Journal of Data Analytics, Information, and Computer Science JDAICS)
Volume 3, No 3 – Juli 2026
e-ISSN : 3032-4696

DESIGN OF AN AUTOMATED CLASSIFICATION SYSTEM USING INDOBERT TRANSFORMERS AND LOCAL NEWS TEXT SUMMARIZATION WITH LLAMA 3 ON RADARTEGAL.COM

Keisya Anazwa Octa Reviandy[1], Sam Farisa Chaerul Haviana[2]
Sultan Agung Islamic University, Indonesia
[1]keisyanazwa2004@gmail.com
[2]sam@unissula.ac.id

Received: 28-04- 2026	Revised: 02-06-2026	Approved: 05-06-2026

ABSTRACT

In the digital news production process, editorial teams face challenges in manually categorizing news articles and generating summaries, which are time-consuming, inefficient, and prone to inconsistencies that affect content management quality and Search Engine Optimization (SEO). This study aims to design and develop an automated system for news classification and summarization on the radartegal.com platform using the Transformer-based IndoBERT model for automatic news category classification and the Large Language Model (LLM) LLaMA 3 for abstractive text summarization. The research methodology consisted of problem identification, literature review, dataset collection, text preprocessing, IndoBERT fine-tuning, LLaMA 3 implementation, pipeline integration, and system evaluation. Classification performance was evaluated using accuracy, precision, recall, and F1-score, while summarization quality was evaluated using ROUGE metrics. Experimental results showed that the IndoBERT model achieved an accuracy of 84.00%, precision of 84.58%, recall of 84.00%, and F1-score of 83.95%. Meanwhile, the LLaMA 3 summarization module achieved ROUGE-1, ROUGE-2, and ROUGE-L scores of 0.4672, 0.2732, and 0.4122, respectively. The integrated system successfully automated editorial workflows, improved categorization consistency, generated informative summaries, and supported SEO optimization. These findings demonstrate that the proposed system can improve editorial efficiency while maintaining content quality in local digital news publishing.

Keywords: *IndoBERT; LLaMA 3; News Classification; Text Summarization; Natural Language Processing; SEO; Local News.*

INTRODUCTION

In the digital news production process, editorial teams play a crucial role in ensuring that every published article meets quality standards, is assigned to the appropriate category, and includes a concise summary before publication. In online news portals, publication speed is a strategic factor because it directly influences audience engagement, content visibility, and website competitiveness. Consistent publication during peak reading hours increases opportunities for attracting readers, extending session duration, and improving user interaction. Furthermore, search engines tend to favor websites that regularly publish relevant and updated content, which can contribute to better visibility in Search Engine Results Pages (SERP) and increased organic traffic (Masruri & Nihayati, 2022; Wijaya et al., 2025).

Despite the importance of efficient content publishing, many editorial activities are still performed manually. Before publication, editors are required to review articles, assign categories, and create summaries or short descriptions.

These activities generally require approximately 3–5 minutes per article. In newsroom environments where 30–50 articles may be processed daily, editorial teams can spend several hours performing repetitive tasks. This condition becomes increasingly challenging when publication schedules must be maintained while ensuring consistency and quality across all content. Similar challenges have been reported in digital journalism environments, where repetitive editorial processes often reduce operational efficiency and increase workload for newsroom staff (Sonni et al., 2024).

In addition to requiring considerable time, manual categorization and summarization may create inconsistencies due to differences in interpretation among editors. Inconsistent category assignment can negatively affect website information architecture and make content organization less structured. From an SEO perspective, inaccurate categorization may reduce topical authority and weaken the relevance signals used by search engines to evaluate content quality. Likewise, summaries produced under time constraints often focus only on extracting portions of the article without considering persuasive language, keyword relevance, or user engagement factors. As a result, generated meta descriptions may become less attractive, potentially reducing Click-Through Rate (CTR) and limiting content discoverability through search engines (Masruri & Nihayati, 2022). These challenges indicate the need for an automated solution capable of supporting editorial workflows while maintaining consistency, efficiency, and SEO quality.

Recent developments in Artificial Intelligence (AI) and Natural Language Processing (NLP) have created significant opportunities for automating language-related tasks. Among the most influential innovations is the Transformer architecture, which has demonstrated superior performance compared to conventional machine learning approaches in various NLP applications. According to Zhang and Shafiq (2024), Transformer-based models provide stronger contextual understanding because they are capable of capturing long-range semantic relationships within textual data. These capabilities have enabled significant improvements in tasks such as text classification, sentiment analysis, machine translation, and document summarization.

