Infoman's : Jurnal Ilmu-ilmu Informatika dan Manajemen Vol.20 No.1 May (2026)
DOI: 10.33481/infomans.vxxixx

Comparative Performance Analysis of SVM, LSTM, and IndoBERT for Sentiment Analysis of User Reviews on the Maganghub Kemnaker Platform

[1]Muhammad Arsayuan Wijaya, [2]Indah Puji Astuti, [3]Dyah Mustikasari
[1,2,3]Universitas Muhammadiyah Ponorogo
[1,2,3]Jl. Budi Utomo No. 10, Ronowijayan, Siman, Ponorogo, East Java, Indonesia, 63471
email : [1] arsayuans@gmail.com , [2]indahpujiastuti@umpo.ac.id, [3]dyahmustika@umpo.ac.id

ABSTRACT
The Maganghub Kemnaker National Internship Program, launched on October 1, 2025, attracted significant public attention while also generating both criticism and praise on the social media platform X. This study compares the performance of three classification algorithms—Support Vector Machine (SVM), Long Short-Term Memory (LSTM), and IndoBERT—for sentiment analysis of public opinions regarding the program. The dataset consists of 6,825 Indonesian-language tweets collected between October and December 2025 after duplicate removal. The data were preprocessed and labeled as positive, neutral, or negative using two approaches: automated labeling with a Hugging Face model and manual annotation. Each algorithm was evaluated under five hyperparameter tuning scenarios using an 80:20 train–test split and assessed with accuracy, precision, recall, and F1-score. The results show that IndoBERT achieved the highest performance, with an accuracy of 85.27% using the Hugging Face–labeled dataset, outperforming SVM (79.05%) and LSTM (78.68%). IndoBERT's superior performance is attributed to its Transformer-based architecture and multi-head self-attention mechanism, which effectively capture bidirectional contextual information, including sarcastic expressions commonly found in social media posts. The findings were further implemented in a web-based dashboard to support interactive public opinion monitoring for policymakers.

Keywords **-** IndoBERT, LSTM, Maganghub Kemnaker, Sentiment Analysis, SVM

1. Introduction

Internship is a structured work-training program that allows individuals to directly apply the competencies gained during their education in the workplace [1]. In Indonesia, the Ministry of Manpower (Kemnaker) introduced Maganghub as an official national platform to facilitate and standardize the implementation of internships in order to expand employment opportunities and strengthen the competitiveness of the workforce [2]. Since its launch on 1 October 2025, Batch 1 of the National Internship Program recorded an extremely high level of interest, with more than 156,159 applicants, 424,743 applications, and the participation of 1,668 companies [3]. The massive scale of the program also triggered turbulence in public opinion in the digital space; on one hand the program received appreciation, while on the other hand complaints emerged regarding technical issues such as frequent website downtime, inaccurate data synchronization, security issues in the form of cross-login account mix-ups, suboptimal user experience, limited quota availability, and registration schedules considered ambiguous. This dynamic of pro and contra sentiment constitutes valuable data that can be used as evaluation material for Kemnaker in refining the program's mechanism in subsequent batches.

Sentiment analysis is a branch of research in natural language processing (NLP) that encompasses sentiment extraction, sentiment classification, opinion summarization, and emotion detection, and has developed rapidly since the early 2000s [4]. In general, public opinion is classified into three main

This is an open article under

classes—positive, neutral, and negative—with approaches that can be grouped into lexicon-based, traditional machine learning, deep learning, and hybrid approaches [5]. Traditional machine learning approaches such as Support Vector Machine (SVM) have proven effective; one study reported that SVM achieved 76.6% accuracy in analyzing public sentiment toward PSSI on platform X [6]. However, this approach has limitations in capturing semantic context and word dependencies [5], so recent research has shifted toward deep learning approaches such as LSTM, which can automatically extract features through neural network architectures. For example, the application of LSTM and Bidirectional LSTM (BiLSTM) variants in analyzing the competency gap of vocational education graduates achieved an accuracy of around 90% and an F1-score of 0.91 [7].

