Journal of Applied Data Sciences
Vol. 7, No. 1, January 2026, pp. 218-231

IndoBERT-SupCon: A Supervised Contrastive Learning Model for Analyzing Public Perception on Halal Tourism

Sri Mona Octafia[1,*], Rio Andika Malik[2], Annisa Weriframayeni[3], Delpa[4]

[1,3,4]Digital Business and Social Sciences, Universitas Perintis Indonesia, Padang, West Sumatera 25173, Indonesia

[2] Computer Science, Universitas Putra Indonesia YPTK, Padang, West Sumatera 25126, Indonesia

(Received: May 25, 2025; Revised: August 1, 2025; Accepted: November 13, 2025; Available online: December 9, 2025)

Abstract

The primary objective of this research is to develop and evaluate a robust deep learning model for accurately analyzing stakeholder perceptions of halal tourism development in Pariaman, West Sumatra, based on qualitative textual data. The main contribution is the introduction of IndoBERT-SupCon, a novel architecture that enhances the Indonesian BERT model with a Supervised Contrastive Learning (SupCon) mechanism. A novel method for producing more discriminative feature representations for complex viewpoints is presented in this paper, which is one of the first to use this sophisticated fine-tuning technique to Indonesian socio-political sentiment analysis. Conceptually, the model is trained to simultaneously minimize classification error while optimizing the feature space, pulling representations of similar sentiments closer together and pushing dissimilar ones further apart. To achieve this, we collected 1,022 primary textual responses through online surveys with tourists and in-depth interviews with key stakeholders, including SME owners and government officials. The SMOTE oversampling technique was employed on the training data to mitigate class imbalance. Experimental results on the test data demonstrate that the IndoBERT-SupCon model achieved outstanding performance, with a final accuracy of 96.59% and a macro F1-score of 0.97. These results significantly surpass the performance of a standard fine-tuned IndoBERT baseline, confirming the effectiveness of the SupCon approach. The findings provide the Pariaman local government with a highly valid, data-driven tool for more responsive and effective policy formulation. This research offers a robust framework that can be applied to other public policy domains, showcasing the value of advanced deep learning in transforming qualitative stakeholder feedback into actionable insights.

Keywords: Halal Tourism, Public Perception, Deep Learning, IndoBERT, Supervised Contrastive Learning

1. Introduction

Halal tourism has emerged as one of the world's fastest-growing tourism businesses [1] Indonesia has a great chance to dominate this industry because it is the nation with the largest Muslim population worldwide [2]. Halal tourism has been deemed a national strategic program by the Indonesian government in an effort to diversify tourism offerings and boost visitor numbers, both of which are anticipated to contribute to local and national economic growth [3]. In addition to offering halal-certified facilities and services [4], the public's and other pertinent stakeholders' favourable attitudes and acceptance are also essential to the strategy's successful execution [5].

Various regions including Pariaman City in West Sumatra Province, are seeking to implement this national policy [6]. Pariaman has a great basis for creating a genuine halal tourism destination because it is a region with a strong Minangkabau culture and strong Islamic beliefs [7]. However, the implementation of this policy is not without potential challenges and complexities. This initiative can spark debate among various stakeholders. For instance, economic conflicts may arise concerning the policy's impact on conventional or non-halal tourism operators, as well as potential socio-cultural friction if the 'halal' branding is perceived as exclusive by parts of the local community or non-Muslim tourists [8]. Therefore, a comprehensive understanding of the opinions, aspirations, and concerns of various stakeholders, such as visiting tourists, Micro, Small, and Medium Enterprises (MSMEs) as service providers, and government officials as regulators, is crucial to the success of this project. Without an accurate understanding of public

*Corresponding author: Sri Mona Octafia (srimonaoctafia@upertis.ac.id)
DOI: https://doi.org/10.47738/jads.v7i1.1045

sentiment, policies risk being ineffective or even counterproductive. Public perception analysis has historically frequently been carried out by manual qualitative procedures, which are laborious and subject to subjective researcher bias [9]. But new developments in Natural Language Processing (NLP) technology provide ways to more effectively and impartially analyse massive amounts of textual data [10]. For a variety of Indonesian-language text processing tasks, such as sentiment analysis, pre-trained language models like IndoBERT have shown remarkable efficacy [11].

Although there are frequently difficulties when using models like IndoBERT with traditional fine-tuning techniques. One of these is the challenge of differentiating between sentiments with nuanced or unclear nuances, like "undecided" and "disagree." Unbalanced class distribution data, which is a prevalent issue in real-world datasets, can cause model performance to deteriorate. These drawbacks point to a need for greater study into creating resilient model architectures that can generate more discriminatory feature representations for sentiment analysis.

This study suggests a novel hybrid paradigm named IndoBERT- SupCon to address these issues. This model combines the advantages of IndoBERT with a sophisticated Supervised Contrastive Learning (SupCon) fine-tuning technique. The goal of the SupCon technique is to maximise the distance between data from various sentiment classes while teaching the model to map data with the same sentiment to neighbouring points in the feature space. The use of the SMOTE oversampling technique to rectify the class imbalance in the training data reinforces this strategy.

This research provides several important contributions. In order to classify sentiment in the Indonesian language, we first propose and validate the IndoBERT- SupCon architecture, which has been shown to perform exceptionally well. Second, we use a cutting-edge deep learning technique to give the first thorough sentiment analysis of Pariaman's halal tourism stakeholders' opinions. Third, the analysis's findings offer data driven insights that local governments and pertinent stakeholders may implement to create strategies that are more successful.

2. Literature Review

2.1. Halal Tourism and Public Perception Analysis

A subset of tourism known as halal or Muslim-friendly travel offers amenities and services that are compliant with Islamic principles. This idea goes beyond halal cuisine to encompass lodging, places of worship, Islamic banking, and Islamically-based travel experiences [12]. The market for halal travel has expanded rapidly in recent years due to the rise in the number of middle-class Muslims around the world [13]. Being the most populous Muslim nation, Indonesia has made halal tourism development a national priority program in order to increase its competitiveness in the global tourism market [14].

Development of halal tourism destinations is largely dependent on stakeholder acceptability and views [15]. These impressions include the opinions of both Muslim and non-Muslim tourists, the backing of local governments, and the participation of MSMEs as the foundation of the tourism industry [16]. To ensure more inclusive and successful policy creation, it is essential to accurately analyse stakeholder sentiment in order to identify strengths, weaknesses, and potential social conflicts [17].

Given the complexity and nuances inherent in stakeholder perceptions, traditional analytical methods like manual surveys or qualitative interviews often face significant limitations. Such processes are not only time-consuming and resource-intensive but are also difficult to scale across large volumes of feedback and remain susceptible to the subjective bias of the researcher. To overcome these challenges, advancements in Natural Language Processing (NLP) offer a more objective, scalable, and efficient solution. Specifically, sentiment analysis, also known as opinion mining, has emerged as a powerful computational approach to systematically identify, extract, and categorize the opinions expressed in textual data.

2.2. Sentiment Analysis in the Tourism Sector

Sentiment analysis, often known as opinion mining, is a subfield of Natural Language Processing (NLP) that uses computing to identify and categorise opinions represented in textual data [18]. Sentiment analysis has been utilised extensively in the tourism industry to glean information from social media, travel forums, and evaluations of lodging in order to assess visitor satisfaction and destination reputation.

Initially, sentiment analysis depended on standard machine learning algorithms like Naïve Bayes, Support Vector Machines (SVM), and Logistic Regression. These approaches are useful for easy tasks, but they have a lot of drawbacks. Model performance is mostly dependent on laborious manual feature engineering procedures and frequently misses sarcasm, irony, or deeper contextual meaning in a sentence.

