Evaluating Hybrid GA-SVM Feature Selection for Indonesian Sentiment Classification Using LSTM

Siti Mujilahwati[1*], Noor Zuraidin Bin Mohd Safar[2]

[1]*Informatics Engineering, Faculty of Science and Technology, Universitas Islam Lamongan, Indonesia*

[2]*Faculty of Computer Science and Information Technology, Universiti Tun Husain Onn Malaysia, Malaysia*

*corr-author: moedjee@unisla.ac.id

Abstract - The high dimensionality and noisy characteristics of Indonesian social media text present significant challenges for sentiment classification models. Redundant and irrelevant features may reduce classification efficiency and negatively affect model generalization performance. This study evaluates a hybrid wrapper-based feature selection approach that integrates Genetic Algorithm (GA) and Support Vector Machine (SVM) to optimize TF-IDF feature representations before classification using Long Short-Term Memory (LSTM). The experiments were conducted on 1,918 Indonesian Twitter comments related to SARS-CoV-2 sentiment, consisting of 1,044 negative and 874 positive labels. The proposed GA-SVM mechanism reduced the feature space from 35,343 to 17,931 selected features. Two evaluation scenarios were employed in this study. Under the hold-out train-test split evaluation, the GA-SVM+LSTM model achieved the best accuracy of 91.41% using a learning rate of 0.0001. Meanwhile, the 10-fold cross-validation evaluation produced an average accuracy of 89.41%, indicating stable generalization performance across different data partitions. The experimental results also show that the proposed feature selection approach improved computational efficiency by reducing training time from 263.64 seconds to 173.54 seconds. Overall, the findings indicate that hybrid GA-SVM feature selection can effectively improve TF-IDF-based sentiment classification performance for Indonesian social media text.

Keywords: Sentiment analysis, Indonesian language, genetic algorithm, SVM, LSTM, feature selection, hybrid model.

I. INTRODUCTION

Sentiment analysis has become an important research area for understanding public opinion expressed through social media platforms [1-4]. In Indonesia, social media data are widely used to analyse public responses toward social, political, and health-related issues due to the large volume of user-generated textual content [5-7] . However, sentiment classification for Indonesian social media text remains challenging because the language used in online communication is often informal, noisy, and unstructured. The frequent use of slang words, abbreviations, inconsistent spelling, and non-standard writing styles may increase data complexity and reduce classification performance [8-11].

Another major challenge in sentiment analysis is the high dimensionality of textual feature representations [12, 13] . Text preprocessing and feature extraction methods such as Term Frequency–Inverse Document Frequency (TF-IDF) often generate large feature spaces containing redundant and irrelevant information [14, 15] . High-dimensional feature representations may negatively affect computational efficiency, increase training complexity, and reduce model generalization capability [13, 16]. Therefore, feature selection plays an important role in identifying representative features while reducing noise and unnecessary attributes from textual datasets [17].

Various feature selection approaches have been investigated in sentiment analysis studies, including filter-based, wrapper-based, and hybrid methods [17-19]. Among these approaches, wrapper-based methods generally provide better feature subset optimization because they evaluate feature combinations directly using classifier performance [20, 21]. Previous studies have demonstrated that Genetic Algorithm (GA) is effective for exploring large feature search spaces and identifying representative feature subsets [22, 23]. In addition, Support Vector Machine (SVM) has frequently been employed as a fitness evaluation mechanism in wrapper-based feature selection due to its strong classification capability in high-dimensional text data [24, 25].

Deep learning models have also been widely applied in sentiment classification tasks because of their ability to learn complex feature interactions from textual representations [26-28] . Among various deep learning approaches, Long Short-Term Memory (LSTM) has shown promising performance in text classification studies [29-31]. Previous work by [32] reported that GA-SVM-based feature selection improved sentiment classification performance under a hold-out evaluation

setting. However, previous studies have primarily focused on performance improvement under limited evaluation configurations and have provided limited discussion regarding model robustness, comparative feature selection evaluation, and generalization performance across different validation scenarios.

In addition, although hybrid wrapper-based feature selection approaches have been explored in several studies [20-21], limited attention has been given to evaluating their effectiveness for Indonesian noisy social media text using multiple validation schemes, particularly under both hold-out and cross-validation settings. Furthermore, the impact of optimized TF-IDF feature representations on LSTM-based sentiment classification performance remains insufficiently discussed in comparative experimental studies.

Based on these considerations, this study evaluates a hybrid GA-SVM feature selection approach for Indonesian sentiment classification using LSTM. In the proposed approach, Genetic Algorithm is employed to explore the feature search space, while Support Vector Machine is used as a fitness evaluation mechanism to identify representative TF-IDF feature subsets. The selected features are subsequently used as input for the LSTM classification model. The experiments were conducted using both hold-out train-test split and 10-fold cross-validation scenarios to evaluate classification performance, robustness, and computational efficiency.

The main contributions of this study are summarized as follows:

1. Evaluating the effectiveness of hybrid GA-SVM feature selection for Indonesian sentiment classification using LSTM under multiple evaluation scenarios.
2. Providing comparative analysis between hybrid GA-SVM and several conventional feature selection approaches, including Information Gain, Chi-square, and standard Genetic Algorithm.
3. Investigating the impact of optimized feature subsets on classification performance and computational efficiency for Indonesian noisy social media text.

The remainder of this paper is organized as follows. Section II describes the research methodology, including preprocessing, feature extraction, hybrid GA-SVM feature selection, and LSTM classification. Section III presents the experimental results and discussion. Finally, Section IV concludes the study and provides recommendations for future work.