One of the most prominent Transformer-based models for Indonesian-language processing is IndoBERT. Developed and evaluated through the IndoNLU benchmark, IndoBERT was specifically trained on Indonesian corpora and demonstrated superior performance across various NLP tasks compared to multilingual language models (Wilie et al., 2020). Several recent studies have further validated the effectiveness of IndoBERT. Federico Tantoro et al. (2025) reported that IndoBERT achieved high performance in automatic news classification tasks by effectively identifying contextual patterns associated with different news categories. Similarly, Tandi et al. (2025) showed that the integration of IndoBERT with machine learning features improved performance in Indonesian textual entailment recognition. In sentiment analysis research, Akhdaan et al. (2024) also demonstrated that IndoBERT combined with Confident Learning significantly improved classification accuracy and F1-score. These findings indicate that IndoBERT possesses strong contextual understanding of Indonesian-language text and is highly suitable for automated news categorization.

In addition to classification, text summarization has become one of the most important applications of NLP in content management systems. Traditional extractive summarization methods generally select important sentences directly

from source documents, often resulting in summaries that lack coherence and readability. Recent advancements in Large Language Models (LLMs) have enabled the development of abstractive summarization approaches that generate new sentences while preserving the original meaning of the text. Among these models, LLaMA has emerged as one of the most influential open-source LLMs. Research conducted by Touvron et al. (2023) demonstrated that LLaMA can achieve competitive performance in text generation tasks despite utilizing fewer computational resources than many proprietary models. Furthermore, Gao et al. (2022) highlighted the capability of large language models to perform zero-shot language generation tasks, including summarization, without requiring extensive task-specific training.

Recent studies have further strengthened the potential of LLMs in summarization applications. Tahmid et al. reported that LLaMA-based models can generate high-quality summaries while maintaining contextual consistency even without domain-specific fine-tuning. Likewise, Bailly et al. (2025) demonstrated that LLM-based summarization approaches are capable of producing coherent and informative summaries for long-form documents by effectively preserving essential information from the original text. These findings suggest that LLMs can generate concise and readable summaries suitable for digital publishing environments, where information must be presented quickly while maintaining relevance and clarity.

Although previous studies have demonstrated the effectiveness of IndoBERT for text classification and LLaMA-based models for abstractive summarization, most research has focused on these tasks independently. Existing studies generally evaluate classification or summarization performance as separate NLP applications without addressing the practical needs of editorial workflows. Meanwhile, modern digital newsrooms require integrated solutions capable of performing multiple editorial tasks simultaneously. According to Sonni et al. (2024), the future of digital journalism increasingly depends on AI-assisted systems that can automate newsroom operations while maintaining content quality and editorial consistency.

Therefore, a significant research gap remains in the development of an end-to-end editorial automation system that integrates Transformer-based news classification and Large Language Model-based summarization within a unified workflow. Such a system has the potential to reduce repetitive editorial work, improve consistency in content organization, and support SEO-oriented publishing practices. To address this gap, this study aims to design and develop an NLP-based automation system that integrates the IndoBERT Transformer model for automatic news classification and the LLaMA 3 Large Language Model for abstractive news summarization. Specifically, this study aims to: (1) develop an automated news classification module using IndoBERT, (2) implement an abstractive summarization module using LLaMA 3, and (3) integrate both modules into a unified editorial automation pipeline capable of supporting efficient, consistent, and SEO-oriented digital news publishing workflows. The proposed system is expected to improve editorial productivity, maintain categorization consistency, generate informative summaries, and enhance the overall competitiveness of local digital news platforms.