Deep learning has brought about a new era in natural language processing (NLP) by enabling transformer-based architectures that can comprehend a sentence's word context in both directions [19]. Bidirectional Encoder Representations from Transformers (BERT), created by Google, is among the most well-known models [9]. BERT achieves state-of-the-art performance on a variety of NLP tasks after being trained on a vast corpus of text data. It can generate word representations that are rich in contextual meaning.

IndoBERT is one of the specialised BERT models created to handle the complexity and distinctiveness of the Indonesian language [20]. Compared to multilingual models, IndoBERT has a superior grasp of syntax, vocabulary, and regional linguistic subtleties because it was trained only on a sizable corpus of data in the Indonesian language. In Indonesia, IndoBERT has demonstrated remarkable efficacy in a range of applications, such as sentiment analysis and text classification.

2.3. Supervised Contrastive Learning for Text Classification

There remains potential for improvement, especially with regard to the calibre of the feature representations that the model produces [21], even if fine-tuning IndoBERT using the conventional Cross-Entropy loss function has produced positive results [21]. The Cross-Entropy loss does not specifically teach the model to differentiate between semantically similar classes [23]. Instead, it concentrates only on minimising prediction error [24].

Supervised Contrastive Learning (SupCon) provides a novel solution in this regard. A training method called SUPCON seeks to improve the feature space that the model has learnt [25]. It operates by "pushing" representations of data from different classes (negative pairs) as far apart as feasible while pulling representations of data from the same class (positive pairs) as alike as possible [26]. SupCon enables the model to acquire more structured and discriminative feature representations by defining these positive and negative pairings with label information [27]. This is especially helpful for classification problems when class borders are not always obvious [28].

According on the literature review, various research gaps can be discovered [29]. First, despite the fact that sentiment analysis of halal travel has been done, a lot of research still uses classic machine learning techniques or unstructured social media data [30]. There is currently little research combining deep learning analysis with primary data from surveys and stakeholder interviews [31]. Second, even if IndoBERT has been used extensively, traditional fine-tuning techniques are frequently the only way to employ it [32]. In order to enhance IndoBERT's performance in the Indonesian setting, more complex fine-tuning methods like Supervised Contrastive Learning are currently rarely explored [33].

While Supervised Contrastive Learning has shown significant promise for Transformer-based models in various languages, its application to Indonesian-specific language models like IndoBERT remains a largely unexplored area. A comprehensive literature search indicates that, to the best of our knowledge, no prior study has proposed or evaluated an architecture combining IndoBERT with a Supervised Contrastive Learning fine-tuning mechanism for any Indonesian NLP task. This highlights a critical research gap in leveraging advanced fine-tuning strategies to enhance the performance of local language models in the Indonesian context. This work proposes and assesses the IndoBERT-SupCon model for sentiment analysis from structured stakeholder data in a particular case study of the growth of halal tourism in Pariaman, West Sumatra.

3. Methodology

This research methodology is systematically designed to develop and evaluate the proposed sentiment analysis model. The research workflow is divided into several main stages, including data collection and preprocessing, class balancing, model training with the IndoBERT-SupCon architecture, and performance evaluation. This research framework, as illustrated in figure 1, presents the overall process flow.

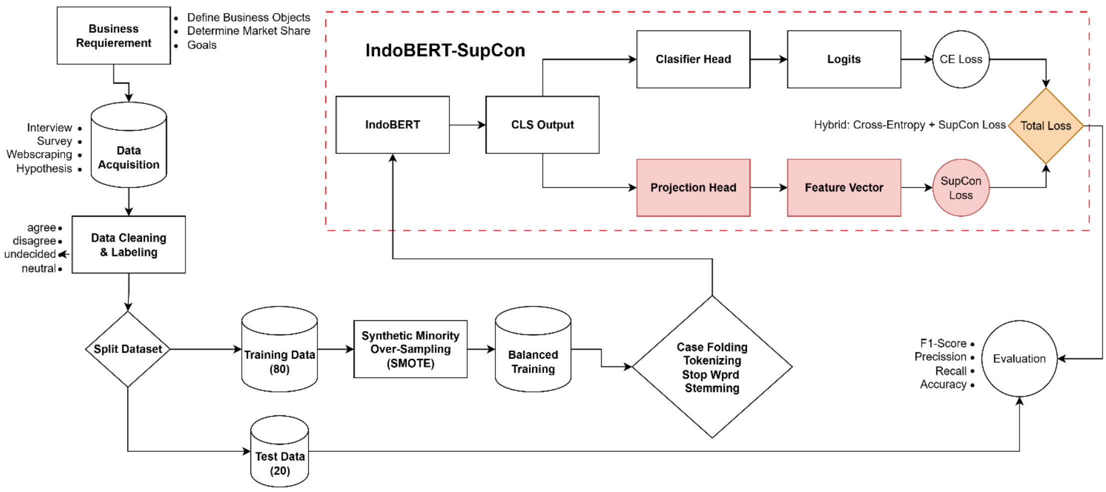


Figure 1. Research Structure for Developing IndoBERT-SupCon

The process begins with the collection of qualitative data from various stakeholders. The collected textual data then undergoes a preprocessing stage to clean and standardize the text. Afterward, the data is divided into training and test datasets. The SMOTE oversampling technique is applied to the training data to address class imbalance. The balanced training data is then used to train the IndoBERT- SupCon model. Finally, the trained model is evaluated using the test data to measure its performance based on accuracy, precision, recall, and F1-score metrics.

3.1. Data Collection

The data source in this study is primary data obtained through a mixed-method approach. The study population comprised three key stakeholder groups: tourists, Micro, Small, and Medium-Sized Enterprises (MSMEs) in the tourism sector, and government officials. A purposive sampling technique was employed with specific inclusion criteria defined for each group to ensure the relevance and diversity of the collected data.

Tourist participants were selected based on the criterion of having visited Pariaman City within the last 12 months to ensure their feedback was based on recent experiences. The selection focused on businesses in the food, beverage, and accommodation sectors (MSMEs) that were located within primary tourist areas. An additional criterion was that the business had been operational for at least two years, ensuring representation from established local enterprises. The last participants were specifically selected from the Pariaman Tourism Office (Government Stakeholders). The criteria for inclusion were direct involvement in the planning, implementation, or evaluation of the city's halal tourism policies. Tourists completed an survey to gather both quantitative and qualitative data, and stakeholders were interviewed to gather detailed qualitative data. From this process 1.022 textual sentiment data were collected, which were then labeled into four categories: 'Agree', 'Disagree', 'Undecided', and 'Neutral'.

3.2. Data Preprocessing

The raw text data underwent several preprocessing stages to enhance data quality and reduce noise before being fed into the model. These steps were carefully chosen to normalize the text while considering the morphological richness of the Indonesian language. These actions consist of Case folding is the process of changing all text to lowercase in order to maintain consistency and steer clear of terms that are identical save for capitalisation. All text was converted to lowercase. This is a standard normalization step to ensure consistency and prevent the model from treating identical words with different capitalization (e.g., 'Pariaman', 'pariaman') as separate entities.

Tokenizing is the process of dividing sentences into discrete word units, or tokens. Sentences were segmented into individual word units, or tokens. This procedure is a fundamental step for the model's text analysis. This procedure is an essential part of the text analysis of the model. To help the model focus on more informative words we use Stopword Removal, common Indonesian words with low semantic sentiment value (such as 'which', 'in', 'and', 'is') were removed. For this task, we utilized the standard Indonesian stopword list provided by the Sastrawi library, a widely recognized toolkit for Indonesian NLP. While modern Transformer models can handle stop words, we found this step beneficial

 ISSN 2723-6471

for reducing noise in our specific, moderately-sized dataset. To reduce vocabulary variation we use Stemming, affixed words were converted to their root form (e.g., 'developing' becomes 'develop'). We employed the Sastrawi stemmer, an algorithm specifically designed and benchmarked for the Indonesian language's complex affixation rules. This normalization helps the model to group semantically related concepts, which can improve generalization on a dataset of this scale.

