

## SENTIMENT ANALYSIS OF EMOJI AND LATINIZED ARABIC IN INDONESIAN YOUTUBE COMMENTS: A LABERT-LSTM MODEL

M. Noer Fadli Hidayat<sup>1\*</sup>, Didik Dwi Prasetya<sup>2</sup>, Triyanna Widiyaningtyas<sup>3</sup>

Department of Electrical Engineering and Informatics, Universitas Negeri Malang, Indonesia<sup>123</sup>

Department of Informatics Engineering, Universitas Nurul Jadid Probolinggo, Indonesia<sup>1</sup>

m.noer.2305349@students.um.ac.id<sup>1</sup>, didikdwi@um.ac.id<sup>2</sup>, triyannaw.ft@um.ac.id<sup>3</sup>

Received: 15 January 2025, Revised: 02 May 2025, Accepted: 05 May 2025

\*Corresponding Author

### ABSTRACT

This study addresses the challenges of sentiment analysis on Indonesian-language YouTube comments, which are complex due to the use of dialects, slang words, emojis, and Latinized Arabic text. The proposed LABERT-LSTM model integrates BERT for deep feature extraction and Bi-LSTM to capture word sequence context effectively. The dataset comprises 24,593 YouTube comments from five renowned Islamic preachers discussing the topic of "tahlilan". After data preprocessing, the model was evaluated using accuracy, precision, recall, and F1-score metrics. The results demonstrate that LABERT-LSTM achieved an accuracy of 0.95756, precision of 0.94014, recall of 0.91815, and an F1-score of 0.92868, outperforming standalone BERT and Bi-LSTM models by reducing misclassification and improving predictions for negative, positive, and neutral sentiment classes. Future research recommendations include expanding the dataset to other social media platforms, adopting advanced NLP techniques, conducting studies in other languages, and optimizing the model for enhanced performance and computational efficiency.

**Keywords :** Sentiment Analysis, Latinized Arabic, Emoji, BERT, Bi-LSTM, LABERT-LSTM

### 1. Introduction

The rapid growth of internet usage has significantly transformed how users interact through social media platforms, particularly on video-based platforms like YouTube, TikTok, and Instagram Reels (Suhaimin et al., 2023). Among these, YouTube remains the most dominant, with over 5 hours of video uploaded every second (Chakma et al., 2022; Wang et al., 2022; Yao et al., 2022). Beyond content consumption, the comment section of YouTube has evolved into a rich source of public sentiment, where users express opinions, emotions, and criticisms (Saikia et al., 2023). These comments, however, present analytical challenges due to informal language, dialects, and a unique characteristic in Indonesian Islamic discourse: the frequent use of Latinized Arabic expressions (e.g., *Alhamdulillah*, *Astaghfirullah*). Despite their semantic richness, such expressions are underrepresented in current sentiment analysis studies, which largely focus on standard Indonesian text, emoticons, or explicit emotional cues. This underutilization represents a significant gap in the field, particularly in capturing culturally embedded sentiments expressed in hybrid language forms. Therefore, addressing this gap is essential for improving sentiment classification accuracy in Indonesian-language datasets.

Several previous studies have explored sentiment analysis (Z. Li & Zou, 2024) in the Indonesian (Yunitasari et al., 2019) context using traditional machine learning (ML) techniques such as Naïve Bayes (Aribowo et al., 2021), Support Vector Machines (Monesa & Jayadi, 2023), or deep learning models like CNN and LSTM (Kusumaningrum et al., 2023; Rahmanti et al., 2021). While these approaches have demonstrated acceptable performance, most are limited in capturing the nuanced semantics found in informal or mixed-language text, particularly involving religious or cultural expressions like Latinized Arabic. Even recent attempts that utilize BERT or hybrid BERT-CNN-LSTM models (Aribowo et al., 2021; Murfi et al., 2024) still overlook the presence of embedded symbolic sentiment (Karo et al., 2022; Yulita et al., 2023) conveyed through Arabic expressions or emoji (Elfajr & Sarno, 2018; Yunitasari et al., 2019), which are prevalent in Indonesian YouTube discourse. These models tend to generalize sentiment polarity based on syntactic features without incorporating cultural or religious cues embedded in the language. Hence, there is a pressing need to develop a model that not only

processes deep contextual relationships within text (via BERT) but also understands sequential emotional cues (via Bi-LSTM)—especially for sentiment-rich environments such as Islamic YouTube content (Chrismanto et al., 2023). This is the core rationale behind the proposed LABERT-LSTM model.

To address the research gap identified above, this study introduces a hybrid deep learning architecture—LABERT-LSTM—which synergistically combines Bidirectional Encoder Representations from Transformers (BERT) and Bidirectional Long Short-Term Memory (Bi-LSTM). BERT is utilized as a pre-trained transformer model capable of generating rich, contextualized word embeddings that account for the bidirectional nature of language (Jia et al., 2023). This capacity is essential for capturing the implicit sentiment conveyed in religious expressions and informal comments that may not follow standard grammatical rules. Meanwhile, Bi-LSTM is incorporated to further capture temporal dependencies and sequential sentiment cues across text, particularly valuable for handling the syntactic flow of mixed-language phrases and the positional impact of emojis. Unlike traditional models that treat tokens as independent or rely on surface-level lexical features, LABERT-LSTM aims to provide a deep semantic and emotional understanding of Indonesian YouTube comments, particularly in cases enriched by Latinized Arabic expressions and emoji usage—features that are frequently neglected in prior sentiment analysis approaches.

Building upon the above rationale, this study aims to develop and evaluate the LABERT-LSTM model for sentiment analysis in Indonesian-language YouTube comments, with a particular focus on text enriched by Latinized Arabic expressions, emojis, and slang. Unlike previous works that treat such elements as noise or exclude them entirely during preprocessing, this study treats them as critical indicators of sentiment and cultural context. To achieve this, the model integrates contextualized embeddings from BERT with sequential modeling via Bi-LSTM, thereby enhancing the model's ability to capture nuanced sentiment patterns. The dataset comprises 24,593 comments collected from YouTube videos featuring five prominent Indonesian preachers discussing the culturally debated topic of *tahlilan*. This topic was selected due to its high engagement and polarity in public discourse, making it an ideal case for testing the model's robustness in handling mixed-language sentiment. By explicitly incorporating non-standard elements—Latinized Arabic, emojis, and slang—into the model's learning process, this study contributes not only to improved sentiment classification accuracy but also to the advancement of context-aware natural language processing (NLP) in low-resource, culturally specific language domains.

Therefore, this study has three main objectives: (1) to investigate the role of Latinized Arabic expressions and emojis in shaping sentiment in Indonesian YouTube comments, (2) to design a hybrid model—LABERT-LSTM—that effectively captures both contextual and sequential sentiment cues, and (3) to empirically evaluate the model's performance across varying data conditions. The novelty of this study lies in its explicit integration of Latinized Arabic as a sentiment-bearing linguistic feature, which to the best of our knowledge, has not been systematically addressed in prior Indonesian sentiment analysis research. Additionally, the use of emoji and slang is not merely tokenized but semantically interpreted via lexicon-based annotation, enhancing emotional recognition. This research contributes to the growing field of cross-lingual and culturally grounded sentiment analysis, particularly in underrepresented languages and informal communication settings. The findings are expected to offer both theoretical implications—by extending sentiment analysis frameworks to hybrid linguistic constructs—and practical benefits in areas such as social media monitoring, online moderation, and localized natural language understanding (NLU) systems.

## 2. Literature Review

Various studies have been conducted to address the challenges of sentiment analysis in Indonesian, a language rich in dialects, borrowed words, and informal usage.

### A. Sentiment Analysis in Indonesian Social Media Platforms

Sentiment analysis on Indonesian-language text has received considerable attention due to the unique linguistic characteristics and social dynamics of Indonesian users. Studies have been

conducted across various platforms, such as Twitter (Rahmanti et al., 2022) (Yunitasari et al., 2019), YouTube (Aribowo et al., 2021), and e-commerce platforms (Murfi et al., 2024). For instance, (Rahmanti et al., 2022) analyzed COVID-19 discussions on Twitter using Naïve Bayes, reaching 90.26% accuracy. Meanwhile, (Aribowo et al., 2021) explored sentiment in YouTube comments using Cross-Domain Sentiment Analysis (CDSA) and achieved 91.19% accuracy by leveraging likes and dislikes as features. These studies highlight the diversity and richness of sentiment across platforms, yet most of them neglect culturally embedded expressions such as Latinized Arabic, which are common in religious and traditional discourse.