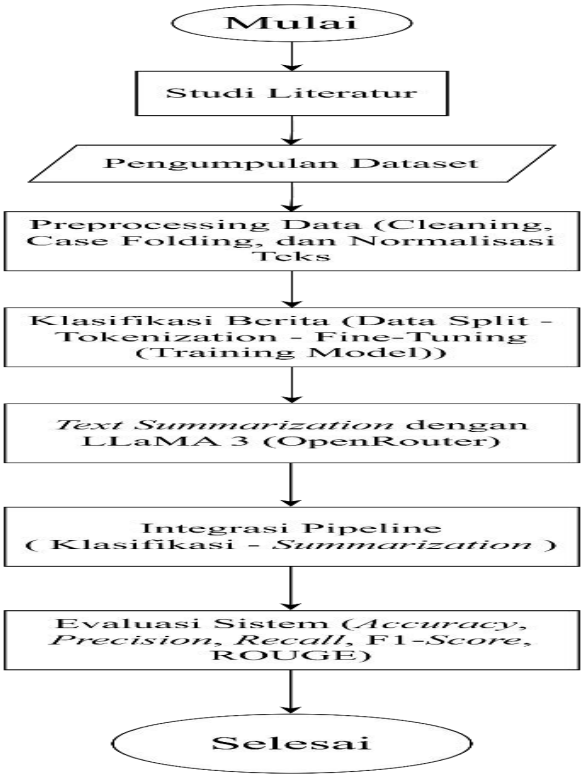
RESEARCH METHODS

This study employed a Research and Development (R&D) approach to

design, develop, and evaluate an automated editorial system based on Natural Language Processing (NLP). The proposed system integrates Transformer-based text classification and Large Language Model (LLM)-based abstractive summarization to support editorial workflows in digital news production. Real newsroom data obtained from the internal Content Management System (CMS) of radartegal.com were utilized throughout the research process. The primary objective was to automate the editorial tasks of news categorization and summary generation, which are traditionally performed manually and require significant editorial effort. Through the integration of classification and summarization into a unified workflow, the proposed system aims to improve editorial efficiency, maintain category consistency, and support Search Engine Optimization (SEO) performance.

The research methodology followed a structured development process consisting of research preparation, literature review, dataset collection, text preprocessing, classification model development, summarization implementation, system integration, interface development, system testing, and comprehensive evaluation. Each stage was conducted sequentially to ensure that raw news articles could be transformed into structured outputs consisting of predicted categories and SEO-oriented summaries. The overall research workflow is illustrated in Figure 1.

The workflow consisted of thirteen main stages: (1) research preparation, (2) literature review, (3) dataset collection, (4) text preprocessing and normalization, (5) dataset splitting, (6) tokenization, (7) IndoBERT training and fine-tuning, (8) classification model evaluation, (9) LLaMA 3 summarization implementation, (10) pipeline integration, (11) user interface development using Gradio, (12) functional testing using manual and Excel-based inputs, (13) quantitative and qualitative evaluation, followed by conclusions and recommendations.



Picture 1 Reserch Workflow Flowchart

Research Preparation

The research began with a preparation stage aimed at identifying practical challenges within editorial workflows of digital news production. Initial observations were conducted on the content publication process at radartegal.com to understand how editorial staff perform category assignment and summary creation before publication. The analysis revealed that both processes were predominantly performed manually and required considerable editorial time. Furthermore, variations in category assignment and summary writing among editors could potentially affect content consistency and SEO performance. These findings served as the foundation for defining system requirements and determining the objectives of the proposed automation system.

1. Literature Review

A comprehensive literature review was conducted to establish the theoretical foundation of the research and identify recent developments related to NLP-based editorial automation. Scientific publications were collected from academic databases including Google Scholar, Scopus, SpringerLink, IEEE Xplore, and conference proceedings. The review primarily focused on studies published between 2020 and 2025.

The literature review covered three main domains. The first domain focused on Transformer-based text classification, particularly studies involving BERT and IndoBERT architectures for Indonesian-language processing. The second domain examined abstractive summarization using Large Language Models such as LLaMA and other generative AI models. The third domain explored automation systems and Artificial Intelligence applications within digital journalism and content management environments.

The literature findings indicated that Transformer-based architectures consistently outperform conventional machine learning approaches in classification tasks due to their contextual understanding capabilities. Likewise, recent studies demonstrated that Large Language Models can generate coherent and informative abstractive summaries. However, most previous research addressed classification and summarization as separate tasks, while limited studies explored their integration into a newsroom-oriented editorial automation pipeline. This research gap motivated the development of the proposed system.

2. Dataset Collection

The dataset used in this study was obtained directly from the internal Content Management System (CMS) of radartegal.com. The use of real editorial data ensured that the developed system reflected actual newsroom conditions and publication workflows.

The collected dataset consisted of news articles published within approximately one year. Each record contained several attributes, including article title, article content, publication date, category label, author information, editorial metadata, and article identifiers. Data extraction was performed using Microsoft Excel format to facilitate further preprocessing and analysis.

For classification purposes, the dataset included articles that had been manually assigned to editorial categories. For summarization purposes, the dataset included editor-generated summaries that served as reference outputs during evaluation. Prior to model development, duplicate records, incomplete entries, and articles with inconsistent labels were removed to ensure data quality and reliability.

3. Text Preprocessing and Data Standardization

Before model implementation, preprocessing was performed to transform raw editorial content into structured textual data suitable for Transformer-based processing. News articles obtained from CMS databases frequently contain formatting inconsistencies, HTML tags, URLs, punctuation artifacts, and irregular spacing that may negatively affect model performance.

