Journal of Soft Computing Exploration

Homepage: https://shmpublisher.com/index.php/joscex

p-ISSN: 2746-7686
e-ISSN: 2746-0991

.

The application of IndoBERT in tourist sentiment analysis: a comparative evaluation with SVM and LSTM

Yoannes Romando Sipayung[1], Sri Mujiyono[2], Anni Malihatul Hawa[3]
[1,2,3] Departement of Informatics Engineering, Faculty of Computer Science and Education, Universitas Ngudi Waluyo, Indonesia

Article Info

Article history:

Received April 4, 2026
Revised April 25, 2026
Accepted April 28, 2026

Keywords:

Sentiment analysis
IndoBERT
SVM
LSTM
Tourism
YouTube

ABSTRACT

YouTube comments provide valuable public opinions about tourist destinations, but their informal and unstructured nature makes sentiment analysis challenging. Therefore, an automatic sentiment classification approach is needed to support tourism evaluation and promotion strategies. This study aims to analyze tourist sentiment toward tourism in the Bangka Belitung Islands based on comments on the YouTube platform. The analysis was conducted using a comparative approach with three models: IndoBERT, SVM, and LSTM. The dataset consisted of 1,000 YouTube comments, which were reduced to 913 valid comments after preprocessing, including data cleaning, case folding, normalization, tokenization, and stopword removal. The sentiment distribution consisted of 434 neutral comments, 333 positive comments, and 146 negative comments, indicating an imbalanced class distribution. Model performance was evaluated using accuracy, precision, recall, and F1-score metrics based on a confusion matrix. The results show that IndoBERT performed best with an accuracy of 0.71 and the highest F1-score compared to the other models. The SVM model demonstrated fairly stable performance with an accuracy of 0.69, while the LSTM achieved an accuracy of 0.68 with lower performance on the minority class. The results indicate that transformer-based models are more effective in understanding linguistic context than machine learning and deep learning models. This study is expected to contribute to the development of sentiment analysis based on social media data in the tourism sector.

Corresponding Author:

Yoannes Romando Sipayung,
Informatics Engineering, Faculty of Computer Science and Education, Universitas Ngudi Waluyo,
186 Diponegoro Street, Gedanganak, East Ungaran, Semarang Regency, Central Java, Indonesia.
Email: yoannesromando@unw.ac.id
https://doi.org/10.52465/joscex.v7i2.42

1. INTRODUCTION

The growth rate of the tourism industry significantly outpaces other economic sectors, indicating that this sector plays a vital role in driving global economic growth at an impressive pace [1] . More than seven decades after Indonesia's independence, the country's tourism industry has begun to experience significant growth and development. The Indonesian government's strategic efforts regarding tourism sector development began approximately forty years ago, marking a long-term commitment to advancing the nation’s tourism potential [2] .

According to data from the Ministry of Tourism and Creative Economy, the number of international tourist arrivals in Indonesia through all entry points in November 2024 totaled 1,092,067, consisting of 940,570 visits, or 86.13%, recorded through immigration, and 151,497 visits, or 13.87%, recorded via Mobile Positioning Data at border entry points (Ministry of Tourism and Creative Economy, 2024). Meanwhile, according to data from the Central Statistics Agency (BPS), the number of domestic tourist visits in December 2024 reached 810,438, indicating that domestic tourism continues to show growth and high enthusiasm among the public for exploring domestic tourist destinations (Central Statistics Agency, 2024). Under UNCLOS 1982, Indonesia is designated as the world’s largest archipelagic nation, surpassing other archipelagic nations such as the Philippines, Japan, and New Zealand [3] . In 2022, the Central Statistics Agency (BPS) recorded that Indonesia has 17,001 islands spread across 38 provinces [4] .

One of the provinces in Indonesia is the Bangka Belitung Islands Province. Previously, this region was part of South Sumatra Province. Then, on November 21, 2000, Bangka Belitung was declared a separate province. Thus, the Bangka Belitung Islands officially became the 31st province in Indonesia [5] . The Bangka Belitung Islands consist of two main islands, namely Bangka and Belitung, as well as hundreds of smaller islands. The total number of named islands reaches 470, but only 50 of them are inhabited. This region is divided into six regencies, including Bangka, West Bangka, Central Bangka, South Bangka, Belitung, and East Belitung Regencies, as well as the city of Pangkalpinang, which serves as the provincial capital [6] .

The main economic pillars of the Bangka Belitung Islands Province include the mining, agriculture, plantation, fisheries, livestock, and service sectors. Within the service sector, the primary focus is on tourism, particularly the development of beach tourism, nature tourism, and the province’s rich cultural and traditional tourism resources [7]. The tourism potential in the Bangka Belitung Islands Province is immense. The development of the tourism sector in this region is strategic because it can have a significant positive impact on the local economy, particularly for the local community, through the optimal management of tourism assets. According to data from the Central Statistics Agency (BPS) for 2024, the number of domestic tourists visiting the Bangka Belitung Islands was recorded at 2,500,106 people [8].

