# Multilabel Aspect-Based Sentiment Analysis with Diverse Embedding Techniques and Finetuned Transformers for Label Detection

Yusril Falih Izzaddien
*Department of Informatics*
*Institut Teknologi Sepuluh Nopember*
Surabaya, Indonesia
6025241026@student.its.ac.id

Riyanarto Sarno
*Department of Informatics*
*Institut Teknologi Sepuluh Nopember*
Surabaya, Indonesia
riyanarto@its.ac.id

Agus Tri Haryono
*Department of Informatics*
*Institut Teknologi Sepuluh Nopember*
Surabaya, Indonesia
7025231020@student.its.ac.id

Abdullah Faqih Septiyanto
*Department of Informatics*
*Institut Teknologi Sepuluh Nopember*
Surabaya, Indonesia
7025221022@student.its.ac.id

Dwi Sunaryono
*Department of Informatics*
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
dwi@its.ac.id

Raihan Adam Handoyo Winarso
*Department of Informatics*
*Institut Teknologi Sepuluh Nopember*
Surabaya, Indonesia
6025241043@student.its.ac.id

2024 Beyond Technology Summit on Informatics International Conference (BTS-I2C) | 979-8-3315-0857-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/BTS-I2C63534.2024.10941997

***Abstract*—The increasing volume of user-generated content on social media has amplified the spread of hate speech, posing significant challenges to online safety and content moderation. This study presents a comprehensive approach for aspect-based sentiment classification using multiple embedding and classification techniques. The dataset, sourced from the ETHOS multi-label hate speech corpus, includes eight distinct labels: violence, directed_vs_generalized, gender, race, national_origin, disability, religion, and sexual_orientation. Preprocessing techniques were applied to remove symbols, numbers, and irrelevant elements, ensuring data uniformity. Aspect category keywords were expanded using a generative language model (ChatGPT) to enhance the keyword corpus for each label. The research compares five methods: Word2Vec + cosine similarity (AC1), Global Vectors for Word Representation (GloVe) + cosine similarity (AC2), Bidirectional Encoder Representations from Transformers (BERT) embedding + cosine similarity (AC3), fine-tuned BERT (AC4), and pretrained Robustly Optimized BERT (RoBERTa) with fine-tuning (AC5). Evaluation was conducted using Repeated K-Fold cross-validation to prevent bias, measuring performance through metrics such as mean accuracy, precision, recall, F1 score, and Hamming loss. Results indicate that AC5, leveraging pretrained RoBERTa fine-tuned on the dataset, achieved superior performance with a mean F1 score of 0.674, mean precision of 0.862, and mean accuracy of 0.890, highlighting the efficacy of transformer-based models in capturing complex, context-specific sentiments. This research underscores the potential of combining extensive pretraining with domain-specific fine-tuning for precise and reliable multilabel sentiment classification.**

***Keywords*— Aspect-Based Sentiment Analysis (ABSA), multilabel classification, hate speech detection, word embeddings, Transformer, cosine similarity, cross-validation.**

## I. INTRODUCTION

The rise of social media has amplified hate speech, reinforcing biases and fueling real-world discrimination and division [1]. Tackling online hate is now crucial for safer digital communities [2]. Researchers are using Aspect-Based Sentiment Analysis (ABSA) to tackle these challenges by analyzing sentiments tied to specific aspects within a text, rather than just overall sentiment [3]. Unlike traditional sentiment analysis, ABSA offers a detailed look at sentiments targeting specific topics within text, making it ideal for analyzing complex opinions in social media and hotel reviews [4], Traditional ABSA uses multiclass classification, assigning each aspect a single label, limiting its ability to capture multiple simultaneous sentiments [5]. ABSA struggles with text containing overlapping sentiments underscoring the need for multilabel classification [2].

According Table I., ABSA research has evolved from rule-based and lexicon methods to advanced techniques, including deep learning models like Long Short-Term Memory (LSTM) and Bidirectional Long Short-Term Memory (BiLSTM) [6], [7] networks and Senticircle [8], earlier models improved sentiment detection with context but lacked attention mechanisms BERT and RoBERTa advanced this by capturing bidirectional context and focusing on key words [9].

