Sentiment Analysis of X Users Toward Electric Motorcycles Using SVM and BERT Algorithms

Calvin Adiwinata[1], Afiyati Afiyati[2*]

[1,2]*Computer Science Faculty, Mercu Buana University, Indonesia*

*corr-author: afiyati.reno@mercubuana.ac.id

Abstract – This study presents a comparative analysis of Support Vector Machine (SVM) and Bidirectional Encoder Representations from Transformers (BERT) for sentiment analysis on electric motorcycles in Indonesia using data from the social media platform X, formerly known as Twitter. The dataset of 128,711 tweets collected between 2015 and 2024 was refined through systematic preprocessing, reducing the corpus to 38,954 entries after data cleaning, tokenization, and feature selection. The objective was to evaluate algorithm performance in classifying public sentiment, with metrics including accuracy, precision, recall, and computational efficiency. Results showed that SVM achieved higher overall accuracy 89.74% with strong precision for positive sentiment 91%, while BERT, specifically the IndoBERT variant, demonstrated superior recall for negative sentiment 91% despite slightly lower accuracy 87.90%, effectively capturing nuanced contextual language, such as sarcasm, informal expressions, and emotionally ambiguous statements that require deeper semantic understanding beyond literal word meanings. Computational analysis revealed that SVM required approximately 53 minutes of CPU training, compared to BERT's 3.3 hours on GPU. The study suggests that SVM is optimal for rapid, resource-constrained applications, whereas BERT excels in detailed contextual analysis. These findings guide stakeholders in selecting algorithms based on analytical priorities, such as monitoring public reception or addressing consumer concerns.

Keywords: electric motorcycles, sentiment analysis, SVM, BERT, social media.

I. INTRODUCTION

The development of electric vehicles, particularly electric motorcycles, has become a primary focus in various countries, including Indonesia [1]. In line with increasing global efforts to reduce carbon emissions and dependence on fossil fuels, electric motorcycles are regarded as an environmentally friendly and efficient transportation solution. The Government of Indonesia has actively supported this transition through various policies and incentives aimed at promoting the adoption of electric vehicles [2]. However, public acceptance of electric motorcycles still faces challenges, as the success of this technology largely depends on societal perceptions and sentiment.

Social media platforms, specifically X (formerly known as Twitter), have emerged as key channels for the public to express opinions and discuss diverse issues. According to the Indonesian Internet Service Providers Association (APJII), the majority of the Indonesian population utilizes the internet to access social media platforms [3], thereby providing a valuable source of public opinion data. Recent studies have demonstrated the significant potential of social media data in understanding public sentiment. For example, one study [4] successfully collected more than 5,800 sentiment data points from social media over a period of three months, while another study [5] analyzed the sentiment towards electric vehicles from X data over a five-year period, achieving an accuracy of 95.79%.

While these studies provide useful insights, most of them tend to focus on a single algorithm or use relatively small datasets, leaving an important gap in comparative studies, especially in the context of Indonesian social media and electric motorcycles [6,7]. This research seeks to address that gap by critically comparing two well-established sentiment analysis algorithms: Support Vector Machine (SVM) and Bidirectional Encoder Representations from Transformers (BERT). This study distinguishes itself by using a significantly larger dataset comprising 128,711 tweets collected between 2015 and 2024, thus offering deeper and more representative coverage across time.

SVM has proven effective in sentiment analysis tasks, particularly when dealing with complex data patterns and limited datasets [8,9]. Its ability to handle non-linear data through mathematical functions and its resistance to overfitting make it highly suitable for analyzing social media text [10,11]. In contrast, BERT offers advantages in understanding the complex nuances of language through its bidirectional approach, which considers both preceding and succeeding contexts during text analysis [12,13]. Its capability to manage informal language variations and contextual subtleties renders it especially appropriate for analyzing social media data [14,15].

This study uses IndoBERT P2, a variant of BERT tailored for Indonesian language tasks. Developed by the IndoNLU team, IndoBERT is trained on a large corpus of Indonesian texts, enabling it to better capture the linguistic and cultural context. Previous research has shown that IndoBERT significantly outperforms both multilingual and general BERT models when applied to Indonesian sentiment datasets [16-18].

