**Advance Sustainable Science, Engineering and Technology (ASSET)**
**Vol. 8, No.1, January 2026, pp. 02601021-01 ~ 02601021-011**
**ISSN: 2715-4211 DOI:** https://doi.org/10.26877/asset.v8i1.2637

# Development and Evaluation of an IndoBERT-Based NLP Model for Automated Clickbait Detection

**Sandy Kurniawan[1*], Adhe Setya Pramayoga[1], Yeva Fadhilah Ashari[1], Muhammad Afrizal Amrustian[2]**

[1]Department of Informatics, Faculty of Science and Mathematics, Universitas Diponegoro, Jl. Prof. H. Soedarto, S.H., Tembalang, Semarang 50275, Central Java, Indonesia

[2]Department of Information and Computer Science, King Fahd University of Petroleum and Minerals, Academic Belt Road, Dhahran 31261, Saudi Arabia

*sandy@live.undip.ac.id

**Abstract**. The rapid growth of digital news platforms necessitates reliable and automated systems for maintaining content quality at scale. This study presents the engineering and evaluation of an IndoBERT-based Natural Language Processing (NLP) framework for automated clickbait detection in Indonesian news headlines. The proposed framework is designed as an end-to-end text classification pipeline, incorporating data preprocessing, tokenization, fine-tuning of a pretrained IndoBERT model, and systematic performance evaluation. Experiments were conducted using the CLICK-ID dataset comprising 15,000 Indonesian news headlines, with an 80:20 stratified train–test split. The fine-tuned model achieved an accuracy of 0.83, with a precision of 0.82, recall of 0.77, and an F1-score of 0.79 for the clickbait class. Further evaluation using threshold-independent metrics yielded a ROC-AUC value of 0.89 and an average precision of 0.88, indicating strong discriminative capability under moderate class imbalance. Comparative analysis shows that the proposed approach outperforms prior CNN, Bi-LSTM, and ensemble-based methods evaluated on the same dataset. These results demonstrate that IndoBERT provides a robust foundation for engineering automated clickbait detection systems tailored to Indonesian-language news streams.

**Keywords**: IndoBERT, NLP system design, clickbait detection, machine learning pipeline, model evaluation

***(Received 2025-08-28, Revised 2025-12-23, Accepted 2025-12-26, Available Online by 2026-01-22)***

## 1. Introduction

Currently, sources of information and news can be accessed easily through the internet. Various platforms, including news websites, social media, and blogs, provide rapid access to information on a wide range of topics [1]. Along with advances in technology and increasingly faster internet connections, news dissemination has become more widespread, enabling users to obtain information from diverse perspectives [2]. However, from an engineering standpoint, this rapid growth of digital news content introduces significant challenges in building automated, scalable, and reliable Natural Language Processing (NLP) systems for content quality control.

In today's media environment, where numerous news outlets compete for audience attention, publishers frequently employ sensational headlines to increase engagement [3]. This strategy, commonly referred to as clickbait, relies on linguistically optimized headlines designed to attract readers but often diverge from the semantic content of the associated articles, thereby misrepresenting the actual information presented [4, 5]. For NLP system development, such semantic divergence poses a non-trivial text classification challenge that requires models capable of capturing contextual and semantic relationships rather than relying on surface-level lexical cues.

Although clickbait strategies can effectively increase website traffic, they raise concerns regarding information quality and reliability. Mismatches between headlines and content may reduce user trust and contribute to the dissemination of sensationalized or misleading information [6, 7]. Consequently, robust and automated clickbait detection mechanisms are required to support large-scale digital news processing pipelines and ensure reliable content filtering. Addressing this problem necessitates engineering-oriented solutions that can operate consistently across diverse linguistic styles and news domains.

Clickbait detection has therefore attracted increasing research attention across multiple languages, including English and Indonesian [8, 9]. Early approaches primarily relied on feature-based methods, extracting handcrafted features such as sensational words, headline length, or syntactic patterns to identify clickbait characteristics [10, 11]. While these methods offered interpretability, their dependence on domain-specific heuristics limited scalability and generalizability. Subsequently, traditional machine learning techniques, including Support Vector Machines (SVMs) [12–14], Naive Bayes [15], decision trees [12, 14, 16], and ensemble methods [17], were introduced to improve classification performance using a combination of handcrafted and automatically learned features.