TABLE I. RELATED WORKS ANALYSIS

| Work | Finding | Limitation |
|---|---|---|
| [4] | Expanded aspect categories with Wikipedia synonyms. | Outdated BiLSTM model, small corpus, single-label classification. |
| [5] | Used ELMo + Senticircle for aspect representation. | Lacked polarity detection, noisy results, accuracy <80%. |
| [6] | Expanded corpus with latent dislocation + SVM. | Ignored aspect context, single-label classification. |
| [7] | Used grammatical rules + Senticircle for aspects. | Struggled with explicit/implicit opinions. |
| [8] | Applied LSTM + TF-ICF for semantic similarity. | Predefined reviews, lacked contextual transformer, single-label. |

TABLE II. KEYWORD ASPECT CATEGORIES

| Aspect Category | Prompt to GPT-4o | Example Keywords |
|---|---|---|
| Violence | "Please provide me 500 keyword hate for aspect {Aspect_category} detailed in one word. generate like you are the dictionary" | stab, assault, abuse, harm, kill, assault, abuse, bloodshed, aggression, etc |
| Directed vs. Generalized | | personal, specific, targeted, group, public , etc |
| Gender | | woman, man, gender, sexism, misogyny , patriarchy, feminism, equality , men etc |
| Race | | race, racial, ethnicity, black, white , nigga, etc |
| National Origin | | immigrant, foreign, nationality, country, border , African, China, etc |
| Disability | | disability, disabled, handicapped, ableist , etc |
| Religion | | Islam, Christian, Hindu, Buddhist, atheist , Muhammad, Yesus, Judas etc |
| Sexual Orientation | | LGBTQ, homosexual, gay, lesbian, queer , LGBT, etc |

Transformer models outperformed earlier models in multiclass, enabling scalable sentiment analysis [10]. ABSA has advanced to deep learning, but challenges remain in handling overlapping sentiments, especially in hate speech [11]. Current research uses BERT-based multilabel classification techniques, like associative feature networks to capture nuanced sentiments [12]. Generative models like GPT enhance ABSA by automating aspect-specific keyword creation, reducing manual effort, and improving keyword alignment with real-world language [13].

This study introduces a novel multilabel classification framework for hate speech detection, addressing the challenge of overlapping sentiments within comments, such as those exhibiting both violence and race labels. By employing GPT to generate aspect-specific keywords, the dataset is enriched with a comprehensive corpus, enhancing alignment with real-world language. Advanced embedding methods, including Word2Vec, GloVe, BERT, and fine-tuned RoBERTa, are evaluated, with RoBERTa achieving superior accuracy through fine-tuning and cross-validation. This contribution advances ABSA by integrating generative models, multilabel detection, and transformer-based architectures, eliminating model confusion and paving the way for scalable, precise content moderation systems.

## II. Related Theory

### A. Preprocessing

Preprocessing is crucial for multilabel ABSA, ensuring data consistency and noise reduction for better model accuracy. This research follows several stages to improve input data quality before classification: The preprocessing steps for text data include punctuation and number removal, which simplifies text by eliminating unnecessary symbols, particularly in social media content. Normalization is then applied to ensure consistency by converting text to lowercase and expanding contractions. Stopword removal follows, eliminating common words that do not contribute significant meaning, thereby enhancing processing efficiency.

### B. Keyword Aspect Categories

In this research, aspect categories pertain to specific multilabel sentiment classifications within social media data, focusing on eight critical categories: violence, directed_vs_generalized, gender, race, national_origin, disability, religion, and sexual_orientation. Each category is identified by a set of distinct keywords that facilitate the targeted detection of hate speech and sentiment analysis related to social topics. To achieve robust and contextually rich keyword identification, GPT-3 was utilized for generating these keyword lists, as illustrated in Table II. This approach ensures comprehensive coverage, leveraging GPT-3's training on an extensive, diverse corpus to capture nuanced language patterns and terms associated with each aspect. The generated keywords enable more accurate categorization and analysis, supporting deeper insights into how each aspect manifests within social media discourse.

