step, the cleaned linguistic tokens were converted into numerical representations. Each token was mapped to its corresponding integer ID based on the pre-trained transformer's vocabulary. Furthermore, attention masks and positional embeddings were generated to retain the sequential context of the sentences.

Preprocessing Output: The culmination of this pipeline transforms qualitative text entries into quantitative matrices or tensors. The final output consists of tokenized sequences, attention masks, and target labels. This structured representation is computationally optimized and serves as the direct input for the feature engineering and machine learning modeling stages.

Idealis: Indonesia Journal Information System
Volume 9, Nomor 2, Juli 2026
ISSN 2684-7280 (*online*)
Halaman 259-267
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

2.3. Feature Engineering and Machine Learning Modeling

Following the preprocessing pipeline, the quantitative tensors were passed into the feature engineering and modeling stage. This phase constitutes the core intelligence of the proposed system through a combination of statistical feature selection and transformer-based deep learning architectures. The modeling process was systematically divided into three specialized Natural Language Processing tasks. Before the data were processed by the transformer models, mutual information was used to identify informative features and reduce redundant tokens, thereby improving computational efficiency and interpretability.

a. Module 1: Text Classification via BERT

To intelligently route incoming documents and user queries, a Text Classification module was developed using a fine-tuned BERT (Bidirectional Encoder Representations from Transformers) architecture. This approach leverages the proven superiority of transformer models for classification tasks within educational data [14]. Process: The contextualized embeddings generated from the preprocessing phase were fed into the BERT encoder. A dense classification head with a Softmax activation function was attached to the output of the special [CLS] token to compute probability distributions. Training was conducted using a maximum sequence length of 128 tokens, batch size of 16, learning rate of 2e-5, AdamW optimization, and early stopping based on validation loss. Output: The model assigns incoming unstructured text into one of five predefined operational classes (academic regulations, student administration, finance, facilities, and general information), serving as the routing component of the one-gate platform.

b. Module 2: Abstractive Auto-Summarization

Lengthy formal documents, such as academic regulations and curriculum guidebooks, often cause information overload for students. To mitigate this, an Auto-Summarization module was implemented using two sequence-to-sequence transformer architectures: BART and PEGASUS. These models were selected because they are capable of generating abstractive summaries while maintaining semantic coherence [15]. Process: Both models used an encoder-decoder structure and were evaluated on the same testing subset against human-written reference summaries. The experiments used a maximum input length of 512 tokens, a maximum output length of 120 tokens, beam search with four beams, and ROUGE-1, ROUGE-2, and ROUGE-L as evaluation metrics. Output: The module produces concise summaries of academic regulations and service documents while preserving key procedural information.

c. Module 3: Generative FAQ Chatbot System

To handle direct, real-time user interactions, a conversational agent was deployed using a transformer-based generative framework. Chatbots have been widely reported as effective tools in higher education because they improve service responsiveness and reduce administrative load [16]. Process: This module receives a user's natural language query, uses the intent recognized by the BERT classification module, and retrieves relevant policy contexts from the summarization module. The response quality was validated through a user survey instrument that measured answer accuracy, response speed, ease of use, and overall satisfaction on a five-point Likert scale. Output: The final output is an interactive dialogue system capable of resolving routine administrative inquiries autonomously.

2.4. Evaluation Metrics

To rigorously assess the performance and reliability of the Intelligent One-Gate System, a comprehensive evaluation framework was established. Evaluation employed both quantitative statistical metrics for the machine learning models and qualitative assessments for user interaction. A comprehensive evaluation using both automatic metrics and human feedback provides balanced insights into system performance and real-world applicability [17].

a. Quantitative Metrics for Classification

The performance of the Text Classification module (BERT) was evaluated using standard multi-class classification metrics derived from a confusion matrix:

Based on Equations (1)-(4), the classification evaluation combines accuracy for overall correctness, precision for prediction reliability, recall for coverage of relevant cases, and F1-score for balancing precision and recall under potential class imbalance. Accuracy is a metric that evaluates the overall effectiveness of a classification model by calculating the ratio of correctly predicted instances—both positives and negatives—against the total number of instances in the dataset. In other words, it measures how frequently the system provides the correct output regardless of the class distribution. This metric is widely adopted in machine learning because of its simplicity and intuitive interpretation[18].

$$Accuracy\ i\ \frac{TP+}{TP+TN+FP+FN} \qquad (1)$$

Idealis: Indonesia Journal Information System
Volume 9, Nomor 2, Juli 2026
ISSN 2684-7280 (*online*)
Halaman 259-267
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

Precision focuses on the reliability of positive predictions by measuring the proportion of correctly predicted positive instances compared to all instances classified as positive. A high precision score indicates that the model generates fewer false positives, making it highly useful in scenarios where incorrect positive predictions can lead to significant consequences, such as academic information retrieval or medical diagnoses.

$$Precision_k \frac{TP_k}{TP_k + FP_k} \quad (2)$$

Recall evaluates the ability of a model to identify all relevant positive cases by comparing the number of correctly predicted positives to the total number of actual positives in the dataset. This metric is particularly important in contexts where missing a positive instance could be more detrimental than producing a false alarm. In academic service automation, for example, recall ensures that students' queries are comprehensively addressed [19].

$$Recall_k \frac{TP_k}{TP_k + FN_k} \quad (3)$$

F1-score provides a balanced measure by combining both precision and recall into a single metric, defined as their harmonic mean. It is especially valuable in situations with imbalanced data distributions, as it accounts for both the correctness and completeness of the model's predictions. A high F1-score indicates that the system achieves a strong balance between minimizing false positives and capturing all true positives, which is crucial for intelligent service platforms like academic chatbots and document classification systems [20].

$$F1_k \frac{2\ Precision_k Recall_k}{Precision_k + Recall_k} \quad (4)$$

b. Quantitative Metrics for Summarization

The Auto-Summarization module (BART and PEGASUS) was evaluated using the ROUGE (Recall-Oriented Understudy for Gisting Evaluation) metric. Specifically, ROUGE-N (unigram and bigram overlap) and ROUGE-L (Longest Common Subsequence) were calculated to mathematically compare the linguistic overlap and structural similarity between the machine-generated abstractive summaries and human-authored reference summaries.

ROUGE-1 measures the overlap of unigrams (single words) between the system-generated summary and the reference summary. It evaluates how many individual words in the generated summary match those in the reference, thereby reflecting the overall lexical similarity. A high ROUGE-1 score indicates that the model effectively captures the important words from the source text.

$$ROUGE - 1 = \frac{\sum w \in Ref\, min(Countsys(w), Countref(w))}{\sum w \in Ref\, Countref(w)} \quad (5)$$

ROUGE-2 extends this evaluation to bigrams (two-word sequences), focusing on the model's ability to preserve word order and local context. By considering bigram overlaps, ROUGE-2 provides a stricter evaluation of summary quality compared to ROUGE-1. This metric is especially useful for assessing whether the generated summary maintains meaningful phrases rather than isolated words.

$$ROUGE\ -2\ \frac{\sum b \in Ref\, min(Countsys(b), Countref(b))}{\sum b \in Ref\, Countref(b)} \quad (6)$$

ROUGE-L evaluates the longest common subsequence (LCS) between the system-generated summary and the reference. Unlike ROUGE-1 and ROUGE-2, which rely on exact n-gram matches, ROUGE-L accounts for sentence-level structure and word order, making it more robust in capturing fluency and coherence. A high ROUGE-L score suggests that the summary maintains a logical sequence of ideas consistent with the reference text [21].

$$ROUGE - L\ \frac{LCS(Ref, Sys)}{Leng\ \ (Ref)} \quad (7)$$

Together, ROUGE-1, ROUGE-2, and ROUGE-L provide a comprehensive assessment of summarization performance by balancing lexical similarity, contextual accuracy, and structural coherence. Based on Equations (5)-(7), ROUGE evaluation in this study was used to compare machine-generated summaries with human-written reference summaries, so higher values indicate stronger overlap and better preservation of important academic information. These metrics are widely adopted in evaluating abstractive and extractive summarization models and remain standard benchmarks for summarization research in natural language processing.