Despite these advances, traditional machine learning models often struggle with complex linguistic structures, as clickbait does not depend solely on individual keywords but also on subtle semantic cues and contextual framing [18–20]. To address these limitations, deep learning approaches have increasingly been adopted as end-to-end learning paradigms for clickbait detection. Neural architectures such as Long Short-Term Memory (LSTM) networks [21], Convolutional Neural Networks (CNNs) [22, 23], and Transformer-based models including BERT, RoBERTa, and IndoBERT have demonstrated superior capability in capturing contextual dependencies and semantic relationships within text [24–27], In particular, Transformer-based models are well suited for clickbait detection due to their ability to model long-range dependencies and nuanced semantic structures [28, 29].

For the Indonesian language context, IndoBERT offers a distinct advantage, as it is pretrained on large-scale Indonesian corpora and is better equipped to capture linguistic nuances and cultural contexts that are difficult for general-purpose language models to represent [30, 31]. While prior studies have reported promising results using BERT-based architectures for clickbait detection, most have focused primarily on classification accuracy, with limited discussion of system design, reproducibility, and deployment feasibility, especially in Indonesian-language settings. Furthermore, challenges such as dataset limitations and the evolving diversity of clickbait techniques remain insufficiently addressed from an engineering perspective.

To address these gaps, this study focuses on the engineering and evaluation of an IndoBERT-based Natural Language Processing (NLP) system for automated clickbait classification in Indonesian news headlines. Rather than treating clickbait solely as a media phenomenon, this research frames the task as a text classification and system design challenge, emphasizing model fine-tuning, data preprocessing pipelines, and reproducible evaluation procedures. By leveraging a pretrained IndoBERT model and systematically benchmarking its performance against established deep learning approaches, this paper contributes to the engineering of an IndoBERT-based text classification system optimized for

Indonesian-language datasets, with an emphasis on reproducibility, performance evaluation, and sustainable AI implementation.

## 2. Methodology

The research on detecting clickbait was carried out through several key phases. These phases included gathering data, preprocessing text, developing the model, evaluating it, and analyzing the outcomes. Each phase was structured to fulfill the research goal: identifying clickbait in Indonesian-language online news headlines using IndoBERT. The general flow of this research is illustrated in Figure 1. These phases collectively form an end-to-end NLP system pipeline designed for automated clickbait detection.

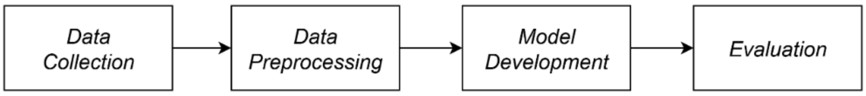



**Figure 1.** IndoBERT Clickbait Detection General Flow

### *2.1. Dataset Collection*

This research utilizes a dataset of Indonesian-language online news headlines, known as the CLICK-ID dataset, which was sourced from a study published by [32]. The main dataset comprises 15,000 headlines from online news, categorized as either clickbait or non-clickbait. The detailed composition of this dataset is shown in Table 1. For experimental reproducibility, the dataset was divided into training and testing subsets using an 80:20 stratified split to preserve class distribution consistency across experiments.

**Table 1.** CLICK-ID dataset distribution

| Label | Training | Testing | Total |
|---|---|---|---|
| Non-clickbait | 6968 | 1742 | 8710 |
| Clickbait | 5032 | 1258 | 6290 |
| Total amount | 12000 | 3000 | 15,000 |

### *2.2. Preprocessing*

During the preprocessing phase, the initial task involves tokenizing the text data, particularly focusing on the title column. This is achieved using the IndoBERT tokenizer. Tokenization serves to transform raw text into a format suitable for model processing. The process involves breaking the text into smaller components known as tokens. Each word or subword is then converted into a numerical form that the model can interpret. This tokenization process enables the model to comprehend words or phrases more effectively within their context.

Following tokenization, the subsequent step involves padding to ensure uniform input text length throughout the dataset. In this research, the maximum text length is designated as max_len=50, which means any text exceeding 50 tokens will be shortened, while shorter text will be padded to maintain a consistent input length. This padding is essential because the model processes data in batches, requiring all data within a batch to be of the same length for efficient parallel processing, thereby enhancing training stability and efficiency.

After tokenization and padding, the training and testing datasets were organized using the TensorDataset and Hugging Face Dataset. These datasets were formatted to be compatible with Hugging Face's Trainer, simplifying and structuring the model training process. The processed dataset includes the tokenized input and an attention mask, which instructs the model on which parts of the input to focus on and which to disregard during training (i.e., the padded sections). This attention mask aids the model in concentrating on relevant tokens while ignoring the padded ones. Thus, the data has been processed into a format that the model can accept, ensuring it is ready for the training phase and can enhance training efficiency. The complete preprocessing workflow, including tokenization, padding, and attention mask generation, is illustrated in Figure 2.