II. METHOD

The overall research framework used in this study is presented in Fig. 1. The proposed approach consists of text preprocessing, TF-IDF feature extraction, hybrid GA-SVM feature selection, LSTM-based classification, and performance evaluation using hold-out and 10-fold cross-validation scenarios.

A. Data Collection

The dataset used in this study consists of Indonesian Twitter comments related to SARS-CoV-2 sentiment. The data were collected from social media platforms and manually labeled into two sentiment categories, namely positive and negative sentiments. In total, the dataset contains 1,918 text samples, consisting of 1,044 negative comments and 874 positive comments. The collected dataset was subsequently prepared for preprocessing and feature extraction stages. Although the dataset size is relatively limited for deep learning experiments, the study focuses primarily on evaluating the impact of feature optimization on classification behaviour rather than developing large-scale language representation models.

B. Text Preprocessing

Text preprocessing was performed to reduce noise and improve the quality of textual data before feature extraction. Several preprocessing steps were applied, including case folding, text cleaning, tokenization, stop word removal, and stemming. Case folding was used to convert all characters into lowercase format, while text cleaning removed unnecessary symbols, punctuation, URLs, and special characters. Tokenization separated sentences into individual words, followed by stop word removal to eliminate irrelevant common words. Finally, stemming was applied to transform words into their root forms to reduce vocabulary variations in Indonesian text data.

C. Frequency-Based Feature Representation

This research uses Term Frequency-Inverse Document Frequency (TF-IDF) to transform textual data into numerical representations. The TF-IDF weight for a term t in document d is calculated by multiplying the frequency of occurrence of the term (Term Frequency) by the inverse of its document frequency (Inverse Document Frequency) [33]. Mathematically, this weighting is formulated as in (1):

$$w_{t,d} = tf_{t,d}\ x\ \log\left(\frac{N}{df_t}\right) \quad (1)$$

Where $w_{t,d}$ is the final weight of feature t in document d. $tf_{t,d}$ is the number of occurrences of term t in document d. N is the total number of documents in the dataset. df_t is the number of documents containing term t [34, 35].

Although LSTM architectures are traditionally designed to capture sequential dependencies from ordered token representations, the objective of this study is different. The primary focus is to evaluate the effectiveness of the proposed GA-SVM feature optimization framework on high-dimensional textual features. Therefore, TF-IDF was intentionally selected as the input representation because it provides an interpretable and sparse feature space that allows the impact of feature selection to be directly observed. In this context, the LSTM model is utilized as a non-linear classifier to learn complex interactions among optimized feature subsets rather than to model linguistic sequence information.

D. Feature Selection Stage: Hybrid Wrapper GA-SVM

The hybrid GA-SVM feature selection stage was employed to identify representative TF-IDF feature subsets before the LSTM classification process. In this approach, the Genetic Algorithm was used to explore candidate combinations of features. Although the final classification model is an LSTM, SVM was selected as the fitness evaluator because wrapper-based optimization requires repeated fitness evaluations over hundreds of generations. Using LSTM as the fitness evaluator would substantially increase computational cost and optimization time. Previous studies have shown that SVM provides reliable discriminative evaluation for high-dimensional sparse text representations and can effectively guide feature subset selection. The optimized feature subset is subsequently validated using LSTM to assess its generalizability across a different classification architecture.

1) Population Initialization: Randomly generates chromosomes, where each gene (bit 0 or 1) represents whether a feature is selected or not.

2) Fitness Function Evaluation (SVM): Each candidate feature subset generated by the Genetic Algorithm was evaluated using an SVM classifier under a train-test split evaluation scheme. The SVM classification accuracy was used as the fitness value to measure the quality of the corresponding feature subset as in (2).

$$Fitness = Accuracy\ (SVM(x_{subset,y})) \quad (2)$$

where $x_{subset,y}$ is the TF-IDF features that have been selected by a particular chromosome in the GA. To reduce computational complexity during the GA optimization process, fitness evaluation was performed using a fixed train-test split rather than nested cross-validation. Since the optimization involved evaluating many candidate feature subsets across 200 generations, repeated cross-validation would significantly increase computational requirements. Model robustness was subsequently assessed using a separate 10-fold cross-validation procedure on the final selected feature subset.

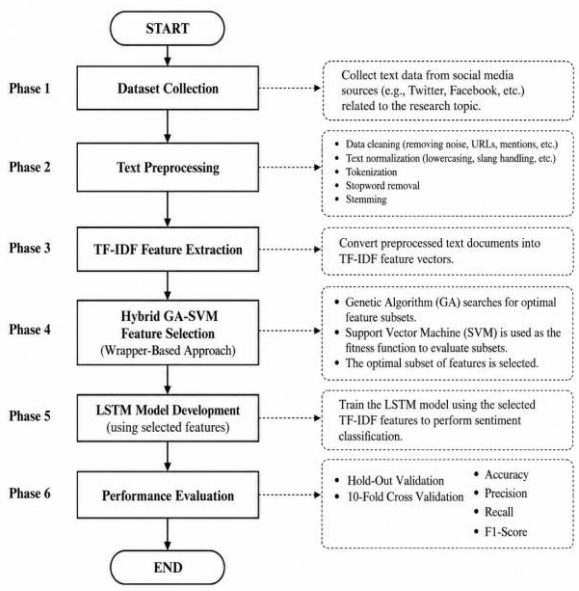


Fig. 1 Research framework of the Evaluation hybrid GA-SVM and LSTM-based sentiment classification model

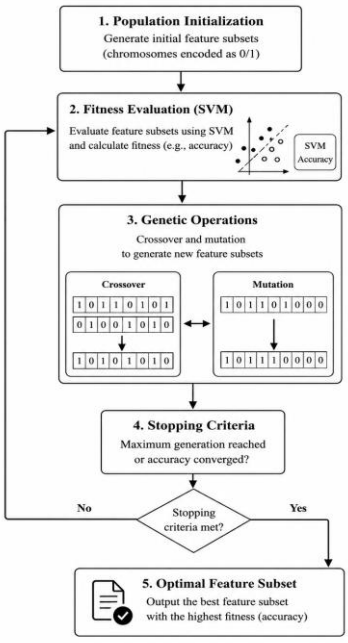


Fig. 2 Workflow of the hybrid GA-SVM feature selection process.