Idealis: Indonesia Journal Information System
Volume 9, Nomor 2, Juli 2026
ISSN 2684-7280 (*online*)
Halaman 259-267
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

c. Qualitative Evaluation Chatbot Module

Beyond algorithmic metrics, the final web-based deliverable outputs, particularly the FAQ Chatbot System, underwent qualitative evaluation. This involved controlled field testing with students and administrative staff to assess usability, response naturalness, and effectiveness in resolving user intent. The questionnaire items were reviewed by two academic information service practitioners to ensure content validity, and the final instrument used a five-point Likert scale covering answer accuracy, response speed, ease of use, and overall satisfaction. The resulting scores were summarized using descriptive statistics to complement the quantitative model evaluation.

2.5. System Integration and Deliverable Outputs

The successful development of the individual NLP modules necessitated a robust deployment strategy to transition the models from an experimental environment into a functional, user-facing application. As depicted in the final stage of the methodology architecture, the modules were integrated into a unified "Intelligent One-Gate System" utilizing a web-based hybrid architecture. This integration is crucial, as cohesively integrated systems allow seamless access and significantly improve user interaction within complex digital campus ecosystems [22].The integrated system was deployed in a university environment. End-user feedback was collected from students and administrative staff to refine usability. Iterative validation in real academic settings ensures that the system aligns with user needs and institutional digital transformation goals[23]. Integration Architecture the system architecture follows a distinct client-server model:

a. Backend Services: The core machine learning models (BERT, BART, PEGASUS, and the conversational framework) were deployed on the server side using Python Flask. Flask was selected for its lightweight micro-framework capabilities, effectively serving the heavy transformer models as RESTful API endpoints. This setup ensures that the computationally intensive tasks (such as tensor processing and text generation) are handled securely on the server without overloading the user's device.
b. Frontend Interface: The client side was designed using responsive web technology to ensure accessibility across various devices (desktops, tablets, and smartphones).

Deliverable Outputs: The integration of these technologies culminated in three primary deliverable outputs accessible via the web platform:

a. Text Classification System Dashboard: An administrative interface where staff can upload unstructured documents. The system automatically tags and routes these documents into the five predefined categories, streamlining document management.
b. Auto-Summarization System: A dedicated tool within the portal where students or staff can input lengthy academic regulations (e.g., PDF links or raw text) and instantly receive a cohesive, abstractive paragraph summary.
c. Intelligent FAQ Chatbot System: A responsive, interactive chat widget embedded directly into the university's main portal, providing users with 24/7 real-time assistance for routine academic and administrative inquiries based on the engineered knowledge base.

3. Results and Discussion

3.1. Preprocessing Results

The initial dataset consisted of 1,000 academic documents and service queries distributed across academic regulations, student administration, finance, facilities, and general information. After cleaning, case folding, tokenization, lemmatization, class balancing, and stratified splitting, the corpus became a standardized input set for the three NLP modules. The stratified split allocated 70% of the data for training, 15% for validation, and 15% for testing, preserving the proportional distribution of each service category. This preprocessing stage reduced syntactic noise, minimized class imbalance, and ensured that subsequent model evaluation was based on comparable and representative data rather than raw administrative text.