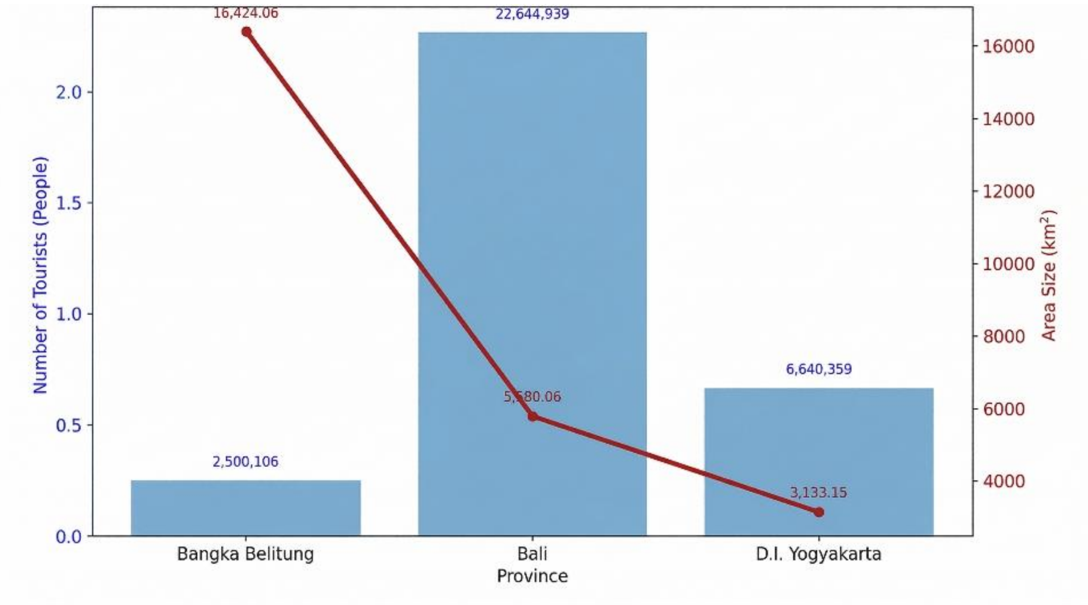

Figure 1. Area and number of domestic tourists in babel, bali, and diy in 2024

Data obtained from the Central Statistics Agency in 2024 shows that Bangka Belitung Province has the largest land area among the three provinces analyzed, at 16,424.06 km². However, despite its significant land area, the number of tourists visiting Bangka Belitung was only 2,500,106. This figure is far lower than that of Bali and Yogyakarta Special Region, which recorded 22,644,939 [9] and 6,640,359 visitors, respectively [10]. This indicates that Bangka Belitung’s tourism potential has not been fully utilized, and promotional strategies as well as destination development need to be enhanced to attract more visitors.

Tourism growth is closely linked to advancements in information technology. Tourists have the opportunity to share their experiences through popular apps and social media platforms to view and share their travels [11] . One social media platform that plays a significant role in the tourism industry is YouTube. As a video-sharing platform, YouTube serves as the primary visual information source for prospective tourists to explore destinations through travel review content (vlogs) and official regional promotions [12] [13] .

The presence of comment sections on the YouTube platform is highly beneficial because they serve as a space for interaction where travelers provide feedback, real-life experiences, and additional information about accessibility, transportation options, and the quality of tourist attractions in real time. These comments often describe personal experiences in greater depth than mere numerical ratings. These opinions from travelers can serve as a valuable source of evaluation data, which can later be applied in the development of programs or strategic promotional action plans, whether related to physical development or the improvement of non-physical services in the tourism sector.

The BERT *transformer* method was chosen to address this issue. BERT was selected due to its ability to understand word context bidirectionally, which allows the model to capture the meaning of a word based on the words before and after it simultaneously [14] [15] . Additionally, the use of the IndoBERT variant offers specific advantages because it has been pre-trained using a very large Indonesian language corpus, making it more robust in recognizing informal sentence structures, slang usage, and linguistic ambiguities that frequently appear in social media comment sections such as YouTube [16] . The *Self-Attention Mechanism* in this architecture also allows the model to assign greater weight to the most relevant words in determining sentiment, resulting in more accurate classification compared to conventional methods such as *Naïve Bayes* or SVM [17] [18] .

This study on sentiment analysis builds upon several previous studies that have addressed similar themes. A number of earlier studies have sought to map tourists' perceptions across various destinations. A previous study [19] conducted a sentiment analysis of visitor reviews of the Dieng tourist area by applying the *K-Nearest Neighbour* (K-NN) algorithm, which achieved an accuracy of 86%. In addition, [20] utilised the social media platform Twitter to analyse tourism sentiment in Lombok using the *Naïve Bayes Classifier* method, which achieved an accuracy of 93.5% with a focus on safety and accommodation factors The use of other conventional methods such as the *Support Vector Machine* (SVM) was also employed [21] to analyse public opinion regarding tourism in Madura, achieving an accuracy rate of 92.59%.

Although the traditional *machine learning* methods mentioned above yield good results, developments in *Natural Language Processing* (NLP) technology are now shifting towards the use of deep learning to capture more complex linguistic contexts. Research previously conducted [20] [22] explored the use of the Bidirectional Encoder Representations from Transformers (BERT) model, specifically IndoBERT, for sentiment analysis of reviews of tourist destinations in Bali, achieving a very high accuracy of 95%.

However, there is *a research gap* in which the application of state-of-the-art models such as BERT remains limited to popular tourist destinations like Bali and has not yet been extensively explored for regions with emerging tourism potential, such as the Bangka Belitung Islands. Additionally, reviews on the YouTube platform exhibit unique linguistic characteristics that require a more sophisticated modeling approach.

Research on sentiment analysis in the tourism sector is now shifting toward implicit sentiment detection capable of capturing hidden emotions such as sarcasm and ambiguity in tourist reviews. A systematic literature review indicates that while multimodal *deep learning* models have advanced rapidly, significant research gaps remain in studies specific to the Indonesian region and the use of datasets more relevant to local tourism conditions[23] .

Therefore, this study aims to fill this gap by implementing a BERT-based *transformer* method to analyze tourist sentiment in the Bangka Belitung Islands, as well as comparing its performance with several *baseline* models, namely *Support Vector Machine* (SVM) and *Long Short-Term Memory* (LSTM).

The dataset used in this study was derived from user comments on the YouTube platform discussing tourism in Bangka Belitung. Specifically, this study uses a variant of IndoBERT, a *pre-trained* model trained on a large Indonesian language corpus, giving it superior capabilities in understanding sentence structure and local dialects. The main differences between this study and previous research lie in the specific research focus, the volume of data used, and the optimization of the IndoBERT model to strengthen tourism promotion strategies in the Bangka Belitung Islands region.

2. METHOD

This study was conducted through several systematic stages, ranging from data collection to model evaluation. The overall research workflow is shown in Figure 2. The research stages were carried out systematically, beginning with the data collection process and ending with the analysis of the final results, as shown in Figure 1. The research workflow began with the extraction of 1,000 comments from YouTube using the YouTube Data API v3, which were then processed through text *preprocessing* stages including *case folding*, *cleaning*, normalization, tokenization, and *stopword removal*. The core stage of the research involved implementing the IndoBERT model through a *fine-tuning* process in the Google Colab environment to obtain an optimal classification model. The model's performance was then evaluated using the *Confusion Matrix* metric. This entire process concluded with an analysis of the results and sentiment interpretation aimed at providing strategic recommendations for tourism promotion in the Bangka Belitung Islands. As a comparison, *baseline* models consisting of *Support Vector Machine* (SVM) and *Long Short-Term Memory* (LSTM) were used, enabling an analysis of each model's performance.