The first preprocessing step involved text cleaning. HTML tags, URLs, numbers, symbols, and unnecessary punctuation were removed using regular expression techniques while preserving meaningful textual information. Subsequently, case folding was applied to convert all characters into lowercase form to reduce token variability.

Text normalization was then performed to address abbreviations and shorthand expressions commonly found in Indonesian news writing. A normalization dictionary was developed to convert abbreviated terms into their complete forms, thereby improving semantic consistency and reducing ambiguity.

Unlike traditional machine learning preprocessing approaches, stopword removal and stemming were intentionally omitted. Transformer-based architectures rely heavily on contextual relationships between tokens. Preserving the original sentence structure allows the model to capture semantic dependencies more effectively and maintain contextual meaning throughout the learning process.

4. Dataset Splitting

Following preprocessing, the dataset was divided into training and testing subsets. An 80:20 split ratio was applied, where 80% of the data were allocated for model training and 20% were reserved for evaluation.

Stratified sampling was implemented to ensure that category distributions remained proportional across both subsets. This approach prevented class imbalance from influencing model performance and ensured fair evaluation of the classification model.

5. Tokenization

Before being processed by the classification model, textual data were converted into numerical representations through tokenization. The HuggingFace tokenizer associated with the indobenchmark/indobert-base-p1 checkpoint was utilized for this purpose.

Tokenization generated input IDs and attention masks required by the Transformer architecture. Padding and truncation techniques were applied to ensure consistent sequence lengths. The maximum sequence length was set to 256 tokens to balance contextual coverage and computational efficiency.

Category labels were subsequently encoded into numerical values to facilitate mathematical processing during model training. Each

category was assigned a unique numerical identifier to maintain consistency during training and inference.

6. System Evaluation

The classification module was developed using IndoBERT, a pretrained Transformer-based language model specifically trained on Indonesian corpora. IndoBERT was selected due to its strong capability in capturing contextual relationships within Indonesian-language text.

Fine-tuning was conducted using the labeled editorial dataset to adapt the pretrained model to newsroom-specific classification requirements. The implementation utilized HuggingFace's AutoModelForSequenceClassification architecture.

Several training parameters were configured to achieve stable convergence. The learning rate was set to 2e-5, batch size to 16, and the model was trained for three epochs. During training, the model continuously updated its parameters through forward propagation, loss calculation, backpropagation, and optimization processes.

The final model generated category predictions along with confidence scores based on softmax probability distributions.

7. Classification Model Evaluation

After fine-tuning, the classification model was evaluated using the testing dataset. Four evaluation metrics were utilized: accuracy, precision, recall, and F1-score.

Accuracy measured the proportion of correctly classified articles. Precision evaluated the relevance of category predictions, while recall measured the model's ability to retrieve articles belonging to the correct category. F1-score provided a balanced assessment between precision and recall.

These metrics were selected because they provide a comprehensive understanding of classification performance and are widely used in text classification research.

8. LLaMA 3 Summarization Implementation

The summarization component was developed using LLaMA 3 integrated through the OpenRouter API. Unlike the classification module, summarization relied on instruction-based generation and therefore did not require supervised training.

The model used in this research was meta-llama/llama-3-8b-instruct. Prior to summarization, lightweight preprocessing was applied to preserve textual consistency while maintaining paragraph structure.

Prompt engineering played a crucial role in controlling summary quality. Prompts were specifically designed to generate concise, informative, and SEO-oriented summaries that captured the key information contained within each article. Additionally, prompts encouraged the inclusion of journalistic elements corresponding to the 5W+1H framework.

Generation parameters included a temperature value of 0.3 and a maximum output length of 100 tokens to maintain output consistency and reduce randomness.

9. **Pipeline Integration**

Following the independent development of classification and summarization modules, both components were integrated into a unified NLP pipeline.

The integrated workflow begins when a news article is submitted to the system. The article first undergoes preprocessing before being passed to the IndoBERT classification model. After category prediction and confidence calculation, the same article is forwarded to the LLaMA 3 summarization module for summary generation.

The final output consists of the predicted category, confidence score, generated summary, and summary length. This integrated architecture enables the system to automate multiple editorial tasks through a single workflow.

10. **User Interface Development Using Gradio**

To facilitate practical deployment and user interaction, a web-based interface was developed using the Gradio framework. Gradio was selected because it enables rapid deployment of machine learning applications while providing an intuitive graphical user interface.

The developed interface allows users to submit individual articles through manual text input or process multiple articles simultaneously through Excel file uploads. Classification results, confidence scores, generated summaries, and summary statistics are displayed automatically within the interface.

The implementation of Gradio improves system accessibility and enables editorial staff to interact with the automation system without requiring technical expertise.