**Figure 2.** Preprocessing workflow

### *2.3. Model Development*

This research employs IndoBERT, a BERT variant tailored for the Indonesian language, to identify clickbait in Indonesian text. IndoBERT was selected due to its superior ability to grasp context and linguistic subtleties in Indonesian, outperforming the original BERT model designed for English. The strength of IndoBERT lies in its capacity to comprehend intricate sentence structures in Indonesian, enhancing its effectiveness in text classification tasks like clickbait detection.

We utilized BertTokenizer and BertForSequenceClassification to set up the model and tokenizer. BertTokenizer is responsible for breaking down text into a format that the model can process. On the other hand, BertForSequenceClassification is a BERT model specifically designed for classification tasks, distinguishing between two categories: clickbait and non-clickbait. This model facilitates text classification by comprehending the context of words and phrases within the text more deeply.

The implementation of the model utilized Hugging Face Transformers, which offers a user-friendly interface for importing IndoBERT and fine-tuning it on custom datasets. By using Hugging Face, we can take advantage of various tools and functions to effectively handle model training and evaluation. Text is tokenized with BertTokenizer before being input into the model, and BertForSequenceClassification is employed to conduct classification based on two specified labels.

During the model training phase, TrainingArguments are employed to set up the training process. Key parameters configured include the number of epochs (num_train_epochs=3), the batch size for both training and evaluation, as well as warmup steps and weight decay for optimizing the model. Logging steps are established to track training progress and to conduct model evaluations at regular intervals. Furthermore, Hugging Face's Trainer is utilized to train the model with the training dataset and assess it using the test dataset. The benefit of using Trainer lies in its ability to streamline the management of the training and evaluation process, allowing users to handle model parameters and results more easily without requiring complex coding.

**Table 2.** Training Configuration

| Parameter | Value |
|---|---|
| Pretrained Model | IndoBERT |
| Training Epochs | 3 |
| Training Batch Size | 16 |
| Evaluation Batch Size | 16 |
| Optimizer | AdamW |
| Weight Decay | 0.01 |
| Warm-up Steps | 500 |
| Evaluation Interval | 500 steps |
| Model Checkpoint Interval | 500 steps |
| Logging Interval | 10 steps |
| Framework | Hugging Face Transformers |

Table 2 summarizes the training configuration used for fine-tuning the IndoBERT model. The selected hyperparameters aim to balance classification performance and computational efficiency, ensuring stable convergence while maintaining feasibility for scalable deployment in automated NLP systems. The configuration was designed to support efficient fine-tuning without excessive computational overhead, aligning with sustainable AI deployment considerations.

### *2.4. Evaluation*
Once the training phase concludes, the model's ability to distinguish between clickbait and non-clickbait is assessed using test data. The trainer.predict() function is employed to generate predictions, yielding probabilities for each category (clickbait and non-clickbait). To translate these probabilities into clear labels (0 or 1), np.argmax() is utilized, selecting the label with the highest probability for each text instance.

To thoroughly evaluate model performance, several metrics were calculated, including accuracy, precision, recall, and the F1 score. Accuracy is measured using the accuracy_score, reflecting the proportion of correct predictions out of the total dataset. Precision measures the model's effectiveness in correctly identifying clickbait, while recall assesses its ability to detect all clickbait instances. The F1 score, which is the harmonic mean of precision and recall, provides a more detailed assessment of model performance, especially in cases of class imbalance.

To achieve a comprehensive understanding of the model's performance, we utilized the classification_report() function from sklearn.metrics. This function provides metrics like precision, recall, f1-score, and support for each class, with support indicating the number of samples per class in the test dataset. For a clearer visualization of the model's performance, a confusion matrix is employed. This matrix, generated using confusion_matrix() from sklearn.metrics and visualized with Seaborn, shows the distribution of correct and incorrect predictions in a matrix format. It clearly highlights false positives (FP), false negatives (FN), true positives (TP), and true negatives (TN), making it an essential tool for evaluating model errors and identifying areas for improvement.

This approach ensures that IndoBERT is effectively utilized to identify clickbait in Indonesian news articles. The model undergoes training and assessment with the processed dataset, and the evaluation outcomes offer insights into the model's ability to detect clickbait, allowing for comparisons with conventional methods. These evaluation metrics collectively provide insight into both predictive performance and system reliability for deployment in automated content monitoring environments.