### C. Generating Aspect Category Keywords Using GPT

To create a comprehensive keyword list for each aspect category like Table II, this research used GPT-3 to expand upon manually selected seed terms. By generating keywords through GPT-3, the study achieved broader vocabulary coverage and reduced bias from human selection, ensuring the terms reflect real-world language use across social platforms. Table II illustrates sample prompts and generated terms for each category, highlighting GPT-3's role in ensuring nuanced, real-world relevance.

### D. Word Embedding and Cosine Similarity

Word embedding converts words into dense vector representations, placing similar words close together in a high-dimensional space. This research uses Word2Vec, GloVe, and BERT embeddings, with cosine similarity to measure the closeness between keyword embeddings and sample text embeddings for multilabel classification. The cosine similarity formula is as follows equation 1:

$$\text{Similarity}(w_i, w_j) = \frac{\sum_{m=1}^{K} w_{im} \cdot w_{jm}}{\sqrt{\sum_{m=1}^{K} (w_{im})^2} \cdot \sqrt{\sum_{m=1}^{K} (w_{jm})^2}} \quad (1)$$

where:

a. $w_i$ and $w_j$ are the word vectors for words $i$ and $j$,

b. $K$ is the dimensionality of the word embeddings,

c. $w_{im}$ and $w_{jm}$ are the components of the vectors $w_i$ and $w_j$ in the $m$-th dimension.

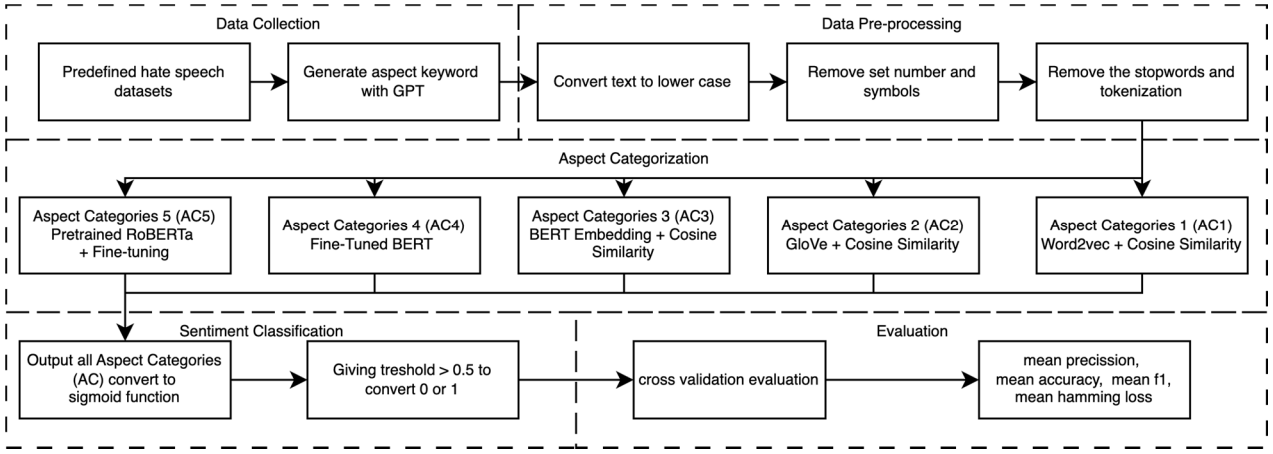


Fig 1. Proposed Method

### E. *Word2Vec Embedding (300d Google)*

Word2Vec, a classic word embedding method developed by Google, employs neural network-based architectures to generate word vectors in a high-dimensional space. It offers two model architectures: Continuous Bag-of-Words (CBOW) and Skip-Gram (SG). CBOW predicts a target word given its surrounding context, while SG predicts the context words for a given target word. These methods leverage large corpora to capture semantic and syntactic relationships between words efficiently. In this research, pre-trained Word2Vec embeddings from the Google News dataset, offering 300-dimensional vectors, are used as input features to establish a robust foundation for sentiment classification tasks [15].