3.3. Class Balancing with SMOTE

The data distribution's preliminary analysis shows that there is an imbalance in the number of samples across sentiment classes [34]. We used the Synthetic Minority Over-sampling Technique (SMOTE) to avoid biassing the model in favour of the majority class [35]. SMOTE generates fresh synthetic samples for the minority class [36]. From samples in the minority class [37], one or more nearest neighbours (k-nearest neighbours) are chosen [38], and new data is generated along the line that connects these samples to form these new samples [39].

While SMOTE is an effective technique for addressing imbalance, we acknowledge its potential drawbacks, such as the risk of creating synthetic samples in overlapping class regions, which could introduce noise or generate examples that are not fully representative of the original data distribution. However, given the severity of the class imbalance in our initial dataset, we determined that the benefit of preventing the model from becoming biased towards the majority class significantly outweighed this potential risk. To prevent data leakage to the test data, this approach is exclusively used on the training data [40].

3.4. Proposed Model Architecture: IndoBERT-SupCon

The core of our study is the proposed IndoBERT-SupCon architecture, which integrates a pre-trained IndoBERT model with a dual-head, hybrid loss fine-tuning strategy. The main component of this model's three primary parts is IndoBERT, which serves as the Base Encoder [41]. The architecture, detailed in figure 2, consists of three primary components (IndoBERT Base Encoder, Parallel Heads for Dual Loss Calculation, and Hybrid Loss Function) as detail as figure 2.

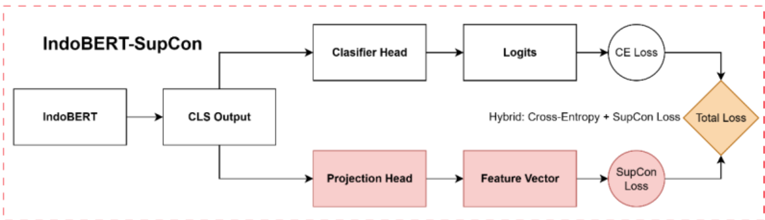


Figure 2. IndoBERT-SupCon Architecture

As the basis encoder, we employ the IndoBERT Base Encode model, which has been pre-trained on a sizable corpus of Indonesian language data [42]. We use the pre-trained indobenchmark/indobert-base-p2 model as our base encoder. For a given input text, IndoBERT generates contextualized embeddings. We specifically utilize the final hidden state vector corresponding to the special CLS output token as the aggregate representation for the entire sentence. IndoBERT transforms input text into contextually rich numeric vector representations, or embeddings [43].

The second component is the Head for Supervised Contrastive Learning (SupCon) [29]. We include a projection head trained with the SUPCON loss function in place of only using the IndoBERT output for direct classification. Refining the feature space is its goal [31]. In the course of training, this loss "pushes" embeddings of distinct sentiments farther away and "pulls" embeddings of the same sentiment closer together [26]. As a result, the classifier can more easily differentiate the more organised feature representation.

The CLS embedding vector is simultaneously fed into two separate "heads" as Parallel Heads for Dual Loss Calculation. A Classification Head, which is a standard single linear layer followed by a softmax function. This head maps the sentence embedding to the probability distribution over the four sentiment classes ('Agree', 'Disagree', 'Undecided', 'Neutral'). The output logits are used to compute the standard Cross-Entropy (CE) Loss.

A Projection Head, which is a Multi-Layer Perceptron (MLP) with one hidden layer and a ReLU activation function. This head projects the high-dimensional CLS embedding into a lower-dimensional space suitable for contrastive

 ISSN 2723-6471

learning. The resulting feature vectors are then used to calculate the Supervised Contrastive (SupCon) Loss. The SupCon loss function works to pull embeddings from the same class (positive pairs) closer together in this space, while pushing embeddings from different classes (negative pairs) further apart.

During the training phase, SupCon and conventional Cross-Entropy losses are combined to maximise the final classification accuracy as well as the quality of the feature representations. The model is optimized using a hybrid loss function that is a weighted sum of the two individual losses.

$$L_{total} = (1-\lambda)L_{CE} + \lambda . L_{SupCon} \quad (1)$$

Where L_{CE} is the Cross-Entropy and L_{SupCon} is is the Supervised Contrastive loss. The hyperparameter λ balances the contribution of each loss component. This hybrid approach ensures that the model not only learns to classify sentiments correctly (via L_{CE}) but also learns a more structured and discriminative feature space (via L_{SupCon}). This layer creates probabilities for each sentiment class ('Agree,' 'Disagree,' 'Undecided,' and 'Neutral') after receiving feature representations that have been tuned by IndoBERT and SupCon.

3.5. Experimentasl Setup and Evaluation Metrics

We used a hold-out validation approach. The dataset was divided into 80% training data and 20% testing data. This division was performed using a stratified split method to ensure that the proportion of each sentiment class ('Agree,' 'Disagree,' 'Undecided,' 'Neutral') remained consistent across both the training and testing datasets. This step is crucial to prevent biased evaluations and ensure that the model is tested on a balanced representation of the data, which is particularly important given the presence of minority classes in the initial data.

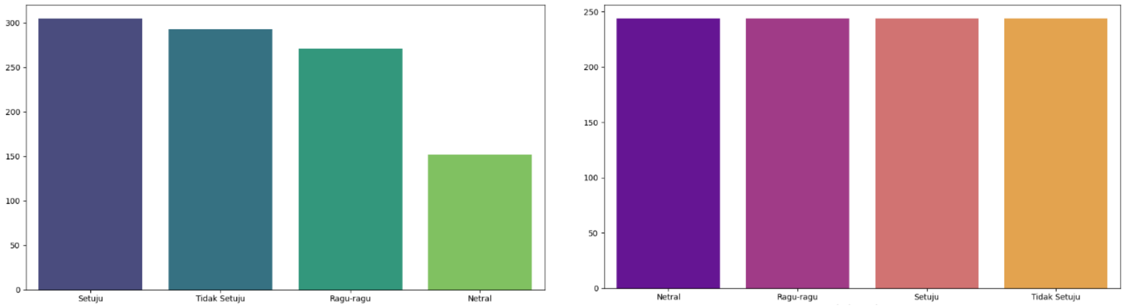
We recognize that k-fold cross-validation is the gold standard for more comprehensively measuring model generalization ability. However, given the primary goal of this study is to propose and evaluate the feasibility of a new architecture (IndoBERT-SupCon) on a specific halal tourism dataset, we focused our in-depth analysis on a representative test set. We note the use of more extensive cross-validation as an important direction for model validation in future research.

The model was trained for 10 epochs using the AdamW optimizer. A batch size of 16 and a learning rate of 2e-5 were selected for the fine-tuning process. These hyperparameters were chosen based on established best practices and recommendations from seminal works on fine-tuning BERT-based models. A small learning rate (2e-5) is commonly advised to prevent catastrophic forgetting of the pre-trained knowledge, while a batch size of 16 was chosen as it provided a good balance between stable gradient estimation and GPU memory constraints. These values were also confirmed to yield stable convergence during our preliminary experiments

Four common measures were used to assess the model's performance. Accuracy is the proportion of all accurate forecasts. Precision describes the model's ability to correctly categorise negative samples as positive. Recall or sensitivity refers to the model's capacity to recognise all true positive samples. The F1-Score provides a single metric that combines precision and recall by taking the harmonic mean of the two.