## **B. Machine Learning and Deep Learning Models for Sentiment Classification**

Traditional machine learning models like SVM, Naïve Bayes, and Random Forest have been widely used for sentiment classification in Indonesian text, but they often struggle with informal, slang-rich, or mixed-language content (Monesa & Jayadi, 2023) (Yunitasari et al., 2019). Recent approaches have adopted deep learning methods, including CNN (Kusumaningrum et al., 2023), LSTM, and hybrid CNN-LSTM or BERT-CNN models (Murfi et al., 2024), showing improvements in F1-score and general accuracy. However, these models tend to rely solely on surface lexical features without fully capturing semantic richness or sequential dependencies, particularly for domain-specific phrases like Latinized Arabic. In contrast, the proposed LABERT-LSTM integrates contextual embedding (BERT) and sequential modeling (Bi-LSTM) to better capture both deep semantics and temporal sentiment flow, especially within informal or symbolic expressions.

## **C. Enriching Sentiment Models with Emojis, Slang, and Demographics**

Several studies have incorporated additional data sources to enhance sentiment classification, such as user demographics CNN (Kusumaningrum et al., 2023), emoticons (Elfajr & Sarno, 2018), and engagement metrics like likes and views (Yulita et al., 2023). (Yunitasari et al., 2019) successfully improved sarcasm detection in tweets using emoticon-weighted sentiment classification. However, most existing models either strip emoji during preprocessing or treat them as irrelevant noise. This study diverges by treating emojis as emotion-rich features, assigning sentiment weights through a dedicated lexicon, and combining them with Latinized Arabic and Indonesian slang. This fusion helps the LABERT-LSTM model handle informal digital expressions holistically—something that remains underexplored in prior research.

## **D. Enriching Sentiment Models with Emojis, Slang, and Demographics**

One of the most persistent challenges in sentiment analysis is handling mixed-language expressions and context-specific religious or cultural terms. Previous studies rarely address this issue directly, often omitting Latinized Arabic words entirely from preprocessing pipelines (Chrismanto et al., 2023). These expressions, while not in standard Indonesian, carry powerful emotional or evaluative connotations, especially in religious communities. Our study builds on this gap by explicitly curating a Latinized Arabic lexicon and treating these terms as first-class semantic features. Unlike prior works, LABERT-LSTM attempts to extract sentiment value not only from standard lexical cues but also from symbolic, religious, and affective markers that are often overlooked.

## **3. Research Methods**

This study adopts an experimental approach to sentiment analysis using a hybrid LABERT-LSTM model. The dataset comprises 24,593 comments retrieved from YouTube videos of five widely followed Indonesian Islamic preachers—KH. Ahmad Bahaiddin Nursalim, Buya Yahya, Ustadz Abdul Somad, Ustadz Adi Hidayat, and Ustadz Khalid Basalamah—discussing the culturally significant topic of *tahlilan*. These preachers were selected due to their high volume of engagement and diverse follower demographics, ensuring a rich and varied sentiment dataset. However, the authors acknowledge the potential for selection bias, which may reflect specific religious leanings. Future studies are recommended to broaden the scope to include preachers or speakers from varying theological orientations for better

generalizability. The proposed LABERT-LSTM model was evaluated against baseline methods, including Naïve Bayes, Bi-LSTM, and standalone BERT, to assess comparative performance. Additionally, given the imbalance in sentiment class distribution (e.g., more positive than neutral), the model employed weighted loss functions during training to mitigate classification bias toward dominant classes.

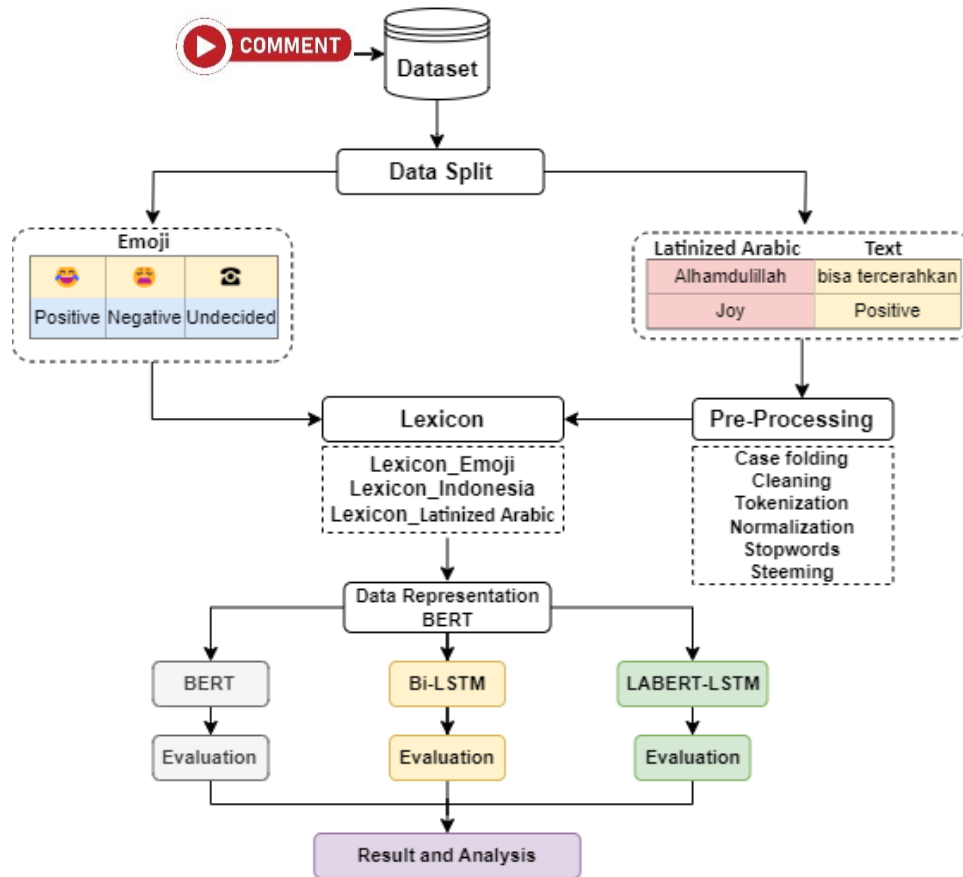


Fig. 1. The workflow of the proposed LABERT-LSTM model

### 3.1 Dataset

The dataset for this study was obtained from the comment sections of YouTube videos featuring five prominent Islamic preachers in Indonesia: KH. Ahmad Bahaiddin Nursalim (Gus Baha), Ustadz Abdul Somad, Ustadz Adi Hidayat, Buya Yahya, and Ustadz Khalid Basalamah. These preachers were selected based on their national popularity, engagement volume, and contrasting perspectives on Islamic traditions, providing a diverse representation of opinions related to the topic of *tahlilan*—a religious practice that continues to generate theological debate in Indonesia. Data collection was conducted using the YouTube Data API, focusing on videos explicitly discussing *tahlilan* and uploaded by verified or credible channels. Only videos with high viewer engagement (comments, likes, and views) were included, ensuring the dataset reflected active public discourse.

While the dataset offers rich sentiment variation, it is acknowledged that limiting data to only five preachers may introduce selection bias, particularly reflecting the audiences of specific religious orientations. To address this, the chosen figures span different Islamic backgrounds, from traditionalist to reformist, allowing the dataset to capture a spectrum of theological sentiment. Future work is encouraged to expand the dataset by including broader speaker diversity and incorporating other platforms such as Facebook, TikTok, or Instagram to enhance generalizability and cross-platform comparison.

### 3.2 Data Split

Following the data collection process, the dataset was categorized based on sentiment polarity into three classes: positive (1), negative (-1), and neutral (0). The labeling process incorporated multiple sentiment indicators, including text polarity from the lexicon, Latinized Arabic expressions, and emojis. Each comment was labeled based on the aggregated sentiment value from these components. To ensure accuracy in emoji interpretation, emojis were extracted and annotated independently using the *Emoji Sentiment Ranking v1.0* (Kralj Novak et al., 2015), which assigns sentiment scores based on large-scale human evaluations. The relative position of emojis within the comment was preserved during tokenization to retain contextual information.

The dataset was then split into training, validation, and test sets using an 80:10:10 ratio, stratified by sentiment class to maintain proportional distribution. However, analysis revealed a mild class imbalance, with negative comments slightly outnumbering neutral and positive ones. To address this, the LABERT-LSTM model was trained using a weighted categorical cross-entropy loss function, giving higher importance to minority classes. This technique helps prevent the model from being biased toward dominant sentiment categories and improves performance in underrepresented classes such as neutral sentiment.

