



Harnessing BERT for Semantic Understanding in Tourism Recommendation Engines

Renita Astri^{1*}, Lai Po Hung², Suaini Binti Sura³, Ahmad Kamal⁴

¹Sistem Informasi, Fakultas Farmasi Sains dan Teknologi, Universitas Dharma Andalas, Padang, Indonesia

^{2,3}Computer Science, Faculty of Computing and Informatics, University Malaysia Sabah, Sabah, Malaysia

⁴Teknik Informatika, Institut Bisnis dan Teknologi Pelita Indonesia, Pekanbaru, Indonesia

¹rethakamal@unidha.ac.id, ²laipohung@ums.edu.my, ³su_sura@ums.edu.my, ⁴ahmad.kamal@lecturer.pelitaindonesia.ac.id

Abstract

It will be necessary for attraction managers within hotels to track guests' lifestyles to keep the business running. Such an understanding may be achieved, for example by analyzing reviews on attractions to capture the attitudes of the visitors towards the services and business within the tourism industry. The approach utilizes web scraping to gather user-generated reviews, using text preprocessing, data pre-processing, and further improvement of the model using labelled sentiment data divided into three sentiment classes: positive, negative, or neutral. The dataset consisting of 908 reviews were divided in 70:15:15 ratio for training, validation and testing. The effectiveness of the model was evaluated using accuracy, precision, recall, and the F1-score. In this study, the BERT deep learning model is used to classify sentiments of Indonesian tourist. Using the SmallBERT variant fine-tuned on 515k reviews for 5 epochs, the model achieved 91.40% accuracy, 90.51% precision, recall, and F1 score. The results indicate a dominance of positive sentiments, visualized using tableau. This research provides a robust foundation for developing intelligent sentiment-based recommendation systems in the tourism sector and suggests future exploration using other transformer-based models such as GPT, T5, or BART for comparative analysis.

Keywords: BERT; classification; deep learning; fine-tuning; NLP

How to Cite: Renita Astri, L. Po Hung, S. Binti Sura, and A. Kamal, "Harnessing BERT for Semantic Understanding in Tourism Recommendation Engines", *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 9, no. 4, pp. 879 - 885, Aug. 2025.

Permalink/DOI: <https://doi.org/10.29207/resti.v9i4.6575>

Received: April 23, 2025

Accepted: August 7, 2025

Available Online: August 19, 2025

This is an open-access article under the CC BY 4.0 License

Published by Ikatan Ahli Informatika Indonesia

1. Introduction

The tourism sector has long been recognized as a vital driver in fostering economic development. Consequently, it requires optimization to enhance community income in tourist areas through the creation of employment opportunities[1]. In the contemporary world, information technology has been widely and effectively utilized across diverse sectors. One of the most prominent examples is the tourism sector, where its application has transformed traditional practices and significantly enhanced the overall experience for travelers. Information technology plays a crucial role in advancing the tourism industry by offering easy access to information, simplifying services for tourists, and providing various other benefits[2].

Today's travelers can effortlessly seek references about tourist attractions through various platforms such as TripAdvisor, Airbnb, TripIt, Google Maps, and others [3]. Prospective visitors can explore information about

tourist destinations based on the experiences of travelers who have visited these places through reviews shared online[4]. This ease of access has stimulated greater enthusiasm for traveling, which is evident from the reviews and feedback posted on social media and digital tourism platforms. Such expressions of opinion can generally be grouped into positive, negative, or neutral categories. To process and organize these varied viewpoints, machine learning techniques can be applied. Sentiment analysis, a subfield of Natural Language Processing (NLP), aims to detect and categorize textual opinions, especially to assess whether an author's viewpoint is positive, negative, or neutral [5]. This sentiment analysis can serve as the foundation for recommendation systems and data visualization.

One of the machine learning methods used for this purpose is the transformer model. This model has proven effective in understanding the context and nuances of language, delivering more accurate results

compared to traditional methods [6]. The transformer model employs an architecture capable of processing large and complex datasets, thereby capturing relationships between words and phrases more effectively [7] and generating deeper representations of textual meanings. Consequently, the application of this model in sentiment analysis can improve prediction quality and provide richer insights into public opinions on various issues.