### F. *GloVe Embedding (300d Wikipedia)*

GloVe (Global Vectors for Word Representation), developed by Stanford researchers, is a classic embedding technique that derives word representations by factorizing a global word-word co-occurrence matrix. Unlike Word2Vec, which is prediction-based, GloVe uses statistical methods to analyze aggregated co-occurrence data, making it particularly adept at capturing both local and global linguistic features. In this study, 300-dimensional GloVe embeddings pre-trained on large datasets like Wikipedia and Gigaword are employed. These embeddings provide subtle semantic insights that are critical for effective multilabel hate speech detection and classification [15].

### G. *BERT Embedding*

BERT embeddings capture words in context, creating dynamic, context-sensitive vectors that adapt to the specific sentence. This capability significantly improves semantic nuance capture, making BERT ideal for fine-grained sentiment detection, especially with overlapping labels. In this research, BERT enhances the model's ability to interpret complex, contextual sentiment in social media text [16].

### H. *RoBERTa Pretrained and Fine-Tuned*

RoBERTa improves upon BERT with optimized pretraining on larger datasets and batch sizes, enhancing downstream task performance. In this research, RoBERTa's pretrained model is further fine-tuned on the specific dataset, boosting its ability in multilabel sentiment classification, especially in distinguishing nuanced sentiments like generalized versus directed hate [17].

### I. *Evaluation*

In assessing aspect categorization and sentiment analysis, three primary evaluation metrics: Precision,Recall, and F1-score. These metrics are derived from the confusion matrix, which serves as the foundation for calculating each measure.

a) Precision measures the accuracy of positive predictions and is calculated as equation 2:

$$\text{Precision} = \frac{TP}{TP+FP} \tag{2}$$

b) Recall assesses the model's ability to identify all actual positives and is defined as equation 3:

$$\text{Recall} = \frac{TP}{TP+FN} \tag{3}$$

c) F1-score represents the harmonic mean of Precision and Recall, providing a balanced metric when there is a trade-off between the two as equation 4:

$$\text{F1} - \text{score} = \frac{\text{2 x Precision x Recall}}{\text{Precision+Recall}} \tag{4}$$

d) Accuracy indicates the overall correctness of predictions and is given as equation 5:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \tag{5}$$

where:

a. TP: True Positives

b. FP: False Positives

c. TN: True Negatives

d. FN : False Negatives

## III. Proposed Method

The Proposed Method according Figure 1 demonstrates the workflow for multilabel hate speech classification using ABSA The process begins with predefined hate speech datasets and the generation of aspect-specific keywords using GPT model. Text preprocessing includes standardization, tokenization, symbol and stopword removal. Five embedding to compare: Word2Vec, GloVe, BERT, fine-tuned BERT, and

TABLE III. DATASET DISTRIBUTION

| Label | Instances | Keyword Expanded GPT |
|---|---|---|
| Violence | 142 | 329 |
| Directed vs Generalized | 135 | 259 |
| Gender | 86 | 300 |
| Race | 76 | 273 |
| National Origin | 74 | 327 |
| Disability | 53 | 300 |
| Religion | 81 | 376 |
| Sexual Orientation | 73 | 238 |

fine-tuned RoBERTa. The outputs undergo sigmoid transformation with a threshold for binary classification. cross-validation evaluates model performance using metrics precision, accuracy, F1-score, and Hamming loss.

### A. Data Collection

The dataset used in this research combines comments collected from YouTube videos and Reddit threads, specifically targeting videos and subreddits likely to attract hate speech content, with data sourced from Hugging Face's ETHOS: onlinE haTe speecH detectiOn dataSet. This dataset, consists of 433 comments annotated across eight distinct labels representing various aspects of hate speech. The label distribution provides insights into the prevalence of each hate speech category within the dataset, enriching the analysis with a comprehensive view of labeled data.

