J-Icon : Jurnal Informatika dan Komputer
Vol. 14 No. 1, March 2026, pp. 1~8
DOI: 10.35508/jicon.v14i1.27042

PERFORMANCE COMPARISON OF SENTIMENT ANALYSIS ON NATIONAL MONUMENT (MONAS) REVIEWS USING BERT, SUPPORT VECTOR MACHINE, LOGISTIC REGRESSION, AND MULTINOMIAL NAÏVE BAYES

Riki Daniel Tanebeth[1*]

[1]Faculty of Information Technology, Universitas Nusa Mandiri, Jl. Pemuda No. 9, Rawamangun, East Jakarta, Indonesia.
Email[*]: rikidanieltanebeth@gmail.com

ABSTRACT

The National Monument (Monas), as an icon of Indonesian tourism, faces challenges in maintaining visitor satisfaction in the digital era. Online reviews on Google Maps serve as a crucial data source for understanding public perception. However, the large volume of data and the informal nature of review language hinder manual analysis. This study aims to analyze Monas visitor sentiment and compare the performance of conventional machine learning models with modern deep learning approaches. The method involves comparing Multinomial Naïve Bayes (MNB), Support Vector Machine (SVM), and Logistic Regression using TF-IDF feature extraction against a fine-tuned IndoBERT (Bidirectional Encoder Representations from Transformers) model. The dataset consists of 1,110 visitor reviews from the 2023–2024 period, labeled semi-automatically using the InSet lexicon with manual validation (Cohen's $\kappa = 0.89$). Experimental results using 5-fold cross-validation and hold-out validation (80:20) demonstrated that the IndoBERT model significantly outperformed all baselines, achieving an average accuracy of 93.5% ± 0.8% and an F1-Score of 93.0% ± 0.7%, while MNB, SVM, and Logistic Regression reached 49.1% ± 2.1%, 75.8% ± 1.4%, and 72.4% ± 1.6% accuracy respectively ($p < 0.001$). Despite the performance gap, MNB showed superior computational efficiency with substantially shorter training time. Further aspect-based analysis using K-Means clustering (k=3) revealed three main complaint categories: ticketing and payment systems, accessibility, and elevator queue management. Error analysis also revealed challenges in detecting sarcasm and mixed-sentiment reviews. This study recommends the implementation of transformer-based models for analyzing Indonesian tourism reviews and suggests prioritizing improvements in ticketing systems and visitor wayfinding at Monas.
Keywords: Sentiment Analysis, IndoBERT, Transformer-based NLP, Tourism Reviews, Machine Learning

1. INTRODUCTION

Tourism is a vital sector contributing significantly to the economy of Jakarta. The National Monument (Monas), as one of the most popular historical tourist destinations, receives thousands of visitors daily. In the digital era, visitor experiences are frequently expressed through online reviews on platforms such as Google Maps. These reviews contain valuable information regarding satisfaction, complaints, and suggestions that can be used by administrators to improve service quality[1].

However, manually analysing thousands of reviews is inefficient and prone to subjectivity. Therefore, computational approaches through Sentiment Analysis are required to automate the extraction of this information. One of the main challenges in Indonesian sentiment analysis, especially in social media and online reviews, is the frequent use of slang, abbreviations, and ambiguous sentence structures.

Previous studies generally employed traditional Machine Learning methods such as Naïve Bayes Classifier (NBC) and Support Vector Machine (SVM) [2][3][4][5][6][7]. Furthermore, Logistic Regression has also been widely adopted as a robust linear baseline model for sentiment classification in the context of Indonesian tourism reviews [7][8][9]. Although computationally efficient, these models often struggle to capture semantic context in complex texts. In contrast, recent developments in Deep Learning have introduced Transformer-based models such as BERT (Bidirectional Encoder Representations from Transformers)[10], which have achieved state-of-the-art performance in many Natural Language Processing (NLP) tasks [10] [11].