4. Results and Discussion

The experimental results of the suggested IndoBERT- SupCon model are shown in this section along with a discussion of their ramifications. Both a quantitative assessment of the model's performance and a qualitative interpretation of the sentiment analysis findings on opinions of halal travel in Pariaman are part of the analysis. To validate every element of the suggested framework, a phased model performance evaluation is carried out. These phases consist of tracking the training process, analysing data distribution, visualising keywords, and doing a final assessment using accepted metrics. Making ensuring the training data is not skewed towards a specific class is the first stage in the evaluation process. The initial data indicates that the quantity of samples varies unevenly among sentiment classifications. The SMOTE technique was used to solve this. The distribution of the training data following the oversampling procedure is shown in figure 3.



(a) (b)

Figure 3. (a) Data Distribution Before SMOTE, (b) Training Data Distribution After SMOTE

Figure 3 visually confirms that the class balancing process was successful. Each sentiment class (‘Agree’, ‘Undecided’, ‘Disagree’, ‘Neutral’) now has an equal number of samples. This step is crucial to ensure the model learns from a fair representation of each sentiment, thus preventing it from becoming overly dominant in predicting the majority class and improving its ability to recognize the minority class. To gain a qualitative understanding of the most dominant terms within each sentiment category, a word cloud was created from the data corpus. Word clouds visually represent word frequency, with larger words indicating more frequent occurrence.

Figure 4 shows frequently occurring keywords and presents a word cloud to offer a high-level visualization of the most frequent terms within the data corpus. For example, in the sentiment 'Agree', words like support, good, amplify and potential might be dominant. Conversely, in the sentiment 'Disagree', words like reduce, conflict, and uncomfortable might be most prominent. While this frequency-based approach is admittedly superficial and cannot capture deep semantic nuances on its own, it provides a useful preliminary validation of our sentiment labels. To gain a more profound understanding of the reasoning behind stakeholder sentiments, we conducted a qualitative analysis of representative comments from each category.

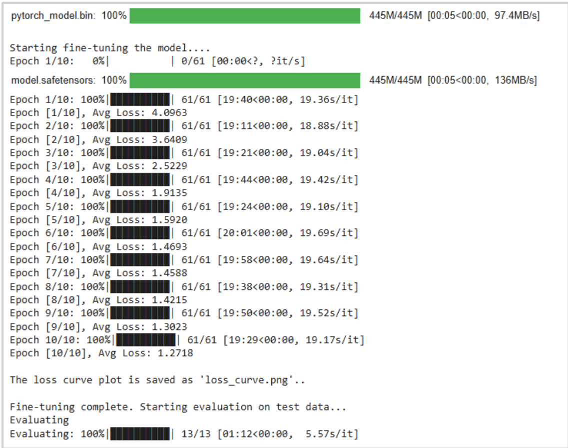
Figure 4. Word Cloud from Perception Class

For instance, the prominence of terms such as support, good, and potential aligns well with the 'Agree' sentiment, pointing towards underlying themes of economic optimism and development support. Conversely, the appearance of words like reduce, conflict, and worried in negative comments suggests that stakeholder concerns often revolve around potential social friction and negative community impacts. This initial qualitative check provides a foundation for our quantitative results. A more in-depth analysis using advanced techniques, such as topic modeling or SHAP explanations, would be a valuable direction for future research to further dissect these underlying themes.

4.1. Model Performance Evaluation

The fine-tuning process of the IndoBERT-SupCon model is closely monitored to ensure the model learns effectively. Figure 5 shows the console output logs during the training process. From the log in figure 5 it can be seen that the Average Loss value (average error) continues to decrease consistently in each epoch, from 4.0963 in Epoch 1 to 1.2718 in Epoch 10. This steady decrease is textual evidence that the model weight optimization process is running well and the model is successfully learning from the training data. For a more intuitive visualization of the training process, the loss curve is presented in figure 6.

Journal of Applied Data Sciences
Vol. 7, No. 1, January 2026, pp. 218-231
ISSN 2723-6471



```
pytorch_model.bin: 100% 445M/445M [00:05<00:00, 97.4MB/s]

Starting fine-tuning the model....
Epoch 1/10:   0%|          | 0/61 [00:00<?, ?it/s]
model.safetensors: 100% 445M/445M [00:05<00:00, 136MB/s]
Epoch 1/10: 100%|██████████| 61/61 [19:40<00:00, 19.36s/it]
Epoch [1/10], Avg Loss: 4.0963
Epoch 2/10: 100%|██████████| 61/61 [19:11<00:00, 18.88s/it]
Epoch [2/10], Avg Loss: 3.6409
Epoch 3/10: 100%|██████████| 61/61 [19:21<00:00, 19.04s/it]
Epoch [3/10], Avg Loss: 2.5229
Epoch 4/10: 100%|██████████| 61/61 [19:44<00:00, 19.42s/it]
Epoch [4/10], Avg Loss: 1.9135
Epoch 5/10: 100%|██████████| 61/61 [19:24<00:00, 19.10s/it]
Epoch [5/10], Avg Loss: 1.5920
Epoch 6/10: 100%|██████████| 61/61 [20:01<00:00, 19.69s/it]
Epoch [6/10], Avg Loss: 1.4693
Epoch 7/10: 100%|██████████| 61/61 [19:58<00:00, 19.64s/it]
Epoch [7/10], Avg Loss: 1.4588
Epoch 8/10: 100%|██████████| 61/61 [19:38<00:00, 19.31s/it]
Epoch [8/10], Avg Loss: 1.4215
Epoch 9/10: 100%|██████████| 61/61 [19:50<00:00, 19.52s/it]
Epoch [9/10], Avg Loss: 1.3023
Epoch 10/10: 100%|██████████| 61/61 [19:29<00:00, 19.17s/it]
Epoch [10/10], Avg Loss: 1.2718

The loss curve plot is saved as 'loss_curve.png'..

Fine-tuning complete. Starting evaluation on test data...
Evaluating
Evaluating: 100%|██████████| 13/13 [01:12<00:00,  5.57s/it]
```