By conducting a performance comparison between SVM and IndoBERT P2 on a comprehensive dataset from the X platform, this study aims to provide practical insights into the strengths and limitations of each approach. Unlike prior studies, this study not only focus on algorithmic performance but also highlight the implications of algorithm choice for real-world applications in sentiment analysis. The results are intended to guide both researchers and practitioners in selecting the most appropriate method for analyzing public sentiment in Indonesian social media contexts.

II. METHOD

This study uses sentiment analysis to understand public opinion about electric motorcycles on the X platform. The research process consists of several key stages, as shown in Fig. 1.

A. Data Collection

Data were collected through a web scraping technique using Tweet Harvest [19]. The initial phase involved defining search parameters that included the keyword *‘motor listrik’* and a data collection period from 2015 to 2024. Tweet Harvest automatically scraped Twitter’s search result pages and extracted relevant information from each tweet, including tweet text, posting time, and user username. A total of 128,711 tweets were collected.

B. Data Preprocessing

The data preprocessing stage consisted of several sub-stages aimed at cleaning and preparing the dataset for analysis.

1) Initial Data Cleaning: Empty records (NULL/NaN) were removed. Only tweets containing the exact phrase "motor listrik" as a continuous term were kept. Tweets that mentioned "motor" and "listrik" separately were excluded, as they may not directly refer to electric motorcycles. Duplicate tweets were also removed, reducing the dataset to 55,031 unique entries.

Text Preprocessing: Text cleaning was performed using regular expressions (regex) to remove unnecessary elements such as usernames, hashtags, retweet tags, links, symbols, numbers, and punctuation. All text was then converted to lowercase. Informal or slang words were normalized using a custom dictionary. Tokenization, the process of splitting text into individual words, was done using the NLTK word_tokenize function. Stopwords (commonly used words like "saya", "yang", "dan") were removed using both English and Indonesian stopword lists, plus additional social media terms. This eliminated 241,347 stopwords across 803 unique types. Stemming, which reduces words to their base form (e.g., kemahalan → mahal), was done using the Sastrawi library. Words that were either too long (more than 20 characters) or appeared very rarely (fewer than 2 times) were also removed, helping reduce noise in the data. After all cleaning steps and another round of duplicate removal, the final dataset contained 38,954 tweets.

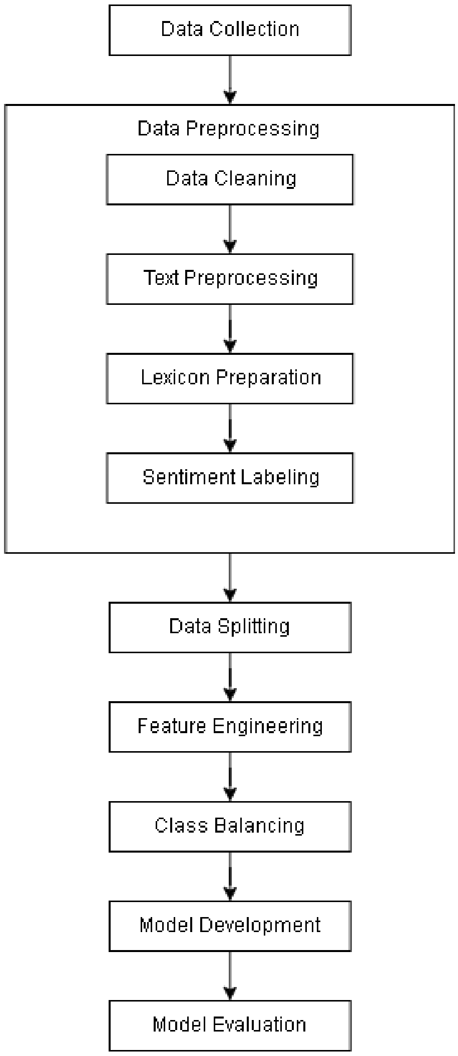

Fig. 1 Research methodology

C. Lexicon Preparation

The lexicon dictionary was compiled from three primary sources: the GitHub repository *MohFahmi27* [20], *InSet* [21], and the *angelmetanosaa* dataset [22]. The weight values in the dictionary ranged from -5 to +5. Manual lexicon additions were performed for specific words such as *bisa*, *buat*, and *lebih*, which were assigned positive weights. After integration and duplicate removal, the final dictionary comprised 9,125 words from an initial total of 30,664 words.