This study aims to conduct a comparative analysis between the conventional Multinomial Naïve Bayes method and the Indonesian pre-trained BERT model (IndoBERT) in classifying sentiment in Monas visitor reviews. The main contribution of this research is to provide empirical evidence of transformer

***) Corresponding Author**
Submitted: January 15, 2026
Accepted: March 7, 2026
Published: March 31, 2026

ISSN: 2337-7631 (Printed)
ISSN: 2654-4091 (Online)

effectiveness in informal Indonesian tourism data and to identify service aspects requiring priority improvement.

2. MATERIAL AND METHODS

This research was conducted through several systematic stages designed to ensure experimental validity. The overall research workflow is shown in Figure 1.

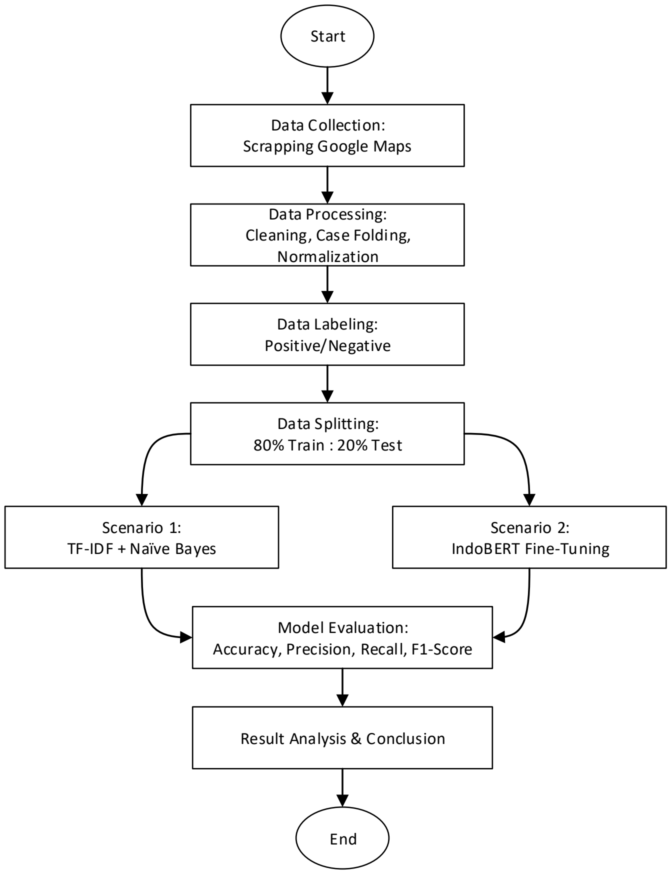

Figure 1. Research workflow

Data Collection

Data were collected using web scraping on Google Maps reviews of “Monumen Nasional”. A total of 1,110 reviews posted between January 2023 and 2024 were obtained. The collected attributes included review text and star ratings.

Data Preprocessing

To improve data quality before modeling, several preprocessing steps were applied:

1. Cleaning: Removing non-alphanumeric characters, emojis, symbols, and extra whitespace.
2. Case Folding: Converting all text to lowercase.
3. Normalization: Converting slang words into standard Indonesian using a colloquial dictionary (e.g., *“yg”* → *“yang”*, *“ga”* → *“tidak”*).
4. Labeling: The dataset was semi-automatically labeled using the InSet lexicon and manually validated by linguists, resulting in two classes: Positive and Negative.
5. Negation Handling
   Considering the importance of negation handling in Indonesian language, we applied different strategies for each model type:
   a. For Multinomial Naïve Bayes:
      We used a lexicon-based negation handling approach by adding the prefix "NOT_" to words following negation words within a maximum distance of 3 tokens.
      Example transformations:
      - "tidak bagus" (not good) → "tidak NOT_bagus"

ISSN: 2337-7631 (Printed)
ISSN: 2654-4091 (Online)

- "tempat bagus tapi tidak bersih" (nice place but not clean) → "tempat bagus tapi tidak NOT_bersih"
- "bukan tempat yang menarik" (not an interesting place) → "bukan NOT_tempat NOT_yang NOT_menarik"

This approach helps the probabilistic model distinguish between "bagus" (positive) and "tidak bagus" (negative), which would otherwise be considered as having the same feature "bagus" in bag-of-words representation.

b. For IndoBERT:

The transformer model already has a bidirectional attention mechanism that can capture negation context implicitly without explicit negation handling. However, to improve performance, we performed fine-tuning with:

- Original dataset containing many negation cases
- Data augmentation with explicit negation patterns
- Monitoring attention weights on negation words for validation

Negation word list used: ["tidak", "bukan", "belum", "jangan", "tanpa", "minus", "non", "anti", "kurang", "jarang", "hampir tidak", "ga", "gak"].