BERT, which stands for Bidirectional Encoder Representations from Transformers, is a deep learning framework that establishes connections between every input and output element, with the associated weights being calculated in real time [8]. Historically, language models could only process textual input sequentially, either from left to right or right to left. However, such models were unable to perform both directions simultaneously. BERT, on the other hand, is uniquely designed to read bidirectionally [9], [10]. By utilizing its bidirectional nature, BERT is pre-trained on tasks in NLP, including Masked Language Modeling and Next Sentence Prediction [11]. In the context of BERT's development, specialized models such as IndoBERT can be utilized. Built on the BERT framework, IndoBERT represents a cutting-edge model for the Indonesian language [12], [13]. IndoBERT has been trained on an extensive dataset of Indonesian text, encompassing over four billion words. In studies involving IndoBERT, the model was thoroughly evaluated against different word embeddings and pre-trained models, including Multilingual BERT and XLM-R, to measure its effectiveness [14]. The results demonstrated that IndoBERT outperformed most existing models, exceeding the performance of Multilingual models in 8 out of 12 tasks. Overall, IndoBERT achieved the highest average score in classification tasks. Monolingual models like IndoBERT have shown a superior ability to capture sentiment-level semantics in both casual and formal language styles compared to multilingual models [15], despite being 40%–60% smaller in size than their multilingual counterparts. These advantages highlight that for applications focused on a specific language, monolingual models like IndoBERT can yield more accurate and contextually relevant results in natural language processing [16]. In this study, the authors implement the IndoBERT model to conduct sentiment analysis on review data of tourist attractions in West Sumatra, Indonesia. Additionally, the study compares the performance of two optimizers, AdamW and AdaFactor, during the fine-tuning of the IndoBERT model. By analyzing the results of this comparison, the study aims to determine the most suitable optimizer to enhance IndoBERT's performance.

One previous study concluded that a tourist destination must encompass four key components of tourism: attraction, accessibility, amenity, and ancillary services [17], [18]. Another study highlighted that the better the quality of a tourist attraction, the higher the satisfaction

of visitors, and the more likely it is to gain recognition among other travelers through positive recommendations [19], [20]. This study aims to analyze visitor sentiments regarding tourist attractions based on reviews to understand the comparative sentiment of the public toward the level of satisfaction with visits to these attractions. The analysis makes use of BERT, a pre-trained deep learning model, with additional adjustments through fine-tuning. It is expected that the outcomes of this study will equip tourism stakeholders, with a focus on the hospitality industry, with enhanced understanding to guide decisions on service delivery and business strategies.

2. Methods

The primary purpose of this study is to evaluate sentiment classification methods for reviews of tourist attractions in West Sumatra using the IndoBERT transformer model. The Indonesian government and various stakeholders in the tourism sector require this sentiment analysis model as a solution to critical challenges in developing tourist destinations by identifying attractions that capture visitors' interest [21]. The deep learning technique known as BERT offers higher accuracy compared to other transformer-based techniques. BERT is a machine learning approach for pre-training NLP. The IndoBERT model was selected because of its proven ability to accurately capture the semantic meaning of text written in the Indonesian language. IndoBERT is a large pre-trained model trained on approximately 4 billion words from a specialized Indonesian-language corpus, comprising over 20 gigabytes of text data [22]. This has resulted in high accuracy in sentiment analysis for the Indonesian language. The research involves several stages, including collecting review text and ratings via web crawling, preprocessing, splitting and formatting data, fine-tuning, optimization, and evaluating the results. The research workflow is illustrated in Figure 1.

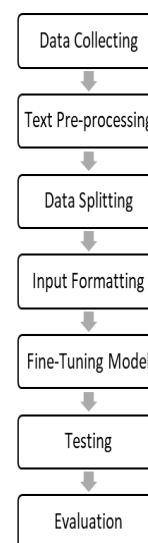


Figure. 1. Research workflow

2.1 Data Collection

The data used in this study was obtained through web scraping tools to collect information from websites. This information was collected from reviews of tourist attractions in West Sumatra, written in Indonesian and sourced from tourism-related websites. To prove the diversity of the results, this study used a dataset sourced from Google Reviews. The data collected consisted of 51,961 visitor comments on eight tourist attractions in West Sumatra, namely Pagaruyung Basa Palace, Kapalo Banda Taram, Jam Gadang, Japanese Tunnel, Padang Beach, Puncak Lawang, Anai Valley, and Gandorhah Beach.

The review data consists of user feedback about tourism sites, categorized into negative, positive, and neutral sentiments. The data was then stored in a CSV file, enabling the sentiment analysis model to be trained.

The model, developed using the IndoBERT framework, was trained and tested on this dataset. Before processing the data, the researchers performed data cleaning, referred to as text preprocessing. This process involved removing special characters, converting all text to lowercase, and eliminating irrelevant stop words to ensure the model could learn from the most significant and meaningful information. Once the cleaning process was complete, the next step was to split the data into training and testing sets to effectively evaluate the model's performance.