## 3. **Results and Discussion**
This section reports the engineering evaluation of the fine-tuned IndoBERT classifier on the held-out test set, focusing on predictive performance and error distribution using standard classification metrics and a confusion matrix. The results are presented to support benchmarking against prior architectures and to assess the model’s reliability for integration into automated NLP-based content monitoring pipelines.

### *3.1. Model Performance*
Once the training phase concluded, IndoBERT's capabilities were evaluated using the test dataset. The model achieved an accuracy rate of 0.83, showcasing its proficiency in distinguishing between clickbait and non-clickbait article titles. To further assess its performance, metrics such as precision, recall, and F1-score were employed. As detailed in the classification report in Table 3, the model attained a precision of 0.82 and a recall of 0.77 for the clickbait category, while for the non-clickbait category, it achieved a precision of 0.84 and a recall of 0.88. This suggests that IndoBERT is fairly adept at identifying both clickbait and non-clickbait articles, with a slight inclination towards more accurately recognizing non-clickbait articles. Such behavior is important in system deployment because it affects the trade-off between filtering strictness (false positives) and missed detections (false negatives) in automated monitoring settings.

**Table 3.** Classification Report on Clickbait Detection

| Label | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| Non-clickbait | 0.84 | 0.88 | 0.86 | 1742 |
| Clickbait | 0.82 | 0.77 | 0.79 | 1258 |
| Accuracy | | | 0.83 | 3000 |

A confusion matrix is employed to visually depict the distribution of model predictions, as illustrated in Figure 3. This matrix displays the count of true positives (TP) and true negatives (TN) that were accurately classified, alongside the false positives (FP) and false negatives (FN) that were incorrectly predicted. From this matrix, it is evident that out of 1,258 clickbait articles, around 965 were accurately identified as such (TP), while 293 were mistakenly classified as non-clickbait (FN). On the other hand, among the 1,742 non-clickbait articles, 1,527 were correctly identified (TN), and 215 were wrongly categorized as clickbait (FP).

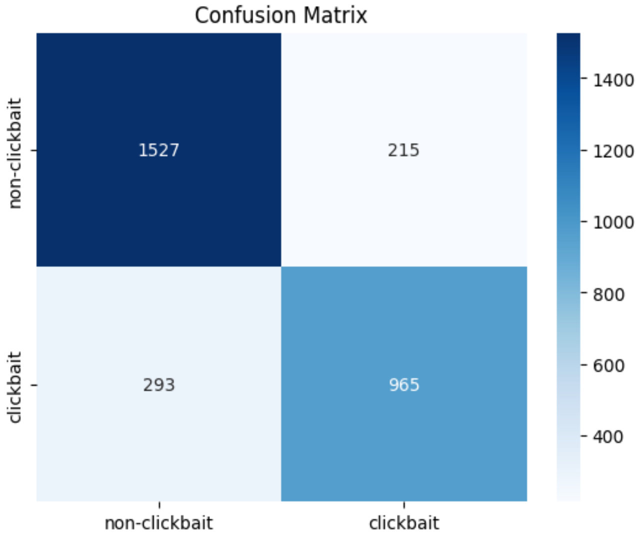
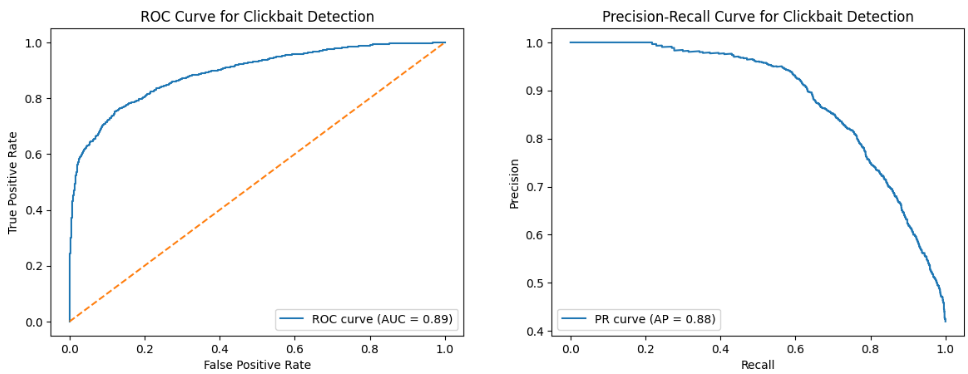



**Figure 3.** IndoBERT Clickbait Detection Confusion Matrix

This error distribution suggests that the model prioritizes precision in identifying non-clickbait content, resulting in a relatively low false positive rate. From a system deployment standpoint, such behavior is advantageous in automated filtering pipelines where minimizing false alarms is critical. However, the presence of false negatives indicates that certain clickbait patterns remain challenging to detect, particularly those with subtle or ambiguous linguistic structures. These findings highlight potential areas for system-level optimization, such as threshold adjustment, dataset enrichment with harder samples, or cost-sensitive training to better balance false positive and false negative trade-offs.