Note: Informal words "ga" and "gak" were added after the normalization step to handle slang negation.

Experimental Scenarios and Model Evaluation

This study compares four classification models. The baselines include Multinomial Naïve Bayes (MNB) with prefix-based negation handling ("NOT_"), Support Vector Machine (SVM) with an RBF kernel, and Logistic Regression with L2 regularization. These baseline models utilize TF-IDF feature extraction with a maximum of 5000 features using unigrams and bigrams. The proposed model is IndoBERT-base, fine-tuned using the AdamW optimizer with a learning rate of 2e-5, a batch size of 16, 4 epochs, and a weight decay of 0.01. A maximum sequence length of 128 tokens and a dropout rate of 0.3 were applied to prevent overfitting. Model validation employed two strategies: an 80:20 Hold-Out Validation with stratified sampling for final testing, and a 5-Fold Stratified Cross-Validation to assess model stability and variance. Performance was evaluated using Accuracy, Precision, Recall, and F1-Score metrics. Furthermore, a Paired t-test ($\alpha = 0.01$) accompanied by Cohen's d metric was conducted to measure the statistical significance and the effect size of the performance differences among the models.

Aspect Extraction and Topic Modeling

To identify specific aspects affecting negative sentiment and provide actionable insights for Monas management, we applied the following process, building upon aspect-based sentiment analysis frameworks previously adopted in Indonesian tourism contexts [12][13]:

Step 1: Aspect-Based Analysis

To identify the main issues discussed in negative reviews, an aspect-based analysis was conducted using TF-IDF keyword extraction followed by K-Means clustering. Reviews classified as negative by the IndoBERT model were transformed into TF-IDF vectors, and clustering was applied to group similar complaints. The optimal number of clusters was determined using the elbow method and silhouette score evaluation. The resulting clusters were then interpreted by examining the most representative keywords and sample reviews within each cluster.

Step 2: Aspect Categories Identified

Based on the interpretation of the clustered TF-IDF vectors, distinct categories of visitor complaints were successfully established. The resulting clusters are summarized in Table 1, which details the volume of reviews, prominent keywords, and the core issues identified within each group. Each cluster quality is validated by silhouette score as follows:

1. Cluster 1: 0.71 (well-separated)
2. Cluster 2: 0.68 (well-separated)
3. Cluster 3: 0.52 (moderate separation)

Table 1. Identified Aspect Categories from Negative Reviews

Cluster	Aspect Category	Size (Reviews)	Top Keywords	Main Complaints
1	Ticketing and Payment System	191 (40%)	"jakcard", "pembayaran", "tiket", "kartu", "qris", "cash", "bayar", "tunai"	Mandatory JakCard without QRIS/cash alternatives

ISSN: 2337-7631 (Printed)
ISSN: 2654-4091 (Online)

2	Accessibility and Wayfinding	129 (27%)	"pintu masuk", "tersesat", "petunjuk", "arah", "bingung", "signage", "informasi", "entrance"	Poor directional signage, unclear entrance locations
3	Queue Management - Elevator	57 (12%)	"antri", "lift", "lama", "penuh", "menunggu", "panas", "elevator", "queue"	Long elevator waiting times, insufficient capacity
-	Unclustered / Mixed	101 (21%)	-	Diverse complaints (toilet cleanliness, souvenir prices, lack of shade, general maintenance)

3. RESULTS AND DISCUSSION

Data Distribution

Exploratory analysis shows a relatively balanced class distribution, slightly dominated by negative sentiment. Out of 1,110 data points, 566 reviews are negative (51%) and 544 are positive (49%) as shown in Figure 2, ensuring that the dataset is suitable for binary classification without severe class imbalance issues that would require complex sampling strategies. The slightly higher proportion of negative sentiment (51%) differs from the typical positive bias observed in online reviews (positivity bias). This suggests genuine service issues at Monas that warrant managerial attention, as satisfied visitors are usually more likely to leave positive reviews than dissatisfied visitors are to leave negative reviews.