Sentiment labeling was then performed using the compiled lexicon, which was manually validated by the researchers. Notwithstanding the manual validation, potential bias inherent in the lexicon-based approach remains due to limitations such as polysemy (for example, the word *ringan* may refer to either motorcycle weight or performance), the inability to detect irony or sarcasm, and the introduction of personal subjectivity through custom lexicon additions.

D. Sentiment Labeling

Based on the compiled lexicon, the sentiment labeling process generated three classes: positive (values greater than 0), negative (values less than 0), and neutral (value equal to 0). The classification yielded 23,571 positive tweets, 10,929 negative tweets, and 4,454 neutral tweets, indicating an imbalanced class distribution.

E. Data Splitting

The dataset was partitioned into training and testing sets using an 80:20 ratio while maintaining class stratification. This stratification ensured a proportional representation of each sentiment class in both subsets.

F. Feature Engineering

To prepare the text for machine learning, it was converted into numerical features using the TF-IDF method (Term Frequency-Inverse Document Frequency), which gives importance to words based on how often they appear in a tweet and how unique they are across the dataset.

For the SVM model, multiple configurations of TF-IDF were explored, including variations in max_features (1000, 3000, 5000) and the word combination range (ngram_range) ((1,1), (1,2)). However, for the BERT model, only max_features of 3000 and 5000 were used, with ngram_range fixed at (1,1). Given the high computational cost of BERT training, the max_features setting was restricted to 3000 and 5000 to manage training time while ensuring a sufficiently rich feature representation.

G. Class Balancing

Because the dataset was imbalanced (more positive tweets), the SMOTE (Synthetic Minority Over-sampling Technique) method was used to generate synthetic examples of the minority classes (negative and neutral). This was applied only to the training set to avoid biasing the model evaluation. The original distribution was preserved in the test set to simulate real-world conditions.

H. Model Development

Model training was executed using two approaches, via CPU and GPU on Google Colab free version. For the CPU-based SVM model, the scikit-learn library was employed, while the GPU-based model utilized cuML. The training process involved testing 144 parameter combinations, including the number of TF-IDF features, ngram_range, C value, kernel, and gamma. To overcome memory limitations on Google Colab for the GPU-based SVM, the training data were processed in chunks, and cross-validation results from each chunk were aggregated to obtain an average score.

For the indoBERT P2 model, training was conducted using a combination of 96 parameter configurations, including learning rate, batch size, number of epochs, optimizer, maximum text length, and number of TF-IDF features.

I. Model Evaluation

Each model was evaluated using suitable criteria. For the SVM model, performance was judged based on cross-validation accuracy, standard deviation, and the gap between training and test scores (overfitting margin). A 5-fold stratified cross-validation was applied to enhance reliability. However, for the BERT model, due to Google Colab's time constraints (maximum 12 hours per session [23]), cross-validation was not used. Instead, validation loss and accuracy were used to assess model performance. In addition to accuracy, both models were compared based on training time and hardware usage (CPU/GPU memory), giving a broader view of the trade-offs involved.

III. RESULT AND DISCUSSION

A. Sentiment Analysis Result

Based on the lexicon employed, the sentiment analysis yielded 23,571 positive, 10,929 negative, and 4,454 neutral sentiments, as illustrated in Fig. 3.

B. Evaluation of the SVM Model

Experiments with the SVM model were conducted using two implementation variants: CPU-based (using scikit-learn) and GPU-based (using cuML), with an exploration of 144 parameter combinations (e.g., variations in the number of TF-IDF features, ngram_range, C, kernel, and gamma) and a 5-fold cross-validation approach.

Table I presents a comparative analysis between the CPU and GPU implementations.

The best CPU-based SVM configuration achieved a mean test accuracy of 88.172%, albeit with a longer training time (3213.135 seconds) and lower peak memory usage (8529 MB). In contrast, the GPU-based SVM model, with its optimal parameter settings, achieved a slightly lower accuracy (86.640%) but required substantially less training time (112.727 seconds), albeit with higher main memory usage (13323 MB) while maintaining moderate GPU memory usage (1367 MB). These results underscore the computational advantage of GPUs in accelerating matrix-intensive operations in SVM, even though they may incur higher memory consumption.

C. Evaluation of the BERT Model

Experiments with the BERT model involved 96 parameter combinations, including adjustments to the learning rate, batch size, number of epochs, optimizer, maximum text length, and number of TF-IDF features (Fig. 2).

Table II compares the performance of the best BERT model configurations on CPU and GPU. However, the CPU-based BERT training could not be completed on Google Colab Free due to the maximum instance runtime limit of 12 hours [23].