2.2. Text Preprocessing

The objective of text preprocessing is to refine the review text data by removing unwanted characters, words, and punctuation marks [23]. The preprocessing steps are as follows: (1). Case Folding: This process refers to transforming every character in a text into lowercase. In this research, case folding was implemented to ensure that all review dataset texts were converted to lowercase letters [24], (2). tokenization: This process refers to dividing text into smaller units, known as tokens, which may consist of words, phrases, symbols, or other meaningful elements [25], (3). Stopword Removal: Stopword removal involves eliminating words that are considered unnecessary for analysis. The purpose of this step is to remove frequently occurring words that do not contribute meaningfully to the analysis, leaving only words that accurately convey the text's semantic meaning[26], (4). Punctuation Removal: Punctuation removal is the process of eliminating symbols within the text document. This includes removing irrelevant characters such as punctuation marks, numbers, HTML tags, and other symbols[27], [28], (5). Padding: Padding involves adding special tokens to ensure that each sequence of text has the same length. As text batches often vary in length, padding is applied to standardize the shorter sequences to match the longest sequence in the batch or the maximum length accepted by the model [29]. In this study, the maximum length used for the BERT model was set to 100 tokens.

2.3 Data Splitting

The dataset splitting process is conducted to evaluate the model's performance after training. The data is divided into three proportions: training data for training and fine-tuning the model, validation data to mitigate overfitting and reduce noise during model training, and testing data to assess the model's accuracy after training[30].

The data proportions are allocated as 70% for training, 15% for validation, and 15% for testing out of the total dataset. Consequently, the amount of data used for the initial fine-tuning totals 636 samples.

2.4 Input Formatting

Although different transformer models vary in design, they share similarities in their input requirements. Most models use a similar input format, necessitating the encoding or formatting of the dataset to align with the specific transformer model's input structure. This process, known as input formatting, includes tokenization, vectorization, and mapping[31].

In the context of BERT, the tokenizer converts a sequence of sentences into segmented tokens, as illustrated in Table 1.

Table 1. BERT tokenization of sentences

Sequence	Token
I dont like this !	"I", "don't", "like", "this", "!"

After combining the steps for tokenization, WordPiece vectorization, adding special tokens, truncating reviews longer than the maximum length, and mapping the ID objects and attention mask into a dictionary, the resulting output is shown in Table 2.

Table 2. Input formatting model BERT

Sequence : "I don't like this !"		
Input ID	Token Type ID	Attention Mask
[101, 2123, 1005, 1056, 2066, 2009, 999, 102, 0, 0]	[0, 0, 0, 0, 0, 0, 0, 0, 0, 0]	[1, 1, 1, 1, 1, 1, 1, 1, 0, 0]

2.5 Fine-Tuning the Model

Training the BERT model does not begin from scratch but utilizes a pre-trained model. The pre-trained BERT model already contains stored knowledge of Language Generation and Understanding for the Indonesian language. Consequently, less time is required to fine-tune the model for this specific task. By fine-tuning the lower layers of the extensively pre-trained model, only minimal adjustments are needed when using its outputs as features for sentiment analysis tasks [32]. Moreover, due to the pre-trained weights, this approach allows the fine-tuning process to be conducted on significantly smaller datasets. Typically, a model trained from scratch would require an extensive dataset to train the network effectively, necessitating considerable time and computational resources for dataset creation. By fine-tuning IndoBERT, the model can achieve high

performance using a much smaller training dataset [33]. During the fine-tuning stage, the IndoBERT model starts with parameters already learned during pre-training, and all parameters are adjusted using labeled data. Throughout the fine-tuning process, every parameter is updated. BERT processes sequences of words as input, which flows through the BERT stack. Each layer applies self-attention, forwards the results through a feed-forward network, and passes them to the next encoder layer.

During the fine-tuning stage, specific tokens are incorporated into every sequence, with [CLS] being a special marker placed at the start of each input instance. The final hidden state corresponding to this token is used as the aggregate sequence representation for classification tasks. Additionally, [SEP] is a token used to separate pairs of sentences. If multiple pairs of sentences exist within one sequence, this token is employed to separate them [34].