3.2. Comparative Model Results

Table 1. The comparative evaluation of the three transformer-based models

Module	Metrics	Results
BERT (Classification)	Accuracy = 0.88 Precision = 0.87 Recall = 0.88 F1-score = 0.875	Reliable categorization of academic documents, with minor confusion in overlapping categories
BART	ROUGE-1 = 0.70	Captured key terms and sentence structure but limited phrase-

Idealis: Indonesia Journal Information System
Volume 9, Nomor 2, Juli 2026
ISSN 2684-7280 (*online*)
Halaman 259-267
available online at http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index

Module	Metrics	Results
(Summarization)	ROUGE-2 = 0.63 ROUGE-L = 0.68	level preservation
PEGASUS (Summarization)	ROUGE-1 = 0.74 ROUGE-2 = 0.66 ROUGE-L = 0.71	Outperformed BART, preserving semantic flow and producing summaries closer to human references
Chatbot (Survey)	Satisfaction = 82% (4.1/5) Accuracy of answers = 4.4/5 Ease of use = 4.1/5 Response speed = 4.0/5	High user acceptance, practical in reducing administrative workload

3.3. BERT Classification Results

As shown in Table 1, the BERT classification module achieved an accuracy of 0.88, precision of 0.87, recall of 0.88, and F1-score of 0.875. From 150 test samples, 132 cases were classified correctly, while 18 cases were misclassified. Most errors occurred between semantically adjacent categories, particularly student administration and academic regulations, because terms such as registration, schedule, and policy appeared in both classes. These results indicate that BERT is reliable for routing institutional documents and queries, although finer domain labels and confidence-based escalation are still needed for ambiguous cases.

3.4. BART and PEGASUS Summarization Results

The summarization experiment compared BART and PEGASUS using ROUGE-1, ROUGE-2, and ROUGE-L against human-written references. BART obtained ROUGE-1 = 0.70, ROUGE-2 = 0.63, and ROUGE-L = 0.68, indicating that it captured key terms and general sentence structure but was weaker in phrase-level preservation. PEGASUS achieved higher scores across all metrics, with ROUGE-1 = 0.74, ROUGE-2 = 0.66, and ROUGE-L = 0.71. The gap-sentence pretraining strategy of PEGASUS helped preserve procedural requirements, deadlines, and eligibility conditions more consistently, making it more suitable for summarizing academic regulations where the loss of key details can reduce service quality.

3.5. Chatbot Evaluation Results

The chatbot module was evaluated through a user satisfaction survey involving 30 students who interacted with the system. The results showed an average score of 4.1/5 or 82%, reflecting high user acceptance. Users gave the highest rating to answer accuracy (4.4/5), followed by ease of use (4.1/5), while response speed received the lowest rating (4.0/5). These findings show that the chatbot is useful for routine academic inquiries and can reduce dependence on administrative staff, but deployment should also address inference latency, server capacity, and response caching to improve real-time performance.

3.6. Discussion

Compared with previous studies that generally focused on one NLP function, such as chatbot consultation, document classification, or text summarization, this study demonstrates the value of integrating classification, summarization, and conversational response generation into a single one-gate platform. The results are consistent with prior research showing that transformer-based models improve semantic understanding in educational services, but the present system extends that work by linking the outputs of multiple modules into an operational academic service workflow. The main performance factors were data quality, semantic overlap between service categories, summarization length control, and system response latency. Therefore, the scientific contribution of this work is not only the use of BERT, BART, PEGASUS, and chatbot technology, but also the integrated architecture that transforms fragmented academic information sources into a unified, accessible, and user-centered service gateway.

4. Conclusion

This study concludes that an integrated transformer-based one-gate system can improve academic information services by combining BERT-based classification, PEGASUS-supported summarization, and chatbot interaction in a unified platform. The main finding is that the system achieved reliable document routing, stronger summarization performance with PEGASUS than BART, and positive user acceptance with an average chatbot satisfaction score of 82%. The contribution of this research lies in the integration of multiple NLP functions into

one practical service architecture, providing both a methodological reference for higher education NLP implementation and a deployable model for reducing fragmented academic service channels.