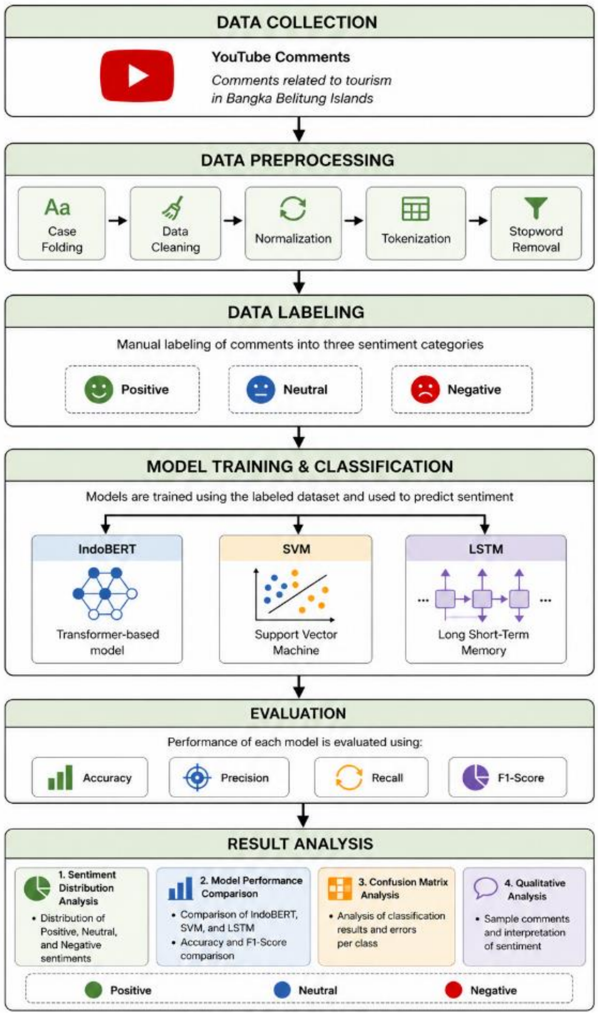

Figure 2. Research flowchart

Data Collection

Data collection for this study was conducted digitally via the YouTube social media platform. YouTube was selected as the data source based on its popularity as a primary visual medium for promoting tourist destinations. Review data or public opinions were extracted from the comment sections of three videos with high relevance and a significant number of viewers related to tourism in the Bangka Belitung Islands. Data collection was performed using the YouTube Data API v3 to ensure the accuracy and structure of the obtained data. Details regarding the video sources and the number of extracted comments are presented in Table 1.

Table 1. Dataset preparation steps

Dataset Stages	Data Size	Description
Dataset Stage	(1000, 4)	Initial dataset containing author, comment, likes, and published_at
Raw YouTube Comment Data	(1000, 1)	Comment text extracted from the initial dataset
Extracted Text Data	(913, 3)	Dataset that has been cleaned and processed for analysis

Table 1 displays the dataset preparation and processing stages used in this study. In the initial stage, 1,000 comments were collected from the YouTube platform related to the research topic. The raw data contained several attributes, including author, comment, likes, and other supporting information. The dataset at this stage was still raw and therefore required further processing before it could be used in the analysis.

Data collection in this study was conducted digitally through the YouTube platform. YouTube was selected as the data source because it provides user-generated comments that reflect public opinions and experiences related to tourist destinations. A total of 1,000 comments related to tourism in the Bangka Belitung Islands were collected using the YouTube Data API v3. The raw dataset contained several attributes, including author, comment, likes, and published date. These data were then used as the initial dataset before entering the preprocessing stage.

Data Preprocessing

Before the modeling stage, the raw comment data were processed through several preprocessing steps to improve data quality and produce a more structured text representation. First, the comment text was extracted from the raw dataset to form a text-based dataset for sentiment analysis. The preprocessing process included data cleaning, case folding, normalization, tokenization, and stopword removal. The data cleaning process was conducted based on several criteria, including removing duplicate comments, comments containing only emojis or symbols, comments with fewer than three words, comments unrelated to the tourism context, comments containing URLs or user mentions, and comments with excessive non-alphanumeric characters. This process was carried out to reduce noise and ensure that the data used in the classification stage contained meaningful textual information.

To provide a clearer illustration of the preprocessing process, examples of text transformation across preprocessing stages are presented in Table X. The examples show how raw comments were gradually transformed through cleaning, case folding, normalization, tokenization, and stopword removal. This helps demonstrate how informal words, URLs, mentions, punctuation, and irrelevant elements were handled before the data were used in the modeling stage.

Table 2. Examples of Text Transformation Across Preprocessing Stages

No.	Original Comment	Cleaning	Case Folding	Normalization	Tokenization	Final Result
1	Kekacauan ini grgr 2 org serakah, udah tua bang...	Kekacauan ini grgr org serakah udah tua bangk...	kekacauan ini grgr org serakah udah tua bangk...	kekacauan ini orang serakah sudah tua bangka...	['kekacauan', 'ini', 'orang', 'serakah', 'sudah', 'tua', 'bangka']	kekacauan orang serakah tua bangka
2	@peeedih Indonesia itu indah, rugi kalo mager...	Indonesia itu indah rugi kalo mager liburan s...	indonesia itu indah rugi kalo mager liburan s...	indonesia itu indah rugi kalau mager liburan s...	['indonesia', 'itu', 'indah', 'rugi', 'kalau', 'mager', 'liburan']	indonesia indah rugi mager liburan
3	@YakultIndonesia @toshibatv_id @AmandaBrownie...	Mending kalian liburan aja yg proper bess...	mending kalian liburan aja yg proper bess...	mending kalian liburan saja yang proper bess...	['mending', 'kalian', 'liburan', 'saja', 'yang', 'proper', 'bess']	mending liburan proper bess
4	Di Bandung pun sama. Pernah tetangga dari Bang...	Di Bandung pun sama Pernah tetangga dari Bangk...	di bandung pun sama pernah tetangga dari bangk...	di bandung pun sama pernah tetangga dari bangk...	['di', 'bandung', 'pun', 'sama', 'pernah', 'tetangga', 'bangka']	bandung tetangga bangka
5	Kangen liburan ke bangka maksudnya	Kangen liburan ke bangka maksudnya	kangen liburan ke bangka maksudnya	kangen liburan ke bangka maksudnya	['kangen', 'liburan', 'ke', 'bangka', 'maksudnya']	kangen liburan bangka maksudnya

After preprocessing, the dataset was reduced from 1,000 raw comments to 913 valid comments. This reduction indicates that some data were not suitable for analysis because they did not provide sufficient semantic information for sentiment classification. The final dataset consisted of cleaned and structured comment text, which was then used for data labeling and sentiment classification using IndoBERT, SVM, and LSTM.