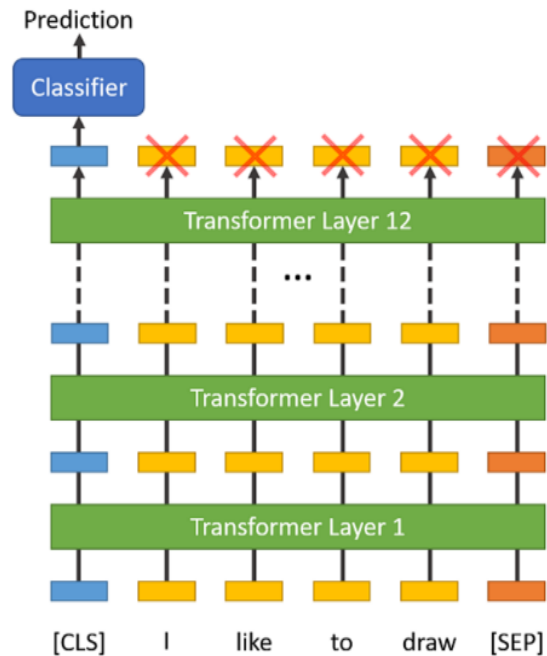


Figure 2. BERT model

In the BERT output, the model focuses solely on the output from the first position, which corresponds to the [CLS] token. The result of this token is passed to the output layer for classification purposes. If additional labels are required for classification, the classifier network can be adjusted to include more output neurons, which are subsequently processed through a softmax layer for centralized training [35].

2.6 Optimizer

Adaptive Moment Estimation (Adam) is an algorithm designed to replace the classical stochastic gradient descent procedure by iteratively updating network weights based on training data [36]. Adam operates by combining the best properties of two stochastic gradient descent extensions: the adaptive gradient algorithm (AdaGrad) and root mean square propagation

(RMSProp). This combination enables Adam to optimize algorithms efficiently, particularly in handling sparse gradients for noisy problems [37]. By leveraging optimization techniques capable of reducing gradient values, this method is highly effective when working with large datasets and a substantial number of parameters. AdamW, a stochastic optimization method, modifies the typical implementation of weight decay in Adam (Adaptive Moment Estimation). L2 regularization in Adam is typically implemented using the following adjustment shown in Equation 1, where w_{t_t} represents the weight decay rate at time t .

$$g_t = \nabla f(\theta_t) + \omega_t \theta_t \quad (1)$$

So AdamW adapted the term weight decay to appear in the gradient update shown in Equation 2.

$$x_t \leftarrow x_{t-1} - \alpha \frac{\beta_1 m_{t-1} + (1 - \beta_1)(\nabla f_t + \omega x_{t-1})}{\sqrt{v_t + \epsilon}} \quad (2)$$

Meanwhile, AdaFactor is an adam-based stochastic optimization method that reduces memory usage while maintaining the empirical benefits of adaptation. AdaFactor reduces memory usage by first proposing to replacing the dismooted squared gradient matrix with a low-rank approximation as shown in Equations 3-6.

$$G_t = \nabla f(X_{t-1}) \quad (3)$$

$$R_t = \beta_2 R_{t-1} + (1 - \beta_2)(G_t^2)1_m \quad (4)$$

$$C_t = \beta_2 C_{t-1} + (1 - \beta_2)1_n^T(G_t^2) \quad (5)$$

$$\hat{V}_t = (R_t C_t / 1_n^T R_t) / (1 - \beta_2^t) \quad (6)$$

Secondly, Adafactor eliminates momentum completely. By increasing the rate of decay over time and truncating the gradient update to the root mean-square as shown in Equation 7.

$$\begin{aligned} RMS(U_t) &= RMS_{x \in X}(U_{xt}) \\ &= \sqrt{Mean_{x \in X}(\frac{g_{xt}^2}{\hat{V}_{xt}}} \end{aligned} \quad (7)$$

And finally, Adafactor multiplies the learning rate by the scale parameter (this is called relative step size) as shown in Equation 8.

$$\alpha_t = \max(\epsilon_2 RMS(X_{t-1}))\rho_t \quad (8)$$

2.7 Assessment and Evaluation

Model performance assessment is a stage to determine the accuracy and performance of the IndoBERT model in conducting sentiment analysis. Evaluation of the two optimizer methods is done by calculating the precision, recall, f-1 score, and accuracy values.