### 3.3 Data Pre-processing

The main challenge in the pre-processing stage lies in text cleaning and categorization, as the text in Indonesian comments exhibits a wide variety of data characteristics. Pre-processing is a crucial step in improving the quality of categorization by eliminating noise or distractions from the text phrases (Ahda et al., 2024). To address this issue, the data pre-processing in this research is conducted through the following steps:

- Case folding  
The process of standardizing all text within a phrase or document to lowercase. The goal is to reduce variation in the text due to differences in the use of uppercase and lowercase letters.
- Cleaning  
Removing unnecessary features for analysis, such as URLs, numbers, punctuation, and non-alphabetic characters. This step is performed to cleanse the data of irrelevant elements.
- Tokenizing  
Splitting the text or phrases into individual words based on the unit of analysis. This process generates word vectors, commonly referred to as a "bag of words" (BOW), where each word is treated as a separate unit for further analysis. For example, "saya senang sekali" becomes ["saya", "senang", "sekali"].
- Normalization  
In the normalization stage, if repeated characters are found in a sequence, they are reduced to a single character. This is because many slang words are written with repeated characters, for example, "amiiiiin" is changed to "amin," "baaiikkk" to "baik," and "baguuuuus" to "bagus." The next step is to remove words consisting of only one character, such as "y" or "t." Then, slang words are converted into standard Indonesian words, such as "udah" into "sudah," "gitu" into "begitu," "emang" into "memang," "liat" into "lihat," and "nggak" into "tidak." For this conversion process, we compiled a slang dictionary containing 15,006 Indonesian slang words.
- Stop-word  
Removing frequently occurring but insignificant words to reduce the corpus size without sacrificing important information contained within the text. Examples include words like "dan", "atau", "di", and "itu".
- Stemming  
Processing words into their root form. This approach reduces word complexity by removing affixes, including prefixes, infixes, and suffixes, so that each word is in its base form. For example, "membahagiakan" becomes "bahagia".

The preprocessing stage in this study was carefully designed to handle the linguistic complexity and informality found in Indonesian YouTube comments, which often include slang,

emojis, and Latinized Arabic expressions. Rather than treating these features as noise, the preprocessing pipeline was constructed to preserve and annotate them for sentiment value. The goal of this stage is to transform raw textual input into a clean, structured, and semantically meaningful format suitable for deep learning input. The following steps were implemented:

- Case Folding: All text was converted to lowercase to standardize word forms and reduce dimensionality caused by case variation (e.g., "Bagus" vs "bagus").
- Cleaning: Irrelevant components such as URLs, numbers, excessive whitespace, and non-alphabetic characters (except emojis and Latinized Arabic tokens) were removed. Emojis and Arabic terms were preserved for semantic analysis.
- Tokenizing: Sentences were split into individual word tokens using whitespace and punctuation delimiters. Emojis were tokenized as individual units to maintain their emotional signal.
- Normalization: Slang and informal variants were mapped to their formal Indonesian equivalents using a custom-built dictionary of 15,006 slang words. Repeated characters were reduced (e.g., "baaaikkk" → "baik"), and one-letter non-meaningful tokens were removed.
- Stop-word Removal: Common functional words (e.g., "yang", "dan", "itu") with minimal semantic weight were removed to reduce feature space size.
- Stemming: Morphological stemming was performed to convert each word to its base/root form (e.g., "membahagiakan" → "bahagia"), using the Sastrawi stemmer for Bahasa Indonesia.

Each comment was also tagged for presence of Latinized Arabic and emoji sequences, which were then passed into lexicon-based sentiment scoring in the next processing stage (see Section 3.4). This approach ensures that all informal and symbolic expressions contribute meaningfully to the sentiment classification process.

### 3.4 Lexicon Construction

A central innovation of this study lies in the construction and integration of two specialized lexicons: (1) a Latinized Arabic lexicon, and (2) an Indonesian emoji-sentiment lexicon, both of which enhance the model's capability to interpret culturally nuanced expressions.

Table 1 - Results of Latinized Arabic Annotation.

Latinized Arabic	Translation	Emotion	Sentiment	Label
Bismillah	In the Name of Allah	Happy	Positive	1
La ilaha illallah	There is no god but Allah	Uncertain	Neutral	0
Insya Allah	If Allah wills	Happy	Positive	1
Tawakkaltu 'Alallah	I put my trust in Allah	Sad	Negative	-1
Alhamdulillah	All Praise be to Allah	Happy	Positive	1
Jazakallah Khairan	May Allah reward you with goodness	Happy	Positive	1
Masya Allah	As per Allah's Will	Sad	Positive	1
Barakallah	May Allah bless	Happy	Positive	1
Hasbiyallah	Allah is sufficient for me	Sad	Negative	-1
Allahu Akbar	Allah is the greatest	Uncertain	Neutral	0
Subhanallah	Glory be to Allah	Uncertain	Neutral	0
Na'udzubillah	We seek refuge in Allah	Sad	Negative	-1
Astaghfirullah	I seek forgiveness from Allah	Sad	Negative	-1
La haula wala Quwwata Illa Billah	There is no power and strength except (with) Allah	Sad	Negative	-1
Fi Amanillah	May you be under Allah's protection	Happy	Positive	1
Inna Lillahi wa Inna Ilaihi Rajiun	Indeed, we belong to Allah and to Him we shall return	Sad	Negative	-1
Amin	Ameen	Uncertain	Neutral	0
Ya Allah	O Allah	Uncertain	Neutral	0

The Latinized Arabic lexicon was curated manually by Arabic language experts at Universitas Nurul Jadid, in consultation with previous sentiment studies (Samreena & Ali,

2024). The lexicon contains 18 commonly used phrases such as *Alhamdulillah*, *Astaghfirullah*, *Subhanallah*, and *Inna lillahi*, each mapped to sentiment values—positive (1), negative (-1), or neutral (0)—based on their typical use in social and religious contexts. For example, *Alhamdulillah* is associated with gratitude and thus labeled positive, while *Astaghfirullah* is often uttered in disapproval, indicating negative sentiment.

Separately, emoji sentiment mapping was performed using the Emoji Sentiment Ranking v1.0 dataset (Novak et al., 2015), which assigns valence scores to 752 emojis based on large-scale user interpretation. These were also mapped into three classes (positive, negative, neutral) and annotated accordingly. Each emoji was extracted from its source comment and stored as a separate token during preprocessing while maintaining its sequential position in the sentence.

Additionally, the study used an Indonesian sentiment lexicon consisting of 10,196 words compiled from YouTube comments and earlier annotated datasets. This lexicon served as the baseline for identifying sentiment polarity in native Indonesian text. In the sentiment scoring phase, each comment was processed using a weighted aggregation function combining scores from the three lexicons (Arabic, emoji, Indonesian) to determine its overall sentiment label.

This lexicon-based architecture provides the LABERT-LSTM model with prior semantic knowledge, enhancing its capacity to interpret mixed-language and symbol-rich input more accurately—particularly in religious or cultural contexts where conventional models often fail.

### 3.5 BERT Representation Data

To represent textual input in a context-aware and semantically rich format, this study employed IndoBERTBASE, a pre-trained BERT model optimized for the Indonesian language (Murfi et al., 2024). BERT was chosen for its proven ability to capture bidirectional contextual relationships in text, making it ideal for interpreting subtle sentiment shifts caused by slang, mixed-language phrases, or symbolic expressions like Latinized Arabic.

The BERT input processing pipeline involved four main steps:

- 1) Tokenization: Using the WordPiece tokenizer, each sentence was broken into subword tokens. Special tokens such as [CLS] (classification marker) were added at the beginning, and [SEP] (separator) at the end. This helped mark the boundary of each comment.
- 2) Padding: To ensure uniform input dimensions, all token sequences were padded to a fixed length of 128 tokens using the [PAD] token. This enabled batch training without truncating meaningful data.
- 3) Numericalization: Each token was converted into an integer index using BERT's vocabulary of 30,522 subword units.
- 4) Embedding: Each indexed token was mapped to a 768-dimensional vector through BERT's embedding layers, which include:
  - a. Token Embeddings (word semantics)
  - b. Segment Embeddings (sequence distinction, not used here due to single-sentence input)
  - c. Positional Embeddings (sequence order)

The final output of the BERT encoder is a matrix of size  $128 \times 768$ , where 128 is the token length and 768 is the hidden size. This matrix was then passed as input to the Bi-LSTM layer, which further modeled the sequential and emotional structure of the sentence.

By leveraging IndoBERT's pretrained knowledge and contextual sensitivity, the model was able to interpret the meaning of Latinized Arabic expressions in varied contexts, differentiating between religious affirmation and emotional critique—something traditional embeddings like Word2Vec or GloVe fail to capture.

### 3.6 Bi-LSTM

After obtaining contextualized token embeddings from the BERT layer, the model passes these representations to a Bidirectional Long Short-Term Memory (Bi-LSTM) network. Unlike traditional unidirectional LSTM, Bi-LSTM processes the sequence in both forward and backward directions, enabling it to capture semantic dependencies that may span across distant

tokens, including long-range emotional cues expressed through slang, emojis, or Latinized Arabic phrases (Sahai et al., 2022; Naga et al., 2023).