IndoBERT Model

This study uses IndoBERT, a pre-trained language model based on the BERT architecture that has been trained on a large Indonesian language corpus. IndoBERT was selected because it is capable of capturing bidirectional linguistic context, allowing the model to understand the meaning of a word based on both its preceding and following words [24] [25] .

In this study, each preprocessed comment was tokenized and converted into input tokens using the IndoBERT tokenizer. The input sequence was added with special tokens, namely [CLS] at the beginning and

[SEP] at the end of the sentence. The representation of the [CLS] token produced by IndoBERT was then used as the aggregate representation of the comment and passed into a classification layer to predict the sentiment class. The sentiment classification process can be formulated as follows:

$$h\_CLS = \mathrm{IndoBERT}(x) \quad (1)$$

$$y = \mathrm{softmax}(W\, h\_CLS + b) \quad (2)$$

where x represents the input comment, h_CLS denotes the contextual representation of the [CLS] token, W and b are trainable parameters in the classification layer, and y represents the probability distribution over the sentiment classes: positive, neutral, and negative. The class with the highest probability is selected as the final prediction. This mechanism enables IndoBERT to capture deeper contextual meaning from Indonesian comments, making it more suitable for sentiment classification compared to traditional models that rely mainly on word frequency or sequential patterns.

Comparison Model

To evaluate the effectiveness of transformer-based models in sentiment detection, this study employs several *baseline* models representing different approaches in natural language processing, namely *machine learning* and *deep learning*. The first model is *the Support Vector Machine* (SVM), which is one of the classic machine learning methods widely used in text classification tasks. SVM utilizes TF-IDF-based feature representation to convert text into numerical vectors, thereby capturing the frequency distribution of words within a document. This model was chosen because it is known to perform stably on small to medium-sized datasets [26] [27] .

Next, the *Long Short-Term Memory* (LSTM) model was used as a representation of a sequential deep learning approach. LSTM has the ability to capture long-term dependencies in text, allowing it to understand the relationships between words in a sentence. This model is widely used in sentiment analysis due to its ability to process sequential data [28] [29] .

Classification and Evaluation

Sentiment in this study is classified into three main categories: positive, neutral, and negative. The classification process was performed using three model approaches: IndoBERT, SVM, and LSTM. These models were used to conduct a comparative evaluation of their performance in sentiment classification. To evaluate the performance of the models, a confusion matrix-based approach was employed. Based on the confusion matrix, four evaluation metrics were calculated, namely accuracy, precision, recall, and F1-score [30].

Accuracy measures the overall proportion of correctly classified instances among all data and is defined as in Equation (3):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

where TP (True Positive) represents correctly predicted positive instances, TN (True Negative) represents correctly predicted negative instances, FP (False Positive) represents incorrectly predicted positive instances, and FN (False Negative) represents incorrectly predicted negative instances. Precision measures the accuracy of positive predictions and is defined as:

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

This metric indicates how many of the predicted positive instances are actually correct. Recall measures the model's ability to identify all relevant instances and is defined as:

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

This metric reflects how well the model captures all actual positive instances. F1-score is the harmonic mean of precision and recall, providing a balanced measure between the two, and is defined as:

$$F1 = 2\,x\,\frac{Precision \times Recall}{Precision + Recall} \quad (6)$$

The use of these four metrics provides a comprehensive evaluation of model performance, particularly in handling imbalanced data distributions, where relying solely on accuracy may lead to misleading conclusion ..

3. RESULTS AND DISCUSSIONS

This section presents the results of sentiment analysis experiments on a dataset of YouTube comments related to tourism in Bangka Belitung. The analysis was conducted using three model approaches: SVM as a *machine learning* model, LSTM as a *deep learning* model, and IndoBERT as a *transformer* model. The discussion covers data distribution, model evaluation results, and sentiment classification analysis.

Sentiment Distribution Analysis

Figure 3 shows the distribution of tourist sentiment generated from the IndoBERT model prediction results. In this study, sentiment labels were not assigned manually. Instead, the preprocessed comments were classified automatically using the fine-tuned IndoBERT model into three sentiment categories: positive, neutral, and negative. The predicted labels were then used to describe the overall sentiment distribution in the dataset. The analysis shows that the neutral category dominates with 434 comments, followed by positive sentiment with 333 comments and negative sentiment with 146 comments. The dominance of neutral sentiment indicates that most user comments tend to be informative or descriptive without containing strong emotional expressions. Meanwhile, the presence of positive and negative sentiments reflects users' evaluations, appreciation, criticism, or concerns regarding tourism in the Bangka Belitung Islands.