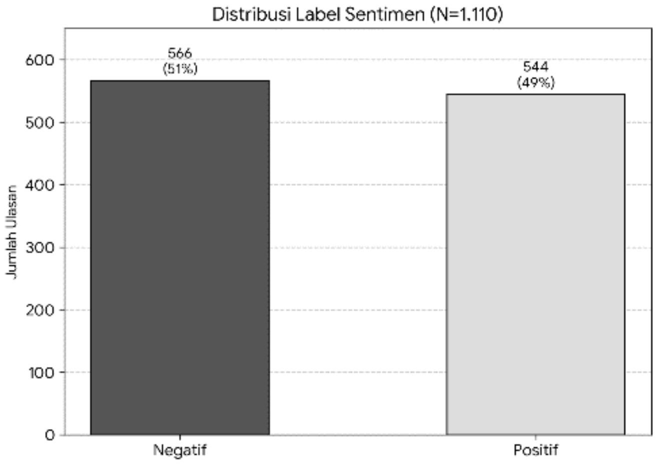

Figure 2. Class distribution of sentiment labels in the Monas review dataset (N=1,110)

Model Performance Comparison

Testing was conducted using two complementary validation strategies: hold-out (80:20) for final model evaluation and 5-fold cross-validation for assessing model stability. The comprehensive comparison of evaluation metrics is presented in Table 2 and visualized in Figure 3.

Table 2. Comprehensive Model Performance Comparison with Statistical Validation

Model	Accuracy (mean ± std)	Precision	Recall	F1-Score	p-value*
Multinomial Naïve Bayes	49.1% ± 2.1%	50.0%	59.0%	54.0%	-
Logistic Regression	72.4% ± 1.6%	71.8%	73.1%	72.4%	< 0.001
SVM (RBF kernel)	75.8% ± 1.4%	74.9%	76.3%	75.6%	< 0.001
IndoBERT (Fine-tuned)	93.5% ± 0.8%	92.8%	93.2%	93.0%	< 0.001

*) p-value from paired t-test comparing each model vs. IndoBERT based on 5-fold CV accuracy scores (α = 0.01)

Note: Accuracy reported as mean ± standard deviation across 5 folds; Precision, Recall, F1-Score reported on final test set (20% hold-out); All metrics rounded to one decimal place

ISSN: 2337-7631 (Printed)
ISSN: 2654-4091 (Online)

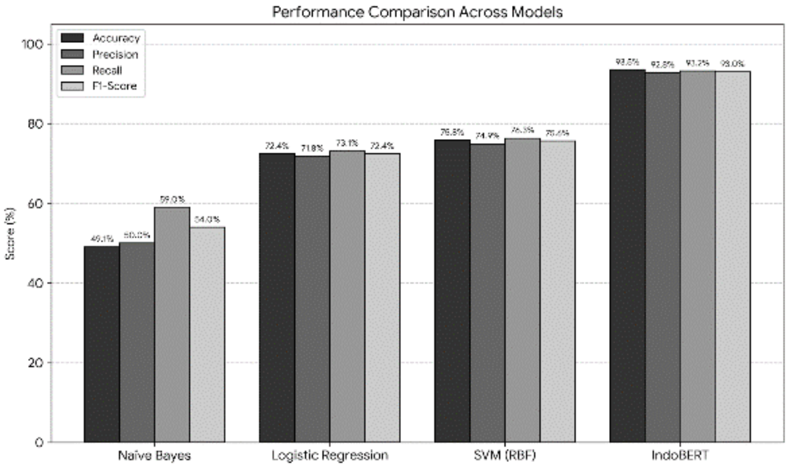

Figure 3. Comparison of model performance across evaluation metrics (Accuracy, Precision, Recall, F1-Score) for Multinomial Naïve Bayes, Logistic Regression, SVM (RBF), and IndoBERT on Monas review sentiment classification