As shown in Fig. 3, the session terminated before completing a single epoch. Consequently, a direct comparison with the GPU-based model could not be performed, and the CPU results are marked as '-' in the table.

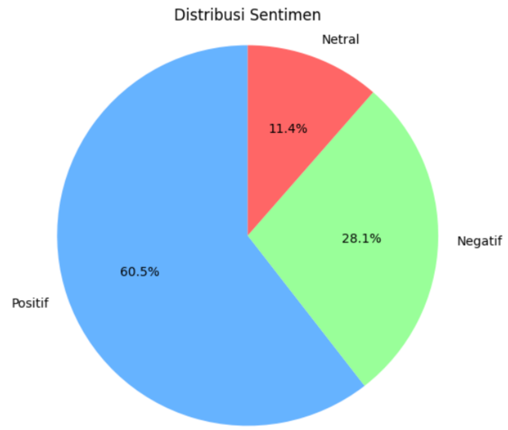

Fig. 2 Sentiment distribution

TABLE I
COMPARISON OF BEST SVM PERFORMANCE ON CPU AND GPU

Model	**Parameter**	**Mean Test Accuracy**	**Training Time (s)**	**Peak Memory Usage (MB)**	**GPU Memory Usage (MB)**
CPU (Scikit-Learn)	{max_features: 5000, ngram_range: (1,1), C: 10, kernel: RBF, gamma: 1}	88.172%	3213.135	8529	-
GPU (CuML)	{max_features: 3000, ngram_range: (1,1), C: 10, kernel: RBF, gamma: scale}	86.640%	112.727	13323	1367

TABLE II
COMPARISON OF BEST BERT PERFORMANCE ON CPU AND GPU

Model	**Parameter**	**Validation Loss**	**Validation Accuracy**	**Best Epoch**	**Training Time (s)**	**Peak Memory Usage (MB)**	**GPU Memory Usage (MB)**
CPU	-	-	-	-	-	-	-
GPU	{learning_rate: 5.00E-05, batch_size: 32, num_epochs: 15, optimizer: sgd, max_length: 512, tfidf_max_features, 5000}	0.3597	88.885%	13	11948.888	2484.398	3203

Conversely, the GPU-based BERT model achieved its best performance using the configuration: {learning_rate: 5.00E-05, batch_size: 32, num_epochs: 15, optimizer: SGD, max_length: 512, tfidf_max_features: 5000}. This configuration yielded a validation loss of 0.3597 and a validation accuracy of 88.885% at the best epoch (13). The GPU-based model required a training time of 11948.888 seconds, with a peak memory usage of 2484.398 MB and GPU memory usage of 3203 MB. Although the training time was considerable, it remained substantially lower than the CPU scenario, which could not complete a full epoch within the prescribed runtime.

D. Comparison between the Best SVM and BERT Models

After determining the optimal configurations for each algorithm, a detailed comparative analysis was conducted between the best CPU-based SVM model and the best GPU-based BERT model. Fig. 4 presents the detailed classification report for the SVM model, including metrics such as precision, recall, and F1-score for each sentiment class.

Meanwhile, Fig. 5 displays the confusion matrix for the SVM model, illustrating the distribution of correctly and incorrectly classified samples across the sentiment classes.

Similarly, Fig. 6 shows the classification report for the BERT model, outlining the performance metrics that reflect its capability in sentiment classification.

Finally, Fig. 7 presents the confusion matrix for the BERT model, depicting the pattern of classification errors and the accuracy in predicting each sentiment category.

The best CPU-based SVM model achieved an overall accuracy of 89.74%, slightly outperforming the GPU-based BERT model in terms of overall accuracy. In the SVM model, the positive class was identified with particularly high performance, demonstrating a precision of 0.91, a recall of 0.96, and an F1-score of 0.94. The confusion matrix for the SVM model indicates that out