Additionally, aspect keywords or terms for each label were expanded using a large language model (ChatGPT), allowing for a comprehensive keyword corpus for each aspect category This enriched vocabulary at Table III

### B. Data Pre-processing

Pre-processing stage in this research is done with the results as follow: 1) convert text to lower case, 2) removes the set of symbols and number , 3) removes the stop words, 4)tokenization.

### C. Aspect Categorization

This research tests embedding techniques (Word2Vec, GloVe, BERT, fine-tuned BERT, and fine-tuned RoBERTa) with cosine similarity to classify comments by aspect categories, using a 0.5 threshold for classification. AC4 and AC5 are fine-tuned to improve detection.

- AC1: Word2Vec + Cosine Similarity:

Word2Vec embeddings (300d Google News vectors) are applied to represent each word in the dataset. Cosine similarity then measures similarity between these vectors and aspect category keywords.

- AC2: GloVe + Cosine Similarity

GloVe embeddings (300d Wikipedia vectors) are employed, followed by cosine similarity to calculate similarity scores against aspect-specific keywords.

- AC3: BERT Embedding + Cosine Similarity

BERT embeddings, which produce context-sensitive representations, are applied to capture nuanced meanings. Cosine similarity evaluates closeness to each aspect category like algorithm at Table IV.

- AC4: Fine-tuned BERT

BERT was fine-tuned with a learning rate of 2e-5, batch size 8, and 5 epochs, improving accuracy by adapting to hate speech categories.

- AC5: Pretrained RoBERTa + Fine-tuning

RoBERTa, fine-tuned with the same hyperparameters, leveraged its optimized architecture and pretraining to enhance detection of nuanced labels like violence and race Table V.

### D. Sentiment Classification

Sentiment classification is similar to aspect categorization but uses a binary approach. Sentiment scores, ranging from 0 to 1, are processed through a sigmoid function. Scores above 0.5 are labeled as positive (1), while those at or below 0.5 are also labeled as 1.

## IV. Experiment and Results

### A. Aspect Categorization Performance

Experiments evaluated each aspect categorization method using Repeated K-Fold cross-validation (5 splits, 2 repeats) and metrics such as accuracy, F1 score, Hamming loss, precision, and recall, with results averaged across folds to ensure robust performance and minimize bias. The model's output is a similarity score, where higher values indicate stronger similarity, while negative values suggest dissimilarity. To mitigate potential bias in label determination, a sigmoid function was applied to transform the similarity scores into a range between 0 and 1. By setting a threshold of 0.5, labels are assigned as relevant or irrelevant, thereby optimizing accuracy in multi-label classification scenarios. This approach ensures that the model better handles the complexities of overlapping categories and improves its overall performance in aspect categorization.

*1)* AC1 Word2Vec + Cosine Similarity:

Word2Vec embeddings with cosine similarity showed moderate performance in aspect categorization, capturing basic semantic relationships but struggling with context nuances.

*2)* AC2: GloVe + Cosine

GloVe with cosine similarity performed slightly worse than Word2Vec, as it struggled with fine-grained categorization in complex contexts.

*3) AC3: BERT Embedding + Cosine Similarity*

BERT embeddings improved F1 score and recall by capturing deep semantic context, though challenges with Hamming loss indicated inconsistent categorizatio

*4) AC4: Fine-Tunned Bert*

Fine-tuning BERT improved all performance metrics, particularly accuracy and precision, leading to more accurate aspect categorizations than unsupervised embeddings.

TABLE IV. PSEUDOCODE MODEL FINETUNNING

| Pseudocode For Model Finetuning |
|---|
| Load dataset and keywords<br>Apply AC model pre-trained, then fine-tune with training dataset labels<br>For each comment in the dataset:<br>a. Generate embeddings with fine-tuned BERT<br>b. Classify each comment based on learned labels using sigmoid output<br>c. Apply a 0.5 threshold to assign labels for each aspect |