In addition to threshold-dependent metrics such as accuracy, precision, recall, and the confusion matrix, the performance of the proposed IndoBERT-based clickbait detection system was further evaluated using threshold-independent metrics. Receiver Operating Characteristic (ROC) and Precision-Recall (PR) curves were employed to analyze the model's discriminative behavior across varying decision thresholds and to assess its robustness under class imbalance conditions.




(a) (b)

**Figure 4.** ROC (a) and Precision–Recall curves (b) of the model

Figure 4 (a) illustrates the ROC curve of the proposed model, achieving an Area Under the Curve (AUC) value of 0.89. This result indicates a strong ability of the model to distinguish between clickbait and non-clickbait headlines across a wide range of classification thresholds. The curve exhibits a steep increase in the true positive rate at relatively low false positive rates, suggesting that the model can effectively capture clickbait patterns while maintaining controlled false alarm levels. Such characteristics are desirable in automated content filtering systems where reliable discrimination is required without excessive misclassification of non-clickbait content.

Figure 4 (b) presents the Precision-Recall curve of the IndoBERT-based classifier, with an average precision (AP) score of 0.88. Given the moderate class imbalance in the dataset, the PR curve provides a more informative assessment of model performance for the clickbait class. The curve demonstrates that high precision is maintained over a broad range of recall values, indicating that the model remains reliable when prioritizing the detection of clickbait headlines. This behavior is particularly important for deployment scenarios where minimizing false positives while maintaining reasonable recall is critical.

Overall, the evaluation results demonstrate that the IndoBERT-based system delivers reliable performance for clickbait detection in Indonesian news headlines. The combined analysis of classification metrics, confusion matrix, and ROC and Precision-Recall curves indicates that the model achieves strong discriminative capability while maintaining balanced performance between clickbait and non-clickbait classes. These results support the suitability of IndoBERT as a core component in engineered NLP systems for automated clickbait detection on Indonesian-language datasets.

### *3.2. Result Analysis*

To further examine the behavior of the proposed IndoBERT-based system, an error pattern analysis was conducted by inspecting false positive (FP) and false negative (FN) predictions. Rather than focusing on linguistic interpretation alone, this analysis aims to identify recurring patterns in misclassified headlines that may inform future improvements to data representation and model robustness. Table 4 and Table 5 summarize the most frequent words observed in FP and FN cases, respectively, with and without stopword removal.

**Table 4.** Frequent Words On False Positives

| Rank | With Stopword | | Without Stopword | |
|---|---|---|---|---|
| | **Word** | **Frequency** | **Word** | **Frequency** |
| 1 | di | 60 | bj | 19 |
| 2 | yang | 24 | habibie | 18 |
| 3 | bj | 19 | indonesia | 10 |
| 4 | habibie | 18 | kpk | 9 |
| 5 | ini | 17 | timnas | 6 |
| 6 | tak | 17 | video | 6 |
| 7 | jadi | 16 | kisah | 6 |
| 8 | untuk | 15 | polisi | 6 |
| 9 | ke | 14 | pakai | 5 |
| 10 | dan | 13 | bikin | 5 |

As shown in Table 4, false positive predictions are influenced by the presence of both high-frequency stopwords and semantically salient terms. When stopwords are retained, common function words such as “di”, “yang”, and “ini” dominate the frequency distribution, reflecting their ubiquitous usage in Indonesian headlines. After stopword removal, more content-specific terms such as “bj”, “habibie”, “indonesia”, and “kpk” become prominent. This indicates that headlines containing references to well-known public figures or institutions are occasionally misclassified as clickbait, likely due to their frequent co-occurrence with sensational news contexts in the training data. Such patterns suggest that entity-related cues alone are insufficient for reliable classification and must be interpreted in conjunction with broader contextual information.