Several limitations remain. The dataset was limited to 1,000 samples, some categories overlapped semantically, informal language variations were not fully covered, and the chatbot response speed still requires optimization. Future studies should expand the corpus, add multilingual and informal-language coverage, improve confidence-based handling for ambiguous queries, and explore more advanced generative models or multimodal inputs to support richer academic service transactions.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to the Institut Teknologi dan Bisnis PalComTech, Palembang, for providing institutional support and access to academic service data that made this research possible. Special appreciation is extended to the students and staff participants who contributed their valuable time in completing the user satisfaction surveys and providing constructive feedback for system improvements. Finally, the authors wish to thank their colleagues in the Information Systems, Informatics and Digital Business study programs for their continuous guidance, insightful discussions, and encouragement throughout the research process.

REFERENCES

[1] D. Firdaus, H. Satria, P. Aliyansyah, W. Haryono, and U. Pamulang, "Pengembangan Aplikasi untuk Monitoring Absensi dan Lembur Karyawan," *Jurnal Komputer Antartika*, vol. 2, no. 4, pp. 147-154, Dec. 2024, doi: 10.70052/JKA.V2I4.640.

[2] R. Kusumawardani, G. M. Trisnandaru, H. Christian, M. H. Manalu, and A. S. Ramadani, "Implementasi Metode Waterfall pada E-Commerce Berbasis Website di Cldg Clothing," *IDEALIS: Indonesia Journal Information System*, vol. 9, no. 1, pp. 11-20, Jan. 2026, doi: 10.36080/IDEALIS.V9I1.3602.

[3] A. Dwiyanto, Manajemen Pelayanan Publik: Peduli, Inklusif, dan Kolaboratif. Yogyakarta, Indonesia: Gadjah Mada University Press, 2018. [Online]. Available: https://books.google.com/books?id=rrtjDwAAQBAJ. Accessed: May 11, 2026.

[4] A. Tamkin, D. Jurafsky, and N. Goodman, "Language Through a Prism: A Spectral Approach for Multiscale Language Representations," *Advances in Neural Information Processing Systems*, vol. 33, pp. 5492-5504, 2020.

[5] G. Vijay Kumar, A. Yadav, B. Vishnupriya, M. Naga Lahari, J. Smriti, and D. Samved Reddy, "Text Summarizing Using NLP," *Advances in Parallel Computing*, vol. 39, pp. 60-67, 2021, doi: 10.3233/APC210179.

[6] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *in Proc. NAACL-HLT*, 2019, pp. 4171-4186, doi: 10.18653/V1/N19-1423.

[7] S. C. Lewis, D. M. Markowitz, and J. B. A. Bunquin, "Journalists, Emotions, and the Introduction of Generative AI Chatbots: A Large-Scale Analysis of Tweets Before and After the Launch of ChatGPT," *Social Media and Society*, vol. 11, no. 1, 2025, doi: 10.1177/20563051251325597.

[8] G. Attigeri, A. Agrawal, and S. V. Kolekar, "Advanced NLP Models for Technical University Information Chatbots: Development and Comparative Analysis," *IEEE Access*, vol. 12, pp. 29633-29647, 2024, doi: 10.1109/ACCESS.2024.3368382.

[9] B. Bhagawanta and C. B. Santoso, "Deteksi Penyakit Daun Kapas dengan Deep Learning Berbasis Convolutional Neural Network (CNN)," *IDEALIS: Indonesia Journal Information System*, vol. 8, no. 2, pp. 220-227, 2025.

[10] Y. E. Hasibuan and A. Zakir, "Intelligent Chatbot for Enhancing Academic Consultation Services in Vocational Schools Using Natural Language Processing," *Jurnal Kecerdasan Buatan dan Teknologi Informasi*, vol. 5, no. 1, pp. 26-33, Jan. 2026, doi: 10.69916/JKBTI.V5I1.389.