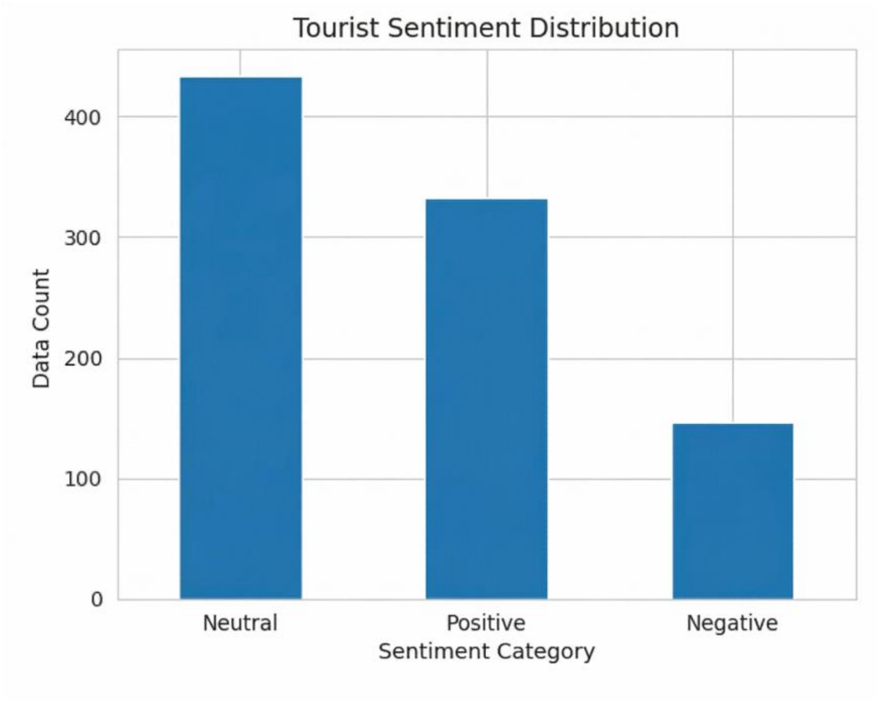

Figure 3. Sentiment distribution

Meanwhile, the significant number of positive sentiments indicates a favorable perception of the tourist attractions discussed, although they do not dominate overall. On the other hand, negative sentiments are the least numerous, suggesting that complaints or criticism from users are relatively lower compared to neutral and positive expressions. This imbalanced data distribution has important implications for the performance of the classification model. The imbalance in data volume across classes, particularly the dominance of the neutral class, has the potential to cause the model to be biased toward the majority class. This condition can affect evaluation metrics such as *precision*, *recall*, and *F1-score* for each class.

In this study, the distribution pattern of prediction results also reflects challenges in the sentiment classification process, particularly in distinguishing between the positive and negative classes, which have fewer data points. Therefore, this sentiment distribution analysis not only provides an overview of the trends in IndoBERT's prediction results but also serves as a foundation for interpreting performance differences between the SVM, LSTM, and IndoBERT models in subsequent evaluation stages.

IndoBERT Model Performance

The implementation of the IndoBERT model in this study demonstrated a fairly good ability to classify tourist sentiment from YouTube comments. Based on tests conducted through a *fine-tuning* process in the Google Colab environment, the model achieved an accuracy rate of 71%. This model's performance reflects the challenges of handling informal and imbalanced social media text data.

As seen in *the confusion matrix* by class, the model showed the most stable performance in the Neutral category with an F1-score of 0.77, followed by the Negative category at 0.71. Meanwhile, the Positive category recorded the lowest F1-score of 0.49. Nevertheless, overall, the IndoBERT model remains superior in capturing the semantic context of the Indonesian language compared to conventional *machine learning* algorithms, particularly in recognizing criticism and suggestions expressed by tourists.

Performance of the Comparison Model

This section discusses the performance of the benchmark models used in this study, namely SVM and LSTM. The evaluation was conducted using the metrics of accuracy, *precision*, *recall*, and *F1-score* to provide an overview of each model's performance in sentiment classification. Based on the test results, the SVM model achieved an accuracy of 0.69, with a *weighted F1-score* of 0.67. These results indicate that SVM performs quite consistently in sentiment classification, particularly for classes with a larger volume of data. However, the relatively low *recall* value for one of the classes suggests that this model still has limitations in identifying all data within that specific class.

Meanwhile, the LSTM model achieved an accuracy of 0.68, with a weighted F1-score of 0.63. Although LSTM is designed to handle sequential data and is capable of capturing relationships between words in a sentence, the experimental results show that this model has not yet been able to deliver optimal performance. This is evident from the very low recall score in one of the classes, indicating that the model struggles to consistently recognize specific sentiment patterns. The difference in performance between SVM and LSTM also shows that an increase in model complexity does not always correlate directly with an increase in accuracy. In this case, LSTM, as a deep learning model, does not demonstrate a significant advantage over SVM. This is likely due to the limited amount of data and an unbalanced class distribution, which affect the model's ability to generalize.

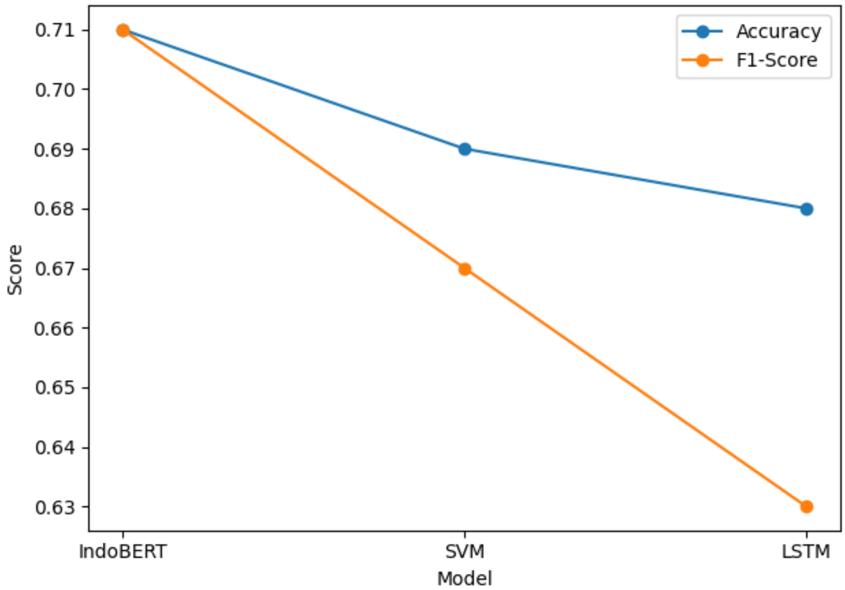

Figure 4. Model performance comparison

The graph in Figure 4 shows a comparison of the performance of the IndoBERT, SVM, and LSTM models based on accuracy and *F1-score*. It can be seen that IndoBERT performs best on both metrics, followed by SVM, while LSTM shows relatively lower performance. This difference indicates that *transformer-based* models are better at understanding linguistic context than *machine learning* or *deep learning* models.