Description:

TP/True Positive is the number of relevant data correctly classified as matches by the model; TN/True Negative is the number of irrelevant data classified as unmatched correctly by the model; FP/False Positive is the number of data that is irrelevant, but classified as a match by the model; FN/False Negative is the number

of data that is relevant, but not classified as a match by the model

3. Results and Discussions

The data processed in this study was collected from user reviews posted on various tourist attractions in West Sumatera over the past six years. The dataset, written in Indonesian, comprises unstructured data obtained through web crawling. The crawling process employed keywords corresponding to the names of tourist destinations, and the text reviews and ratings were extracted using the BeautifulSoup library. A total of 908 data entries were collected. For this dataset, a scoring system was applied: ratings of 1–2 were categorized as negative, 3 as neutral, and 4–5 as positive. These raw data were then prepared for subsequent stages to ensure they were clean, structured, and ready for analysis.

The dataset was divided into three proportions, with the best accuracy achieved using a 70:15:15 split. Training accuracy reached 97%, validation accuracy was 82%, and testing accuracy achieved 95%. In addition to accuracy metrics, a confusion matrix was utilized to evaluate the predictive performance of the IndoBERT model. This research provides valuable insights for developers working on programs to analyze sentiment in reviews of tourist destinations.

3.1 Analysis

The IndoBERT model was trained using a batch size of 32, for 3 epochs, with an optimizer learning rate of 0.00003. The fine-tuning process completed in 36 minutes. The results demonstrate the model's ability to classify sentiment accurately: Positive labels (correctly predicted 725, misclassified 25); Neutral labels (correctly predicted 76, misclassified 5); Negative labels (correctly predicted 58, misclassified 19).

This indicates that the percentage of correctly classified positive labels was 79.8%, negative labels were 6.38%, and the remaining labels constituted 13.96%.

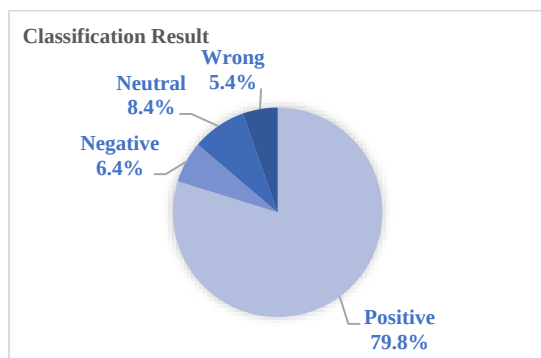


Figure 3. Diagram of classification results distribution

3.2 Optimizer Evaluation

The test evaluation results from IndoBERT for sentiment analysis using AdamW Optimizer on testing data by getting accuracy, precision, recall, and f-1 score values can be seen as follows. The results can be seen

in the table and confusion matrix which includes the actual and predicted values of the includes the actual value and predicted value of the following sentiment analysis.

Table 3. Comparison of AdamW result evaluation

	Precision	Recall	F-1
Positive	0.98	0.91	0.97
Negative	0.92	0.98	0.97
Accuracy : 0.97			

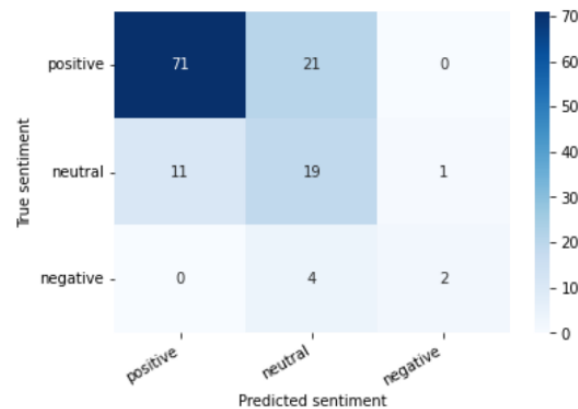


Figure 4. Confusion matrix of AdamW classification results

Table 4 is the test evaluation results from IndoBERT for sentiment analysis using AdaFactor Optimizer against testing data by get accuracy, precision, recall, and f-1 score values. The results can be seen in Table 4 and confusion matrix shown in Figure 5.

Table 4. Comparison of AdaFactor Result Evaluation

	Precision	Recall	F-1
Positive	0.99	0.95	0.98
Negative	0.95	0.99	0.98
Accuracy : 0.98			