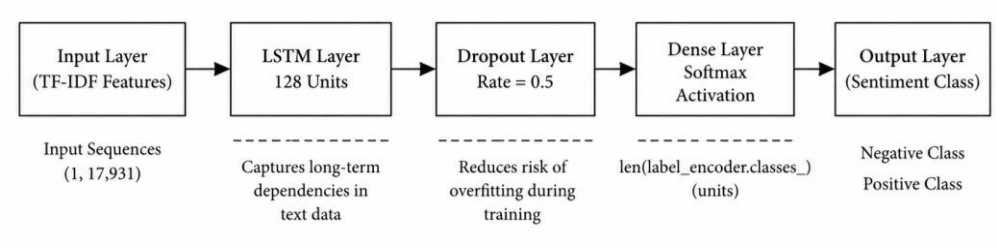


Fig. 3 Architecture LSTM proposed

3) Genetic Operations: The crossover and mutation operations were applied to generate new offspring populations and increase feature subset diversity during the optimization process.

4) Stopping Criterion: The process repeats until the maximum number of generations is reached or the fitness value converges. The result is the Optimal Feature Subset.

The parameter configuration used in the hybrid GA-SVM feature selection process is presented in Table I.

The optimal feature subset obtained from the hybrid GA-SVM process was subsequently used as input for the LSTM classification model in the final evaluation stage.

E. LSTM Classification

This subsection describes the architectural design of the LSTM model proposed in this research. The model architecture Fig. 3 is designed to capture long-term dependencies in text data through the utilization of LSTM layers and is optimized to reduce the risk of overfitting and improve the model's generalization ability in classification tasks [36].

Description of the LSTM architecture used in this research:

1) Input Sequences: Text data is represented as a sequence of feature vectors of size (1, 17,931), where each element represents a feature extracted from the text (e.g., TF-IDF).

2) LSTM Layer (128 Units): An LSTM layer with 128 memory units is used to learn sequential patterns and capture long-term dependencies in the text data, thus better understanding the context between features.

3) Dropout (0.5): A dropout layer with a 50% rate is applied to reduce the risk of overfitting during the training process by randomly deactivating some units.

4) Dense Layer (Fully Connected): The dense layer maps the LSTM feature representations to the output class space, with the number of units adjusted to the number of target classes.

5) SoftMax Activation: The SoftMax activation function is used to generate probability values for each output class.

6) Output Layer: The model produces a final prediction of two classes, namely the positive and the negative class, based on the highest probability generated.

F. Performance and Robustness Evaluation

Model performance evaluation was conducted to assess the extent to which the proposed LSTM model can accurately classify text data into positive and negative classes. This evaluation process aims to measure the model's effectiveness in generalizing to test data and ensure that the model not only produces accurate predictions overall but also consistently within each class [37]. Therefore, the evaluation was conducted using several relevant measurement metrics to provide a comprehensive picture of the model's performance. For this reason, in this research an evaluation model will be created, including:

1) The model was evaluated using standard metrics for text classification:

2) Accuracy (3), Precision (4), Recall (5), and F1-Score (6).

$$\text{Accuracy} = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \quad (3)$$

$$\text{Precision} = \frac{T_p}{T_p + F_p} \quad (4)$$

$$\text{Recall} = \frac{T_p}{T_p + F_n} \quad (5)$$

$$\text{F1} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

T_p: True Positive, instances where the model correctly identifies the positive class.

T_n: True Negative, example where the model correctly identifies the negative class.

F_p: False Positive, an event where the model mistakenly predicted the positive class.

F_n: False Negative, an event where the model mistakenly predicted a negative class.

TABLE I
PARAMETER OF GENETIC ALGORITHM (GA-SVM)

No	Parameter	Value
1	Initial TF-IDF Features	35,343
2	Population Size	150
3	Parent Size (n_parent)	30
4	Mutation Rate	0.01
5	Number of Generation (n_gen)	200
6	Fitness Evaluation Classifier	SVM
7	SVM Kernel	Linear

3) K-Fold Cross Validation: To ensure model stability across different data subsets. For each fold, the metrics calculated include accuracy, precision, recall, and F1-score, based on the error matrix of the model prediction results. The average value of these ten folds provides a comprehensive picture of the accuracy, consistency, and generalization ability of the tested model. Mathematically, the average accuracy value can be formulated as follows:

$$Accuracy_{avg} = \frac{1}{K} \sum_{i=1}^{K=10} Accuracy_i \quad (7)$$

The use of 10-Fold Cross Validation in this research helps ensure that the model's performance does not depend solely on a single data distribution but reflects the model's true ability in classifying texts with various language features and contexts.

4) Significance Test: To compare the performance of the hybrid GA-SVM + LSTM model against a baseline model (LSTM without feature selection or LSTM with a filtering method such as Information Gain [38], Chi-Square [39], and GA [22]).

III. RESULT AND DISCUSSION

This section presents the experimental results of the proposed hybrid GA-SVM feature selection approach for Indonesian sentiment classification using the LSTM model. The analysis focuses on the effectiveness of the feature selection process, the impact of learning rate configurations on classification performance, and robustness evaluation using both hold-out validation and 10-fold cross-validation schemes.