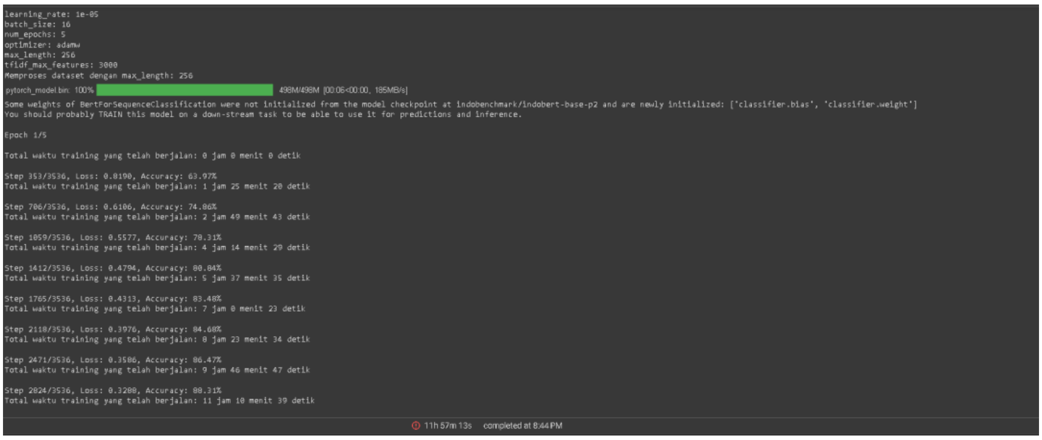

Fig. 3 BERT using CPU on Google Colab Free

	precision	recall	f1-score	support
Negatif	0.90	0.87	0.88	2186
Netral	0.78	0.64	0.70	891
Positif	0.91	0.96	0.94	4714
accuracy			0.90	7791
macro avg	0.86	0.82	0.84	7791
weighted avg	0.89	0.90	0.89	7791

Fig. 4 Classification report SVM

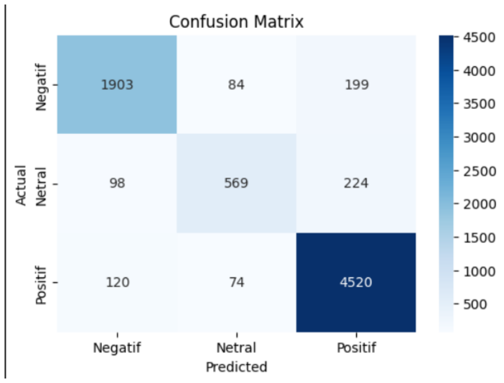

Fig. 5 Confusion matrix SVM

	precision	recall	f1-score	support
Negatif	0.82	0.91	0.87	2186
Netral	0.67	0.75	0.71	891
Positif	0.96	0.89	0.92	4714
accuracy			0.88	7791
macro avg	0.82	0.85	0.83	7791
weighted avg	0.89	0.88	0.88	7791

Fig. 6 Classification report BERT

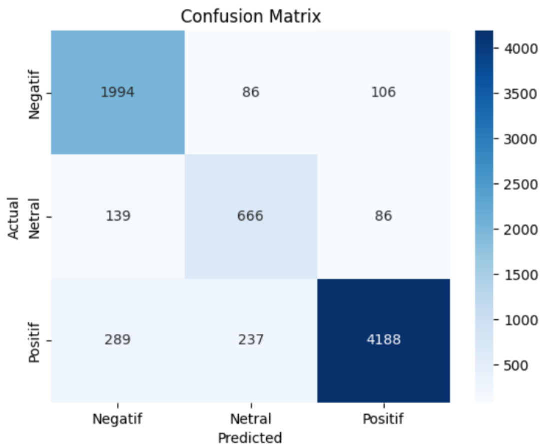

Fig. 7 Confusion matrix BERT

of 4,714 positive samples, 4,520 were correctly classified, with only 194 misclassified. This suggests a strong capability in recognizing positive sentiment.

In contrast, the BERT model exhibited a distinct advantage in detecting negative sentiments. It achieved a recall of 0.91 for the negative class, though its precision was lower at 0.82 compared to the SVM model. The confusion matrix for BERT reveals that 1,994 negative samples were correctly identified, while 292 negative samples were erroneously classified as either positive or neutral. Both models, however, encountered challenges with neutral sentiment classification, achieving F1-scores of 0.70 for SVM and 0.71 for BERT.

These performance differences may be attributed, in part, to the inherent imbalance in the dataset, 61% of the samples were positive, 28% negative, and 11% neutral. The high recall observed in the positive class for the SVM model suggests a bias toward the dominant class, potentially inflating its overall accuracy without ensuring balanced generalization across all sentiment classes.