Several previous studies have compared machine learning and deep learning algorithms with the IndoBERT model for Indonesian-language sentiment analysis. Siva et al. compared IndoBERT and SVM on digital education platform reviews and found that IndoBERT was superior in handling class imbalance, but did not involve LSTM siva[8]. Purwanto evaluated IndoBERT and LSTM on Kenjeran Park tourism reviews, with IndoBERT achieving 87.50% accuracy, outperforming LSTM's 77.93%, without involving SVM [9]. The study most closely related to this topic was conducted by Daniati et al., who compared SVM, LSTM, and fine-tuned IndoBERT on 1,517 comments related to electric cars and reported IndoBERT as the best model with 91% accuracy [10]; however, that research object differs in data characteristics and sentiment distribution from a public-policy context such as Maganghub Kemnaker.

Based on the description above, it is evident that the performance of each algorithm is specific to the context and domain of the data used, so no single algorithm is universally superior across all sentiment analysis scenarios. This study aims to fill that gap by comparing the performance of SVM, representing the traditional machine learning approach; LSTM, representing the sequential deep learning approach; and IndoBERT, representing the modern transformer-based approach optimized for Indonesian, in a new domain: public opinion toward Maganghub Kemnaker. The contributions of this study include a performance comparison of the three algorithms on two different data-labeling methods—Hugging Face-based automatic labeling and manual labeling—with five hyperparameter tuning scenarios for each algorithm, as well as the implementation of the analysis results in a web-based dashboard that can be used by Kemnaker and related policymakers as an informative public-opinion monitoring tool.

2. Research Method

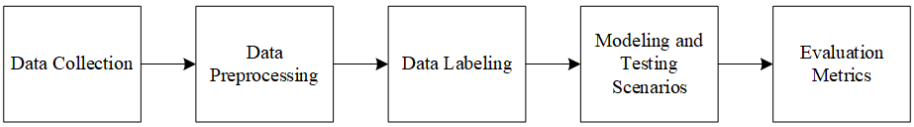


Figure 1. Research Stages

This research was conducted through five main stages as shown in Figure 1, namely data collection, data preprocessing, data labeling, modeling and scenario testing, and model performance evaluation. Each stage is described in the following subsections.

2.1. Data Collection

The research data consists of Indonesian-language tweets discussing the Maganghub Kemnaker program, collected from the social media platform X during the period of 1 October–31 December 2025. The scraping process was carried out in the Google Colaboratory environment using the open-source tweet-harvest library version 2.6.1 with the search parameter "lang:id" to filter Indonesian-language

DOI: 10.33481/infomans.vxxixx

tweets. This process produced a raw dataset of 6,983 rows of data with 15 attributes, including full_text, created_at, favorite_count, retweet_count, and username.

2.2. Data Preprocessing

The raw data was first cleaned of duplicates using Microsoft Excel based on matching tweet URL values, reducing the number of unique data points from 6,983 to 6,825 rows. Text cleaning was then carried out in the Google Colab environment using the pandas and re (Regular Expression) libraries through the following sequential steps: (1) case folding to standardize letters to lowercase, (2) URL removal using the regex pattern http\S+|www\S+, (3) mention removal using the pattern @\w+, (4) hashtag removal using the pattern #\w+, and (5) removal of excess whitespace resulting from the previous steps.

2.3. Data Labeling

The cleaned data were labeled into three sentiment classes (positive, neutral, and negative) using two approaches: automatic labeling with the pre-trained mdunur/indobert-sentiment-classification model via the Hugging Face transformers library and manual annotation by the researcher. Manual labeling served as the ground truth to address limitations of automatic labeling, particularly in detecting sarcasm and irony. The sentiment distribution obtained from both methods is presented in Table 1.