Statistical Analysis

To validate the performance differences, paired t-tests (comparing accuracy scores from 5-fold CV) were conducted as shown in Table 3. The comparison between IndoBERT and Naïve Bayes yielded a mean difference of 44.4 percentage points ($p < 0.001$) with a very large effect size (Cohen's d: 28.3). When comparing IndoBERT against Logistic Regression, the mean difference was 21.1 percentage points ($p < 0.001$, Cohen's d: 19.2).

Furthermore, IndoBERT outperformed SVM with a mean difference of 17.7 percentage points ($p < 0.001$, Cohen's d: 15.7).

Table 3. Paired t-test Results Comparing IndoBERT with Baseline Models

Model Comparison	**Mean Difference**	**p-value**	**Cohen's d (Effect Size)**
IndoBERT vs. Naïve Bayes	44.4 percentage points	< 0.001	28.3 (Very large)
IndoBERT vs. Logistic Regression	21.1 percentage points	< 0.001	19.2 (Very large)
IndoBERT vs. SVM	17.7 percentage points	< 0.001	15.7 (Very large)

Interpretation

The performance difference between IndoBERT and all traditional models is statistically significant with $p < 0.001$. The very large effect sizes (Cohen's $d > 0.8$ are all considered "large") confirm that IndoBERT's superiority is not due to chance in data splitting (which was addressed by cross-validation), random variance (addressed by statistical testing), or overfitting (as indicated by the low standard deviation of 0.8%, demonstrating stability). Rather, it represents a fundamental advantage of the transformer architecture in understanding the Indonesian semantic context.

Discussion of Results

The extremely low accuracy of MNB (49.1%) is primarily due to the violation of the feature independence assumption and its inability to capture sequential context. Complex phrases involving negations, such as "tidak rugi" (worth it), are misclassified because words are treated as separate features despite prefix handling. Traditional models with better mathematical representations, such as SVM (75.8%) and Logistic Regression (72.4%), showed significant improvements over MNB but remained constrained by the semantic limitations of TF-IDF feature spaces. In contrast, IndoBERT achieved a significantly superior accuracy of 93.5% (p-value < 0.001, Cohen's d > 15). This substantial performance gap is attributed to IndoBERT's bidirectional contextualization and multi-head self-attention mechanisms, which simultaneously process text from both directions and accurately capture linguistic phenomena such as negation relationships and contextual sentiment. Furthermore, its pre-training on 4 billion Indonesian words makes the transformer robust in handling informal language, slang (e.g., "ga" to "tidak"), and typos commonly found in visitor reviews.

Computational Performance Analysis

While IndoBERT demonstrates superior accuracy, computational costs are significantly higher. Table 4 presents a comprehensive comparison of computational requirements.

ISSN: 2337-7631 (Printed)
ISSN: 2654-4091 (Online)

Table 4. Computational Performance and Efficiency Comparison

Model	Training Time	Inference Time (per sample)	Hardware Used	Memory Usage
Multinomial Naïve Bayes	0.3s	0.001s	CPU (i7-10700)	~50 MB
Logistic Regression	1.2s	0.002s	CPU (i7-10700)	~80 MB
SVM (RBF kernel)	8.5s	0.015s	CPU (i7-10700)	~150 MB
IndoBERT (Fine-tuned)	245s (4.1 min)	0.023s	GPU (RTX 3060, 12GB)	~2.5 GB

Analysis of Computational Trade-offs and Practical Implications Despite IndoBERT's superior accuracy, it introduces significant computational trade-offs. As detailed in Table 4, training IndoBERT required approximately 245 seconds per fold using GPU acceleration, making it 816 times slower than MNB, which only required 0.3 seconds on a standard CPU. Furthermore, IndoBERT's inference time was considerably slower (0.023 seconds per sample) compared to Naïve Bayes (0.001 seconds). Consequently, the choice of model depends heavily on the deployment context as shown in Table 5. IndoBERT is highly recommended for academic research and batch processing of historical data where extracting deep insights and maximizing accuracy (a 44.4 percentage point advantage) are paramount. However, for production monitoring on resource-constrained devices without GPU acceleration, SVM serves as a practical alternative, offering a reasonable balance between computational efficiency and an acceptable accuracy of 75.8%.