Furthermore, the error patterns differ between the two models. The SVM model misclassified 224 neutral samples as positive, whereas the BERT model made fewer such errors, misclassifying only 86 neutral samples as positive. For the negative class, the SVM model produced 199 false positives (negative samples wrongly classified as positive), while the BERT model recorded 106 similar errors. These findings indicate that although the SVM model achieves higher overall accuracy, the BERT model demonstrates a more nuanced ability to differentiate between negative and positive sentiments, particularly in the context of informal social media language.

Although the SVM model achieved the highest overall accuracy, this result may be influenced by the imbalanced dataset, where the majority of samples (61%) were labeled as positive. This dominance likely caused the model to become biased toward positive sentiment, which can inflate accuracy but reduce performance on minority classes. This tendency is consistent with findings [24], which stated that SVM tends to prioritize the majority class. Even after the application of SMOTE, challenges in predicting the neutral class persist and require additional strategies for improvement.

In contrast, the BERT model, built on a transformer-based architecture, is more adept at understanding context and the semantic meaning of words, even in informal or ambiguous sentences. This strength allowed BERT to perform significantly better in classifying negative sentiments, as evidenced by its higher recall value. BERT can capture subtle cues or tones that traditional models like SVM might overlook, especially in short and noisy social media text. This observation is supported by [25], who evaluated IndoBERT, IndoBERTweet, and IndoGPT, transformer models specifically adapted to the informal Indonesian language domain, and found they outperformed other models in detecting sarcasm and informal expressions on Twitter.

Moreover, BERT has shown greater effectiveness in classifying emotionally ambiguous or neutral tweets compared to both traditional machine learning and other deep learning approaches. In a comparative study by [26] demonstrated that BERT outperformed CNN and LSTM in sentiment analysis related to emotionally charged topics like the Ebola outbreak. Ref. [27] revealed that BERT surpassed models such as SVM, CNN, RNN, and BiLSTM in a swarm-intelligence-based sentiment classification task on social media data.

Therefore, although BERT had slightly lower accuracy overall, it demonstrated a more balanced and context-aware performance. The difference in behavior between the two models highlights how SVM tends to favor majority sentiment, while BERT provides more nuanced classification across all sentiment classes, particularly in identifying negative expressions.

IV. CONCLUSION

This study demonsrates that both SVM and BERT exhibit distinct strengths in sentiment analysis of electric motorcycle discourse on X, with the choice of algorithm contingent on analytical objectives, computational resources, and contextual requirements. The CPU-based SVM achieved marginally higher overall accuracy 89.74% and excelled in detecting positive sentiments with a precision of 0.91 and a recall of 0.96, making it ideal for applications prioritizing rapid analysis of public enthusiasm, such as monitoring marketing campaigns or product launches. In contrast, the GPU-based BERT, though slightly lower in overall accuracy 87.90%, outperformed SVM in identifying negative sentiments with a recall 0.91, proving valuable for early detection of public complaints or critical feedback. However, BERT incurred significantly higher computational costs, requiring 11948.888 seconds of GPU training time and 3203 MB of GPU memory, compared to SVM’s 3213.135 seconds on CPU with 8529 MB RAM. This highlights a critical trade-off: SVM offers computational efficiency and faster processing for resource-constrained environments, while BERT provides deeper contextual understanding at the expense of hardware demands. Methodologically, the study acknowledges limitations in cross-validation consistency, as BERT evaluation was restricted to a single data split due to Google Colab’s 12-hour runtime limit, whereas SVM utilized robust 5-fold cross-validation. A key limitation of the lexicon-based labeling approach is lexical bias, where certain words are given sentiment scores regardless of context. For example, a word like 'ringan' can mean "lightweight" in a technical sense or "not serious" in a different tone, yet the lexicon treats it the same in all cases. Furthermore, both models struggle with sarcasm or ironic expressions, which require deeper contextual understanding that neither TF-IDF nor BERT in its current form fully captures. This aligns with the findings of [28], who highlighted that sarcasm remains difficult to detect automatically due to its reliance on implicit contextual and pragmatic cues. Future research should address dataset imbalance through advanced techniques like hybrid sampling and expand the lexicon to include domain-specific slang and technical terms. Real-time sentiment monitoring systems and cross-platform analyses such as Instagram, TikTok could enhance policymaking relevance. Additionally, exploring lightweight BERT variants or distributed training frameworks may mitigate computational barriers. For practitioners, the decision hinges on operational priorities, SVM for efficiency and positive sentiment tracking, or BERT for nuanced negative feedback analysis where contextual subtleties and sarcasm detection are critical. These insights advance the strategic implementation of sentiment analysis in Indonesia’s evolving electric vehicle sector, balancing technical capabilities with pragmatic resource allocation.