Table 1. Data Labeling Distribution

Sentiment Class	Hugging Face	Manual
Positive	558	852
Neutral	3,825	4,633
Negative	2,442	1,340
Total	6,825	6,825

2.4. Modeling and Testing Scenarios

The labeled data was split with an 80:20 ratio for training and testing data. Each algorithm was tested using five different hyperparameter tuning scenarios, and all scenarios were tested separately on both labeled dataset versions (Hugging Face and manual).

For the SVM algorithm, the optimized parameters included kernel type (linear and rbf), penalty value C (0.1; 1; 10), TF-IDF max features limit (5,000 and 10,000), and ngram range ((1,1) and (1,2)), as presented in Table 2.

Table 2. SVM Testing Scenarios

Scenario	Kernel	C	TF-IDF Max Features	Ngram Range
S1	linear	1	5000	(1,1)
S2	linear	0.1	5000	(1,1)
S3	linear	10	5000	(1,1)
S4	linear	1	10000	(1,2)
S5	rbf	1	5000	(1,1)

For the LSTM algorithm, optimization focused on the neural network architecture, including embedding dimension (64 and 128), number of LSTM units (64 and 128), dropout rate (0.2 and 0.5), batch size (32 and 64), and number of epochs (10 and 15), as presented in Table 3.

This is an open article under

Table 3. LSTM Testing Scenarios

Scenario	Embedding Dim	LSTM Units	Dropout	Batch Size	Epochs
S1	128	64	0.2	32	10
S2	64	64	0.2	32	10
S3	128	128	0.2	32	10
S4	128	64	0.5	32	10
S5	128	64	0.2	64	15

For the IndoBERT algorithm, optimization was carried out through a fine-tuning approach with the parameters learning rate (2.00E-05, 5.00E-05, 1.00E-05), batch size (16 and 32), and number of epochs (3 and 5), while max length was kept constant at 128 for all scenarios, as presented in Table 4.

Table 4. IndoBERT Testing Scenarios

Scenario	Learning Rate	Batch Size	Epochs	Max Length
S1	2.00E-05	16	3	128
S2	5.00E-05	16	3	128
S3	1.00E-05	16	3	128
S4	2.00E-05	32	3	128
S5	2.00E-05	16	5	128

2.5. Evaluation Metrics

The performance of each scenario was evaluated using four standard classification metrics, namely accuracy, precision, recall, and F1-score, calculated based on the model's prediction results on the test data for each sentiment class (positive, neutral, negative).

3. Result and Analysis

3.1. SVM Model Evaluation Result

The evaluation results of the five SVM scenarios on the Hugging Face-labeled and manually labeled datasets are presented in Table 5 and Table 6.

Table 5. SVM Evaluation Result (Hugging Face Labeling)

Scenario	Accuracy	Precision	Recall	F1-Score
S1	0.7890	0.7823	0.7890	0.7781
S2	0.7487	0.6906	0.7487	0.7107
S3	0.7612	0.7552	0.7612	0.7560
S4	0.7824	0.7794	0.7824	0.7746
S5	0.7788	0.7797	0.7788	0.7623

Table 6. SVM Evaluation Result (Manual Labeling)

Scenario	Accuracy	Precision	Recall	F1-Score
S1	0.7905	0.7870	0.7905	0.7780
S2	0.7172	0.7657	0.7172	0.6339
S3	0.7458	0.7472	0.7458	0.7458
S4	0.7868	0.7827	0.7868	0.7741
S5	0.7868	0.7924	0.7868	0.7646

The best SVM scenario for Hugging Face labeling was obtained in S1 with an accuracy of 78.90%, while for manual labeling, the best scenario was also obtained in S1 with an accuracy of 79.05%. The relatively small performance difference between the two labeling methods in this scenario indicates that the combination of a linear kernel with basic parameters (C=1) already provides sufficiently optimal TF-IDF feature representation for SVM in forming a hyperplane that separates sentiment classes.