The Bi-LSTM architecture in this study consists of 128 hidden units, split equally into forward and backward LSTM cells. For each token in the sequence, the forward LSTM computes a hidden state from left to right ( $h \rightarrow$ ), while the backward LSTM processes it from right to left ( $h \leftarrow$ ). These two states are then concatenated to form a complete representation:

$$h_t = [h_t^{\rightarrow}; h_t^{\leftarrow}] \tag{1}$$

This bidirectional processing is especially useful in detecting contextual shifts and subtle negations, such as in sentences where emojis modify sentiment ("*Bagus banget 😊*") or when Latinized Arabic expresses sarcasm or lamentation ("*Alhamdulillah sih, tapi sedih.*"). The Bi-LSTM effectively captures such variations by modeling token interactions across the sentence, both from the beginning and the end.

The output from Bi-LSTM is passed to a fully connected dense layer, followed by a softmax layer to predict the final sentiment class (positive, negative, or neutral). Dropout regularization (rate = 0.5) was applied after the Bi-LSTM layer to prevent overfitting during training. The internal structure of each LSTM unit is shown in Figure 2, detailing how input gates, forget gates, and output gates interact to manage memory flow and preserve long-term dependencies.

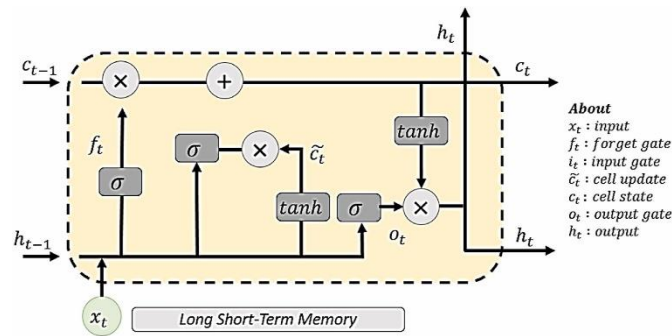


Fig. 2. LSTM Unit

### 3.7 LABERT-LSTM

This study develops a sentiment analysis model for Latinized Arabic in YouTube comments using a combined predictive model of BERT and Bi-LSTM, referred to as the Latinized Arabic Bidirectional Encoder Representation of Bidirectional Long Short-Term Memory Transition (LABERT-BiLSTM). In this model, BERT is utilized as the initial module, while Bi-LSTM serves as the subsequent module. This approach enhances the model initially trained with BERT, which is then integrated into BiLSTM for further training.

The input and output data are structured within the BERT model.  $E = \{e_1, e_1, \dots, e_N\}$  represents the embedding vectors of the words in a comment, generated by the BERT model. If  $E[CLS]$  is the special token indicating the beginning of the comment, and  $E[EMO]$  is the embedding vector for the tokenized emoji:

$$T_i = \text{BERT}(e_i) \text{ for } i = 1, 2, \dots, N \tag{6}$$

where  $T_i$  is the BERT output vector for the  $i$ -th token in the comment.

Next, the Bi-LSTM model is used to extract contextual features from the sequence of BERT output vectors  $T = \{T_0, T_1, T_2, \dots, T_N\}$ . The Bi-LSTM model relies on two LSTM layers to capture context from both directions in the data: one flowing forward and the other backward.

$$\overrightarrow{h}_i = \overrightarrow{LSTM}(T_i, \overrightarrow{h}_{i-1}) \text{ and } \overleftarrow{h}_1 = \overleftarrow{LSTM}(T_i, \overleftarrow{h}_{i+1}) \tag{7}$$



where  $\vec{h}_t$  and  $\overleftarrow{h}_1$  are the hidden states from the forward and backward LSTM, respectively.

The contextual feature vector  $h_i$  for each token is the concatenation of the forward and backward directions:

$$h_i = \vec{h}_i; \overleftarrow{h}_1 \quad (8)$$

The contextual features from the entire comment are then averaged to obtain the overall representation:

$$H = \frac{1}{N} \sum_{i=1}^N h_i \quad (9)$$

This feature vector  $H$  is then used for classification with the Softmax function:

$$y = \text{softmax}(W \cdot H + b) \quad (10)$$

where  $W$  is the weight matrix,  $b$  is the bias, and  $y$  is the probability vector of sentiment classes.

#### 4. Results and Discussions

This section presents and analyzes the experimental results of sentiment classification using the proposed LABERT-LSTM architecture, which integrates Latinized-Arabic BERT with a Bi-LSTM network. The primary objective of this evaluation is to assess the model's effectiveness in handling informal, mixed-language input—particularly YouTube comments enriched with Latinized Arabic phrases and emojis. The analysis is structured into five subsections: (1) data distribution and feature composition, (2) lexicon and emoji sentiment annotation, (3) embedding and data representation, (4) model performance comparison, and (5) error analysis and interpretation. The findings are discussed in relation to the study's initial objectives, highlighting both strengths and limitations of the model in capturing culturally embedded sentiment expressions.

##### 4.1 Dataset

This study utilizes a dataset comprising 24,593 Indonesian-language comments collected from YouTube videos of five prominent preachers in Indonesia, from the time of the video uploads until March 2024. Data collection was carried out using the Python programming language and scraping techniques. YouTube comment data was obtained after acquiring an API key from the Google Developer Console by activating the YouTube plugin in the Google project created. Table 2 displays the number of comments for each preacher. Each comment has been classified based on its sentiment into positive, negative, or neutral categories, with labels 1 for positive, -1 for negative, and 0 for neutral. The distribution of sentiment polarity is divided into three types of comment data: text with emoji and Latinized Arabic, text with Latinized Arabic without emoji, and text without emoji or Latinized Arabic. Table 3 presents the proportion of data with labels across each comment type.

Table 2 - Dataset count.

Name of Preacher	Number of Comments
KH. Ahmad Bahauddin Nursalim	2.842
Buya Yahya	14.379
Ustadz Abdul Somad	2.631
Ustadz Adi Hidayat	3.342
Ustadz Khalid Basalamah	1.399
Total Dataset	24.593

Table 3 - Proportion of data with labels for each comment type.

Sentiment Polarity	Positive	Negative	Neutral
Text, Emoji, and Latinized Arabic	9822	10685	10685
Text and Latinized Arabic (No Emoji)	9962	10011	10011
Text Only (No Emoji, No Latinized Arabic)	8741	12122	12122

For text containing both emojis and Latinized Arabic, positive sentiment was recorded at 9,822 comments (39.9%), negative sentiment at 10,685 comments (43.4%), and neutral

sentiment at 4,086 comments (16.6%). In the text with Latinized Arabic but without emojis, positive sentiment slightly increased to 9,962 comments (40.5%), negative sentiment decreased to 10,011 comments (40.7%), and neutral sentiment increased to 4,620 comments (18.8%). Meanwhile, for text without emojis or Latinized Arabic, positive sentiment dropped to 8,741 comments (35.5%), negative sentiment increased to 12,122 comments (49.3%), and neutral sentiment decreased to 3,730 comments (15.2%). This data indicates that the presence of emojis and Latinized Arabic in the text can influence sentiment distribution. Text containing these elements tends to exhibit fewer negative sentiments and more positive and neutral sentiments compared to text lacking these features.

The emoji annotation process utilized the Emoji Sentiment Ranking version 1.0, sourced from [https://kt.ijs.si/data/Emoji\\_sentiment\\_ranking/index.html](https://kt.ijs.si/data/Emoji_sentiment_ranking/index.html), which includes a dataset of 752 emojis labeled as 1 for positive, 0 for neutral, and -1 for negative. During this stage, a total of 4,187 emojis were identified: 3,102 positive emojis, 957 negative emojis, and 128 neutral emojis. Table 4 shows the number of emojis with their corresponding labels.

Table 4 - Total number of annotated emojis

Label	Total	Percentage
Positive	3102	74,1%
Negative	957	22,8%
Neutral	128	3,1%
Total	4187	100%

After determining the sentiment values for comments containing Latinized Arabic, a filtering process was carried out based on the annotations shown in Table 2. The lexicon method was used to calculate the polarity value of each word in the comment, with negative (-1) and positive (1) values that were then summed to determine the sentiment of each comment. Positive and negative words were identified by matching words in the sentences with those found in the lexicon dictionary available at <https://github.com/commitunuja/analisis-sentimen-naive-bayes-tf-idf>.

Table 5 provides examples of the calculation results for Latinized Arabic words to determine the sentiment used in a sentence. In Example 1, positive sentiment is detected because words like 'bismillah' and 'barakallahfikum' match the positive lexicon, highlighted in green. In Example 2, negative sentiment is identified due to the presence of words that align with the negative lexicon. The negative sentiment is calculated by summing the total positive words (6) and adding the total negative words (-9), resulting in a score of -3, indicating negative sentiment. In Example 3, neutral sentiment is produced because the total positive words (3) and negative words (3) sum to 0, indicating neutral sentiment.