11. **Functional Testing Using Manual and Excel Inputs**

Functional testing was conducted to verify the operational reliability of the integrated system under realistic newsroom conditions.

Two testing scenarios were implemented. The first scenario involved manual article submission through the Gradio interface. The second scenario evaluated batch processing functionality through Excel file uploads. Testing activities focused on validating input handling, preprocessing execution, category prediction, summary generation, export functionality, and overall system stability.

The objective of this stage was to ensure that all system components operated correctly and consistently across different usage scenarios.

12. **System Evaluation**

The final stage of the research involved comprehensive system evaluation using both quantitative and qualitative approaches.

Quantitative evaluation focused on classification and summarization performance. Classification performance was measured using accuracy, precision, recall, and F1-score metrics. Summarization performance was evaluated using ROUGE-1, ROUGE-2, and ROUGE-L metrics, which measure lexical and structural similarity between generated summaries and editor-generated reference summaries.

Qualitative evaluation was conducted through user assessments involving editorial staff. Participants interacted directly with the system and evaluated various aspects using a five-point Likert scale. Evaluation

criteria included category relevance, summary readability, SEO suitability, workflow efficiency, and overall usability.

The combination of quantitative and qualitative evaluation ensured that the proposed system was not only technically effective but also practically applicable in real newsroom environments.

13. **Conclusions and Recommendations**

Based on the results of classification evaluation, summarization assessment, and user testing, conclusions were formulated regarding the effectiveness of the proposed system. Recommendations for future improvements were also identified, including the expansion of training datasets, implementation of additional news categories, and exploration of more advanced language models to further enhance editorial automation performance.

RESULTS AND DISCUSSION

The results and discussion section presents the implementation outcomes of the proposed automated editorial system integrating IndoBERT for news classification and LLaMA 3 for abstractive summarization. The developed system was evaluated using quantitative and qualitative approaches to assess classification performance, summarization quality, system functionality, and practical applicability within newsroom workflows. The evaluation process followed the methodology described in the previous section, including classification model evaluation, summarization assessment, pipeline integration testing, and system implementation.

1. **Dataset Characteristics and Experimental Setup**

The dataset used in this study was collected from the internal Content Management System (CMS) of radartegal.com. The dataset consisted of Indonesian-language news articles categorized according to editorial standards. Each record contained article titles, news content, category labels, publication information, and editorial metadata.

The news articles were classified into five editorial categories, namely Lifestyle, Local, National, Government, and Edukreasi. Prior to model development, the dataset underwent preprocessing stages including text cleaning, case folding, and normalization. Subsequently, the dataset was divided into training and testing subsets using an 80:20 ratio. The training dataset was used to fine-tune the IndoBERT classification model, while the testing dataset was utilized to evaluate classification performance.

The summarization component used the complete article content as input and generated summaries using LLaMA 3 through the OpenRouter API. The implementation environment consisted of Python, HuggingFace Transformers, PyTorch, and OpenRouter integration.

2. **News Classification Results Using IndoBERT**

The classification module was developed using IndoBERT and fine-tuned using newsroom-specific datasets. The model was designed to automatically classify news articles into one of the predefined editorial categories.

The evaluation was conducted using Accuracy, Precision, Recall, and F1-Score metrics. These metrics provide a comprehensive

assessment of model performance by measuring prediction correctness, relevance, completeness, and balance between precision and recall.

Table 1 Classification Performance Metrics

Metrik	Value
Accuracy	0.8400
Precision	0.8458
Recall	0.8400
F1-*Score*	0.8395
Loss	0.6847

The results indicate that the IndoBERT model achieved strong classification performance, with an accuracy of 84.00% and an F1-score of 83.95%. The precision score of 84.58% suggests that the majority of predicted categories were relevant to the actual article categories. Similarly, the recall value of 84.00% demonstrates the model's ability to correctly identify news articles belonging to their respective categories.

The obtained performance indicates that IndoBERT successfully captures contextual and semantic relationships within Indonesian-language news articles. Categories containing distinctive thematic characteristics generally produced higher classification confidence because the contextual vocabulary was more specific and consistent.

However, several classification challenges remained. Articles discussing overlapping topics occasionally produced lower confidence values because multiple categories shared similar contextual information. For example, local government news may contain terminology commonly found in national policy discussions, creating ambiguity during prediction.

3. Abstractive Summarization Results Using LLaMA 3

The summarization module was implemented using LLaMA 3 through OpenRouter. The objective of this component was to generate concise, informative, and SEO-oriented summaries suitable for use as news descriptions.