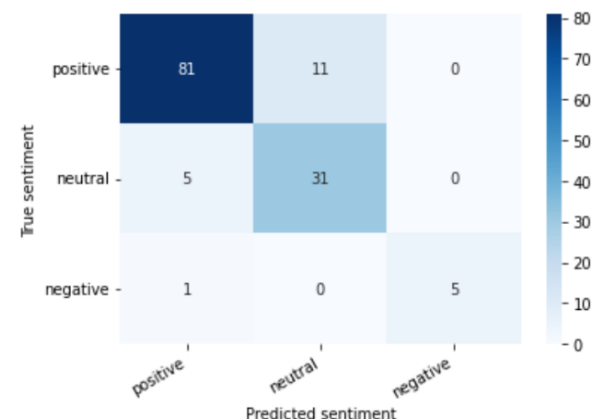


Figure 5. Confusion matrix of AdaFactor classification results

Compared to previous studies in the field of sentiment analysis for Indonesian-language tourism reviews, the method used in this research—specifically, the fine-tuning of the IndoBERT model using AdaFactor and AdamW optimizers—demonstrates superior performance. While earlier works often relied on conventional machine learning algorithms such as

Naïve Bayes, SVM, or basic deep learning models like LSTM and CNN, which typically yielded accuracy scores in the range of 70% to 85%, this study achieved a significantly higher accuracy of up to 98.2% using the IndoBERT model with the AdaFactor optimizer.

This notable improvement is attributed to the use of transformer-based architecture, which captures contextual and semantic nuances in Indonesian text more effectively. Additionally, the application of advanced optimization techniques further enhances the model's learning efficiency and generalization capability. These results indicate that the methodology proposed in this research not only outperforms prior approaches but also sets a new benchmark for sentiment classification tasks in the Indonesian tourism domain.

4. Conclusions

Based on the evaluation results obtained, the fine-tuning of the BERT model for sentiment analysis of tourist destination reviews in West Sumatera demonstrates that the IndoBERT model achieved an accuracy of 95% with a dataset split ratio of 70:15:15. In terms of overall prediction performance, the model produced a recall of 95%, indicating that IndoBERT effectively retrieves all relevant predictive information at a high level. Additionally, the precision accuracy reached 93%, signifying the model's strong ability to predict sentiments accurately. From the overall results of the IndoBERT modeling, it can be concluded that IndoBERT is highly capable of classifying and predicting sentiments with outstanding accuracy, as all accuracy metrics exceed 90%. This study demonstrates that the optimization techniques applied can accurately classify reviews of tourist attractions, categorizing sentiments as positive or negative. Moreover, the results from the AdamW optimizer achieved an accuracy of 97%, while the AdaFactor optimizer provided the best performance with an accuracy of 98.2%. This superior performance can be attributed to memory optimization and smoothing techniques implemented by AdaFactor. It is recommended that future studies compare their results with this study by employing different pre-training techniques and other deep learning algorithms. The dataset of West Sumatera tourism reviews for sentiment analysis tasks can also be enriched further. Future research could explore the performance of Natural Language Generation (NLG) models such as GPT, T5, and BART in similar tasks.

Although the proposed model in this study has achieved high accuracy in sentiment analysis of Indonesian-language tourism reviews, there are several directions that future research can explore. First, incorporating multimodal data such as images or videos alongside text could provide a richer context for sentiment interpretation, especially in tourism-related content. Second, expanding the dataset to include reviews from more diverse regions and tourism platforms could improve the model's generalizability.

Additionally, future studies may consider implementing more advanced transformer architectures such as RoBERTa, DeBERTa, or multilingual models like XLM-RoBERTa to compare performance across different pre-trained models. The integration of explainable AI (XAI) techniques is also recommended to increase transparency in model predictions and enhance trust in automated sentiment analysis systems. Finally, deploying the model into a real-time sentiment monitoring system could further validate its practical effectiveness and support decision-making in the tourism industry.

Acknowledgements

There are no conflicts of interest to disclose. This study was conducted independently and did not receive financial assistance from public, private, or non-profit sectors