Figure 5. Log Proses Fine-Tuning Model

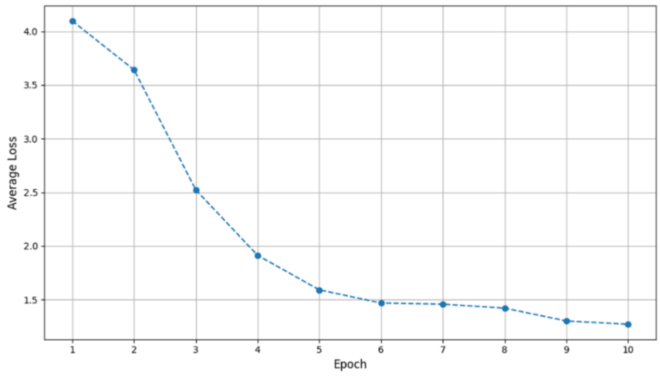
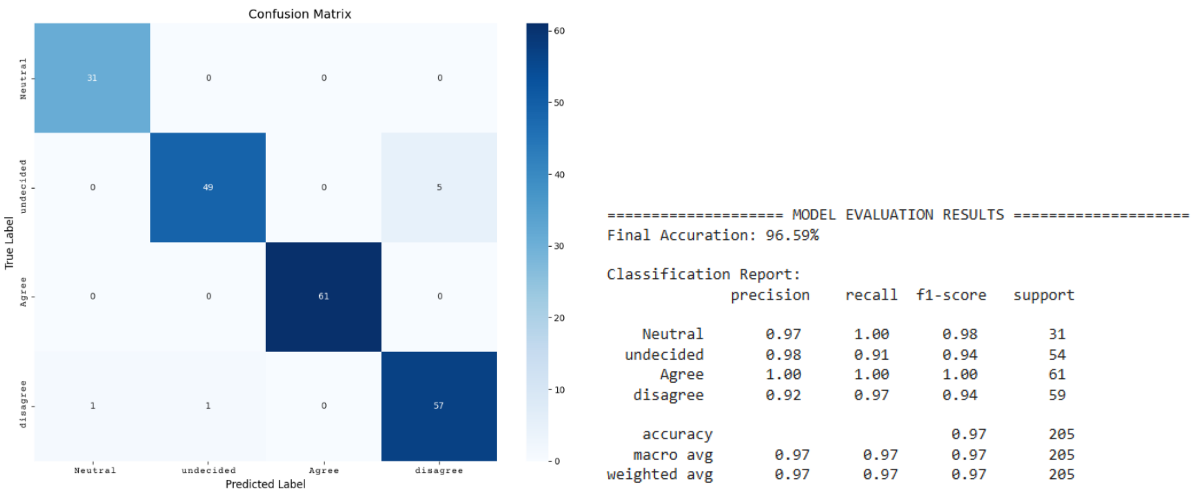


Figure 6. Model Fine-Tuning Process Log

The curve in figure 6 illustrates a smooth and steady downward trend in loss over 10 epochs. The absence of sharp spikes or fluctuations indicates that training is proceeding steadily. The curve's gradual flattening in the final epochs indicates that the model has reached a point of convergence, where it has optimally learned patterns from the training data. A final evaluation was performed on the test data to measure the model's generalization ability. Figure 7 presents a confusion matrix that visualizes the model's classification performance.




```
=================== MODEL EVALUATION RESULTS ===================
Final Accuration: 96.59%

Classification Report:
              precision    recall  f1-score   support

     Neutral       0.97      1.00      0.98        31
   undecided       0.98      0.91      0.94        54
       Agree       1.00      1.00      1.00        61
    disagree       0.92      0.97      0.94        59

    accuracy                           0.97       205
   macro avg       0.97      0.97      0.97       205
weighted avg       0.97      0.97      0.97       205
```

Figure 7. Confusion Matrix IndoBERT-SupCon

The confusion matrix in figure 7 shows that the majority of predictions (numbers on the diagonal from top left to bottom right) were correct. The number of incorrect predictions (numbers outside the diagonal) was minimal. For example, of the 61 data points with the original label 'Agree', the model successfully predicted all of them correctly (61 values in the [Agree, Agree] cells). This visually confirms the model's high accuracy and ability to distinguish between sentiment classes. The classification report further demonstrates the model's robustness, showing strong and relatively balanced performance across all classes, with F1-scores ranging from 0.94 to a perfect 1.00. The model achieved excellent performance (precision, recall, and F1-score of 1.00) on the 'Agree' class, the second-most sampled class in the test set. Near-perfect performance was also demonstrated on the 'Neutral' class (F1-score of 0.98). It is noteworthy that for the 'undecided' class, the most linguistically ambiguous, the model still managed to maintain a very high F1-score of 0.94. The recall metric of 0.91 for this class indicates that the model successfully identified 91% of all 'undecided' sentiments, a very strong result considering its semantic complexity. This consistent performance across all classes further confirms the robustness of the proposed Supervised Contrastive Learning approach.

4.2. Discussion

The comprehensive evaluation presented in Section 4.1 validates the performance of the IndoBERT-SupCon model across multiple aspects, from successful data balancing and convergence of the training process to highly accurate classification capabilities on the test data. These results collectively demonstrate that the proposed architecture is not only stable but also highly effective. To quantify the significance of these achievements and contextualize its superiority, the next step is to compare the model's performance with several baseline architectures commonly used in similar research. The success of the IndoBERT-SupCon model with an accuracy of 96.59% significantly outperforms various commonly used baseline models for sentiment analysis, as shown in table 1.

Table 1. Model Comparison

Model	Accuracy	Classification	Precision	Recall	F1-Score
Stacked Bi-LSTM + Attention	0.893	Neutral	0.91	0.97	0.94
		Undecided	0.92	0.87	0.90
		Agree	0.92	0.90	0.91
		Disagree	0.84	0.86	0.85
BI-GRU + FastText + Attention	0.868	Neutral	0.79	1.00	0.89
		Undecided	0.91	0.78	0.84
		Agree	0.89	0.95	0.92
		Disagree	0.85	0.82	0.82
BI-GRU + Fine-Tuning Embedding	0.8732	Neutral	0.84	1.00	0.91
		Undecided	0.90	0.81	0.85
		Agree	0.89	0.95	0.92
		Disagree	0.85	0.78	0.81
Fine-Tuning + IndoBERT	0.927	Neutral	0.94	1.00	0.97
		Undecided	0.96	0.81	0.88
		Agree	0.92	1.00	0.96
		Disagree	0.90	0.92	0.91
IndoBERT- SupCon (Our Proposed Model)	**0.966**	Neutral	0.97	1.00	0.98
		Undecided	0.98	0.91	0.94
		Agree	1.00	1.00	1.00
		Disagree	0.92	0.97	0.94

*the research used the same dataset to test the entire model.

From table 1, it is clear that our proposed model (IndoBERT-SupCon) provides substantial performance improvements compared to other deep learning architectures such as Bi-LSTM and Bi-GRU. More importantly, our model also outperforms IndoBERT using standard fine-tuning methods. This improvement is directly attributable to the effectiveness of the Supervised Contrastive Learning (SupCon) mechanism. Models based on Recurrent Neural Network (RNN) architectures such as Stacked Bi-LSTM and Bi-GRU, even using attention mechanisms or pre-trained embeddings like FastText, only achieved accuracy in the range of 86.8% to 89.3%. A significant performance jump was seen when using Transformer-based architectures. The standard IndoBERT Fine-Tuning model achieved 92.7% accuracy, demonstrating the superiority of pre-trained models specific to the Indonesian language. However, the pinnacle of this comparison was the proposed model, IndoBERT-SupCon, which achieved 96.6% accuracy. This represents a substantial performance improvement of approximately 3.9% over the standard IndoBERT.

The improvement extends beyond accuracy. The average F1-score, a more holistic metric, also exhibits a similar pattern, with IndoBERT-SupCon achieving a score of 0.97 compared to the standard IndoBERT's 0.93. This notable advantage provides strong empirical support for the research hypothesis: that the addition of the Supervised Contrastive Learning mechanism effectively refines IndoBERT's feature representations. While we identify the SupCon mechanism as the primary driver for the performance lift over the standard IndoBERT baseline, it is important to acknowledge that the model's overall success is likely multi-faceted. The effectiveness of the SMOTE technique in providing a balanced training distribution, alongside the high-quality and context-specific nature of our primary dataset, are also considered significant contributing factors to the final robust performance. By forcing the model to learn more discriminatory representations where similar sentiments are pulled closer together and dissimilar sentiments are pushed further apart, the decision boundary between classes becomes clearer.