Unlike extractive approaches, LLaMA 3 generated abstractive summaries by paraphrasing the original content while preserving the essential meaning of the article. This approach produced summaries that were more natural and readable from an editorial perspective.

To evaluate summarization quality, ROUGE metrics were applied using editor-written summaries as reference outputs.

Table 2 Summarization Performance Using ROUGE Metrics

Hasil Penilaian ROUGE	
ROUGE-1	0.4672
ROUGE-2	0.2732
ROUGE-L	0.4122

The ROUGE-1 score of 0.4672 indicates a moderate overlap of important words between generated summaries and editorial references. The ROUGE-2 score of 0.2732 demonstrates the model's ability to

preserve phrase-level information, while the ROUGE-L score of 0.4122 reflects structural similarity and contextual consistency.

These results indicate that the generated summaries successfully retained key information from the original articles while remaining concise and readable. The summaries generally preserved the main event, key actors, and central message of the news content.

4. Output of Classification and Summarization

To evaluate the practical performance of the proposed system, validation testing was conducted using several news articles representing the five editorial categories used by radartegal.com, namely Lifestyle, Local, National, Government, and Edukreasi. The testing process aimed to verify whether the integrated system could correctly classify news articles and generate concise summaries that remained relevant to the original content.

Table 3 Original Category and Editorial Summary

Judul	Kategori	Ringkasan
Gonta-Ganti Skincare Tapi Belum Glowing? Mungkin Kamu Lupa Merawat Kulit dari Dalam!	*Lifestyle*	Sudah pakai skincare tapi kulit tetap kusam? Simak cara merawat kulit dari dalam dengan Nutricoll B ERL, minuman kolagen terbaik untuk kulit kenyal dan glowing.
Kaesang Pangarep, Putra Bungsu Jokowi Kunjungi Korban Banjir di Brebes	Lokal	Musibah banjir di Brebes kemarin menarik perhatian putra bungsu Jokowi, ini yang dilakukannya
THR 13.077 PPPK Paruh Waktu Pemprov Jateng Cair, Ini Tanggal Pencairannya	Nasional	THR bagi PPPK Paruh Waktu Pemprov Jateng segera cair. Gubernur Ahmad Luthfi menyebut THR akan dicairkan pada hari dan tanggal ini.
Gubernur Jateng Ahmad Luthfi Tekankan Transparansi dan Akuntabilitas dalam Pengadaan Barang dan Jasa	Pemerintahan	Gubernur Jateng menekankan pentingnya transparansi dan akuntabilitas dalam pengadaan barang dan jasa
Perdana! 838 Mahasiswa Universitas Harkat Negeri Tegal Diwisuda	Edukreasi	Sebanyak 838 mahasiswa Universitas Harkat Negeri (UHN) mengikuti persiapan prosesi wisuda di Hotel Bahari Inn, Kota Tegal, Senin 4 Mei 2026.

Table 3 presents examples of outputs generated by the proposed system. The results show that the classification module successfully

identified the appropriate category for each news article based on its contextual content. Articles related to skincare and personal health were classified as *Lifestyle*, articles discussing local events were classified as *Local*, policy and public administration topics were assigned to *Government*, broader public affairs were categorized as *National*, and educational events were classified as *Edukreasi*. These results indicate that the IndoBERT model effectively captured semantic relationships and contextual patterns within Indonesian-language news articles.

In addition to category prediction, the LLaMA 3 summarization module successfully generated concise summaries that preserved the essential information of each article. The generated summaries contained the main event, important actors, and key information while remaining shorter than the original news content. Furthermore, the summaries maintained readability and contextual relevance, making them suitable for use as short descriptions in digital publishing environments.

The validation results demonstrate that the integrated system is capable of transforming raw news articles into structured editorial outputs consisting of category predictions and informative summaries. The generated outputs can support editorial staff by reducing manual categorization and summary-writing activities while maintaining consistency across published content. This capability is particularly valuable in newsroom environments that process large volumes of news articles on a daily basis and require efficient content management workflows.

5. Discussion of Findings

The experimental results demonstrate that the integration of IndoBERT and LLaMA 3 provides an effective solution for newsroom automation. The classification model achieved an accuracy of 84.00%, indicating that Transformer-based architectures can effectively capture semantic relationships and contextual information within Indonesian-language news articles.

The summarization module also demonstrated satisfactory performance, achieving ROUGE scores that indicate meaningful similarity between generated summaries and editorial references. The generated summaries successfully preserved essential information while maintaining concise and readable structures.