A. Convergence Analysis of the GA-SVM Optimization

Fig. 4 illustrates the convergence behavior of the Genetic Algorithm during the feature optimization process across 200 generations.

The convergence curve demonstrates that the GA optimization process progressively improved the fitness score throughout the evolutionary iterations. Rapid improvement occurred during the early generations, indicating effective exploration of the feature search space. After approximately generation 174, the fitness values became relatively stable, suggesting that the optimization process had reached convergence and successfully identified near-optimal feature subsets.

The absence of significant performance degradation after convergence also indicates that the GA-SVM mechanism was relatively resistant to premature convergence. Minor fluctuations observed in later generations were caused by stochastic crossover and mutation operations, which helped maintain population diversity during optimization.

The optimization process reduced the original feature space from 35,343 features to 17,931 selected features, corresponding to approximately 49.24% dimensionality reduction. Despite the substantial reduction, the retained features were still able to preserve important semantic information required for sentiment classification.

The relatively large number of retained features suggests that sentiment information in Indonesian social media text is distributed across various lexical forms, including slang expressions, abbreviations, informal writing styles, and context-dependent word combinations. Unlike structured textual corpora, social media texts contain diverse linguistic variations that contribute to sentiment discrimination. Therefore, aggressive dimensionality reduction may remove useful semantic information and negatively affect classification performance.

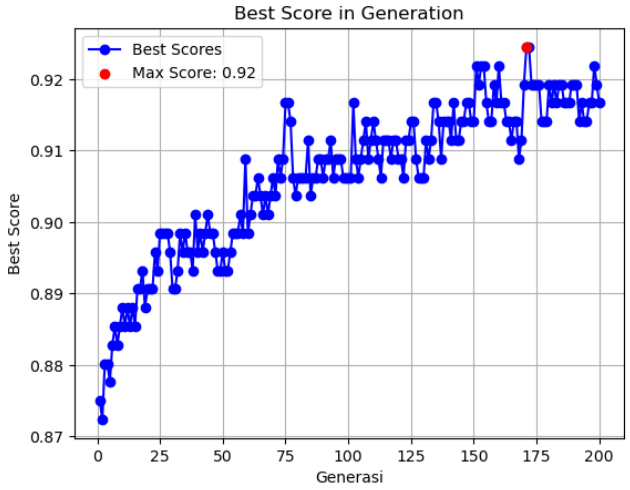


Fig. 4 Best fitness score progression during the GA-SVM feature optimization process over 200 generations.

B. Results of Sentiment Classification Using LSTM (Hold-Out Evaluation)

This section presents the sentiment classification results obtained using the LSTM model with two learning rate configurations, namely 0.0001 and 0.001. The selected feature subsets generated from the hybrid GA-SVM feature selection process were used as input representations for the LSTM classification model. The evaluation was performed using a hold-out train-test split validation scheme as in Table II.

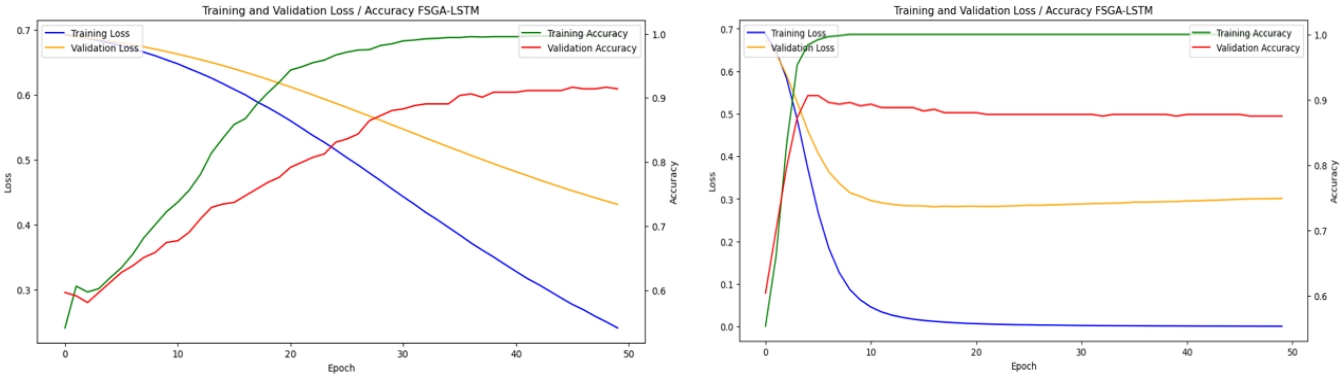
Based on Table II, the learning rate of 0.0001 achieved the best classification performance, with an accuracy of 91.41%, precision of 91.41%, recall of 91.41%, and F1-score of 91.39%. In contrast, the learning rate of 0.001 resulted in lower performance across all evaluation metrics.

Fig. 5 presents the training and validation loss and accuracy curves of the GA-SVM-LSTM model under both learning rate configurations. At a learning rate of 0.0001, the training and validation curves exhibit a more stable convergence pattern with relatively small differences between training and validation performance. This indicates that the optimization process was able to achieve more stable classification behaviour and improved model generalization on unseen data.

Conversely, the learning rate of 0.001 produced larger fluctuations between training and validation performance, indicating less stable optimization behaviour during the learning process. A larger learning rate tends to update model weights more aggressively, potentially reducing model stability during training and negatively affecting classification performance.