Confusion Matrix Analysis

Confusion matrix analysis is used to evaluate a model's performance in classifying sentiment into three categories: positive, neutral, and negative. Through this matrix, the distribution of correct predictions and classification errors across each class can be determined, providing a more detailed picture than relying solely on accuracy scores. Based on the evaluation results, the IndoBERT model showed the best performance with an accuracy of 0.71, followed by the SVM model at 0.69, and the LSTM model at 0.68. This indicates that IndoBERT has better overall capability in classifying sentiment.

When examining the distribution of predictions in *the confusion matrix*, it is evident that most classification errors occur in the positive and negative classes, which are frequently predicted as neutral. This indicates a tendency for the model to classify data into the majority class, which aligns with the dataset's distribution, which is dominated by neutral sentiment. In the SVM model, classification errors primarily occur due to the model's limitations in deeply understanding sentence context. SVM relies on word-based feature representations, making it difficult to capture meanings not explicitly stated in the text. Consequently, some comments that should fall into the positive or negative categories are classified as neutral.

Meanwhile, in the LSTM model, although it is capable of capturing word sequences in sentences, it still exhibits weaknesses in recognizing minority classes. This is evident from the low *recall* score for one of the classes, indicating that most of the data in that class was not successfully identified. This condition suggests that LSTM is not yet optimal in handling data imbalance. Unlike these two models, IndoBERT is able to reduce the classification error rate for each class. This capability is supported by a *transformer* architecture that utilizes a *self-attention* mechanism, allowing the model to understand bidirectional contextual relationships between words. Consequently, IndoBERT is more effective at identifying sentiment.

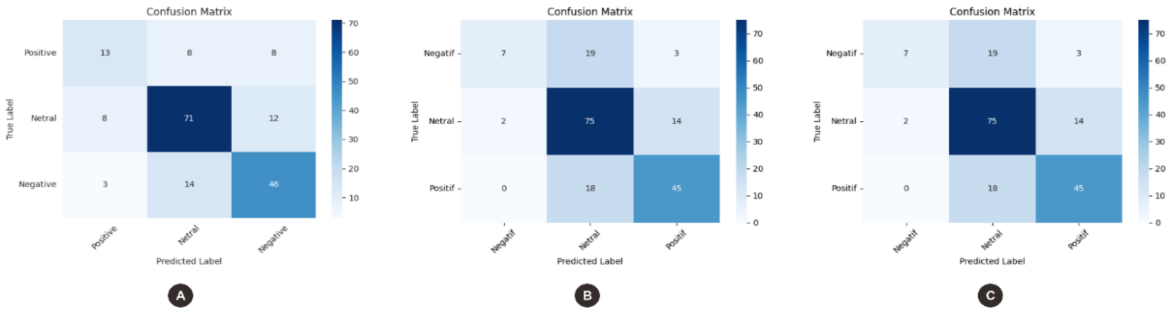

Figure 5. *Confusion Matrix*: IndoBERT model (A), SVM model (B), & LSTM model (C)

Overall, the confusion matrix analysis shows that performance differences among models are influenced not only by accuracy levels but also by each model's ability to handle imbalanced data distributions and the complexity of natural language. Therefore, the use of transformer-based models such as IndoBERT proves to be superior in the context of sentiment analysis on social media data. Overall, Figure 5 shows that the performance differences between models lie not only in accuracy scores but also in classification error patterns. *Transformer-based* models like IndoBERT have proven superior in handling imbalanced data distributions and the complexity of natural language compared to *machine learning* and *deep learning* models.

The graph in Figure 6 shows a comparison of the performance of the IndoBERT, SVM, and LSTM models in classifying sentiment across each category—positive, neutral, and negative—based on the *F1-score*. It is evident that IndoBERT performs best in nearly all categories, particularly for positive sentiment, where it demonstrates a significant improvement compared to the other models. In the neutral category, all three models demonstrate relatively high performance, with *F1-scores* that are not significantly different. This suggests that the neutral class is easier to identify due to its dominant presence in the dataset. Meanwhile, in the negative category, the performance of the three models tends to be balanced, although IndoBERT still delivers competitive results.

The most striking difference is seen in the positive category, where the LSTM model performs the worst. This indicates that the model struggles to recognize patterns of positive sentiment. Conversely, IndoBERT is better at identifying positive sentiment, demonstrating the superiority of transformer-based models in understanding linguistic context.

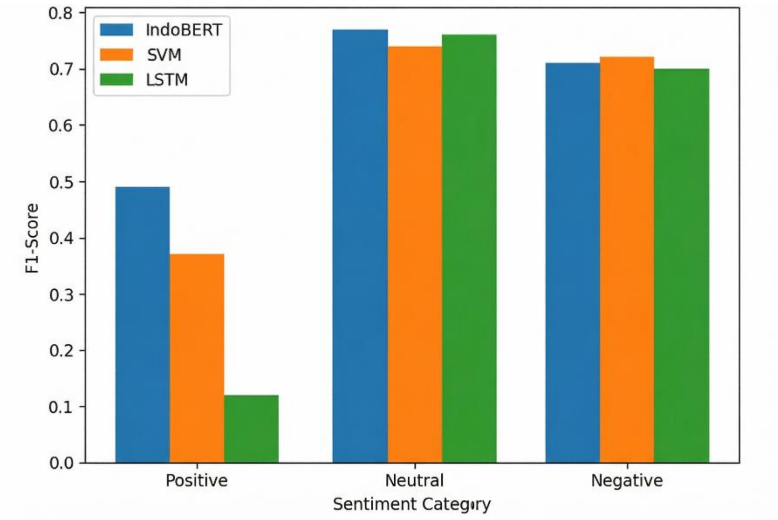

Figure 6. Model performance comparison

Representation of Comment Data and Sentiment Labels

This section presents a qualitative analysis of several sample comments along with the sentiment labels generated by the IndoBERT model, which performed best in this study. This presentation aims to provide a deeper understanding of how the model interprets various forms of opinion expression in the data. Based on Table 2, it is evident that comments with positive sentiment generally contain expressions of appreciation for natural beauty, environmental cleanliness, and the friendliness of the local community. Comments such as "Belitung Island has a very serene atmosphere, with clean beaches and a friendly community" demonstrate explicit positive evaluations of the tourist destination.