DOI: 10.33481/infomans.vxxixx

3.2. LSTM Model Evaluation Result

The evaluation results of the five LSTM scenarios on the Hugging Face-labeled and manually labeled datasets are presented in Table 7 and Table 8.

Table 7. LSTM Evaluation Result (Hugging Face Labeling)

Scenario	Accuracy	Precision	Recall	F1-Score
S1	0.7824	0.7831	0.7824	0.7777
S2	0.7788	0.7770	0.7788	0.7741
S3	0.7788	0.7769	0.7788	0.7756
S4	0.7868	0.7812	0.7868	0.7795
S5	0.7714	0.7672	0.7714	0.7594

Table 8. LSTM Evaluation Result (Manual Labeling)

Scenario	Accuracy	Precision	Recall	F1-Score
S1	0.7648	0.7623	0.7648	0.7632
S2	0.7560	0.7497	0.7560	0.7504
S3	0.7692	0.7627	0.7692	0.7632
S4	0.7670	0.7617	0.7670	0.7607
S5	0.7626	0.7574	0.7626	0.7592

The best LSTM scenario for Hugging Face labeling was obtained in S4 with an accuracy of 78.68%. Meanwhile, for manual labeling, the best scenario was obtained in S3 with an accuracy of 76.92%. These results indicate that LSTM consistently achieved better performance on the Hugging Face-labeled dataset compared to manual labeling.

3.3. IndoBERT Model Evaluation Result

The evaluation results of the five IndoBERT scenarios on the Hugging Face-labeled and manually labeled datasets are presented in Table 9 and Table 10.

Table 9. IndoBERT Evaluation Result (Hugging Face Labeling)

Scenario	Accuracy	Precision	Recall	F1-Score
S1	0.8527	0.8532	0.8527	0.8529
S2	0.8410	0.8431	0.8410	0.8419
S3	0.8505	0.8514	0.8505	0.8509
S4	0.8520	0.8522	0.8520	0.8519
S5	0.8491	0.8485	0.8491	0.8487

Table 10. IndoBERT Evaluation Result (Manual Labeling)

Scenario	Accuracy	Precision	Recall	F1-Score
S1	0.8066	0.8075	0.8066	0.8069
S2	0.8081	0.8077	0.8081	0.8078
S3	0.8183	0.8205	0.8183	0.8192
S4	0.8168	0.8185	0.8168	0.8174
S5	0.8088	0.8087	0.8088	0.8086

The best IndoBERT scenario for Hugging Face labeling was obtained in S1 with an accuracy of 85.27%. For manual labeling, the best scenario was obtained in S3 with an accuracy of 81.83%. Compared to SVM and LSTM, IndoBERT consistently achieved the highest evaluation scores across all metrics and both labeling methods.

This is an open article under

3.4. IndoBERT Model Evaluation Result

A summary comparison of the best scenario from each of the three algorithms is presented in Table 11.

Table 11. Best Model for Each Algorithm

Algorithm	Scenario	Labeling	Accuracy
SVM	S1	Manual	79.05%
LSTM	S4	Hugging Face	78.68%
IndoBERT	S1	Hugging Face	85.27%

Based on the overall testing of 30 scenarios, IndoBERT with the fine-tuning approach on the Hugging Face-labeled dataset Scenario 1 proved to be the best-performing model, with an accuracy of 85.27%, higher than SVM (79.05%) and LSTM (78.68%). This absolute superiority of IndoBERT stems from its Transformer architecture with a multi-head self-attention mechanism that allows the model to read sentence structure bidirectionally, enabling it to capture the full context of a sentence, including sarcastic language style, unlike SVM, which relies solely on word frequency (TF-IDF), or LSTM, which processes text sequentially in one direction.

Another interesting pattern found is that LSTM and IndoBERT consistently achieved peak accuracy on the Hugging Face-labeled dataset, which is presumed to be because model-based labeling has a more structured semantic classification pattern and is relatively free from human subjectivity bias. Conversely, SVM achieved its best performance on manual labeling, indicating that the combination of manual labeling with TF-IDF feature representation in that scenario produced a more optimal class separation for margin-based algorithms such as SVM.