Table 5 - Examples of sentiment calculation results

True Label	Score	Word importance
Positive	1	Bismillah, barokallahufikum ustadz, tim, serta seluruh kaum muslimin. Amin ya rabbal alamin.
Negative	-1	Bismillah, astaghfirullahaladzim. Sadarlah bahwa pasukan Dajjal telah mempengaruhi hati dan pikiran kita di akhir zaman ini. Jangan biarkan kita, umat Islam, terhasut untuk saling membenci dan merasa paling benar atau paling mengikuti Sunnah. Allah menciptakan kita berbeda-beda, tetapi perbedaan dalam pemahaman tidak membuat kita lebih benar dari orang lain. Hati-hati, Dajjal sedang berusaha memecah belah umat Islam dalam misinya
Neutral	0	Setuju dengan pendapat bahwa membaca Yasin, tahlilan seperti bid'ah, jenggotan, celana cingkrangan.

From this process, a total of 12 Latinized Arabic words were identified, used in 984 comments, with 260 negative sentences, 611 positive sentences, and 113 neutral sentences, as shown in Table 6.

Table 6 - Sentiment analysis results for Latinized Arabic in the comments.

Word	Negative	Positive	Neutral	Total
Bismillah	18	48	1	67
La ilaha illallah	51	41	9	101
Insya Allah	52	75	3	130
Alhamdulillah	2	26	1	29
Jazakallah Khairan	6	33	12	51
Masya Allah	7	148	2	157
Barakallah	11	31	10	52
Allahu Akbar	3	24	0	27
Subhanallah	35	121	52	208
Na'udzubillah	17	9	0	26
Astaghfirullah	39	36	16	91
Inna Lillahi wa Inna Ilaihi Rajiun	19	19	7	45
Total	260	611	113	984

**4.2 Experimental Parameter Settings.**

The BERT model used in this study is IndoBERTBASE, developed by Murfi et al. (Murfi et al., 2024). This model is equipped with 12 encoder layers, 12 attention heads, and a hidden size of 768. After the text data is input in the form of embeddings into the pre-trained BERT model, the stack of encoder layers processes this data. The first encoder layer computes the representation of each token, and its output becomes the input for the next encoder layer. This process continues until the 12th encoder layer, the final layer, which produces a contextually-aware embedding vector representation for each token. The final output is a matrix with dimensions  $128 \times 768$ , where 128 represents the number of tokens in the text and 768 is the hidden size. The stages of the text data representation process with BERT are shown in Figure 3.

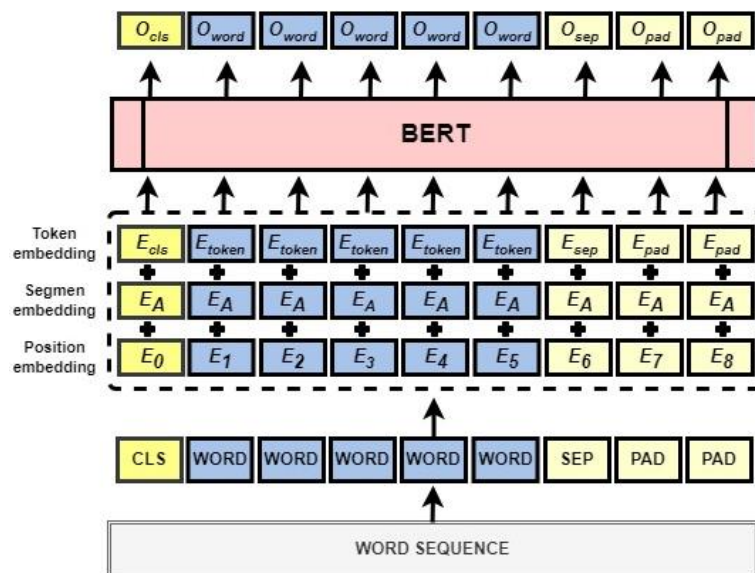


Fig. 3. Stages of Text Data Representation Process with BERT

BERT, as a text representation, performs better compared to traditional embedding methods. This is possible because BERT identifies hidden patterns within the data by reducing the dimensionality of the dataset to a lower dimension. In this study, PCA of BERT embeddings was used to analyze or reduce the dimensions of the embeddings generated by the BERT model. These embeddings are vector representations of words in a text, which are used to gain an understanding of the meaning and relationships between words in the text. With PCA of BERT

embeddings, the dimensionality of the vector representation can be reduced without losing important information, allowing for more efficient analysis or faster computation (Subakti et al., 2022). Figure 4 shows that BERT is able to effectively cluster positive and negative sentiment classes, which is consistent with the superior performance of the model when using BERT as a text representation compared to traditional embedding methods.

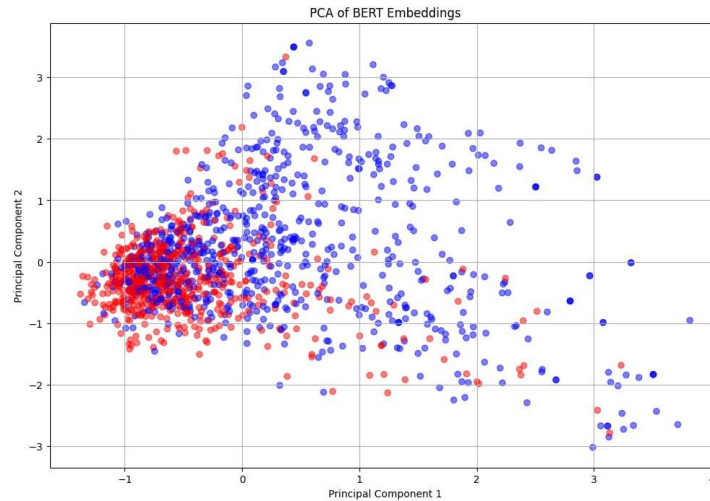


Fig. 4. Visualization of PCA of BERT Embedding from Text Data Representations

#### 4.3 Evaluation Metrics

This article uses precision, recall, and F1-score as evaluation metrics. Precision represents the proportion of correctly predicted positive samples. The recall rate indicates the model's ability to recognize positive samples. If the positive and negative datasets consist of unbalanced text, there may be discrepancies in the calculation of precision and recall. The F1-score combines the two metrics, precision and recall, to better reflect classification performance. The closer the F1-score is to 1, the better the classifier's performance. The specific formulas are detailed as follows:

Precision is defined as

$$Precision = \frac{TP}{TP + FP}$$

Recall is defined as

$$Recall = \frac{TP}{TP + FN}$$

F1-score is defined as:

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} = \frac{2 \times TP}{2 \times TP + FP + FN}$$

#### 4.4 Experimental Analysis

The proposed LABERT-LSTM model in this study combines the pre-training BERT model and the BiLSTM model to enhance sentiment analysis performance on Latinized Arabic text in Indonesian-language comments. Experimental results show that the combined LABERT-BiLSTM model yields better performance in emotion analysis than either model individually. To demonstrate its superiority, the LABERT-LSTM combination model was compared with the performance of BERT and BiLSTM models separately in sentiment classification for different data configurations: text with emoji and Latinized Arabic, text with Latinized Arabic without emoji, and text without emoji and Latinized Arabic, as shown in Table 7. The parameters used in this comparison are consistent with the previous experiments.

Table 7 - Evaluation matrix for the BERT, Bi-LSTM, and LABERT-LSTM models.

Data Configuration	Model	Accuracy	Precision	Recall	F1-Score
Text with Emoji and Latinized Arabic	BERT	0,957	0,948	0,909	0,928
	BiLSTM	0,951	0,944	0,888	0,915
	LABERT-BiLSTM	0,957	0,940	0,918	0,928
Text and Latinized Arabic (No Emoji)	BERT	0,817	0,779	0,768	0,771
	BiLSTM	0,811	0,772	0,781	0,774
	LABERT-BiLSTM	0,821	0,765	0,794	0,777
Text Only (No Emoji, No Latinized Arabic)	BERT	0,765	0,794	0,606	0,687
	BiLSTM	0,767	0,785	0,618	0,691
	BERT-BiLSTM	0,785	0,778	0,629	0,694

In the text data with emoji and Latinized Arabic, both the BERT model and the combined BERT-BiLSTM model achieved the highest accuracy of 0.957, with F1-Scores of 0.928 and 0.928, respectively, while Bi-LSTM performed slightly lower with an accuracy of 0.951 and an F1-Score of 0.915. For text data with Latinized Arabic without emoji, overall performance decreased, with the BERT model achieving an accuracy of 0.817 and an F1-Score of 0.771. Meanwhile, Bi-LSTM and the BERT-BiLSTM combination had slightly lower accuracy, 0.811 and 0.821, with F1-Scores of 0.774 and 0.777, respectively. In text data without emoji and Latinized Arabic, model performance further declined, with the highest accuracy achieved by the BERT-BiLSTM combination at 0.785 and an F1-Score of 0.694, while BERT and Bi-LSTM models had accuracies of 0.765 and 0.767, and F1-Scores of 0.687 and 0.691, respectively. Figure 5 illustrates that the LABERT-BiLSTM combination model achieved the best performance across various data configurations, with higher accuracy and F1-Scores compared to BERT-LSTM and the individual BERT and Bi-LSTM models, particularly on more complex data.