[11] J. Sandhan, R. Singha, N. Rao, S. Samanta, L. Behera, and P. Goyal, "Revisiting Transformer-based Models for Long Document Classification," *in Findings of the Association for Computational Linguistics: EMNLP*, 2022, pp. 7212-7230, doi: 10.18653/V1/2022.FINDINGS-EMNLP.534.

[12] K. Khaled, "Natural Language Processing and Its Use in Education," Int. J. Adv. Comput. Sci. Appl., vol. 5, no. 12, 2014, doi: 10.14569/IJACSA.2014.051210.

[13] W. S. El-Kassas, C. R. Salama, A. A. Rafea, and H. K. Mohamed, "Automatic Text Summarization: A Comprehensive Survey," *Expert Systems with Applications*, vol. 165, 2021, doi: 10.1016/J.ESWA.2020.113679.

[14] X. Chen, H. Xie, D. Zou, L. Xu, and F. L. Wang, "Fine-Tuned BERT Model for Sentiment Classification of Chinese MOOCs," *in Lecture Notes in Computer Science*, vol. 15721, pp. 267-278, 2025, doi: 10.1007/978-981-96-8430-4_21.

[15] B. Gana, H. Allende-Cid, S. Ruping, M. Becerra-Rozas, and J. Zamora, "A Systematic Review of Long Document Summarization Methods: Evaluation Metrics and Approaches," *Neurocomputing*, vol. 655, p. 131287, 2025, doi: 10.1016/J.NEUCOM.2025.131287.

[16] E. Adamopoulou and L. Moussiades, "Chatbots: History, Technology, and Applications," *Machine Learning with Applications*, vol. 2, p. 100006, 2020, doi: 10.1016/J.MLWA.2020.100006.

[17] J. Tang, S. Han, J. Wang, B. He, and J. Peng, "A Comparative Analysis of Performance-Based Resilience Metrics via a Quantitative-Qualitative Combined Approach," *International Journal of Disaster Risk Science*, vol. 14, no. 5, pp. 736-750, 2023, doi: 10.1007/S13753-023-00519-5.

[18] J. C. Obi, "A Comparative Study of Several Classification Metrics and Their Performances on Data," *World Journal of Advanced Engineering Technology and Sciences*, vol. 8, no. 1, pp. 308-314, 2023, doi: 10.30574/WJAETS.2023.8.1.0054.
[19] S. Sravani and P. R. Karthikeyan, "Detection of Cardiovascular Disease Using KNN in Comparison With Naive Bayes to Measure Precision, Recall and F-Score," *AIP Conf. Proc.*, vol. 2821, no. 1, Nov. 2023, doi: 10.1063/5.0177014.
[20] B. H. Reddy, A. S. Vickram, and P. R. Karthikeyan, "Classification of Fire and Smoke Images Using Random Forest Algorithm in Comparison with Decision Tree to Measure Accuracy, Precision, Recall, F-score," *AIP Conf. Proc.*, vol. 2853, 2024, doi: 10.1063/5.0198489.
[21] R. Sh. Othman and I. M. Ibrahim, "Deep Learning for Natural Language Processing: A Review of Models and Applications," *International Journal of Scientific World*, vol. 11, no. 2, pp. 71-78, 2025, doi: 10.14419/T1XNAQ87.
[22] S. Khan, "An Efficient Human Resource Management System Model Using Web-Based Hybrid Technique," *Problems and Perspectives in Management*, vol. 20, no. 2, pp. 220-235, 2022, doi: 10.21511/PPM.20(2).2022.18.
[23] B. Novirahman, H. B. Santoso, and R. Y. K. Isal, "Usability Evaluation and User Interface Design of University Staffing Information System," *in Proc. 5th Int. Conf. Informatics and Computing (ICIC)*, 2020, doi: 10.1109/ICIC50835.2020.9335917.