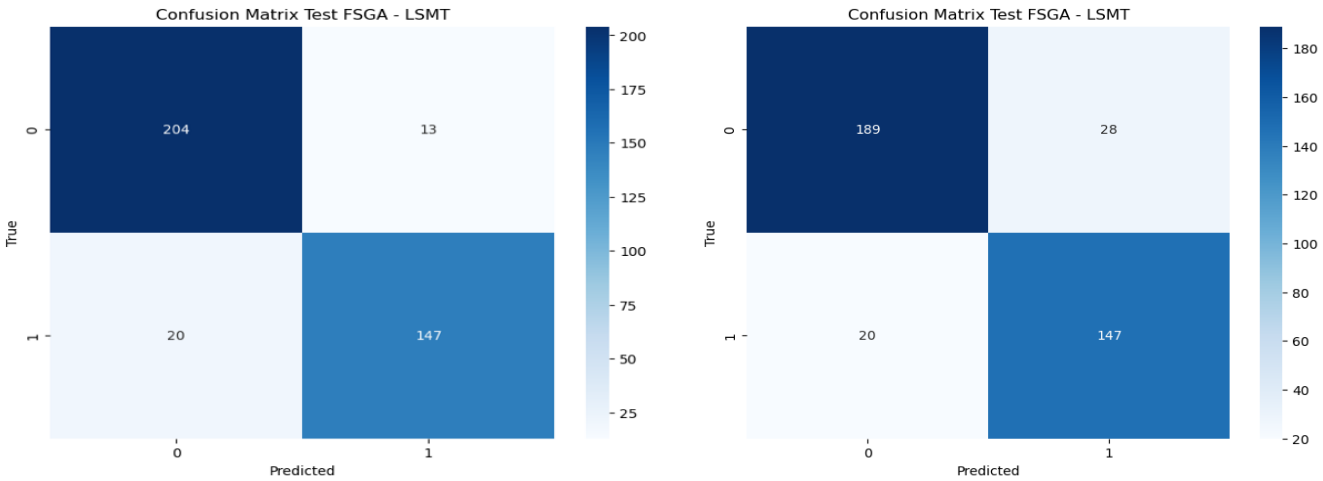
TABLE II
RESULT EVALUATION LSTM USING HYBRID FEATURE SELECTION (GA-SVM)

Learning Rate	Time Training	Accuracy	Recall	Precision	F-Score
0.0001	138.23140358924866	91.41	91.41	91.41	91.39
0.001	147.23239493370056	84.11	84.11	84.24	84.15



(a) Lr 0.0001 **(b) Lr 0.001**

Fig. 5 Graph training and validation loss/accuracy GA-SVM-LSTM



(a) Lr 0.0001 **(b) Lr 0.001**

Fig. 6 Confusion matrix GA-SVM-LSTM

The confusion matrices shown in Fig. 6 further demonstrate the classification behaviour of the GA-SVM-LSTM model. Using a learning rate of 0.0001, the model produced fewer misclassifications for both positive and negative sentiment classes compared to the learning rate of 0.001. These results confirm that a smaller learning rate provides a more stable optimization process for the proposed classification model.

The superior performance achieved using a learning rate of 0.0001 indicates that smaller weight updates enable more stable optimization when training LSTM models on high-dimensional TF-IDF feature representations. The confusion matrix further reveals that the model successfully classified most sentiment instances with relatively few false predictions, suggesting that the selected feature subset preserved discriminative sentiment information required by the classifier.

C. Results of 10-Fold Cross-Validation Evaluation

The robustness of the proposed LSTM model was evaluated using a 10-fold cross-validation scheme. This evaluation provides a more reliable estimate of model generalization by testing performance across multiple data partitions. The following Table III shows the evaluation results with 10-fold cross validation.

Fig. 7 illustrates the distribution of accuracy across 10 folds. The results show minor fluctuations with most folds concentrated around 0.89, indicating stable model performance and good generalization ability."

The LSTM model achieved an average accuracy of 89.41% with a standard deviation of 0.0129, indicating stable performance across different folds. The relatively low variance suggests that the model is not overly sensitive to variations in training and testing splits. Accuracy values across folds range from 0.8715 to 0.9237, showing minor fluctuations in performance. Most folds produce results close to the mean, indicating consistent generalization behaviour. Similar stability is also observed in precision, recall, and F1-score metrics.

The close balance between precision and recall indicates that the model does not exhibit strong bias toward any specific class and is capable of handling both positive and negative sentiment effectively. Overall, these results confirm that the LSTM model trained with the GA-SVM-selected feature subset is robust and generalizable for Indonesian sentiment classification tasks.

TABLE III
10-FOLD CROSS-VALIDATION EVALUATION RESULTS

Metrik	Mean	Std Dev	Minimum	Maximum
Accuracy	0.894103	0.012854	0.871486	0.923695
Precision	0.900659	0.011713	0.881989	0.927686
Recall	0.889221	0.014435	0.863645	0.919671
F1-Score	0.892012	0.013548	0.867905	0.922413

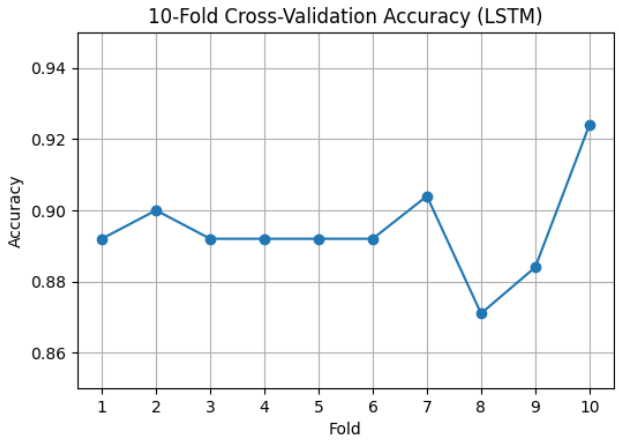


Fig. 7 Distribution of accuracy across 10 folds

Error analysis revealed that most misclassifications occurred among sentiment instances containing overlapping emotional characteristics and ambiguous linguistic expressions. Short texts with implicit sentiment cues were particularly difficult to classify because they provided limited contextual information. These findings are consistent with challenges commonly observed in sentiment analysis of social media data, where informal writing styles and context-dependent meanings often increase classification complexity.