These findings are consistent with the study by Daniati et al., which also reported fine-tuned IndoBERT as the best model compared to SVM and LSTM in the domain of public opinion regarding electric cars, as well as the studies by Siva et al. and Purwanto, which respectively reported the superiority of IndoBERT compared to SVM and LSTM in different domains [8][9][10]. This reinforces the indication that Transformer architecture, particularly IndoBERT optimized for Indonesian, is generally superior in capturing the semantic complexity of Indonesian-language text on social media data compared to traditional machine learning or sequential deep learning approaches, regardless of variations in domain and dataset characteristics.

3.5. IndoBERT Model Evaluation Result

The sentiment prediction results from the three algorithms were implemented in a Flask-based web application with a MySQL database. The application consists of three main pages: Home for data search, Results, which displays a table of analysis results with a majority-voting mechanism among algorithms, and Evaluation, which presents visualizations of model performance metrics. This website was designed so that Kemnaker and related policymakers can monitor public perception of Maganghub Kemnaker interactively and based on data. The following are the details of the functionality and interface of the developed website.

DOI: 10.33481/infomans.vxxixx

A. Home Page

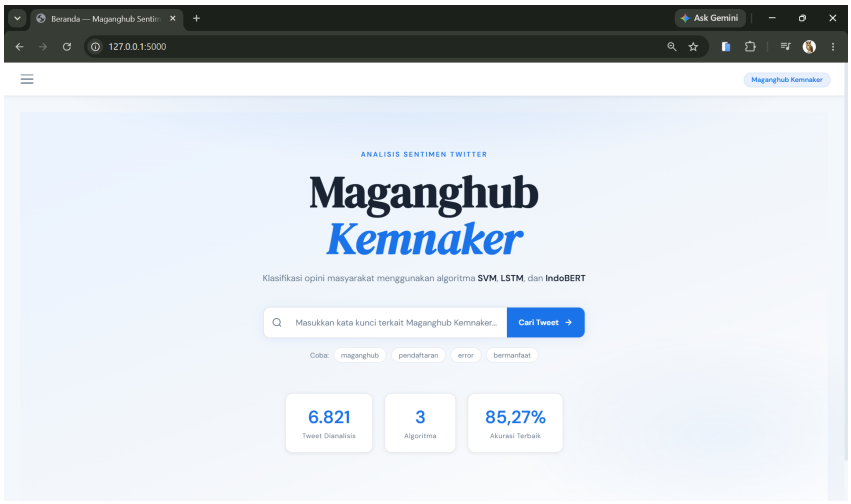


Figure 2. Home Page

Figure 2 shows the interface of the Home Page, which functions as the initial interaction center. This page includes a search bar feature that allows users to enter specific keywords to filter tweets related to Maganghub Kemnaker. In addition, this page also contains information cards displaying basic system statistics, such as the total data analyzed, the number of classification algorithms, and the system's highest accuracy achievement.