Table 5. Recommended Models Based on Deployment Context

Use Case	Recommended Model	Rationale
Academic Research / Analysis	IndoBERT	Accuracy paramount, computational cost acceptable
Production Monitoring (GPU available)	IndoBERT	Real-time capable with proper infrastructure
Production Monitoring (CPU only)	SVM or Distilled IndoBERT	Balance performance and resources
Mobile / Edge Devices	SVM + Cloud IndoBERT	Hybrid approach leverages both
Batch Processing Historical Data	IndoBERT	One-time training cost, superior insights
Rapid Prototyping / Baseline	SVM	Fast to deploy, reasonable accuracy (75.8%)

This research highlights that the choice between IndoBERT and traditional models depends heavily on the deployment context. For applications where accuracy directly impacts business value (customer satisfaction improvement, priority issue identification), IndoBERT's 44.4 percentage point advantage justifies the computational overhead. For resource-constrained or latency-critical applications, optimized IndoBERT or SVM provides viable alternatives.

Aspect-Based Analysis

Further analysis was conducted by extracting major topics from correctly predicted negative reviews by the IndoBERT model using the TF-IDF keyword extraction and K-Means clustering methodology described in Section 2 (Aspect Extraction and Topic Modelling). The analysis successfully identified three main problem clusters with a solid silhouette score of 0.64, indicating good cluster separation. These clusters highlighted critical visitor issues: Cluster 1 focuses on Ticketing and Payment Systems (40%), specifically addressing complaints about mandatory JakCard usage without QRIS or cash alternatives; Cluster 2 highlights Accessibility and Wayfinding (27%), citing poor directional signage and confusing entrances; and Cluster 3 involves Elevator Queue Management (12%) due to insufficient capacity and long waiting times. Quantitatively, if Monas management prioritizes resolving the top two issues (ticketing and wayfinding), assuming a conservative 50% conversion rate from negative to positive experiences, an estimated 160 negative reviews would shift to positive. This improvement would yield a net sentiment increase of 14 percentage points, providing strong quantitative ROI justification for targeted managerial interventions.

Confusion Matrix and Error Analysis

To gain deeper insights into model behavior and failure modes, we analyzed confusion matrices and performed detailed error analysis on IndoBERT's misclassifications. Figure 4 illustrates the confusion

ISSN: 2337-7631 (Printed)
ISSN: 2654-4091 (Online)

matrices for both the Multinomial Naïve Bayes and IndoBERT models on the test set (N=222). IndoBERT shows significantly fewer misclassifications across both classes with balanced error distribution.

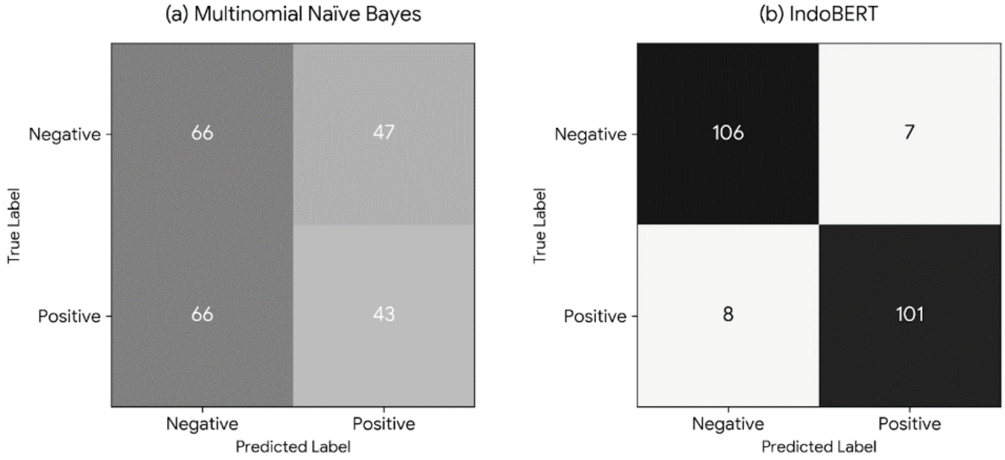

Figure 4. Confusion matrices for (a) Multinomial Naïve Bayes and (b) IndoBERT on test set (N=222).