While standard fine-tuning focuses solely on minimizing classification errors, SupCon actively structures the model's internal feature space. By forcing the model to learn more discriminatory representations, where similar sentiments are pulled closer together and dissimilar sentiments are pushed further apart, the decision boundary between classes becomes clearer. This has proven highly useful for validating stakeholder sentiments, which can provide robust data-driven insights for local governments to design more targeted and effective halal tourism development strategies. Finally, it is essential to consider the ethical implications of deploying such a model in a real-world policy context. While powerful, AI-driven sentiment analysis carries a responsibility to be used wisely. There is a significant risk of inadvertently marginalizing minority voices if their perspectives are underrepresented in the dataset, or perpetuating biases from the human labeling process. To mitigate these risks, this model should be employed not as an automated decision-maker, but as a supportive tool within a "human-in-the-loop" framework. The insights it provides should complement, rather than replace, direct stakeholder engagement, ensuring that policy formulation remains an inclusive and equitable proces

To validate and position the contribution of this research within the broader landscape of sentiment analysis, the performance of the IndoBERT-SupCon model is compared with results from several recent studies. It is crucial to note that this comparison, presented in table 2, is intended to provide general context and illustrate the competitiveness of our model's performance, rather than to serve as a direct, formal benchmark. As the datasets, languages, and specific tasks differ substantially across these studies, a direct claim of superiority over these models would be inappropriate**.** The comparison covers a range of models, from hybrid architectures to Transformer-based models, applied to diverse datasets.

Table 2. Model Comparison

Researcher	**Model**	**Dataset**	**Accuracy**
Bharti et al. [44]	CNN+BiGRU+SVM	ISEAR, WASSA, and Emotion-stimulus	80.11%
Anam et al. [45]	SMOTE+GRU+BiLSTM	Twitter	89.00%
Putra et al. [46]	XGBoost + RF + LSTM meta learner	Kaggle	88.00%
Owen et al. [47]	SBERT+ TextConvonet	Customized Data e-wallet service providers	86.90%
This Research	IndoBERT + Supervise Contrastive Learning	Customized Survey on Tourism	**96.59%**

From the comparison in table 2, several key points can be drawn to justify the superiority of this study. First and foremost, our proposed model, IndoBERT-SupCon, achieved an accuracy of 96.59%, significantly outperforming all the comparative studies presented. This represents a substantial performance difference, approximately 7.59% higher than the study with the second-highest accuracy (Anam et al., 89.00%). In terms of model architecture, previous studies tended to use a combination of hybrid models (e.g., CNN+BiGRU) or ensemble models (XGBoost+RF+LSTM). While these approaches are powerful, our study shows that a more effective strategy is to fine-tune a single, robust pre-trained transformer model (IndoBERT) using a sophisticated fine-tuning technique (Supervised Contrastive Learning), rather than combining multiple older architectures. This indicates that the quality of the feature representations produced by IndoBERT-SupCon is superior.

Dataset factors also play a significant role. Several comparable studies used common public datasets (ISEAR, Kaggle) or social media data (Twitter), which tend to be highly noisy. The study by Owen et al. used custom data, similar to this study, but in a different domain (e-wallet services) and with a different model (SBERT). Our model's success in achieving the highest accuracy on custom datasets from surveys and interviews demonstrates that the IndoBERT-SupCon architecture is highly reliable in handling specific and nuanced data, which is often more complex than general review text. Overall, this comparison confirms the contribution of our research: the combination of a local context-specific language model (IndoBERT), an innovative training mechanism (SupCon), and application to a high-quality primary dataset has proven to produce state-of-the-art performance that surpasses previous approaches.

5. Conclusion

This study was driven by the need for a reliable and accurate model to examine the multifaceted opinions of stakeholders on the growth of halal tourism in Pariaman, as mentioned in the Introduction. It was proposed that adding a more sophisticated learning process to a conventional fine-tuning method for IndoBERT could enhance it. In order

 ISSN 2723-6471

to improve feature representation, we developed the IndoBERT-SupCon model, which incorporates Supervised Contrastive Learning. According to the results in the Results and Discussion chapter, this expectation has been fulfilled. The suggested IndoBERT-SupCon model outperformed a number of baseline models, including a typical fine-tuned IndoBERT, with a high accuracy of 96.59%. This outcome offers compelling proof that the incorporation of Supervised Contrastive Learning, as proposed, successfully generates a more discriminative feature space, resulting in a more precise categorisation of complex emotions. This shows that the results obtained using the suggested methodology and the problem mentioned in the introduction are clearly compatible.

The success of this study brings up a number of opportunities for future development and applications. The IndoBERT-SupCon architecture can be further evaluated on larger and more varied datasets across many domains to confirm its generalisability and robustness for the advancement of the research findings. Additionally, the incorporation of aspect-based sentiment analysis may be investigated in subsequent research. The model should be improved to pinpoint certain elements (such as "facilities," "certification," and "social impact") that influence favourable or unfavourable opinions rather than categorising sentiment in general. This would yield more detailed information.It is important to acknowledge several limitations in this study. First, our comparative analysis relies on a single, stratified 80:20 data split rather than a more robust k-fold cross-validation scheme. Consequently, while the performance improvements of our proposed IndoBERT-SupCon model over the baselines are substantial and consistent across multiple metrics, we did not perform formal statistical significance testing. The primary focus of this study was the empirical demonstration of the proposed architecture's effectiveness. Therefore, a rigorous statistical validation of the model's superiority is noted as an important direction for future work.

Furthermore, this study does not include a formal analysis of computational efficiency. Our objective was centered on establishing predictive accuracy, and as such, a benchmark of the practical trade-offs between accuracy, training time, and resource requirements was considered beyond our initial scope. This remains a crucial next step for evaluating real-world deployment scenarios. Finally, limitations related to the dataset itself should be considered. The use of purposive sampling, while effective for targeting specific stakeholders, may limit the generalizability of our findings to the entire population. Additionally, while the SMOTE technique was essential for mitigating model bias due to class imbalance , it carries an inherent risk of generating synthetic samples or artifacts that may not perfectly capture the nuances of the original data distribution. These factors should be taken into account when interpreting the results.

In terms of potential applications, the Pariaman local government can use this model as a useful tool to track public opinion over time and make data-driven policy changes. A strong technique for examining stakeholder input and making sure that strategic objectives are in line with public opinion. The framework developed in this study, from data collection to analysis, can also be adopted as a blueprint for other regional development studies in Indonesia, offering a robust method for analyzing stakeholder feedback and ensuring strategic initiatives align with public sentiment.

6. Declarations

6.1. Author Contributions

Conceptualization: S.M.O., and R.A.M; Methodology: R.A.M; Software: R.A.M; Validation: S.M.O., D, and A.W; Formal Analysis: D., and A.W.; Investigation: S.M.O.; Resources: D.; Data Curation: R.A.M.,; Writing Original Draft Preparation: S.M.O., and R.A.M; Writing Review and Editing: S.M.O., and R.A.M..; Visualization: A.W.; All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The dataset generated and analyzed for this study is not publicly available due to privacy and confidentiality restrictions, as it contains qualitative responses from survey and interview participants. However, the data can be made available from the corresponding author upon reasonable request and subject to a data use agreement to ensure the protection of participant anonymity.

6.3. Funding

This research was supported by a grant from the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia under the 2025 Fundamental Research scheme.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] L. Y. Arnakim, M. F. Karim, T. N. Mursitama, and M. P. Tahalele, "Tourism in ASEAN: A Catalyst for Regional Integration," *J. ASEAN Stud.*, vol. 11, no. 1, pp. 1-12, 2023, doi: 10.21512/jas.v11i1.10506.

[2] A. Sutono, Sumaryadi, W. Rahtomo, F. F. Puksi, and O. Permana, "Re-modelling Muslim-friendly tourism value chain ecosystem: the case study of West Sumatera, Indonesia," *Cogent Soc. Sci.*, vol. 10, no. 1, pp. 1-14., 2024, doi: 10.1080/23311886.2024.2361536.