TABLE V. PSEUDOCODE MODEL EMBEDDING

| Pseudocode For Model Embedding |
|---|
| # AC 1,2,3 Model Embedding<br>Load dataset and keywords<br>Apply AC embedding to convert words to vectors<br>For each comment in the dataset:<br>a. Compute cosine similarity between comment vectors and keyword vectors for each aspect<br>b. Apply sigmoid to cosine similarity scores and set threshold at 0.5<br>c. If similarity score >= 0.5 for an aspect, assign label for that aspect |

TABLE VI. SENTIMENT CLASSIFICATION RESULT

| Aspect Categorization | Multilabel Prediction |
|---|---|
| AC 1 | Gender, Sexual Orientation, Religion, Disability |
| AC 2 | Gender, Sexual Orientation, Religion |
| AC 3 | Gender, Disability |
| AC 4 | Genderm Sexual Orientation |
| AC 5 | Gender, Sexual Orientation, Generalized |

TABLE VII. RESULT COMPARISON

| Metrics | AC1 | AC2 | AC3 | AC4 | AC5 |
|---|---|---|---|---|---|
| Label-based Accuracy | 0.734 | 0.712 | 0.707 | 0.816 | **0.941** |
| Mean Label-based Accuracy | 0.739 | 0.719 | 0.670 | 0.822 | **0.890** |
| Mean F1 Score | 0.292 | 0.081 | 0.218 | 0.221 | **0.674** |
| Mean Precision | 0.411 | 0.118 | 0.223 | 0.445 | **0.862** |
| Mean Recall | 0.270 | 0.137 | 0.256 | 0.163 | **0.587** |
| Mean Hamming Loss | 0.260 | 0.280 | 0.329 | 0.177 | **0.110** |

*5) AC5: Pretrained RoBERTa + Fine-Tuning.*

Pretrained RoBERTa, fine-tuned on the dataset, achieved the highest performance metrics, including label-based accuracy and mean F1 score. Its superior performance, particularly in accuracy and mean precision, is attributed to its initial training on a large, relevant corpus, making it highly effective for nuanced sentiment detection..

TABLE VIII. RESULT COMPARISON EXTENDED

| Model / Method | Runtime (sec) | Runtime (min) |
|---|---|---|
| GloVe-Embed+Cosine | 25.98 | ~0.43 |
| Word2Vec-Embed+Cosine | 687.66 | ~11.46 |
| BERT-Embedding+Cosine | 323.96 | ~5.40 |
| BERT Finetunning | 2898.27 | ~48.30 |
| RoBERTa Finetunning | 1708.91 | ~28.48 |

### *B. Comparisan of Performance Across AC Methods*

As shown in Table VII, AC5 (RoBERTa + Fine-Tuning) outperformed all other methods, achieving the highest mean accuracy of 0.89 with a Hamming loss of 0.1. This suggests that the fine-tuning of pretrained RoBERTa effectively captures task-specific features, leading to high classification accuracy. In contrast, AC4 (Fine-Tuned BERT) also showed strong performance, surpassing unsupervised embeddings in all metrics. However, despite the high accuracy, the F1 score remains suboptimal at 0.67, indicating the model's struggle with multi-label classification, particularly in distinguishing between overlapping categories like as we can see Table VI confuse multi label classification. This is further reflected in the precision and recall trade-offs. The Hamming loss of 0.1, while relatively low, points to occasional misclassifications, suggesting room for improvement in label differentiation. These results highlight the strength of fine-tuning large pretrained models, but also reveal the challenges in handling label overlap. Future work should focus on refining the model's ability to resolve ambiguities in multi-label settings, potentially through advanced techniques such as label-wise attention or increasing dataset diversity

### *C. Computational Complexity and Efficiency*

The computational complexity of each model was evaluated based on the total runtime for training, as summarized in Table VII. Pretrained embedding models such as Word2Vec and GloVe with cosine similarity methods demonstrated the shortest training times, with GloVe-Embed+Cosine requiring just ~26 seconds and Word2Vec-Embed+Cosine completing in ~688 seconds. These methods are lightweight due to their reliance on pre-computed embeddings and lack of fine-tuning requirements.