Table 2. Comments and sentiment types

Comment	Sentiment Type
Sangat disayangkan bahwa kondisi danau menjadi keruh dan tercemar akibat aktivitas pertambangan.	Negative
Pantai tersebut memiliki kondisi yang sangat baik dan menarik.	Positive
Bangka Belitung dikenal sebagai daerah penghasil timah, dan beberapa tokoh publik Indonesia juga berasal dari wilayah tersebut.	Neutral
Eksploitasi sumber daya timah di wilayah Bangka, termasuk pada area danau alami, telah menyebabkan kerusakan lingkungan yang signifikan dan terjadi secara masif.	Cons
Pulau Belitung memiliki suasana yang sangat asri, dengan pantai yang bersih serta masyarakat yang ramah.	Positive
Kondisi hutan mengalami kerusakan yang signifikan akibat eksploitasi sumber daya secara berlebihan.	Negative

Conversely, comments with negative sentiment tend to highlight environmental damage caused by mining activities. For example, statements regarding polluted lakes or forest destruction due to resource exploitation indicate criticism of the resulting environmental impacts. These comments generally carry strong evaluative meaning, whether explicit or implicit. Meanwhile, comments with neutral sentiment are more informative and do not contain a clear opinion. Examples include statements about Bangka Belitung as a tin-producing region or the hometown of a particular public figure, which focus more on presenting facts without dominant emotional expression.

The results of this analysis indicate that variations in the form of expression within comments, whether explicit or implicit, pose a challenge in the sentiment classification process. Nevertheless, the IndoBERT model is capable of identifying sentiment patterns quite effectively through its understanding of linguistic context. Therefore, this qualitative analysis serves as a complement to the quantitative evaluation, while also validating the classification results obtained. Although comparison models such as SVM and LSTM were also used in this study, the qualitative analysis focused on IndoBERT because this model demonstrated superior performance in the quantitative evaluation. Therefore, the results presented in this section represent the best model's ability to understand the context and meaning of sentiment in the comment data.

4. CONCLUSION

This study aims to analyze tourist sentiment toward tourism in the Bangka Belitung Islands based on YouTube comments using a comparative approach between *machine* learning, *deep learning*, and *transformer* models. The results show that the sentiment distribution is dominated by the neutral category, reflecting that most comments are informative and do not contain explicit opinions. Based on the evaluation results, the IndoBERT model demonstrated the best performance with higher accuracy and *F1-score* values compared to the SVM and LSTM models. The SVM model, as a *machine learning* approach, showed fairly stable performance but has limitations in understanding linguistic context. Meanwhile, the LSTM model, as a *deep learning* representation, has not yet been able to provide a significant performance improvement, particularly in identifying minority classes.

Confusion matrix analysis shows that classification errors generally occur in the positive and negative classes, which tend to be predicted as neutral, influenced by the imbalanced data distribution. Additionally, qualitative analysis indicates that variations in language expression pose a major challenge in the sentiment classification process. Overall, the results of this study indicate that *transformer-based* models such as IndoBERT have an advantage in understanding linguistic context more deeply, making them more effective at identifying sentiment. Thus, this study contributes to the development of context-based sentiment analysis on social media data, particularly in the tourism domain.

REFERENCES

[1] I. M. Sudiarta, N. Suharsono, L. E. Tripalupi, and M. R. Irwansyah, “Analisis Dampak Perkembangan Pariwisata Terhadap Kondisi Sosial Ekonomi Masyarakat,” *Bus. Account. Educ. J.*, vol. 2, no. 1, pp. 22–31, 2021, doi: 10.15294/baej.v0i1.42765.

[2] F. Titing Koerniawati, “Destinasi Wisata, Sumber Daya Manusia Pariwisata Dan Pariwisata Berkelanjutan,” *Siwayang J. Publ. Ilm. Bid. Pariwisata, Kebudayaan, Dan Antropol.*, vol. 1, no. 1, pp. 39–50, 2022, doi: 10.54443/siwayang.v1i1.52.

[3] R. Anindita, “Jurnal Kajian Stratejik Ketahanan Nasional Indonesia Sebagai Negara Kepulauan Indonesia Sebagai Negara Kepulauan,” *J. Kaji. Strat.*, vol. 2, no. 2, 2019, doi: 10.7454/jkskn.v2i2.10022.

[4] Badan Pusat Statistik, “Luas Daerah dan Jumlah Pulau Menurut Provinsi 2023.” [Online]. Available: https://www.bps.go.id/id/statistics-table/3/VUZwV01tSlpPVlpsWlRKbmMxcFhhSGhEVjFoUFFUMDkjMw==/luas-daerah-dan-jumlah-pulau-menurut-provinsi--2023.html?year=2023

[5] Nuraini, *Sejarah Terbentuknya Provinsi Kepulauan Bangka Belitung*. Tanjungpinang: Departemen Kebudayaan dan Pariwisata Balai Pelestarian Sejarah dan Nilai Tradisional Tanjungpinang, 2007.

[6] Y. Megawandi, “Pembangunan Pariwisata di Provinsi Kepulauan Bangka Belitung dalam Pendekatan Whole of Government,” vol. 1, no. 2, pp. 108–119, 2020.

[7] Badan Kesatuan Bangsa dan Politik Provinsi Bangka Belitung, “Potensi Daerah.” [Online]. Available: https://kesbangpol.babelprov.go.id/content/potensi-daerah

[8] B. P. S. P. K. B. Belitung, “Jumlah Perjalanan Wisatawan Nusantara Menurut Kabupaten/Kota Tujuan (Perjalanan) 2024.” [Online]. Available: https://babel.bps.go.id/id/statistics-table/2/MTE3OSMy/jumlah-perjalanan-wisatawan-nusantara-menurut-kabupaten-kota-tujuan--perjalanan-.html

[9] B. P. S. P. Bali, “Jumlah Perjalanan Wisatawan Nusantara Menurut Kabupaten/Kota Tujuan (Perjalanan) 2024.” [Online]. Available: https://bali.bps.go.id/id/statistics-table/2/NDEzIzI=/jumlah-perjalanan-wisatawan-nusantara-menurut-kabupaten-kota-tujuan--perjalanan-.html

[10] B. Yogyakarta, “Jumlah Perjalanan Wisatawan Nusantara Menurut Kabupaten/Kota Tujuan (Perjalanan) 2024.” [Online]. Available: https://yogyakarta.bps.go.id/id/statistics-table/2/NDEzIzI=/jumlah-perjalanan-wisatawan-nusantara-menurut-kabupaten-kota-tujuan--perjalanan-.html

[11] M. S. Syahlan, D. Irmayanti, and S. Alam, “Analisis Sentimen Terhadap Tempat Wisata Dari Komentar Pengunjung Dengan Menggunakan Metode Support Vector Machine (Svm),” *Simtek J. Sist. Inf. dan Tek. Komput.*, vol. 8, no. 2, pp. 315–319, 2023, doi: 10.51876/simtek.v8i2.281.