D. Comparative Performance of Feature Selection Methods

Table IV presents the comparative classification accuracy achieved by different feature selection methods evaluated using the same preprocessing pipeline and LSTM architecture. All feature selection methods were evaluated using identical preprocessing procedures, training-testing partitions, LSTM architecture, optimizer settings, batch size, epoch configuration, and learning rate to ensure a fair comparison.

Fig. 8 shows that the hybrid GA-SVM method achieved the highest classification accuracy of 91.41%, outperforming conventional feature selection approaches including PCA, Chi-Square, GA, and RFE.

Among the evaluated methods, PCA produced the lowest performance because dimensionality reduction was performed based on variance representation rather than class discriminative capability. Consequently, several informative textual features may have been removed during transformation. Meanwhile, RFE demonstrated relatively better performance because the feature elimination process considered iterative feature contribution during model training.

The superior performance of the hybrid GA-SVM method can be attributed to its ability to balance exploration and exploitation during feature subset optimization. The Genetic Algorithm explored various combinations of textual features, while SVM evaluated the discriminative quality of each subset. This interaction enabled the model to retain semantically informative terms while eliminating redundant features that contributed minimally to classification performance.

Qualitative examination of the selected features suggests that the GA-SVM mechanism consistently retained emotionally informative terms, repetitive emphasis patterns, and context-specific sentiment expressions. By combining the global search capability of Genetic Algorithms with the discriminative evaluation power of SVM, the proposed approach was able to identify feature subsets that improved classification performance while maintaining important semantic information. This explains the substantial performance improvement observed over PCA, Chi-Square, conventional GA, and RFE approaches.

IV. CONCLUSION

This study evaluated the effectiveness of a hybrid Genetic Algorithm–Support Vector Machine (GA-SVM) feature selection approach for Indonesian sentiment classification using an LSTM model on high-dimensional textual datasets. The experimental results demonstrated that the proposed optimization strategy successfully reduced the original feature space from

TABLE IV
COMPARATIVE PERFORMANCE OF DIFFERENT FEATURE SELECTION

Feature Selection Method	Accuracy (%)
PCA	71.0
Chi-Square	74.0
GA	72.0
RFE	78.0
GA-SVM	91.41

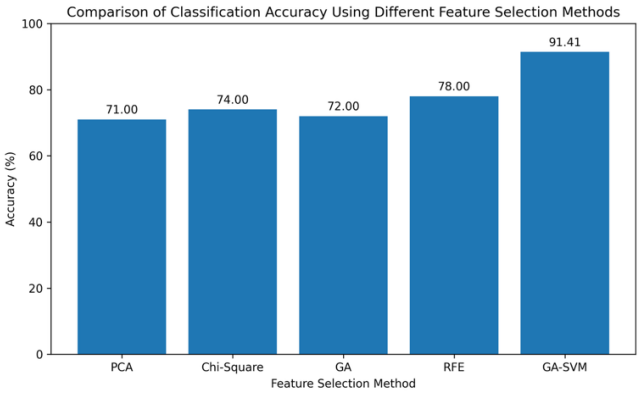


Fig. 8 Comparison of feature selection methods based on accuracy

5,343 to 17,931 representative features while preserving essential semantic information required for sentiment classification. The integration of the GA-SVM-selected feature subset with the LSTM classifier achieved the highest classification accuracy of 91.41% under the hold-out evaluation scheme and an average accuracy of 89.41% under 10-fold cross-validation, outperforming conventional feature selection methods including PCA, Chi-Square, GA, and RFE. The convergence analysis further revealed that the optimization process progressively identified increasingly discriminative feature subsets throughout the evolutionary search process, indicating the effectiveness of combining the global search capability of Genetic Algorithms with the discriminative evaluation power of SVM. These findings suggest that the proposed hybrid feature selection mechanism can improve feature representation quality and classification performance for noisy Indonesian social media text. Nevertheless, several limitations should be acknowledged. The experiments were conducted on a relatively limited dataset, which may restrict the generalizability of the findings across broader domains and textual contexts. In addition, the study employed TF-IDF feature representations and evaluated only the LSTM architecture, whereas alternative contextual embeddings and more advanced deep learning models such as BiLSTM, GRU, Transformer-based architectures, and IndoBERT may exhibit different optimization behaviors and classification performance. Furthermore, the computational cost associated with the GA-SVM optimization process remains a practical challenge when applied to large-scale datasets. Therefore, future research should investigate the proposed feature optimization strategy using larger and more diverse Indonesian textual datasets, explore its integration with contextual embedding techniques and advanced deep learning architectures, and examine approaches for improving computational efficiency, feature interpretability, and cross-domain generalization to further enhance the robustness and applicability of the proposed method.

REFERENCES

[1] K. S. Er. P. Balpreet Kaur, "SENTIMENT ANALYSIS FOR PRODUCT REVIEW," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 05, no. 04, pp. 4582–4592, Apr. 2023, doi: 10.56726/irjmets36776.