References

- [1] K. Lei, C. Wen, and X. Wang, "RETRACTED: Research on the coordinated development of tourism economy based on embedded dynamic data," *Microprocess. Microsyst.*, vol. 82, p. 103933, 2021, doi: <https://doi.org/10.1016/j.micpro.2021.103933>.
- [2] G. E. Bayram, "Impact of Information Technology on Tourism," in *The Emerald Handbook of ICT in Tourism and Hospitality*, A. Hassan and A. Sharma, Eds., Emerald Publishing Limited, 2020, pp. 243–257. doi: 10.1108/978-1-83982-688-720201015.
- [3] J. Ahn and D. Im, "Automatic Recommendation of Nearby Tourist Attractions related to Events," vol. 21, no. 3, pp. 407–413, 2020, doi: <https://doi.org/10.5762/KAIS.2020.21.3.407>.
- [4] Y. Luo, X. Zhang, Y. Qin, Z. Yang, and Y. Liang, "Tourism Attraction Selection with Sentiment Analysis of Online Reviews Based on Probabilistic Linguistic Term Sets and the IDOCRIW-COCOSO Model," *Int. J. Fuzzy Syst.*, vol. 23, no. 1, pp. 295–308, 2021, doi: 10.1007/s40815-020-00969-9.
- [5] D. G. Sainger, "Sentiment Analysis - An Assessment of Online Public Opinion: A Conceptual Review," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 5, pp. 1881–1887, 2021, doi: 10.17762/turcomat.v12i5.2266.
- [6] M. Qian, C. Newton, and D. Qian, "Cultural Understanding Using In-context Learning and Masked Language Modeling BT - HCI International 2021 - Late Breaking Papers: Multimodality, eXtended Reality, and Artificial Intelligence," C. Stephanidis, M. Kurosu, J. Y. C. Chen, G. Fragomeni, N. Streitz, S. Konomi, H. Degen, and S. Ntoa, Eds., Cham: Springer International Publishing, 2021, pp. 500–508.
- [7] L. Huang, W. Chen, and H. Qu, "Accelerating Transformer for Neural Machine Translation," *ACM Int. Conf. Proceeding Ser.*, pp. 191–197, 2021, doi: 10.1145/3457682.3457711.
- [8] N. Q. K. Le, Q. T. Ho, T. T. D. Nguyen, and Y. Y. Ou, "A transformer architecture based on BERT and 2D convolutional neural network to identify DNA enhancers from sequence information," *Brief. Bioinform.*, vol. 22, no. 5, pp. 1–7, 2021, doi: 10.1093/bib/bbab005.
- [9] R. K. Kaliyar, "A Multi-layer Bidirectional Transformer Encoder for Pre-trained Word Embedding: A Survey of BERT," in *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 2020, pp. 336–340. doi: 10.1109/Confluence47617.2020.9058044.
- [10] Y. Higuchi, B. Yan, S. Arora, T. Ogawa, T. Kobayashi, and S. Watanabe, "BERT Meets CTC: New Formulation of End-to-End Speech Recognition with Pre-trained Masked Language Model," *Find. Assoc. Comput. Linguist. EMNLP 2022*, pp. 5515–5532, 2022, doi: 10.18653/v1/2022.findings-emnlp.402.
- [11] B. Li, H. Zhou, J. He, M. Wang, Y. Yang, and L. Li, "On the sentence embeddings from pre-trained language models,"