**Table 5.** Frequent Words On False Negatives

| Rank | With Stopword | | Without Stopword | |
|---|---|---|---|---|
| | **Word** | **Frequency** | **Word** | **Frequency** |
| 1 | di | 80 | bj | 21 |
| 2 | dan | 30 | habibie | 19 |
| 3 | bj | 21 | indonesia | 17 |
| 4 | habibie | 19 | kpk | 17 |
| 5 | tak | 18 | vs | 10 |
| 6 | yang | 18 | menteri | 9 |
| 7 | kpk | 17 | orang | 8 |
| 8 | indonesia | 17 | kota | 8 |
| 9 | ke | 16 | liga | 8 |
| 10 | dengan | 15 | aksi | 8 |

On the other hand, Table 5 presents the most frequent words associated with false negative cases, where clickbait headlines were incorrectly classified as non-clickbait. Similar to the FP analysis, stopwords dominate the frequency distribution when they are not removed, obscuring more informative lexical patterns. Once stopwords are excluded, terms such as *"bj"*, *"habibie"*, *"indonesia"*, *"kpk"*, and *"menteri"* emerge as recurrent features. These findings indicate that certain clickbait headlines adopt a more neutral or formal wording style, reducing the presence of overtly sensational cues and making them harder to distinguish from legitimate news headlines. As a result, the model may fail to capture subtle clickbait strategies that rely on implicit curiosity rather than explicit provocation.

A notable observation from both false positive and false negative analyses is the recurrence of similar content-related terms across error categories. The overlap of words such as "bj", "habibie", "indonesia", and "kpk" suggests that misclassification is not driven by individual keywords alone, but by the surrounding semantic structure of the headline. This highlights a key limitation in relying on surface-level lexical signals and underscores the importance of deeper contextual modeling. From an engineering standpoint, these results point toward potential system-level enhancements, such as expanding the dataset with more diverse headline structures or incorporating additional contextual signals to improve robustness against ambiguous or borderline cases.

### *3.3. Result Comparison*

This subsection compares the performance of the proposed IndoBERT-based system with prior clickbait detection approaches that were evaluated using the same CLICK-ID dataset. The comparison is intended to position the proposed model within the existing engineering landscape and to assess its effectiveness relative to established deep learning and ensemble-based architectures.

**Table 6.** Result Comparison with Other Similar Research

| No | Author | Method | Results |
|---|---|---|---|
| 1 | [32] | CNN<br>Bi-LSTM | Highest using CNN = 0.7639<br>Highest using Bi-LTSM = 0.7779 |
| 2 | [17] | Ensemble | Average using Voting Classifier = 0.7763<br>Average using Stacking Classifier = 0.7728 |
| 3 | This Study | IndoBERT | IndoBERT achieved = 0.8306 |

As summarized in Table 6, the IndoBERT-based model achieved an accuracy of 0.8306, outperforming convolutional neural network (CNN), Bi-LSTM, and ensemble-based methods reported in previous studies. Earlier approaches using CNN and Bi-LSTM architectures achieved maximum accuracies of 0.7639 and 0.7779, respectively, while ensemble methods such as voting and stacking classifiers reported average accuracies below 0.78. The observed performance gain suggests that contextualized representations produced by IndoBERT are more effective in capturing semantic dependencies within Indonesian news headlines than architectures relying on static embeddings or sequential feature extraction alone.

Beyond numerical accuracy improvements, the comparison highlights the role of pretrained Transformer-based models in simplifying feature engineering and enabling end-to-end learning pipelines. Unlike traditional deep learning approaches that often require careful feature selection or architecture-specific tuning, IndoBERT leverages large-scale pretraining to provide richer contextual representations, which contributes to more stable classification behavior across diverse headline structures. This characteristic is particularly relevant for system development scenarios where robustness and generalizability are prioritized.

Overall, the comparative results indicate that the proposed IndoBERT-based framework offers a competitive and reliable solution for automated clickbait detection in Indonesian-language news. While accuracy remains the primary basis for comparison in this study, the results support the use of Transformer-based models as a strong foundation for engineering scalable NLP systems, especially when combined with reproducible training configurations and standardized evaluation protocols.

## 4. Conclusion

This study presented the engineering and evaluation of an IndoBERT-based Natural Language Processing framework for automated clickbait detection in Indonesian news headlines. Using the CLICK-ID dataset, the proposed system achieved an accuracy of 0.83, with a precision of 0.82, recall of 0.77, and an F1-score of 0.79 for the clickbait class, demonstrating balanced classification performance across both clickbait and non-clickbait categories.