An important finding of this study is the successful integration of classification and summarization into a single editorial workflow. Unlike previous studies that treated these tasks independently, the proposed system combines both functionalities into an end-to-end automation pipeline. This integration allows editorial staff to obtain category predictions and article summaries simultaneously, thereby reducing manual workload and improving workflow consistency.

The generated summaries also support SEO-oriented publishing practices because they remain concise, informative, and within the recommended character limit for digital descriptions. Consistent categorization further contributes to better content organization and may support topical authority within news portals.

Overall, the findings indicate that NLP and LLM technologies can be effectively integrated into newsroom environments to improve editorial efficiency, consistency, and content management quality..

CONCLUSION

This study successfully designed and developed an NLP-based editorial automation system that integrates the IndoBERT Transformer model for automatic news classification and the LLaMA 3 Large Language Model for abstractive news summarization. The proposed system was developed to address the challenges of manual categorization and summary generation within digital newsroom workflows.

The first objective of this research was to develop an automated news classification module using IndoBERT. Based on the experimental results, the classification model achieved an accuracy of 84.00%, a precision of 84.58%, a recall of 84.00%, and an F1-score of 83.95%. These results indicate that IndoBERT is capable of effectively capturing contextual and semantic information from Indonesian-language news articles and accurately classifying them into the five editorial categories used by radartegal.com, namely Lifestyle, Local, National, Government, and Edukreasi.

The second objective was to implement abstractive news summarization using LLaMA 3. The evaluation results showed that the summarization module achieved ROUGE-1, ROUGE-2, and ROUGE-L scores of 0.4672, 0.2732, and 0.4122, respectively. Furthermore, the generated summaries were able to preserve the essential information contained in the original articles while remaining concise, readable, and suitable for SEO-oriented news descriptions. The abstractive approach also enabled the system to produce more natural summaries compared to conventional extractive methods.

The third objective was to integrate both components into a unified editorial automation pipeline. The validation results demonstrated that the integrated system successfully transformed raw news articles into structured editorial outputs consisting of predicted categories and generated summaries. The system was able to process news articles through both manual input and Excel-based batch processing, supporting practical newsroom operations and reducing repetitive editorial tasks.

Overall, the findings demonstrate that the integration of Transformer-based classification and Large Language Model-based summarization can effectively support editorial workflows by improving efficiency, maintaining category consistency, and generating informative summaries. Therefore, the proposed system provides a practical solution for local digital news organizations seeking to enhance content management processes and support SEO-oriented publishing activities.

For future work, the system may be improved through the use of larger and more diverse datasets, the inclusion of additional news categories, and the exploration of more advanced language models to further improve classification accuracy and summarization quality.

DAFTAR PUSTAKA

[1] M. Adita Widianti and M. Budi Srikandi, “Netizen journalism and the democratization of local news: A case study of @abouttng on Instagram in Tangerang, Indonesia,” Channel, vol. 13, no. 1, pp. 62–75, 2025, doi:

10.12928/channel.
[2] G. Airlangga, "Comparative analysis of machine learning models for chronic disease indicator classification using U.S. chronic disease indicators dataset," MALCOM: Indonesian Journal of Machine Learning and Computer Science, vol. 4, no. 3, pp. 1034–1042, 2024, doi: 10.57152/malcom.v4i3.1403.
[3] D. Al Akhdaan, T. E. Sutanto, and M. Liebenlito, "Confident Learning pada IndoBERT: Peningkatan kinerja klasifikasi sentimen," The Indonesian Journal of Computer Science, vol. 13, no. 5, 2024, doi: 10.33022/ijcs.v13i5.4401.
[4] A. Auriemma Citarella, M. Barbella, M. G. Ciobanu, F. De Marco, L. Di Biasi, and G. Tortora, "Assessing the effectiveness of ROUGE as unbiased metric in extractive vs. abstractive summarization techniques," Journal of Computational Science, vol. 87, 2025, doi: 10.1016/j.jocs.2025.102571.
[5] A. Bailly, A. Saubin, G. Kocevar, and J. Bodin, "Divide and summarize: Improve SLM text summarization," Frontiers in Artificial Intelligence, vol. 8, 2025, doi: 10.3389/frai.2025.1604034.
[6] T. B. Brown et al., "Language models are few-shot learners," arXiv preprint arXiv:2005.14165, 2020.
[7] J. Federico Tantoro and I. D. M. B. A. Darmawan, "Klasifikasi berita berdasarkan kategori menggunakan convolutional neural network dengan IndoBERT," JNATIA, vol. 3, no. 4, 2025.
[8] L. Gao, X. Ma, J. Lin, and J. Callan, "Precise zero-shot dense retrieval without relevance labels," arXiv preprint arXiv:2212.10496, 2022.
[9] N. M. Gardazi, A. Daud, M. K. Malik, A. Bukhari, T. Alsahfi, and B. Alshemaimri, "BERT applications in natural language processing: A review," Artificial Intelligence Review, vol. 58, no. 6, 2025, doi: 10.1007/s10462-025-11162-5.
[10] A. Grattafiori et al., "The Llama 3 herd of models," arXiv preprint arXiv:2407.21783, 2024.
[11] N. H. Masruri and D. A. Nihayati, "Optimasi SEO on-page pada long-tail keyword untuk meningkatkan visibilitas website di dalam SERP," JRST (Jurnal Riset Sains dan Teknologi), vol. 6, no. 2, pp. 141–150, 2022, doi: 10.30595/jrst.v6i2.13071.
[12] Y. Maulana, "Analisis dampak penggunaan content management system (CMS) open source terhadap industri portal berita media online," Jurnal Algoritma, vol. 21, no. 1, pp. 1–8, 2024, doi: 10.33364/algoritma/v.21-1.1141.
[13] M. Mazeika, B. Li, and D. Forsyth, "How to steer your adversary: Targeted and efficient model stealing defenses with gradient redirection," arXiv preprint arXiv:2206.14157, 2022.
[14] G. Michel, E. V. Epure, R. Hennequin, and C. Cerisara, "Evaluating LLMs for quotation attribution in literary texts: A case study of LLaMA3," arXiv preprint arXiv:2406.11380, 2025.
[15] T. Mshvidobadze, "Python for automating machine learning tasks," JINAV: Journal of Information and Visualization, vol. 2, no. 2, pp. 77–82, 2021, doi: 10.35877/454ri.jinav373.
[16] F. N. A. Ionendri, F. Candra, and A. Rizal, "News classification using natural language processing with TF-IDF and multinomial naïve Bayes," Journal of Applied Computer Science and Technology, vol. 6, no. 1, pp. 37–45, 2025, doi: 10.52158/jacost.v6i1.1099.
[17] A. F. Sonni, H. Hafied, I. Irwanto, and R. Latuheru, "Digital newsroom transformation: A systematic review of the impact of artificial intelligence on journalistic practices, news narratives, and ethical challenges," Journalism

and Media, vol. 5, no. 4, pp. 1554–1570, 2024, doi: 10.3390/journalmedia5040097.

[18] M. Tahmid, R. Laskar, X.-Y. Fu, C. Chen, and S. Bhushan, "Building real-world meeting summarization systems using large language models: A practical perspective."

[19] T. Y. Tandi, T. F. Abidin, and H. Riza, "Incorporation of IndoBERT and machine learning features to improve the performance of Indonesian textual entailment recognition," Journal of Information Systems Engineering and Business Intelligence, vol. 11, no. 2, pp. 173–186, 2025, doi: 10.20473/jisebi.11.2.173-186.

[20] H. Touvron et al., "LLaMA: Open and efficient foundation language models," arXiv preprint arXiv:2302.13971, 2023.

[21] R. Uddin, A. Basit, Y. Khan, S. J. Shazib, and S. Hossain, "BERT-based fake news detection: A transformer-driven approach for misinformation classification on Twitter," International Journal on Science and Technology.

[22] R. F. Wijaya, Z. Syahputra, Khairul, I. Purnama, and R. S. Hardinata, "Strategi pengelolaan konten dengan search engine optimization pada website," Jurnal Komputer Teknologi Informasi Sistem Informasi (JUKTISI), vol. 4, no. 2, pp. 1417–1422, 2025, doi: 10.62712/juktisi.v4i2.680.

[23] B. Wilie et al., "IndoNLU: Benchmark and resources for evaluating Indonesian natural language understanding," arXiv preprint arXiv:2009.05387, 2020.

[24] H. Zhang and M. O. Shafiq, "Survey of transformers and towards ensemble learning using transformers for natural language processing," Journal of Big Data, vol. 11, no. 1, 2024, doi: 10.1186/s40537-023-00842-0.

[25] J. Zhang, Y. Zhao, M. Saleh, and P. J. Liu, "PEGASUS: Pre-training with extracted gap-sentences for abstractive summarization," arXiv preprint arXiv:1912.08777, 2020.

[26] X. Zhang, X. Xie, L. Ma, X. Du, Q. Hu, Y. Liu, J. Zhao, and M. Sun, "Towards characterizing adversarial defects of deep learning software from the lens of uncertainty," in Proceedings of the International Conference on Software Engineering, 2020, pp. 739–751, doi: 10.1145/3377811.3380368.