[2] P. Arsi, B. A. Kusuma, and A. Nurhakim, "Analisis Sentimen Pindah Ibu Kota Berbasis Naive Bayes Classifier," *Jurnal Informatika Upgris*, vol. 7, no. 1, 2021, doi: 10.26877/jiu.v7i1.7636.

[3] V. Chandradev, I. M. A. D. Suarjaya, and I. P. A. Vayupati, "Analisis Sentimen Review Hotel Menggunakan Metode Deep Learning BERT," *Jurnal Buana Informatika*, vol. 14, no. 2, 2023.

[4] A. Perdana, A. Hermawan, and D. Avianto, "Analisis Sentimen Terhadap Isu Penundaan Pemilu di Twitter Menggunakan Naive Bayes Clasifier," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 11, no. 2, pp. 195–200, Jul. 2022, doi: 10.32736/sisfokom.v11i2.1412.

[5] S. Jurnalis Pipin and H. Kurniawan, "Analisis Sentimen Kebijakan MBKM Berdasarkan Opini Masyarakat di Twitter Menggunakan LSTM," *Jurnal SIFO Mikroskil*, vol. 23, no. 2, pp. 197–208, Oct. 2022, doi: 10.55601/jsm.v23i2.900.

[6] E. Y. Hidayat, R. W. Hardiansyah, and A. Affandy, "Analisis Sentimen Twitter untuk Menilai Opini Terhadap Perusahaan Publik Menggunakan Algoritma Deep Neural Network," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 7, no. 2, pp. 108–118, Sep. 2021, doi: 10.25077/teknosi.v7i2.2021.108-118.

[7] R. A. Pramunendar, D. P. Prabowo, and R. A. Megantara, "Metode Recurrent Neural Network (RNN) Dengan Arsitektur LSTM untuk Analisis Sentimen Opini Publik Terkait Vaksin COVID-19," *JURNAL INFORMATIKA UPGRIS*, vol. 8, no. 1, pp. 44–48, Jun. 2022.

[8] S. Khomsah and A. S. Aribowo, "Text-Preprocessing Model Youtube Comments in Indonesian," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 4, no. 4, pp. 648–654, Aug. 2020, doi: 10.29207/resti.v4i4.2035.

[9] G. Angiani M. Ferrari, R. Fontanini, P. Fornacciari, E. Iotti, F. Magelli, and S. Manicardi, "A comparison between preprocessing techniques for sentiment analysis in Twitter," in *CEUR Workshop Proceedings*, G. Armano, A. Bozzon, M. Cristani, and A. Giuliani, Eds., Italy: Creative Commons CC0, Sep. 2016. Accessed: Sep. 04, 2025. [Online]. Available: https://ceur-ws.org/Vol-1748/paper-06.pdf

[10] R. Setiabudi, N. M. S. Iswari, and A. Rusli, "Enhancing text classification performance by preprocessing misspelled words in Indonesian language," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 19, no. 4, pp. 1234–1241, Aug. 2021, doi: 10.12928/TELKOMNIKA.v19i4.20369.

[11] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 2, p. 406, Apr. 2021, doi: 10.30865/mib.v5i2.2835.

[12] M. Buyukkececi and M. C. Okur, "A Comprehensive Review of Feature Selection and Feature Selection Stability in Machine Learning," 2023. doi: 10.35378/gujs.993763.

[13] H. T. Pham, J. Awange, and M. Kuhn, "Evaluation of Three Feature Dimension Reduction Techniques for

Machine Learning-Based Crop Yield Prediction Models," *Sensors*, vol. 22, no. 17, 2022, doi: 10.3390/s22176609.

[14] T. Karpagalingam and M. Karuppaiah, "Optimal feature subset selection based on combining document frequency and term frequency for text classification," *Computing and Informatics*, vol. 39, no. 5, 2021, doi: 10.31577/CAI_2020_5_881.

[15] Siti Mutmainah, Fathir, and Erin Eka Citra, "Improving the Accuracy of Social Media Sentiment Classification with the Combination of TF-IDF Method and Random Forest Algorithm," *Journix: Journal of Informatics and Computing*, vol. 1, no. 1, pp. 30–40, Apr. 2025, doi: 10.63866/journix.v1i1.2.

[16] A. Madasu and S. Elango, "Efficient feature selection techniques for sentiment analysis," *Multimed. Tools Appl.*, vol. 79, no. 9–10, 2020, doi: 10.1007/s11042-019-08409-z.

[17] J. Khan, A. Alam, and Y. Lee, "Intelligent Hybrid Feature Selection for Textual Sentiment Classification," *IEEE Access*, vol. 9, pp. 140590–140608, 2021, doi: 10.1109/ACCESS.2021.3118982.

[18] P. H. Prastyo, I. Ardiyanto, and R. Hidayat, "A Combination of Query Expansion Ranking and GA-SVM for Improving Indonesian Sentiment Classification Performance," *Procedia Comput. Sci.*, vol. 189, pp. 108–115, 2021, doi: 10.1016/j.procs.2021.05.074.

[19] R. S. Sudeep K. Hase, "Hybrid Feature Selection on Social Media Dataset for Sentiment Classification using Deep Learning Techniques," *Communications on Applied Nonlinear Analysis*, vol. 32, no. 9s, pp. 1899–1918, Mar. 2025, doi: 10.52783/cana.v32.4362.

[20] P. H. Prastyo, I. Ardiyanto, and R. Hidayat, "A Review of Feature Selection Techniques in Sentiment Analysis Using Filter, Wrapper, or Hybrid Methods," in *2020 6th International Conference on Science and Technology (ICST)*, IEEE, Sep. 2020, pp. 1–6. doi: 10.1109/ICST50505.2020.9732885.