[3] Sunarmo, M. Thurisaina, A. Amri, and A. Sagita, "Effects of Halal Awareness, Service Quality and Promotion on Visitor Decisions Through Destination Imagery as a Mediator," *Indones. J. Halal Res.*, vol. 6, no. 1, pp. 13–25, 2024, doi: 10.15575/ijhar.v6i1.19489.

[4] S. Pongsakornrungsilp, P. Pongsakornrungsilp, T. Chouykaew, H. Niyomdecha, S. Chansamran, and Y. He, "Integrating Sustainability and Cultural Sensitivity: Clustering Muslim Tourist Lifestyles in the Andaman Coastal Cluster, Thailand," *Sustain.*, vol. 16, no. 21, pp. 1–15, 2024, doi: 10.3390/su16219542.

[5] Supriono, Sudarmiatin, E. T. Djatmika, and A. Hermawan, "Muslim tourist perceived value as determinant on destination brand loyalty," *Tour. Hosp. Manag.*, vol. 30, no. 3, pp. 417–431, 2024, doi: 10.20867/thm.30.3.10.

[6] D. Rusyaida, "Local wisdom as Halal tourism model in West Sumatera: an `Urf perspective," *Indones. J. Islam Muslim Soc.*, vol. 14, no. 2, pp. 259–287, 2024, doi: 10.18326/ijims.v14i2.259-287.

[7] S. Abbasian, G. Onn, and D. Nordberg, "The perception of halal concept of hoteliers in the light of social exchange theory: a Swedish study," *J. Hosp. Tour. Insights*, vol. 7, no. 4, pp. 2403–2421, 2024, doi: 10.1108/JHTI-01-2023-0038.

[8] R. Pratiwi, M. Takhim, W. N. R. Wardhani, Ragimun, A. Sonjaya, A. Rahman, E. Basmar, and B. Pambudi. "The Collaboration of Penta Helix to Develop Halal Tourism Villages in Batang, Cental Java," *Int. J. Sustain. Dev. Plan.*, vol. 19, no. 7, pp. 2753–2761, 2024, doi: 10.18280/ijsdp.190732.

[9] M. K. Anam, P. P. Putra, R. A. Malik, T. A. Putra, and Y. Elva, R. A. Mahessya, M. B. Firdaus, Ikhsan, and C. R. Gunawan. "Enhancing the Performance of Machine Learning Algorithm for Intent Sentiment Analysis on Village Fund Topic." J. Appl. Data Sci., vol. 6, no. 2, pp. 1102–1115, 2025, doi: 10.47738/jads.v6i2.637.

[10] N. R. P. Syallya, A. A. Pravitasari, and A. Helen, "NLP-Based Intent Classification Model for Academic Curriculum Chatbots in Universities Study Programs," *J. RESTI*, vol. 9, no. 1, pp. 111–117, 2025, doi: 10.29207/resti.v9i1.6276.

[11] A. S. Ekakristi, A. F. Wicaksono, and R. Mahendra, "Intermediate-task transfer learning for Indonesian NLP tasks," *Nat. Lang. Process. J.*, vol. 12, no. May, pp.1-16 100161, 2025, doi: 10.1016/j.nlp.2025.100161.

[12] I. M. Permana and F. Adam, "Impact of Islamic attributes and destination image on tourist satisfaction and revisit intentions in halal tourism: A study in Banda Aceh," *Int. J. Adv. Appl. Sci.*, vol. 11, no. 4, pp. 118–127, 2024, doi: 10.21833/ijaas.2024.04.014.

[13] A. Alam, R. T. Ratnasari, M. N. H. Ryandono, A. Prasetyo, Y. Syahidah, and F. A. Bafana, "A comparative systematic literature review between Indonesia and Malaysia Halal tourism studies (2010-2022)," *Multidiscip. Rev.*, vol. 7, no. 3, pp. 1-16 2024, doi: 10.31893/multirev.2024061.

[14] A. Ariyanto, M. Adam, S. Musnadi, and M. A. Djalil, "Empowering and Influencing Factors of Intellectual and Spiritual Capital in Improving Organizational Performance at Tourism Institutions," *J. Ecohumanism*, vol. 3, no. 7, pp. 5074–5085, 2024, doi: 10.62754/joe.v3i7.4616.

[15] F. M. Suud, K. S. Rouzi, M. Al-Fairusy, T. Na'imah, H. Harinawati, and R. B. Yasa, "Tourism Education Development on the Coastline: Challenges Empirical Study on Islamic Law Province," *J. Ilm. Peuradeun*, vol. 12, no. 3, pp. 1289–1310, 2024, doi: 10.26811/peuradeun.v12i3.1172.

[16] L. Alfarajat, "Halal Food and Medical Tourism: Muslim Patients' Experiences and Satisfaction in South Korea," *J. Relig.*

Health, vol. 63, no. 5, pp. 3274–3290, 2022, doi: 10.1007/s10943-022-01727-x.

[17] W. Listiani, A. E. M. Ningdyah, and A. J. Rohaeni, "Desire to Revisit: Memorable Experiences Drive Domestic Tourists to Return to Bali," *J. Kaji. Bali*, vol. 14, no. 1, pp. 147–168, 2024, doi: 10.24843/JKB.2024.v14.i01.p07.

[18] K. Mnassri, R. Farahbakhsh, and N. Crespi, "Multilingual Hate Speech Detection: A Semi-Supervised Generative Adversarial Approach," *Entropy*, vol. 26, no. 4, pp.1-19 2024, doi: 10.3390/e26040344.

[19] G. Kątek, R. Kozik, A. Pawlicka, M. Pawlicki, and M. Choraś, "In depth analysis for securing the truth: Addressing the fake news challenge with graph neural networks," *Neurocomputing*, vol. 654, no. June, pp. 1-10, 2025, doi: 10.1016/j.neucom.2025.131327.

[20] G. Z. Nabiilah, S. Y. Prasetyo, Z. N. Izdihar, and A. S. Girsang, "BERT base model for toxic comment analysis on Indonesian social media," *Procedia Comput. Sci.*, vol. 216, no. 2022, pp. 714–721, 2022, doi: 10.1016/j.procs.2022.12.188.

[21] M. Ahmad, M. Waqas, A. Hamza, S. Usman, I. Batyrshin, and G. Sidorov, "UA-HSD-2025: Multi-Lingual Hate Speech Detection from Tweets Using Pre-Trained Transformers," *Computers*, vol. 14, no. 6, pp. 1–24, 2025, doi: 10.3390/computers14060239.

[22] H. Honda and M. Hagiwara, "Context-dependent neuro-symbolic AI through self-supervised learning with large language models," *Neurocomputing*, vol. 654, no. August, pp. 1-16, 2025, doi: 10.1016/j.neucom.2025.131269.

[23] H. Xiao, F. Dornaika, J. Charafeddine, and J. Bi, "Metric learning-enhanced semi-supervised Graph Convolutional Network for multi-view learning," *Inf. Fusion*, vol. 124, no. October 2024, pp. 1-14, 2025, doi: 10.1016/j.inffus.2025.103420.

[24] F. Dornaika, Z. Ibrahim, J. Charafeddine, and A. Bosaghzadeh, "Towards dynamic self-training for scalable semi-supervised learning on graphs," *Neurocomputing*, vol. 654, no. August, pp. 1-14, 2025, doi: 10.1016/j.neucom.2025.131223.

[25] He, Z., Wang, H., and Li, L. Fake News Detection Based on Contrastive Learning and Cross-Modal Interaction. Symmetry 2025, vol 17, no.8, pp. 1-22, 2025, doi: 10.3390/sym17081260.