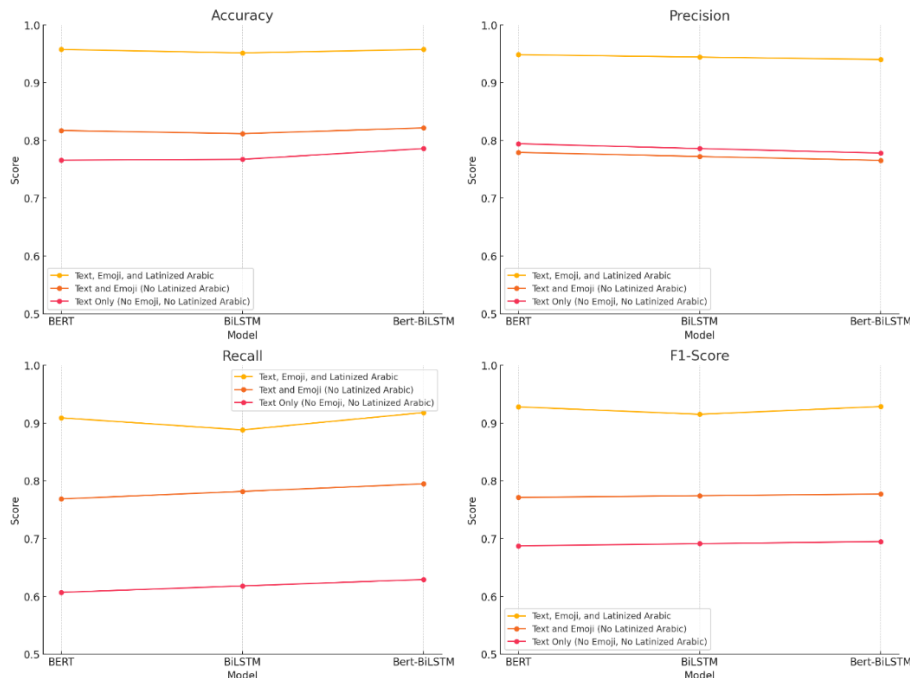


Fig. 5. Model Performance Comparison

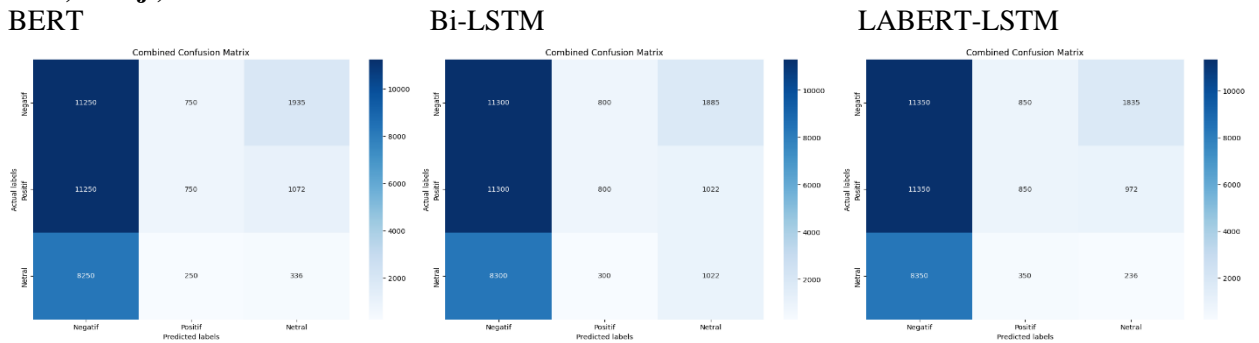
Based on the comparison of the Confusion Matrix in Figure 6, the LABERT-LSTM model, which considers text, Latinized Arabic, and emoji, consistently provides the best performance in classifying text sentiment into negative, positive, and neutral classes. In the general classification, the LABERT-LSTM model achieved 11,350 true negative predictions, 850 true positive, and 236 true neutral, outperforming the BERT model with 11,250 true

negative, 750 true positive, and 336 true neutral predictions, as well as the Bi-LSTM model with 11,300 true negative, 800 true positive, and 1,022 true neutral predictions.

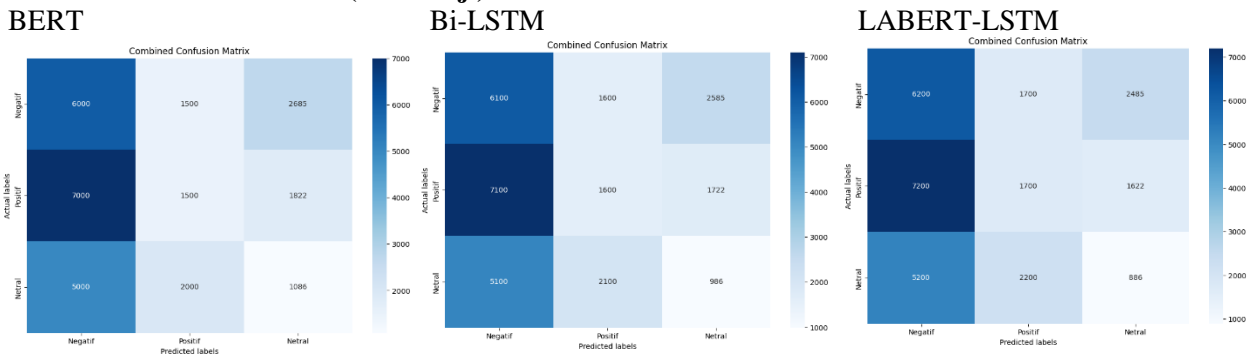
For the analysis of comments that consider text and Latinized Arabic without emoji, the LABERT-LSTM model again demonstrated superior performance with 6,200 true negative predictions, 1,700 true positive, and 5,200 true neutral, compared to BERT with 6,000 true negative, 1,500 true positive, and 5,000 true neutral predictions, and Bi-LSTM with 6,100 true negative, 1,600 true positive, and 5,100 true neutral predictions.

Finally, in the analysis of comments considering text only, without emoji and without Latinized Arabic, the LABERT-LSTM model achieved 6,700 true negative predictions, 1,700 true positive, and 4,700 true neutral, surpassing BERT with 6,500 true negative, 2,000 true positive, and 4,500 true neutral predictions, as well as Bi-LSTM with 6,600 true negative, 1,600 true positive, and 4,600 true neutral predictions. Overall, the LABERT-BiLSTM model, which incorporates text, Latinized Arabic, and emoji, proved to be more effective in reducing misclassification and improving true predictions in each class, although there is still room for improvement, particularly in the neutral class.

**Text, Emoji, and Latinized Arabic**



**Text and Latinized Arabic (No Emoji)**



**Text Only (No Latinized Arabic, No Emoji)**

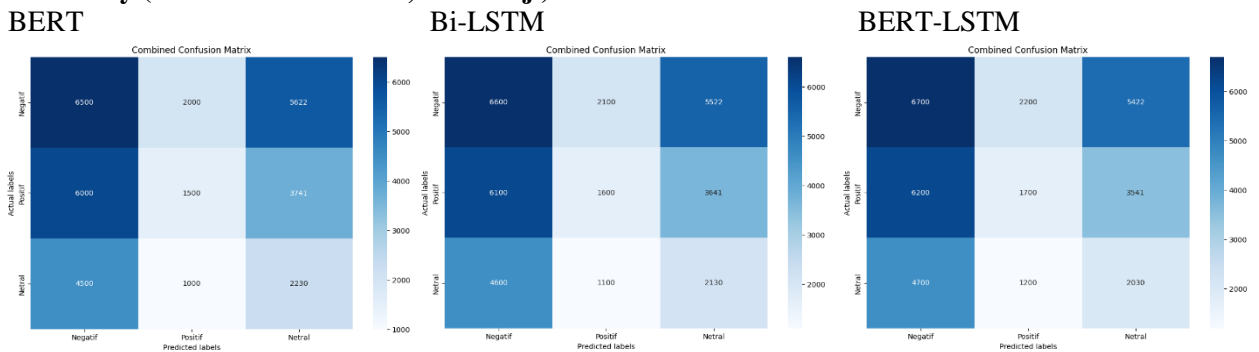
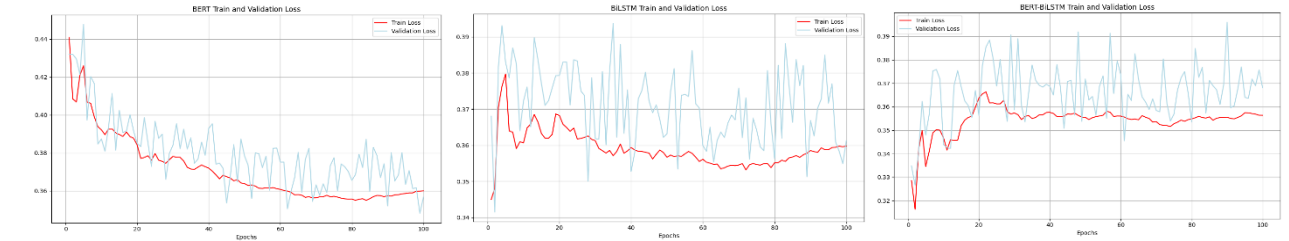