ACKNOWLEDGEMENT

Gratitude is extended to all individuals and organizations that provided support and participated in this research. Sincere thanks are offered for the invaluable advice, constructive criticism, and input that contributed significantly to the study's success.

REFERENCES

[1] V. Tulus Pangapoi Sidabutar, “Kajian pengembangan kendaraan listrik di Indonesia: prospek dan hambatannya,” *Jurnal Paradigma Ekonomika*, vol. 15, no. 1, pp. 21–38, May 2020, doi: 10.22437/paradigma.v15i1.9217.

[2] A. F. Riyadi, F. R. Rahman, M. A. Nofa Pratama, M. K. Khafidli, and H. Patria, “Pengukuran Sentimen Sosial Terhadap Teknologi Kendaraan Listrik: Bukti Empiris di Indonesia,” *EXPERT: Jurnal Manajemen Sistem Informasi dan Teknologi*, vol. 11, no. 2, p. 141, Dec. 2021, doi: 10.36448/expert.v11i2.2171.

[3] APJII, “Survey Internet APJII 2024,” 2024.

[4] Y. Sartika Sari, “Sentiment Analysis of Public Service Performance in Jakarta Using the Naïve Bayes Algorithm and Support Vector Machine,” *International Journal of Multidisciplinary Research and Publications (IJMRAP)*, vol. 7, no. 3, pp. 49–55, 2024, [Online]. Available: http://site.com

[5] G. N. Hendrawan and H. Kusniyati, “Evaluasi Performa Naive Bayes dan SVM dalam Analisis Sentimen Kendaraan Listrik di Media Sosial Twitter Evaluating the Performance of Naive Bayes and SVM in Sentiment Analysis of Electric Vehicles on Twitter Social Media,” *Journal of Computing Engineering, System and Science) e-ISSN*, vol. 9, no. 1, pp. 299–313, 2024, [Online]. Available: www.jurnal.unimed.ac.id

[6] A. Nata, D. Wicaksono, and D. Jayus, “Analisis Sentimen Publik Indonesia Terhadap Motor Listrik pada Media Sosial Twitter,” *THETA OMEGA: Journal of Electrical Engineering*, vol. 4, no. 1, p. 2023, 2023, doi: 10.31002/jeecit.v4i1.7526.

[7] R. Merdiansah, Siska, and A. Ali Ridha, “Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT,” *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI*, vol. 7, no. 1, pp. 221–228, 2024, doi: 10.55338/jikomsi.v7i1.2895.

[8] M. H. Ali Al-Abyadh *et al.*, "Deep Sentiment Analysis of Twitter Data Using a Hybrid Ghost Convolution Neural Network Model," *Comput Intell Neurosci*, vol. 2022, 2022, doi: 10.1155/2022/6595799.

[9] T. S. Sabrila, Y. Azhar, and C. S. K. Aditya, "Analisis Sentimen Tweet Tentang UU Cipta Kerja Menggunakan Algoritma SVM Berbasis PSO," *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 7, no. 1, pp. 10–19, Jan. 2022, doi: 10.14421/jiska.2022.7.1.10-19.

[10] K. Ahmed *et al.*, "Exploiting Stacked Autoencoders for Improved Sentiment Analysis," *Applied Sciences (Switzerland)*, vol. 12, no. 23, Dec. 2022, doi: 10.3390/app122312380.

[11] Q. Yang and C. Liu, "Application of twin objective function SVM in sentiment analysis," in *Frontiers in Artificial Intelligence and Applications*, IOS Press BV, Dec. 2020, pp. 221–228. doi: 10.3233/FAIA200786.

[12] A. Bello, S. C. Ng, and M. F. Leung, "A BERT Framework to Sentiment Analysis of Tweets," *Sensors*, vol. 23, no. 1, Jan. 2023, doi: 10.3390/s23010506.

[13] N. Liu and J. Zhao, "A BERT-Based Aspect-Level Sentiment Analysis Algorithm for Cross-Domain Text," *Comput Intell Neurosci*, vol. 2022, 2022, doi: 10.1155/2022/8726621.