In contrast, models involving fine-tuning, such as BERT-Embedding+Cosine-Similarity, required a moderate runtime of ~324 seconds, reflecting the computational cost of combining contextual embeddings with similarity calculations. The fully fine-tuned transformer models, BERT and RoBERTa, exhibited significantly higher runtimes of ~2898 seconds (~48.3 minutes) and ~1709 seconds (~28.5 minutes), respectively, due to their parameter-intensive optimization during fine-tuning.

While RoBERTa demonstrated the highest classification accuracy, its computational demand is approximately 40% lower than BERT, highlighting its efficiency despite being a

more advanced architecture. This trade-off between runtime efficiency and classification performance is critical for selecting the appropriate model depending on the application's needs.

## V. CONCLUSIONS

The prevalence of hate speech on digital platforms has become a critical issue, posing significant challenges for content moderation and online safety. Traditional methods of classifying hate speech, primarily reliant on keyword-based or rule-based approaches, have proven inadequate in capturing nuanced forms of harmful language, particularly across diverse categories. These methods often fail to address the context and semantic variations inherent in subtle or indirect hate speech. The emergence of deep learning techniques, especially transformer-based models like BERT and RoBERTa, offers a promising alternative. By leveraging contextual embeddings, these models enhance classification precision and enable sophisticated analysis of complex linguistic patterns. However, the application of these advanced models to diverse hate speech categories, such as violence, gender, race, and national origin, remains a challenge, requiring further exploration.

This study introduced a novel approach for hate speech classification, integrating semantic cosine similarity with a comparative analysis of transformer-based models such as BERT and RoBERTa, fine-tuned for hate speech detection. The results, as detailed in Tables VII, demonstrate the clear superiority of fine-tuned transformers over traditional embedding methods. The fine-tuned RoBERTa model achieved an outstanding accuracy of 96%, with an average k-fold cross-validation score of 0.890. Meanwhile, the fine-tuned BERT model reached 81.66% accuracy, with a mean score of 82.25%. In contrast, pre-trained embeddings like Word2Vec and GloVe, used without fine-tuning, exhibited lower classification performance, with Word2Vec outperforming GloVe due to its broader vocabulary and richer vector representations.

These findings underscore the critical trade-off between accuracy and computational cost. While fine-tuning transformer models significantly improves accuracy, it demands substantially higher computational resources, as shown by training times of ~1709 seconds for RoBERTa and ~2898 seconds for BERT. In comparison, traditional embeddings like Word2Vec and GloVe, with training times of ~688 and ~26 seconds respectively, offer computational efficiency but at the cost of reduced performance. This highlights the importance of selecting the appropriate approach based on resource availability and application needs. Future work could focus on integrating multiple embeddings, such as GloVe, Word2Vec, and BERT, to harness their combined strengths and further enhance performance. Additionally, leveraging pre-trained models specifically tailored for hate speech detection and fine-tuning them with larger, more diverse datasets could significantly improve accuracy and robustness. The dataset used in this study, limited to 433 samples, represents a key constraint, and expanding this dataset would enable better generalization. By refining these models and incorporating additional training data, future research has the potential to revolutionize hate speech detection and moderation, contributing to a safer and more inclusive digital environment.


## ACKNOWLEDGEMENT

This research was funded by Institut Teknologi Sepuluh Nopember (ITS) under Penelitian Kolaborasi Pusat and under the Project Scheme of the Publication Writing and Intellectual Property Rights (IPR) Incentive (Penulisan Publikasi dan Hak Kekayaan Intelektual/PPHKI) Program 2024.


## REFERENCES


[1] H. S. Wicaksana, R. Kusumaningrum, and R. Gernowo, "Determining community happiness index with transformers and attention-based deep learning," *IAES Int. J. Artif. Intell.*, vol. 13, no. 2, pp. 1753–1761, 2024, doi: 10.11591/ijai.v13.i2.pp1753-1761.

[2] Z. Wang, "ABSA for Chinese review with fine-tuned BERT and zero-shot classification," *Appl. Comput. Eng.*, vol. 49, no. 1, pp. 229–235, 2024, doi: 10.54254/2755-2721/49/20241196.