[21] O. M. Alyasiri, Y. N. Cheah, and A. K. Abasi, "Hybrid Filter-Wrapper Text Feature Selection Technique for Text Classification," in *International Conference on Communication and Information Technology, ICICT 2021*, 2021. doi: 10.1109/ICICT52195.2021.9567898.

[22] S. Belkarkor, I. Hafidi, and M. Nachaoui, "Feature Selection for Text Classification Using Genetic Algorithm," in *Lecture Notes in Networks and Systems*, vol. 656 LNNS, 2023, pp. 69–80. doi: 10.1007/978-3-031-29313-9_7.

[23] N. Azhar, P. P. Adikara, and S. Adinugroho, "Analisis Sentimen Ulasan Kedai Kopi Menggunakan Metode Naive Bayes dengan Seleksi Fitur Algoritme Genetika," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 8, no. 3, pp. 609–618, Jun. 2021, doi: 10.25126/jtiik.2021834436.

[24] B. Ghaddar and J. Naoum-Sawaya, "High dimensional data classification and feature selection using support vector machines," *Eur. J. Oper. Res.*, vol. 265, no. 3, 2018, doi: 10.1016/j.ejor.2017.08.040.

[25] D. Patel, A. Saxena, and J. Wang, "A Machine Learning-Based Wrapper Method for Feature Selection," *International Journal of Data Warehousing and Mining*, vol. 20, no. 1, 2024, doi: 10.4018/IJDWM.352041.

[26] C. H. Lin and U. Nuha, "Sentiment analysis of Indonesian datasets based on a hybrid deep-learning strategy," *J. Big Data*, vol. 10, no. 1, 2023, doi: 10.1186/s40537-023-00782-9.

[27] S. Sarkar, S. Vinay, C. Djeddi, and J. Maiti, "Classification and pattern extraction of incidents: a deep learning-based approach," *Neural Comput. Appl.*, vol. 34, no. 17, 2022, doi: 10.1007/s00521-021-06780-3.

[28] N. C. Dang, M. N. Moreno-García, and F. De la Prieta, "Sentiment analysis based on deep learning: A comparative study," *Electronics (Switzerland)*, vol. 9, no. 3, 2020, doi: 10.3390/electronics9030483.

[29] L. Kurniasari and A. Setyanto, "Sentiment analysis using recurrent neural network-lstm in bahasa Indonesia," *Journal of Engineering Science and Technology*, vol. 15, no. 5, 2020.

[30] N. Hanafiah, Y. Setiawan, A. Buntaran, and M. Reynaldi, "Sentiment Analysis of Tourism Objects on Trip Advisor Using LSTM Method," *Journal of Computer Science and Technology Studies*, vol. 4, no. 2, 2022, doi: 10.32996/jcsts.2022.4.2.1.

[31] T. Crippa Praha and M. Nugraheni, "Indonesian Fake News Classification Using Transfer Learning in CNN and LSTM," Indonesia, Sep. 2024. doi: http://dx.doi.org/10.62527/joiv.8.2.2126.

[32] S. Mujilahwati, N. Z. Mohd Safar, K. M. N. Ku Khalif, and N. Ghazalli, "Optimizing Sentiment Analysis of Indonesian Texts: Enhancing Deep Learning Models with Genetic Algorithm-Based Feature Selection," *Journal of Soft Computing and Data Mining*, vol. 5, no. 2, Dec. 2024, doi: 10.30880/jscdm.2024.05.02.016.

[33] M. T. Razaq, D. Nurjanah, and H. Nurrahmi, "Analisis Sentimen Review Film Menggunakan Naive Bayes Classifier dengan Fitur TF-IDF," *e-Proceeding of Engineering*, vol. 10, no. 2, 2023.

[34] A. Aziz, "Analisis Sentimen Identifikasi Opini Terhadap Produk, Layanan dan Kebijakan Perusahaan Menggunakan Algoritma TF-IDF dan SentiStrength," 2022.

[35] D. Sugiarto, E. Utami, and A. Yaqin, "Perbandingan Kinerja Model TF-IDF dan BOW untuk Klasifikasi Opini Publik Tentang Kebijakan BLT Minyak Goreng," *Jurnal Teknik Industri* , vol. 12, no. 3, 2022.

[36] F. C. Zegarra, J. Vargas-Machuca, and A. M. Coronado, "A COMPARATIVE STUDY OF CNN, LSTM, BiLSTM, AND GRU ARCHITECTURES FOR TOOL WEAR PREDICTION IN MILLING PROCESSES,"

Journal of Machine Engineering, vol. 23, no. 4, 2023, doi: 10.36897/jme/174019.

[37] V. R. Prasetyo, M. F. Naufal, and K. Wijaya, "Sentiment Analysis of ChatGPT on Indonesian Text using Hybrid CNN and Bi-LSTM," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 9, no. 2, pp. 327–333, Apr. 2025, doi: 10.29207/resti.v9i2.6334.

[38] N. M. G. Dwi Purnamasari, M. A. Fauzi, I. Indriati, and L. S. Dewi, "Cyberbullying identification in twitter using support vector machine and information gain based feature selection," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 18, no. 3, p. 1494, Jun. 2020, doi: 10.11591/ijeecs.v18.i3.pp1494-1500.

[39] E. Hokijuliandy, H. Napitupulu, and F. Firdaniza, "Analisis Sentimen Menggunakan Metode Klasifikasi Support Vector Machine (SVM) dan Seleksi Fitur Chi-Square," *SisInfo : Jurnal Sistem Informasi dan Informatika*, vol. 5, no. 2, 2023, doi: 10.37278/sisinfo.v5i2.670.