Further evaluation using threshold-independent metrics confirmed the robustness of the proposed system. The ROC analysis yielded an AUC value of 0.89, while the Precision–Recall curve achieved an average precision of 0.88, indicating strong discriminative capability and reliable performance under moderate class imbalance conditions. Comparative benchmarking against CNN, Bi-LSTM, and ensemble-based approaches showed that the IndoBERT-based framework consistently outperformed previous methods evaluated on the same dataset.

Overall, these results support the suitability of IndoBERT as a core component for engineering automated clickbait detection systems tailored to Indonesian-language datasets. While this study focused on predictive performance and error pattern analysis, future work may extend the framework by incorporating computational efficiency measurements, larger and more diverse datasets, and deployment-oriented evaluations to further strengthen its applicability in scalable NLP-based content monitoring systems.

## References


[1] Tejedor S, Portalés-Oliva M, Carniel-Bugs R, et al. Journalism Students and Information Consumption in the Era of Fake News. *Media Commun* 2021; 9: 338–350.

[2] Ardia DS, Ringel E, Ekstrand V, et al. Addressing the Decline of Local News, Rise of Platforms, and Spread of Mis- and Disinformation Online: A Summary of Current Research and Policy Proposals. *SSRN Electronic Journal*. Epub ahead of print 22 December 2020. DOI: 10.2139/SSRN.3765576.

[3] Garde-Eransus E, Llamas Saíz C. Discursive Strategies and Linguistic Marks in Clickbait Headlines. *Estudios Sobre el Mensaje Periodistico* 2025; 31: 1–13.

[4] Pelau C, Pop MI, Stanescu M, et al. The Breaking News Effect and Its Impact on the Credibility and Trust in Information Posted on Social Media. *Electronics (Switzerland)*; 12. Epub ahead of print 1 January 2023. DOI: 10.3390/ELECTRONICS12020423.

[5] Molyneux L, Coddington M. Aggregation, Clickbait and Their Effect on Perceptions of Journalistic Credibility and Quality. *Journalism Practice* 2020; 14: 429–446.

[6] Carcioppolo N, Lun D, Mcfarlane SJ. Exaggerated and Questioning Clickbait Headlines and Their Influence on Media Learning. *J Media Psychol* 2022; 34: 30–41.

[7] Naeem B, Khan A, Beg MO, et al. A deep learning framework for clickbait detection on social area network using natural language cues. *J Comput Soc Sci* 2020; 3: 231–243.

[8] Dastidar AK, Khairnar A, Anand M, et al. Deep Dive into Clickbait Secrets: Integrating Multi-modal Features and Leveraging Deep Learning Architectures. *Lecture Notes in Networks and Systems* 2024; 1020 LNNS: 161–173.

[9] Jung AK, Stieglitz S, Kissmer T, et al. Click me. . .! The influence of clickbait on user engagement in social media and the role of digital nudging. *PLoS One*; 17. Epub ahead of print 1 June 2022. DOI: 10.1371/JOURNAL.PONE.0266743.

[10] Liu T, Yu K, Wang L, et al. WCD: A New Chinese Online Social Media Dataset for Clickbait Analysis and Detection. *Proceedings of 2021 7th IEEE International Conference on Network Intelligence and Digital Content, IC-NIDC 2021* 2021; 368–372.

[11] Coste CI, Bufnea D, Niculescu V. A New Language Independent Strategy for Clickbait Detection. *2020 28th International Conference on Software, Telecommunications and Computer Networks, SoftCOM 2020*. Epub ahead of print 17 September 2020. DOI: 10.23919/SOFTCOM50211.2020.9238342.

[12] Yadav KK, Bansal N. A Comparative Study on Clickbait Detection using Machine Learning Based Methods. *2023 International Conference on Disruptive Technologies, ICDT 2023* 2023; 661–665.

[13] Brogly C, Rubin VL. Detecting Clickbait: Here's How to Do It. *Canadian Journal of Information and Library Science* 2018; 42: 154–175.

[14] Santoso Hadi P, Muljono, Fanani AZ, et al. Using Extra Weight in Machine Learning Algorithms for Clickbait Detection of Indonesia Online News Headlines. *Proceedings - 2021 International Seminar on Application for Technology of Information and Communication: IT Opportunities and Creativities for Digital Innovation and Communication within Global Pandemic, iSemantic 2021* 2021; 37–41.