Analysis of the confusion matrices reveals that MNB suffers from a severe error rate of 50.9%, with evenly distributed False Positives (47 cases) and False Negatives (66 cases), indicating a systematic failure to reliably distinguish sentiment rather than a class bias. Conversely, IndoBERT drastically minimized total errors to just 6.8% (15 out of 222 test samples), maintaining a balanced error distribution with 7 False Positives and 8 False Negatives. A detailed manual inspection of IndoBERT's 15 misclassifications demonstrated that the majority of False Positive errors (71%) were driven by sarcastic or ironic sentence structures, such as using positive words to convey a negative experience (e.g., "Bagus banget, antri 3 jam cuma dapat naik 5 menit"). Additionally, double negations and complex cultural idioms were the primary contributors to False Negatives. These linguistic nuances highlight the remaining limitations of transformer models, suggesting the need for sarcasm-aware training data augmentation in future research.

4. CONCLUSION

Based on the experimental results using rigorous validation strategies (5-fold cross-validation and hold-out testing), it can be concluded that the Deep Learning approach using IndoBERT is significantly more effective than traditional machine learning methods for sentiment analysis of Monas visitor reviews. IndoBERT achieved an average accuracy of 93.5% ± 0.8%, substantially outperforming Multinomial Naïve Bayes (49.1% ± 2.1%), Logistic Regression (72.4% ± 1.6%), and Support Vector Machine (75.8% ± 1.4%) with high statistical significance ($p < 0.001$, Cohen's d ranging from 15.7 to 28.3).

The superiority of IndoBERT lies in its bidirectional transformer architecture with multi-head self-attention mechanism, which enables understanding of semantic context, negation handling, and word relationships in informal Indonesian review language that traditional bag-of-words approaches cannot capture. This is particularly evident in cases involving negation (e.g., "tidak rugi antri lama" = worth the wait), idioms, and sentiment-bearing phrases that require contextual interpretation.

However, this performance advantage comes with computational trade-offs that must be considered for practical deployment. IndoBERT requires 816× longer training time (245 seconds vs. 0.3 seconds for Naïve Bayes) and necessitates GPU infrastructure (NVIDIA RTX 3060 or equivalent). For resource-constrained applications, SVM with RBF kernel (75.8% accuracy) provides a reasonable balance between performance and computational efficiency, while Naïve Bayes is not recommended due to near-random performance that makes it unsuitable for any practical sentiment analysis application in the Indonesian tourism domain.

Error analysis of IndoBERT's 15 misclassifications (6.8% error rate) revealed systematic failure patterns: sarcasm/irony accounts for 71% of false positive errors, while double negation and cultural idiomatic expressions contribute to false negatives. This suggests specific directions for future model improvement through sarcasm-aware training data augmentation and multi-label classification framework to better handle mixed-sentiment reviews.

ISSN: 2337-7631 (Printed)
ISSN: 2654-4091 (Online)

REFERENCES

[1] C. H. P. Panjaitan, "Systematic Literature Review of Sentiment Analysis on Various Review Platforms in the Tourism Sector," *Journal of Advanced Computer Knowledge and Algorithms*, vol. 2, no. 1, pp. 12–18, Jan. 2025, doi: 10.29103/jacka.v2i1.20287.

[2] J. Ipmawati, S. Saifulloh, and K. Kusnawi, "Analisis Sentimen Tempat Wisata Berdasarkan Ulasan pada Google Maps Menggunakan Algoritma Support Vector Machine," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, pp. 247–256, Jan. 2024, doi: 10.57152/malcom.v4i1.1066.