[14] R. Duan, Z. Huang, Y. Zhang, X. Liu, and Y. Dang, "Sentiment Classification Algorithm Based on the Cascade of BERT Model and Adaptive Sentiment Dictionary," *Wirel Commun Mob Comput*, vol. 2021, 2021, doi: 10.1155/2021/8785413.

[15] M. S. Sayeed, V. Mohan, and K. S. Muthu, "BERT: A Review of Applications in Sentiment Analysis," Jun. 01, 2023, *Ital Publication*. doi: 10.28991/HIJ-2023-04-02-015.

[16] A. S. P. Braja and A. Kodar, "Implementasi Fine-Tuning BERT untuk Analisis Sentimen terhadap Review Aplikasi PUBG Mobile di Google Play Store," *J I M P - Jurnal Informatika Merdeka Pasuruan*, vol. 7, no. 3, pp. 120–128, 2023, doi: 10.51213/jimp.v7i3.779.

[17] H. K. Putra, M. A. Bijaksana, and A. Romadhony, "Deteksi Penggunaan Kalimat Abusive Pada Teks Bahasa Indonesia Menggunakan Metode IndoBERT," *e-Proceeding of Engineering*, vol. 8, no. 2, pp. 3028–3038, 2021, [Online]. Available: https://api.semanticscholar.org/CorpusID:237062279

[18] M. Syahputra, A. Putera Kemala, and D. Ramdhan, "Clickbait Detection in Indonesia Headline News Using IndoBERT and RoBERTa," *Jurnal Riset Informatika*, vol. 5, pp. 425–430, Jun. 2023, doi: 10.34288/jri.v5i3.556.

[19] H. Satria, "Tweet Harvest," Github. Accessed: Dec. 23, 2024. [Online]. Available: https://github.com/helmisatria/tweet-harvest

[20] M. Fahmi, "Kamus Lexicon 1," Github. Accessed: Dec. 23, 2024. [Online]. Available: https://github.com/MohFahmi27/Sentiment-Analysis-for-Bahasa-using-Lexicon-Based-Approach/blob/main/data/datasetAnalysis/lexicon-word-dataset.csv

[21] F. Koto, "Inset (Indonesia Sentiment Lexicon)," Github. Accessed: Dec. 23, 2024. [Online]. Available: https://github.com/fajri91/InSet

[22] A. M. Afinda, "Kamus Lexicon 2," Github. Accessed: Dec. 23, 2024. [Online]. Available: https://github.com/angelmetanosaa/dataset/blob/main/lexicon_negative.csv

[23] C. K. Vu, "How can I prevent Google Colab from disconnecting?," stackoverflow. Accessed: Jan. 15, 2025. [Online]. Available: https://stackoverflow.com/questions/57113226/how-can-i-prevent-google-colab-from-disconnecting?page=2&tab=scoredesc#tab-top

[24] D. Andriyani, A. Faqih, and S. E. Permana, "The Effect of SMOTE Application on Support Vector Machine Performance in Sentiment Classification on Imbalanced Datasets," *Journal of Artificial Intelligence and Engineering Applications*, vol. 4, no. 2, pp. 752–757, 2025, doi: 10.59934/jaiea.v4i2.742.

[25] F. Rahman and A. S. Girsang, "IndoBERTweet for Sarcasm: Evaluating Domain-Adapted Transformers for Indonesian Twitter Sarcasm Classification," *Journal of Logistics, Informatics and Service Science*, vol. 11, no. 2, pp. 155–164, 2024, doi: 10.33168/JLISS.2024.0210.

[26] A. Mirugwe *et al.*, "Sentiment Analysis of Social Media Data on Ebola Outbreak Using Deep Learning Classifiers," *Life*, vol. 14, no. 6, Jun. 2024, doi: 10.3390/life14060708.

[27] N. Sureja, N. Chaudhari, P. Patel, J. Bhatt, T. Desai, and V. Parikh, "Hyper-tuned Swarm Intelligence Machine Learning-based Sentiment Analysis of Social Media," *Engineering, Technology and Applied Science Research*, vol. 14, no. 4, pp. 15415–15421, Aug. 2024, doi: 10.48084/etasr.7818.

[28] A. Afiyati, A. Azhari, A. K. Sari, and A. Karim, "Challenges of Sarcasm Detection for Social Network: a Literature Review," vol. 8, no. 2, pp. 169–178, Nov. 2020, doi: 10.30595/juita.v8i2.8709.