[26] S. Dehghan and M. F. Amasyali, "SupMPN: Supervised Multiple Positives and Negatives Contrastive Learning Model for Semantic Textual Similarity," *Appl. Sci.*, vol. 12, no. 19, pp. 1-20 2022, doi: 10.3390/app12199659.

[27] S. Kang, "LLM-Guided Weighted Contrastive Learning with Topic-Aware Masking for Efficient Domain Adaptation: A Case Study on Pulp-Era Science Fiction," Electron. MDPI, vol. 14, no. 17, pp. 1–24, 2025, doi: 10.3390/electronics14173351.

[28] E. Hashmi, S. Y. Yayilgan, M. M. Yamin, M. Abomhara, and M. Ullah, "Self-supervised hate speech detection in Norwegian texts with lexical and semantic augmentations," *Expert Syst. Appl.*, vol. 264, pp. 1-13, 2024, doi: 10.1016/j.eswa.2024.125843.

[29] M. M. Akhtar, N. S. Bhuiyan, R. Masood, M. Ikram, and S. S. Kanhere, "BotSSCL: Social Bot Detection with Self-Supervised Contrastive Learning," *Online Soc. Networks Media*, vol. 48, no. April, p. 1-16, 2025, doi: 10.1016/j.osnem.2025.100318.

[30] X. Huang, F. Schmelter, M. T. Irshad, A. Piet, M. A. Nisar, C. Sina, and M. Grzegorzek, "Optimizing sleep staging on multimodal time series: Leveraging borderline synthetic minority oversampling technique and supervised convolutional contrastive learning," *Comput. Biol. Med.*, vol. 166, August, pp. 1-17, 2023, doi: 10.1016/j.compbiomed.2023.107501.

[31] L. Kong, W. Yang, and F. Wei, "Implicit Emotion Analysis Based on Improved Supervised Contrastive Learning," *Electron.*, vol. 12, no. 10, pp. 1–12, 2023, doi: 10.3390/electronics12102172.

[32] Q. Wang, W. Zhang, T. Lei, and D. Peng, "Grouped Contrastive Learning of Self-Supervised Sentence Representation," *Appl. Sci.*, vol. 13, no. 17, pp. 1-17, 2023, doi: 10.3390/app13179873.

[33] Y. Huang, J. Liu, and C. Lv, "Chains-BERT: A High-Performance Semi-Supervised and Contrastive Learning-Based Automatic Question-and-Answering Model for Agricultural Scenarios," *Appl. Sci.*, vol. 13, no. 5, pp.1-16, 2023, doi: 10.3390/app13052924.

[34] M. Ali, "Classification of imbalanced travel mode choice dataset with SMOTE and prediction using interpretable machine learning," *Sustain. Futur.*, vol. 10, no. August, pp. 1-13, 2025, doi: 10.1016/j.sftr.2025.101119.

[35] A. Kummer, T. Ruppert, T. Medvegy, and J. Abonyi, "Machine learning-based software sensors for machine state monitoring - The role of SMOTE-based data augmentation," *Results Eng.*, vol. 16, no. 1, pp. 1-10, 2022, doi: 10.1016/j.rineng.2022.100778.

[36] D. Gonzalez-Cuautle, A.Hernandez-Suarez, G. Sanchez-Perez, L.K. Toscano-Medina, J. Portillo-Portillo, J. Olivares-Mercado H.M. Perez-Meana, and A.L. Sandoval-Orozco "Synthetic minority oversampling technique for optimizing classification tasks in botnet and intrusion-detection-system datasets," *Appl. Sci.*, vol. 10, no. 3, pp. 1-19 2020, doi: 10.3390/app10030794.

[37] M. Umer, S. Sadiq, M. M. S. Missen, Z. Hameed, Z. Aslam, M. A. Siddique, and M. NAPPI, "Scientific papers citation analysis using textual features and SMOTE resampling techniques," *Pattern Recognit. Lett.*, vol. 150, no. 1, pp. 250–257, 2021, doi: 10.1016/j.patrec.2021.07.009.

[38] M. Yusoff, Y. Mahmud, P. A. R. Azmi, and M. T. M. Sallehud-din, "The improvement of SMOTE-ENN-XGBoost through Yeo Johnson strategy on Dissolved Gas Analysis dataset," *Energy Reports*, vol. 13, no. April, pp. 6281–6290, 2025, doi: 10.1016/j.egyr.2025.05.013.

[39] J. Fonseca and F. Bacao, "Geometric SMOTE for imbalanced datasets with nominal and continuous features," *Expert Syst. Appl.*, vol. 234, no. July, pp. 1-9, 2023, doi: 10.1016/j.eswa.2023.121053.

[40] I. Aruleba and Y. Sun, "Enhanced credit risk prediction using deep learning and SMOTE-ENN resampling," *Mach. Learn. with Appl.*, vol. 21, no. 1, pp. 1-10, 2025, doi: 10.1016/j.mlwa.2025.100692.

[41] R. Pramana, M. Jonathan, H. S. Yani, and R. Sutoyo, "A Comparison of BiLSTM, BERT, and Ensemble Method for Emotion Recognition on Indonesian Product Reviews," *Procedia Comput. Sci.*, vol. 245, no. C, pp. 399–408, 2024, doi: 10.1016/j.procs.2024.10.266.

[42] M. T. Uliniansyah M. T. Uliniansyah, I. Budi, E. Nurfadhilah, D. I. N. Afra, A. Santosa, A. D. Latief, A. Jarin, G. Gunarso, M. A. Jiwanggi, N. N. Hidayati, R. Fajri, R. R. Suryono, S. Pebiana, S. Shaleha, T. W. Ramdhani, and T. Sampurno, "Twitter dataset on public sentiments towards biodiversity policy in Indonesia," *Data Br.*, vol. 52, no. 1, pp. 1-14, 2024, doi: 10.1016/j.dib.2023.109890.

[43] A. Jazuli, Widowati, and R. Kusumaningrum, "Optimizing Aspect-Based Sentiment Analysis Using BERT for Comprehensive Analysis of Indonesian Student Feedback," *Appl. Sci.*, vol. 15, no. 1, pp. 1–28, 2025, doi: 10.3390/app15010172.

[44] S. K. Bharti S. Varadhaganapathy, R. K. Gupta, P. K. Shukla, M. Bouye, S. K. Hingaa, and A. Mahmoud, "Text-Based Emotion Recognition Using Deep Learning Approach," *Comput. Intell. Neurosci.*, vol. 2022 special issues, no. 1, pp.1-8, doi: 10.1155/2022/2645381.

[45] M. K. Anam M. Munawir, L. Efrizoni, N. Fadillah, W. Agustin, I. Syahputra, T. P. Lestari, M. B. Firdaus, L. Lathifah, and A. K. Sari, "Improved Performance of Hybrid GRU-BiLSTM for Detection Emotion on Twitter Dataset," *J. Appl. Data Sci.*, vol. 6, no. 1, pp. 354–365, 2025, doi: 10.47738/jads.v6i1.459.

[46] P. P. Putra, M. K. Anam, A. S. Chan, A. Hadi, and N. Hendri, "Optimizing Sentiment Analysis on Imbalanced Hotel Review Data Using SMOTE and Ensemble Machine Learning Techniques," *J. Appl. Data Scie*, vol. 6, no. 2, pp. 936–951, 2025, doi: 10.47738/jads.v6i2.618.

[47] C. Owen and D. Suhartono, "Intent Classification in Artificial Intelligence-Based Customer Service Chatbot for E-Wallet Service Providers," *Int. J. Comput. Digit. Syst.*, vol. 17, no. 1, pp. 1–9, 2025, doi: 10.12785/ijcds/1571046212.