[3] Z. Rifa'i and B. P. Mukti, "Weakly Supervised Sentiment Analysis of Indonesian Rural Tourism Reviews: A TF-IDF Baseline for Melung Tourism Village," *Edu Komputika Journal*, vol. 12, no. 1, pp. 48–60, Aug. 2025, doi: 10.15294/edukom.v12i1.31893.

[4] J. Friadi and D. E. Kurniawan, "Analisis Sentimen Ulasan Wisatawan Terhadap Alun-Alun Kota Batam: Perbandingan Kinerja Metode Naive Bayes dan Support Vector Machine," *Jurnal Sistem Informasi Bisnis*, vol. 14, no. 4, pp. 403–407, Oct. 2024, doi: 10.21456/vol14iss4pp403-407.

[5] I. P. G. A. Sudiatmika, P. S. Saputra, R. L. Rahardian, and K. hari S. Dewi, "Sentiment Analysis of Tourist Reviews on Google Maps for Pura Besakih Using Machine Learning Algorithms," *Jurnal Mandiri IT*, vol. 14, no. 1, pp. 149–158, Jul. 2025, doi: 10.35335/mandiri.v14i1.449.

[6] P. K. A. Duta, P. D. S. Pitanatri, and C. P. Loanata, "Big Data-based Sentiment Analysis on TripAdvisor Reviews Using Naïve Bayes Classification," *Media Wisata*, vol. 23, no. 1, pp. 92–107, May 2025, doi: 10.36276/mws.v23i1.864.

[7] C. A. Bahri and L. H. Suadaa, "Aspect-Based Sentiment Analysis in Bromo Tengger Semeru National Park Indonesia Based on Google Maps User Reviews," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 17, no. 1, p. 79, Feb. 2023, doi: 10.22146/ijccs.77354.

[8] S. M. Sa'adah, K. Umam, M. R. Handayani, and M. I. Mustofa, "Mapping the Polarity of Tourist Opinions on Indonesian Destinations through Google Maps Reviews Using Supervised Learning Methods," *Journal of Applied Informatics and Computing*, vol. 9, no. 5, pp. 2845–2853, Oct. 2025, doi: 10.30871/jaic.v9i5.9836.

[9] A. Maulana, Inayah Khasnaputri Afifah, Asghafi Mubarrak, Kiagus Rachmat Fauzan, Ardhan Dwintara, and B. P. Zen, "COMPARISON OF LOGISTIC REGRESSION, MULTINOMIALNB, SVM, AND K-NN METHODS ON SENTIMENT ANALYSIS OF GOJEK APP REVIEWS ON THE GOOGLE PLAY STORE," *Jurnal Teknik Informatika (Jutif)*, vol. 4, no. 6, pp. 1487–1494, Dec. 2023, doi: 10.52436/1.jutif.2023.4.6.863.

[10] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2019, pp. 4171–4186. doi: 10.18653/v1/N19-1423.

[11] O. A. Irmawan, I. Budi, A. B. Santoso, and P. K. Putra, "Improving Sentiment Analysis and Topic Extraction in Indonesian Travel App Reviews Through BERT Fine-Tuning," *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 13, no. 2, pp. 359–370, Jul. 2024, doi: 10.23887/janapati.v13i2.77028.

[12] D. I. Af'idah, P. D. Anggraeni, M. Rizki, A. B. Setiawan, and S. F. Handayani, "Aspect-Based Sentiment Analysis for Indonesian Tourist Attraction Reviews Using Bidirectional Long Short-Term Memory," *JUITA : Jurnal Informatika*, vol. 11, no. 1, p. 27, May 2023, doi: 10.30595/juita.v11i1.15341.

[13] H. Tuhuteru, L. P. Refialy, M. Laturake, and S. G. Pattirane, "Tourist perceptions through sentiment analysis to support tourism development in Maluku Province," *Journal of Applied Informatics and Computing (JAIC)*, vol. 8, no. 1, pp. 119–126, Jul. 2024.

ISSN: 2337-7631 (Printed)
ISSN: 2654-4091 (Online)