Fig. 6. Confusion Matrix

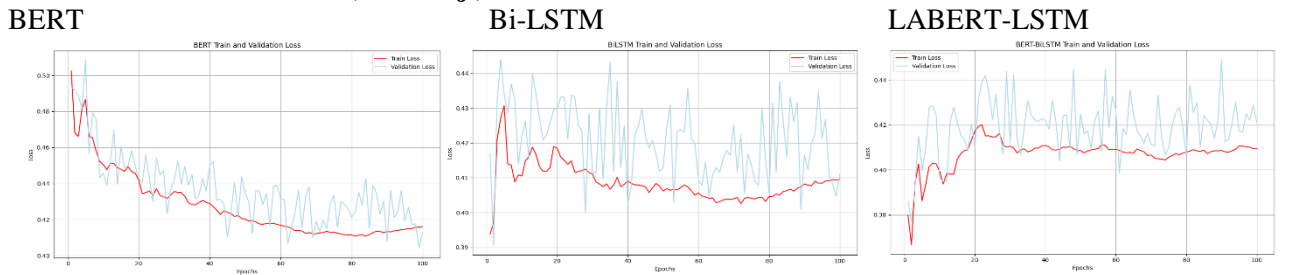
The comparison of the three sentiment analysis models through the Loss Curve on training and validation data, as shown in Figure 7, reveals that the BERT model consistently delivers more stable and consistent performance compared to the Bi-LSTM and LABERT-BiLSTM models for comment data that includes text, Latinized Arabic, and emoji. In sentiment classification, the BERT model shows a steady decline in training loss from around 0.44 to 0.36 after 100 epochs, with fluctuating but decreasing validation loss from 0.38 to 0.36. The Bi-LSTM model experiences a rapid drop in training loss from around 0.39 to 0.35 within the first 20 epochs, but its validation loss fluctuates significantly between 0.37 and 0.35 without a clear downward trend. The LABERT-BiLSTM model also shows a rapid decline in training loss from around 0.39 to 0.34, but its validation loss remains volatile between 0.38 and 0.36.

For comment data that includes text and Latinized Arabic but without emoji, the BERT model shows a decline in training loss from 0.50 to 0.42, with a fluctuating validation loss that drops from 0.50 to 0.42. The Bi-LSTM model decreases from 0.44 to 0.41 within the first 20 epochs and stabilizes at 0.40, though its validation loss remains volatile. The LABERT-BiLSTM model shows a rapid decline in training loss from 0.42 to 0.39 within the first 20 epochs and stabilizes at 0.40, but its validation loss remains fluctuating.

**Text, Emoji, and Latinized Arabic**



**Text and Latinized Arabic (No Emoji)**



**Text (No Latinized Arabic, No Emoji)**

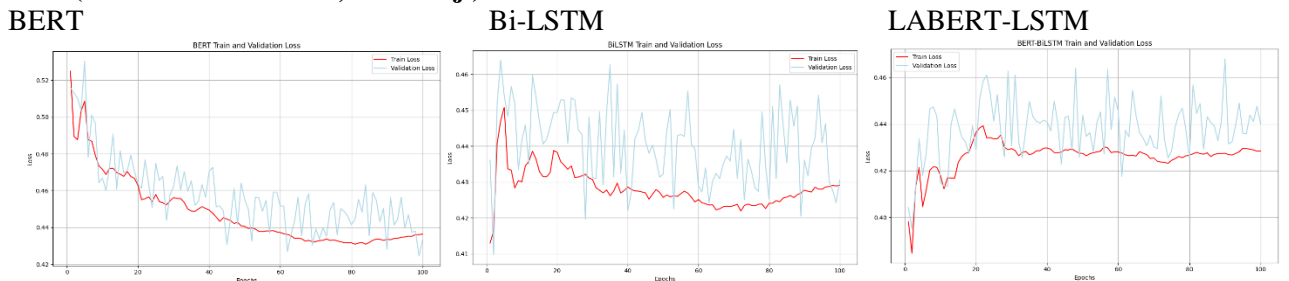


Fig. 7. Train and Validation Loss Curve

For comment data considering text only, without Latinized Arabic and emoji, the BERT model shows a decline in training loss from 0.52 to 0.44, with a fluctuating validation loss that drops from 0.52 to 0.44. The Bi-LSTM model decreases from 0.46 to 0.43 within the first 20 epochs and stabilizes at 0.42, but the validation loss remains volatile. The BERT-BiLSTM model shows a rapid drop in training loss from 0.44 to 0.40 within the first 20 epochs and stabilizes at 0.42, but the validation loss remains fluctuating throughout the training. Overall,

the BERT model demonstrates more stable and consistent performance in reducing training and validation loss compared to Bi-LSTM and the BERT-BiLSTM combination. Although the combination model shows a rapid initial drop in loss, it exhibits significant fluctuations in validation loss.

## 5. Conclusion

This study presents significant advancements in sentiment analysis of Latinized Arabic in Indonesian YouTube comments, introducing the LABERT-LSTM model, which outperforms traditional approaches. The results demonstrate that incorporating Latinized Arabic text, emojis, and slang into the analysis substantially enhances the system's ability to detect emotions in informal language and symbolic expressions. The LABERT-LSTM model achieved superior performance across all configurations, with accuracy improvements of up to 0.95756 and an F1-Score of 0.92868 on complex datasets containing emojis and Latinized Arabic. Notably, the model consistently outperformed BERT and Bi-LSTM in reducing classification errors and improving true predictions in negative, positive, and neutral classes.

Moreover, LABERT-LSTM proves highly applicable for deployment in social media and digital platform technologies in Indonesia, providing a robust framework for understanding user sentiment in localized, informal communication. The comparative evaluation of confusion matrices highlights LABERT-LSTM's dominance in accurately classifying sentiments, particularly on datasets containing rich symbolic and linguistic diversity.

While LABERT-LSTM demonstrated rapid loss reduction in training phases, validation loss fluctuations indicate potential areas for further optimization. Additionally, despite its superior performance in complex configurations, neutral sentiment classification remains a challenge requiring future enhancements. Nonetheless, the LABERT-LSTM model serves as a critical foundation for advancing sentiment analysis in multilingual and culturally nuanced contexts, offering a practical tool for real-world applications in social media analytics and beyond.

## Acknowledgement - optional

The authors sincerely express their gratitude to the Department of Electrical Engineering and Informatics, Universitas Negeri Malang, Indonesia, for the support and facilities provided during the research. This invaluable assistance greatly contributed to the successful completion of this study.

## References

- Abid, F., Li, C., & Alam, M. (2020). Multi-source social media data sentiment analysis using bidirectional recurrent convolutional neural networks. *Computer Communications*, 157, 102–115. <https://doi.org/10.1016/j.comcom.2020.04.002>
- Ahanin, Z., & Ismail, M. A. (2022). A multi-label emoji classification method using balanced pointwise mutual information-based feature selection. *Computer Speech and Language*, 73, 101330. <https://doi.org/10.1016/j.csl.2021.101330>
- Ainin, S., Feizollah, A., Anuar, N. B., & Abdullah, N. A. (2020). Sentiment analyses of multilingual tweets on halal tourism. *Tourism Management Perspectives*, 34, 100658. <https://doi.org/10.1016/j.tmp.2020.100658>
- Al-Azani, S., & El-Alfy, E. S. M. (2021). Early and Late Fusion of Emojis and Text to Enhance Opinion Mining. *IEEE Access*, 9, 121031–121045. <https://doi.org/10.1109/ACCESS.2021.3108502>
- Ahda, F. A. I., Wibawa, A. P., Prasetya, D. D., & Sulistyono, D. A. (2024). Comparison of Adam Optimization and RMS prop in Minangkabau-Indonesian Bidirectional Translation with Neural Machine Translation. *JOIV: International Journal on Informatics Visualization*, 8(1), 231-238. <http://dx.doi.org/10.62527/joiv.8.1.1818>
- Aribowo, A. S., Basiron, H., Yusof, N. F. A., & Khomsah, S. (2021). Cross-domain sentiment analysis model on Indonesian YouTube comment. *International Journal of Advances in Intelligent Informatics*, 7(1), 12–25. <https://doi.org/10.26555/ijain.v7i1.554>