[12] A. Manthiou, P. Klaus, and V. H. Luong, “Slow tourism : Conceptualization and interpretation – A travel vloggers ’ perspective,” *Tour. Manag.*, vol. 93, no. November 2021, p. 104570, 2022, doi: 10.1016/j.tourman.2022.104570.

[13] A. A. Mirzamurodova and K. Seweryn, “Tourist Destination Image Projected by YouTube Travel Videos : The Case of Seoul , Republic of Korea,” 2025, doi: 10.1002/jtr.70049.

[14] H. Murfi, T. Gowandi, G. Ardaneswari, and S. Nurrohmah, “BERT-based combination of convolutional and recurrent neural network for indonesian sentiment analysis,” *Appl. Soft Comput.*, vol. 151, no. June 2023, p. 111112, 2024, doi: 10.1016/j.asoc.2023.111112.

[15] R. L. Mustofa, C. E. Widodo, U. Diponegoro, and P. Korespondensi, “ANALISIS SENTIMEN BERBASIS ASPEK PADA APLIKASI ELEKTRONIK SURVEI KEPUASAN MASYARAKAT (E-SKM) JAWA TENGAH MENGGUNAKAN ASPECT-BASED SENTIMENT ANALYSIS ON THE CENTRAL JAVA PUBLIC SATISFACTION SURVEY ELECTRONIC APPLICATION (E-SKM) USING INDOBERT,” vol. 12, no. 4, pp. 867–874, 2025.

[16] F. Koto and T. Baldwin, “IndoLEM and IndoBERT : A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP,” *Proc. 28th Int. Conf. Comput. Linguist.*, pp. 757–770, 2020.

[17] D. Chi, T. Huang, Z. Jia, and S. Zhang, “Research on sentiment analysis of hotel review text based on BERT-TCN-BiLSTM-attention model,” vol. 25, no. February, pp. 1–10, 2025.

[18] G. Z. Nabiilah, S. Y. Prasetyo, Z. N. Izdihar, and A. S. Girsang, “ScienceDirect ScienceDirect 7th International Conference on Computer Science and Computational Intelligence 2022 BERT base model for toxic comment analysis on BERT base model for

toxic comment analysis on Indonesian social media Indonesian social media," *Procedia Comput. Sci.*, vol. 216, no. 2022, pp. 714–721, 2023, doi: 10.1016/j.procs.2022.12.188.

[19] H. Sibyan and N. Hasanah, "Analisis Sentimen Ulasan Pada Wisata Dieng Dengan Algoritma K-Nearest Neighbor (K-Nn)," *J. Penelit. dan Pengabdi. Kpd. Masy. UNSIQ*, vol. 9, no. 1, pp. 38–47, 2022, doi: 10.32699/ppkm.v9i1.2218.

[20] T. Bey and I. K. Ari, "Implementasi BERT pada Analisis Sentimen Ulasan Destinasi Wisata Bali," *J. Elektron. Ilmu Komput. Udayana*, vol. 12, no. 2, pp. 409–420, 2023.

[21] Doni Abdul Fatah, Eka Mala Sari Rochman, Fajrul Ihsan Kamil, and Ahmad Su'ud, "Sentiment Analysis of Madura Tourism Opinion Using Support Vector Machine (SVM)," *Tech. Rom. J. Appl. Sci. Technol.*, vol. 16, pp. 243–249, 2023, doi: 10.47577/technium.v16i.9988.

[22] E. B. Ramezani, "Sentiment analysis applications using deep learning advancements in social networks: A systematic review," *Neurocomputing*, vol. 634, p. 129862, 2025, doi: https://doi.org/10.1016/j.neucom.2025.129862.

[23] Y. R. Sipayung, M. A. Wibowo, and R. Sanjaya, "Multimodal Implicit Sentiment Analysis for Tourism Development : A Systematic Literature Review," *J. Inf. Syst. Informatics*, vol. 8, no. 1, pp. 781–808, 2026, doi: 10.63158/journalisi.v8i1.1436.

[24] F. Koto and T. Baldwin, "IndoLEM and IndoBERT : A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," 2020.

[25] J. Vincent *et al.*, "Extractive Indonesian News Text Summarization using Extractive Indonesian News Text Summarization usi," *10th Int. Conf. Comput. Sci. Comput. Intell. 2025*, vol. 00, 2025.

[26] R. Setiyawan and Z. Mustofa, "Comparison of the performance of naive bayes and support vector machine in sirekap sentiment analysis with the lexicon- based approach," pp. 122–132, 2024.

[27] A. S. Maurya, "A COMPARATIVE ANALYSIS OF TEXT CLASSIFICATION ALGORITHMS FOR POS AMBIGUITY A COMPARATIVE ANALYSIS OF TEXT CLASSIFICATION ALGORITHMS FOR POS AMBIGUITY USING WEKA," *Comp. Anal. TEXT Classif. ALGORITHMS POS AMBIGUITY USING WEKA*, vol. 12, no. 6, pp. 406–416, 2020.

[28] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM : A Search Space Odyssey," *IEEE Trans. NEURAL NETWORKS Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, 2017.

[29] P. Utami, M. R. Ningsih, D. Ananda, and A. Pertiwi, "Sentimen based-emotion classification using bidirectional long," pp. 281–289, 2024.

[30] M. Cherradi and A. El Haddadi, "Comparative Analysis of Machine Learning Algorithms for Sentiment Analysis in Film Reviews," pp. 137–147, 2024.