B. Sidebar

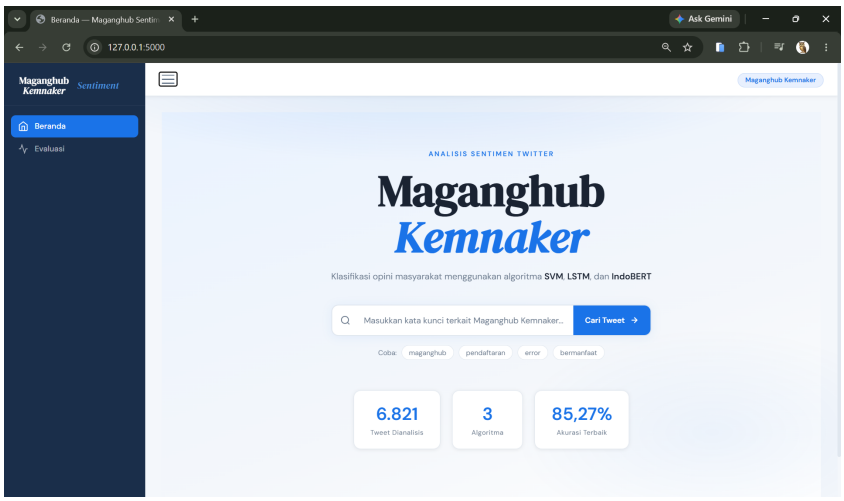


Figure 3. Sidebar

Figure 3 shows the main navigation module in the form of a sidebar placed on the left side of the interface. This component is designed to be responsive in order to maximize the data display area. Through this sidebar, users can easily switch between the search module on the "Home" page and the analytics module on the "Evaluation" page.

This is an open article under

C. Search Result Page

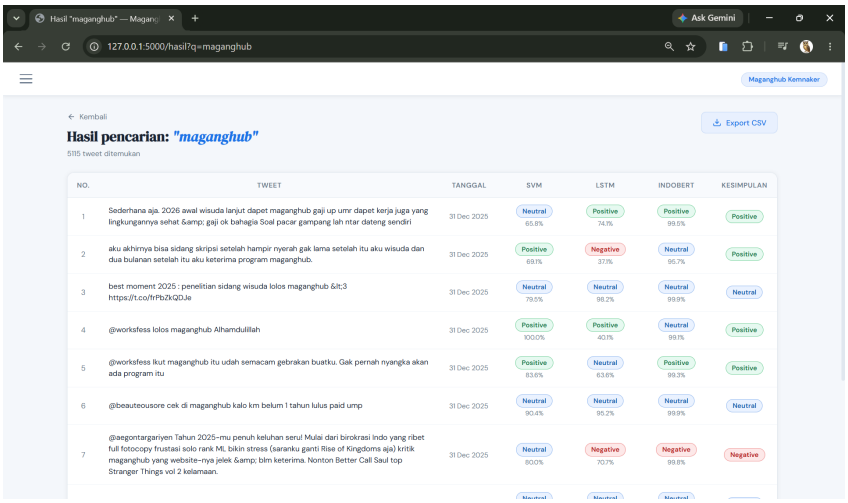


Figure 4. Tweet Search Result Page

Figure 4 represents the Results Page, which renders the output of the keywords searched by the user. This page presents a comprehensive table containing the tweet text, publication date, and a head-to-head comparison of sentiment predictions from the SVM, LSTM, and IndoBERT algorithms. This table also displays the prediction probability percentage from each model along with a final conclusion using a majority-voting mechanism. The system also facilitates downloading the classification results through the "Export CSV" feature.

D. Evaluation Page

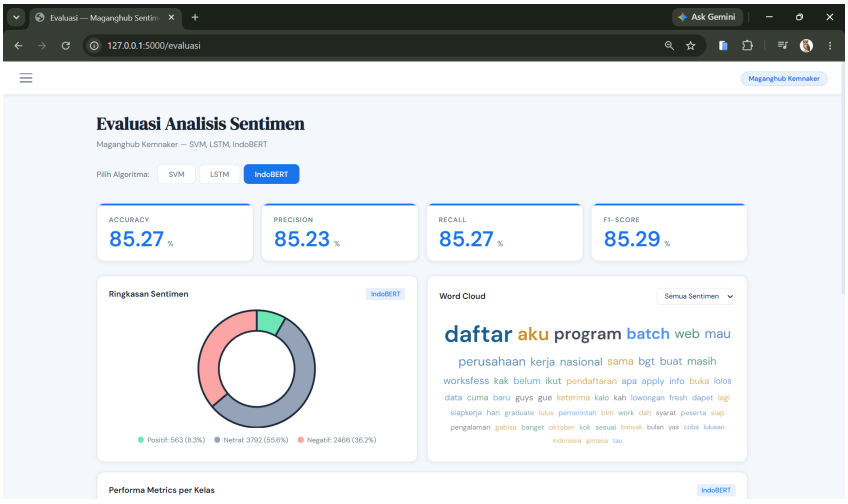


Figure 5. System Evaluation Page (1/2)