- Chakma, K., Ruba, U. B., & Riya, S. D. (2022). YouTube as an information source of floating agriculture: analysis of Bengali language contents quality and viewers' interaction. *Heliyon*, 8(9). <https://doi.org/10.1016/j.heliyon.2022.e10719>
- Chrismanto, A. R., Sari, A. K., & Suyanto, Y. (2023). Enhancing Spam Comment Detection on Social Media With Emoji Feature and Post-Comment Pairs Approach Using Ensemble Methods of Machine Learning. *IEEE Access*, 11, 80246–80265. <https://doi.org/10.1109/ACCESS.2023.3299853>
- de Andrade, C. M., Belem, F. M., Cunha, W., França, C., Viegas, F., Rocha, L., & Gonçalves, M. A. (2023). On the class separability of contextual embeddings representations—or “the classifier does not matter when the (text) representation is so good!”. *Information Processing & Management*, 60(4), 103336. <https://doi.org/10.1016/j.ipm.2023.103336>
- Elfajr, N. M., & Sarno, R. (2018, September). Sentiment analysis using weighted emoticons and SentiWordNet for Indonesian language. In *2018 International Seminar on Application for Technology of Information and Communication* (pp. 234-238). IEEE.
- Habbat, N., Nouri, H., Anoun, H., & Hassouni, L. (2023). Sentiment analysis of imbalanced datasets using BERT and ensemble stacking for deep learning. *Engineering Applications of Artificial Intelligence*, 126, 106999. <https://doi.org/10.1016/j.engappai.2023.106999>
- Haque, R., Islam, N., Tasneem, M., & Das, A. K. (2023). Multi-class sentiment classification on Bengali social media comments using machine learning. *International Journal of Cognitive Computing in Engineering*, 4, 21–35. <https://doi.org/10.1016/j.ijcce.2023.01.001>
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Jia, K., Meng, F., Liang, J., & Gong, P. (2023). Text sentiment analysis based on BERT-CBLBGA. *Computers and Electrical Engineering*, 112, 109019. <https://doi.org/10.1016/j.compeleceng.2023.109019>
- Karmagatri, M., Utama, I. D., Yogie, H., Lemena, J., & Chen, H. H. (2023). Naive Bayes Sentiment Analysis on Perceptions of Halal Certification: A Case Study on Mixue Indonesia. *Journal of Theoretical and Applied Information Technology*, 101(17).
- Karo, I. M. K., Fudzee, M. F. M., Kasim, S., & Ramli, A. A. (2022). Karonese sentiment analysis: a new dataset and preliminary result. *JOIV: International Journal on Informatics Visualization*, 6(2-2), 523-530. <http://dx.doi.org/10.30630/joiv.6.2-2.1119>
- Kusumaningrum, R., Nisa, I. Z., Jayanto, R., Nawangsari, R. P., & Wibowo, A. (2023). Deep learning-based application for multilevel sentiment analysis of Indonesian hotel reviews. *Heliyon*, 9(6). <https://doi.org/10.1016/j.heliyon.2023.e17147>
- Li, D., Rzepka, R., Ptaszynski, M., & Araki, K. (2020). HEMOS: A novel deep learning-based fine-grained humor detecting method for sentiment analysis of social media. *Information Processing and Management*, 57(6). <https://doi.org/10.1016/j.ipm.2020.102290>
- Li, Z., & Zou, Z. (2024). Punctuation and lexicon aid representation: A hybrid model for short text sentiment analysis on social media platform. *Journal of King Saud University - Computer and Information Sciences*, 36(3). <https://doi.org/10.1016/j.jksuci.2024.102010>
- Monesa, F., & Jayadi, R. (2023). Sentiment Analysis On Investment Topic In Indonesia Using Machine Learning Algorithms Approach. *Journal of Theoretical and Applied Information Technology*, 101(13).
- Murfi, H., Theresia Gowandi, S., Ardaneswari, G., & Nurrohmah, S. (2024). BERT-based combination of convolutional and recurrent neural network for Indonesian sentiment analysis. *Applied Soft Computing*, 151. <https://doi.org/10.1016/j.asoc.2023.111112>
- Naseem, U., Razzak, I., Musial, K., & Imran, M. (2020). Transformer based Deep Intelligent Contextual Embedding for Twitter sentiment analysis. *Future Generation Computer Systems*, 113, 58–69. <https://doi.org/10.1016/j.future.2020.06.050>
- Peng, D., & Zhao, H. (2021). Seq2Emoji: A hybrid sequence generation model for short text emoji prediction. *Knowledge-Based Systems*, 214. <https://doi.org/10.1016/j.knosys.2020.106727>
- Peng, S., Cao, L., Zhou, Y., Ouyang, Z., Yang, A., Li, X., ... & Yu, S. (2022). A survey on deep learning for textual emotion analysis in social networks. *Digital Communications and Networks*, 8(5), 745-762. <https://doi.org/10.1016/j.dcan.2021.10.003>

- Rahmanti, A. R., Chien, C. H., Nursetyo, A. A., Husnayain, A., Wiratama, B. S., Fuad, A., Yang, H. C., & Li, Y. C. J. (2022). Social media sentiment analysis to monitor the performance of vaccination coverage during the early phase of the national COVID-19 vaccine rollout. *Computer Methods and Programs in Biomedicine*, 221. <https://doi.org/10.1016/j.cmpb.2022.106838>
- Rahmanti, A. R., Ningrum, D. N. A., Lazuardi, L., Yang, H. C., & Li, Y. C. (2021). Social Media Data Analytics for Outbreak Risk Communication: Public Attention on the “New Normal” During the COVID-19 Pandemic in Indonesia. *Computer Methods and Programs in Biomedicine*, 205. <https://doi.org/10.1016/j.cmpb.2021.106083>
- Ren, Z., Shen, Q., Diao, X., & Xu, H. (2021). A sentiment-aware deep learning approach for personality detection from text. *Information Processing and Management*, 58(3), 102532. <https://doi.org/10.1016/j.ipm.2021.102532>
- Saikia, S., Kumar, V., & Verma, M. K. (2023). Analyzing user sentiments toward selected content management software: a sentiment analysis of viewer’s comments on YouTube. *Information Discovery and Delivery*, 52(4), 378-387. <https://doi.org/10.1108/IDD-01-2023-0009>
- Samreen, A., & Ali, S. A. (2024). Dataset construction to detect human behavior with the help of emotions, sentiments and mood for Roman Urdu. *Data in Brief*, 52, 109906. <https://doi.org/10.17632/d5j9fgbdcn.1>
- Subakti, A., Murfi, H., & Hariadi, N. (2022). The performance of BERT as data representation of text clustering. *Journal of big Data*, 9(1), 15. <https://doi.org/10.1186/s40537-022-00564-9>
- Suhaimin, M. S. M., Hijazi, M. H. A., Mounq, E. G., Nohuddin, P. N. E., Chua, S., & Coenen, F. (2023). Social media sentiment analysis and opinion mining in public security: Taxonomy, trend analysis, issues and future directions. *Journal of King Saud University-Computer and Information Sciences*, 35(9), 101776. <https://doi.org/10.1016/j.jksuci.2023.101776>
- Wang, Z., Gao, P., & Chu, X. (2022). Sentiment analysis from Customer-generated online videos on product review using topic modeling and Multi-attention BLSTM. *Advanced Engineering Informatics*, 52, 101588. <https://doi.org/10.1016/j.aei.2022.101588>
- Yao, Q., Li, R. Y. M., & Song, L. (2022). Construction safety knowledge sharing on YouTube from 2007 to 2021: Two-step flow theory and semantic analysis. *Safety science*, 153, 105796. <https://doi.org/10.1016/j.ssci.2022.105796>
- Yulita, I. N., Wijaya, V., Rosadi, R., Sarathan, I., Djuyandi, Y., & Prabuwo, A. S. (2023). Analysis of government policy sentiment regarding vacation during the COVID-19 pandemic using the bidirectional encoder representation from transformers (BERT). *Data*, 8(3), 46. <https://doi.org/10.3390/data8030046>
- Yunitasari, Y., Musdholifah, A., & Sari, A. K. (2019). Sarcasm Detection For Sentiment Analysis in Indonesian Tweets. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 13(1), 53-62. <https://doi.org/10.22146/ijccs.41136>
- Zhang, T., Yang, K., Alhuzali, H., Liu, B., & Ananiadou, S. (2023). PHQ-aware depressive symptoms identification with similarity contrastive learning on social media. *Information Processing & Management*, 60(5), 103417. <https://doi.org/10.1016/j.ipm.2023.103417>
- Zhao, G., Liu, Z., Chao, Y., & Qian, X. (2021). CAPER: Context-Aware Personalized Emoji Recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 33(9), 3160–3172. <https://doi.org/10.1109/TKDE.2020.2966971>
- Zou, H., & Xiang, K. (2022). Sentiment Classification Method Based on Blending of Emoticons and Short Texts. *Entropy*, 24(3), 19854–19863. <https://doi.org/10.3390/e24030398>