[15] Adrian FHN, Handradika NN, Prasojo AE, et al. Clickbait Detection on Online News Headlines Using Naive Bayes and LSTM. *International Conference on Artificial Intelligence and Mechatronics System, AIMS 2024*. Epub ahead of print 2024. DOI: 10.1109/AIMS61812.2024.10512986.

[16] Pujahari A, Sisodia DS. Clickbait detection using multiple categorisation techniques. *J Inf Sci* 2021; 47: 118–128.

[17] Kurniawan S, Pramayoga AS, Ashari YF. An Ensemble-Based Approach for Detecting Clickbait in Indonesian Online Media. *Jurnal Masyarakat Informatika* 2025; 16: 104–118.

[18] Zheng J, Yu K, Wu X. A deep model based on Lure and Similarity for Adaptive Clickbait Detection. *Knowl Based Syst*; 214. Epub ahead of print 28 February 2021. DOI: 10.1016/J.KNOSYS.2020.106714.

[19] Satpute RS, Agrawal A. Machine Learning Approach for Ambiguity Detection in Social Media Context. *2023 International Conference on Communication, Security and Artificial Intelligence, ICCSAI 2023* 2023; 516–522.

[20] Wang S, Luo J, Luo L. Large-scale Text Multiclass Classification Using Spark ML Packages. *J Phys Conf Ser*; 2171. Epub ahead of print 24 January 2022. DOI: 10.1088/1742-6596/2171/1/012022.

[21] Adrian FHN, Handradika NN, Prasojo AE, et al. Clickbait Detection on Online News Headlines Using Naive Bayes and LSTM. *International Conference on Artificial Intelligence and*

*Mechatronics System, AIMS 2024*. Epub ahead of print 2024. DOI: 10.1109/AIMS61812.2024.10512986.
[22] Kongyoung S, Rugchatjaroen A, Kaothanthong N. Automatic feature extraction and classification model for detecting Thai clickbait headlines using convolutional neural network. *Frontiers in Artificial Intelligence and Applications* 2019; 312: 184–194.
[23] Alhanaya R, Alqarawi D, Alharbi B, et al. Mushakkal: Detecting Arabic Clickbait Using CNN with Various Optimizers. *Journal of Information Technology Management* 2024; 16: 64–78.
[24] Wei F, Nguyen UT. An Attention-Based Neural Network Using Human Semantic Knowledge and Its Application to Clickbait Detection. *IEEE Open Journal of the Computer Society* 2022; 3: 217–232.
[25] Jiayi G, Ke Y, Zhou H, et al. Clickbait Analysis and Detection Method on Chinese Social Media. *Proceedings - 2022 8th International Conference on Big Data Computing and Communications, BigCom 2022* 2022; 342–349.
[26] Suryanto TLM, Wibawa AP, Hariyono, et al. Comparative Performance of Transformer Models for Cultural Heritage in NLP Tasks. *Advance Sustainable Science Engineering and Technology* 2025; 7: 02501015–02501015.
[27] Rupa MC, Ramani K. Hybrid Approaches for Advanced Medical Text Summarization: Combining TF-IDF, BERT, and Seq2Seq Models. *Advance Sustainable Science Engineering and Technology* 2025; 7: 0250301–0250301.
[28] Sirusstara J, Alexander N, Alfarisy A, et al. Clickbait Headline Detection in Indonesian News Sites using Robustly Optimized BERT Pre-training Approach (RoBERTa). *2022 3rd International Conference on Artificial Intelligence and Data Sciences: Championing Innovations in Artificial Intelligence and Data Sciences for Sustainable Future, AiDAS 2022 - Proceedings* 2022; 248–253.
[29] Fakhruzzaman MN, Gunawan SW. CekUmpanKlik: an artificial intelligence-based application to detect Indonesian clickbait. *IAES International Journal of Artificial Intelligence* 2022; 11: 1232–1238.
[30] Karen A, Christopher M, Qomariyah NN, et al. Clarifact-AI: Detecting Fake News in Indonesian Language with Natural Language Processing Using BiLSTM and IndoBERT Models. *10th International Conference on ICT for Smart Society, ICISS 2023 - Proceeding*. Epub ahead of print 2023. DOI: 10.1109/ICISS59129.2023.10291714.
[31] Koto F, Rahimi A, Lau JH, et al. IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP. *COLING 2020 - 28th International Conference on Computational Linguistics, Proceedings of the Conference* 2020; 757–770.
[32] William A, Sari Y. CLICK-ID: A novel dataset for Indonesian clickbait headlines. *Data Brief* 2020; 32: 106231.