DOI: 10.33481/infomans.vxxixx

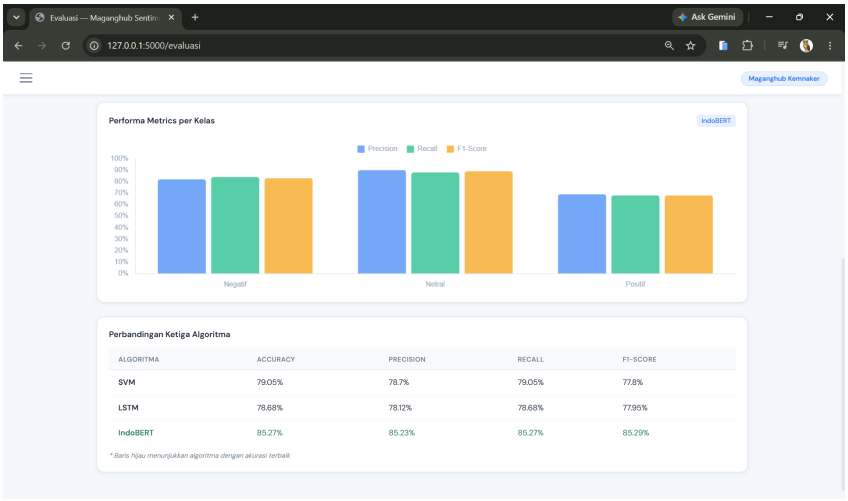


Figure 6. System Evaluation Page (2/2)

Figures 5 and 6 visualize the interface of the Evaluation Page, which focuses on comprehensively presenting algorithm performance metrics. At the top of the page, users are provided with interactive buttons to dynamically display four main metrics: Accuracy, Precision, Recall, and F1-Score for each model. These metrics are accompanied by a Donut Chart visualization representing a summary of sentiment class proportions, as well as a Word Cloud displaying the words with the highest frequency of occurrence in the dataset. Furthermore, at the bottom of the page, the system presents advanced performance analysis through a Bar Chart to evaluate metrics based on each sentiment class (Positive, Neutral, Negative). As a closing feature, this page is equipped with a "Comparison of the Three Algorithms" table that summarizes the overall metric performance of SVM, LSTM, and IndoBERT side by side, complete with visual highlighting of the model that achieved the best accuracy to make it easier for users to draw conclusions.

4. Conclusion

This study successfully compared the performance of the SVM, LSTM, and IndoBERT algorithms in performing sentiment analysis on public opinion regarding Maganghub Kemnaker on the social media platform X. Testing results using the accuracy, precision, recall, and F1-score metrics across 30 scenarios show that IndoBERT achieved the best performance with an accuracy of 85.27%, outperforming SVM at 79.05% and LSTM at 78.68%. IndoBERT's superiority lies in its Transformer architecture based on multi-head self-attention, which is able to understand Indonesian-language context more comprehensively compared to traditional machine learning and sequential deep learning approaches. These sentiment analysis results have been implemented in a web-based dashboard that can be used as an informative public-opinion monitoring tool for Kemnaker and related parties.

Future research is recommended to expand data sources from other social media platforms such as Instagram, Facebook, and TikTok; to compare more advanced Transformer-based algorithms such as RoBERTa, XLM-RoBERTa, or DeBERTa; to add handling for data class imbalance as well as sarcasm detection and emotion analysis; and to develop the dashboard into a system capable of real-time data retrieval.