[3] B. C. Jie Shao, Man Lung Yiu, Masashi Toyoda, Dongxiang Zhang, Wei Wang, *Web and Big Data*. 2019.

[4] A. T. Haryono, R. Sarno, and R. Abdullah, "Aspect-Based Sentiment Analysis of Financial Headlines and Microblogs Using Semantic Similarity and Bidirectional Long Short-Term Memory," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 3, pp. 233–241, 2022, doi: 10.22266/ijies2022.0630.20.

[5] F. Nurifan, R. Sarno, and K. R. Sungkono, "Aspect based sentiment analysis for restaurant reviews using hybrid ELMo-wikipedia and hybrid expanded opinion lexicon-senticircle," *Int. J. Intell. Eng. Syst.*, vol. 12, no. 6, pp. 47–58, 2019, doi: 10.22266/ijies2019.1231.05.

[6] M. D. Pratama, R. Sarno, and R. Abdullah, "Sentiment Analysis User Regarding Hotel Reviews by Aspect Based Using Latent Dirichlet Allocation, Semantic Similarity, and Support Vector Machine Method," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 3, pp. 514–524, 2022, doi: 10.22266/ijies2022.0630.43.

[7] Suhariyanto, R. Abdullah, R. Sarno, and C. Fatichah, "Aspect-Based Sentiment Analysis for Sentence Types with Implicit Aspect and Explicit Opinion in Restaurant Review Using Grammatical Rules, Hybrid Approach, and SentiCircle," *Int. J. Intell. Eng. Syst.*, vol. 14, no. 5, pp. 177–187, 2021, doi: 10.22266/ijies2021.1031.17.

[8] R. A. Priyantina and R. Sarno, "Sentiment analysis of hotel reviews using Latent Dirichlet Allocation, semantic similarity and LSTM," *Int. J. Intell. Eng. Syst.*, vol. 12, no. 4, pp. 142–155, 2019, doi: 10.22266/ijies2019.0831.14.

[9] Q. Xu, L. Zhu, T. Dai, and C. Yan, "Aspect-based sentiment classification with multi-attention network," *Neurocomputing*, vol. 388, pp. 135–143, 2020, doi: 10.1016/j.neucom.2020.01.024.

[10] J. Wang, W. Wu, and J. Ren, "BERT-PG: a two-branch associative feature gated filtering network for aspect sentiment classification," *J. Intell. Inf. Syst.*, vol. 60, no. 3, pp. 709–730, 2023, doi: 10.1007/s10844-023-00785-1.

[11] T. Van Dang, D. Hao, and N. Nguyen, "Vi-AbSQA: Multi-task Prompt Instruction Tuning Model for Vietnamese Aspect-based Sentiment Quadruple Analysis," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, 2024, doi: 10.1145/3676886.

[12] I. Perikos, "Explainable Aspect-Based Sentiment Analysis Using Transformer Models," 2024.

[13] V. Vorakitphan, M. Basic, and G. L. Meline, "Deep Content Understanding Toward Entity and Aspect Target Sentiment Analysis on Foundation Models," 2024.

[14] G. M. Shafiq, T. Hamza, M. F. Alrahmawy, and R. El-Deeb, "Enhancing Arabic Aspect-Based Sentiment Analysis Using End-to-End Model," *IEEE Access*, vol. 11, no. November, pp. 142062–142076, 2023, doi: 10.1109/ACCESS.2023.3342755.

[15] C. Wang, P. Nulty, and D. Lillis, "A Comparative Study on Word Embeddings in Deep Learning for Text Classification," in Proceedings of the 4th International Conference on Natural Language Processing and Information Retrieval (NLPIR '20), Seoul, Republic of Korea, 2021, pp. 37–46. doi: 10.1145/3443279.3443304.

[16] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, vol. 1, no. Mlm, pp. 4171–4186, 2019.

[17] Y. Liu *et al.*, "RoBERTa: A Robustly Optimized BERT Pretraining Approach," no. 1, 2019.