References

[1] A. F. Aksa, "Program Peningkatan Kemampuan Mahasiswa Menghadapi Dunia Kerja Melalui Kegiatan Magang di Kantor Imigrasi dan Koperasi Sangosay Atambua," *Jurnal Umum Pengabdian Masyarakat*, vol. 2, no. 4, pp. 50–56, Dec. 2023, doi: 10.58290/jupemas.v2i4.164.

This is an open article under

[2] Kementerian Sekretariat Negara Republik Indonesia, " Produk Hukum Berita & Artikel Informasi & Layanan Publik Pendaftaran Peserta Magang Nasional Batch II Resmi Dibuka, Berikut Tahapannya," setneg.go.id. Accessed: Nov. 08, 2025. [Online]. Available: https://www.setneg.go.id/baca/index/pendaftaran_peserta_magang_nasional_batch_ii_resmi_dibuka_berikut_tahapannya

[3] Kementerian Ketenagakerjaan Republik Indonesia, "Pendaftaran Program Magang Nasional Batch I Ditutup, Kemnaker Apresiasi Antusiasme Pendaftar dan Perusahaan," kemnaker.go.id. Accessed: Nov. 14, 2025. [Online]. Available: https://kemnaker.go.id/news/detail/pendaftaran-program-magang-nasional-batch-i-ditutup-kemnaker-apresiasi-antusiasme-pendaftar-dan-perusahaan

[4] J. Cui, Z. Wang, S.-B. Ho, and E. Cambria, "Survey on sentiment analysis: evolution of research methods and topics," *Artif. Intell. Rev.*, vol. 56, no. 8, pp. 8469–8510, Aug. 2023, doi: 10.1007/s10462-022-10386-z.

[5] Y. Mao, Q. Liu, and Y. Zhang, "Sentiment analysis methods, applications, and challenges: A systematic literature review," *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 4, p. 102048, Apr. 2024, doi: 10.1016/j.jksuci.2024.102048.

[6] S. Imaddudin, I. Astuti, and S. Ruhama, "Studi Sentimen Masyarakat terhadap PSSI di Era Erick Thohir menggunakan Algoritma Support Vector Machine (SVM) pada Media Sosial X," *Jurnal Penelitian Multidisiplin Bangsa*, vol. 1, no. 8, pp. 1003–1013, Jan. 2025, doi: 10.59837/jpnmb.v1i8.193.

[7] S. Yahya, Z. Sitorus, M. Iqbal, D. Nasution, and R. F. Wijaya, "Analisis Sentimen Penerapan Deep Learning dan Analisis Sentimen terhadap Gap Kompetensi Lulusan Lembaga Pendidikan dan Pelatihan Vokasi terhadap Dunia Kerja dengan Metode Long Short-Term Memory (LSTM)," *Bulletin of Information Technology*, vol. 6, no. 2, pp. 161–172, Jun. 2025.

[8] A. B. Siva, R. Robet, and L. Hoki, "Comparison of IndoBERT and SVM Performance in Sentiment Analysis of Digital Education Platforms," *sinkron*, vol. 10, no. 1, pp. 64–74, Jan. 2026, doi: 10.33395/sinkron.v10i1.15472.

[9] D. D. Purwanto, "Empirical Evaluation of IndoBERT and LSTM for Sentiment Analysis of Tourism Reviews: A Data-Driven Study on Kenjeran Park," *Jurnal Teknik Informatika (Jutif)*, vol. 7, no. 1, pp. 463–474, Feb. 2026, doi: 10.52436/1.jutif.2026.7.1.4901.

[10] E. Daniati, A. Nugroho, A. Ristyawan, and H. Utama, "Perbandingan Kinerja Algoritma SVM, LSTM, dan Fine-tuned IndoBERT dalam Analisis Sentimen Opini Masyarakat Indonesia terhadap Mobil Listrik," *The Indonesian Journal of Computer Science Research*, vol. 5, no. 1, pp. 1–13, Jan. 2026, doi: 10.59095/ijcsr.v5i1.245.