- EMNLP 2020 - 2020 Conf. Empir. Methods Nat. Lang. Process. Proc. Conf., pp. 9119–9130, 2020, doi: 10.18653/v1/2020.emnlp-main.733.
- [12] A. Chowanda and Y. Muliono, "Indonesian Sentiment Analysis Model from Social Media by Stacking BERT and Bi-LSTM," in *2022 3rd International Conference on Artificial Intelligence and Data Sciences (AiDAS)*, 2022, pp. 278–282. doi: 10.1109/AiDAS56890.2022.9918717.
- [13] F. Koto, J. H. Lau, and T. Baldwin, "INDOBERTWEET: A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization," *EMNLP 2021 - 2021 Conf. Empir. Methods Nat. Lang. Process. Proc.*, pp. 10660–10668, 2021, doi: 10.18653/v1/2021.emnlp-main.833.
- [14] I. Ramli, A. A. Krisnadhi, and R. E. Prasoj, "IndoKEPLER, IndoWiki, and IndoLAMA: A Knowledge-enhanced Language Model, Dataset, and Benchmark for the Indonesian Language," in *2022 7th International Workshop on Big Data and Information Security (IWBIIS)*, 2022, pp. 19–26. doi: 10.1109/IWBIIS56557.2022.9924844.
- [15] J. He, A. Wumaier, Z. Kadeer, W. Sun, X. Xin, and L. Zheng, "A Local and Global Context Focus Multilingual Learning Model for Aspect-Based Sentiment Analysis," *IEEE Access*, vol. 10, pp. 84135–84146, 2022, doi: 10.1109/ACCESS.2022.3197218.
- [16] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," *COLING 2020 - 28th Int. Conf. Comput. Linguist. Proc. Conf.*, pp. 757–770, 2020, doi: 10.18653/v1/2020.coling-main.66.
- [17] G. Pérez-Tapia, F. Almeida-García, and P. Mercadé-Melé, "The 'Four Core Elements' as a Measuring Instrument: From Simplicity to Complexity in Tourist Destination," *Economies*, vol. 9, no. 2, 2021. doi: 10.3390/economies9020053.
- [18] M. Dini, I. Curina, B. Francioni, S. Hegner, and M. Cioppi, "Tourists' satisfaction and sense of belonging in adopting responsible behaviors: the role of on-site and social media involvement in cultural tourism," *TQM J.*, vol. 35, no. 9, pp. 388–410, 2023, doi: 10.1108/TQM-03-2023-0085.
- [19] Z.-A. Torabi, A. A. Shalbafian, Z. Allam, Z. Ghaderi, B. Murgante, and A. R. Khavarian-Garmsir, "Enhancing Memorable Experiences, Tourist Satisfaction, and Revisit Intention through Smart Tourism Technologies," *Sustainability*, vol. 14, no. 5, 2022. doi: 10.3390/su14052721.
- [20] A. Kamal, S. B. Sura, L. P. Hung, R. Astri, and T. Kurniati, "Empowerment of MSME crafts recommendation system with content-based filtering method on social commerce after the pandemic," *E3S Web Conf.*, vol. 464, 2023, doi: 10.1051/e3sconf/202346412001.
- [21] R. Astri, L. P. Hung, S. B. Sura, A. Kamal, and R. Yuliet, "Sentiment analysis using naive bayes for reviews of visitors to Padang City beach tourism after the COVID-19 pandemic," *E3S Web Conf.*, vol. 464, 2023, doi: 10.1051/e3sconf/202346406002.
- [22] S. Cahyawijaya et al., "IndoNLG: Benchmark and Resources for Evaluating Indonesian Natural Language Generation," *EMNLP 2021 - 2021 Conf. Empir. Methods Nat. Lang. Process. Proc.*, pp. 8875–8898, 2021, doi: 10.18653/v1/2021.emnlp-main.699.
- [23] A. Kathuria, A. Gupta, and R. K. Singla, "A Review of Tools and Techniques for Preprocessing of Textual Data BT - Computational Methods and Data Engineering," V. Singh, V. K. Asari, S. Kumar, and R. B. Patel, Eds., Singapore: Springer Singapore, 2021, pp. 407–422.
- [24] Rianto, A. B. Mutiara, E. P. Wibowo, and P. I. Santosa, "Improving the accuracy of text classification using stemming method, a case of non-formal Indonesian conversation," *J. Big Data*, vol. 8, no. 1, p. 26, 2021, doi: 10.1186/s40537-021-00413-1.
- [25] L. A. Mullen, K. Benoit, O. Keyes, D. Selivanov, and J. Arnold, "Fast, Consistent Tokenization of Natural Language Text," *J. Open Source Softw.*, vol. 3, no. 23, p. 655, 2018, doi: 10.21105/joss.00655.
- [26] S. Sarica and J. Luo, "Stopwords in technical language processing," *PLoS One*, vol. 16, no. 8 August, pp. 1–13, 2021, doi: 10.1371/journal.pone.0254937.
- [27] M. Siino, I. Tinnirello, and M. La Cascia, "Is text preprocessing still worth the time? A comparative survey on the influence of popular preprocessing methods on Transformers and traditional classifiers," *Inf. Syst.*, vol. 121, p. 102342, 2024, doi: <https://doi.org/10.1016/j.is.2023.102342>.
- [28] M. Nesca, A. Katz, C. K. Leung, and L. M. Lix, "A scoping review of preprocessing methods for unstructured text data to assess data quality," *Int. J. Popul. Data Sci.*, vol. 7, no. 1, pp. 1–15, 2022, doi: 10.23889/ijpds.v7i1.1757.
- [29] F. Alrasheedi, X. Zhong, and P. C. Huang, "Padding Module: Learning the Padding in Deep Neural Networks," *IEEE Access*, vol. 11, pp. 7348–7357, 2023, doi: 10.1109/ACCESS.2023.3238315.
- [30] M. Xiao et al., "Addressing Overfitting Problem in Deep Learning-Based Solutions for Next Generation Data-Driven Networks," *Wirel. Commun. Mob. Comput.*, vol. 2021, 2021, doi: 10.1155/2021/8493795.
- [31] L. Sartran, S. Barrett, A. Kuncoro, M. Stanojević, P. Blunsom, and C. Dyer, "Transformer Grammars: Augmenting Transformer Language Models with Syntactic Inductive Biases at Scale," *Trans. Assoc. Comput. Linguist.*, vol. 10, pp. 1423–1439, Dec. 2022, doi: 10.1162/tacl_a_00526.
- [32] E. Radiya-Dixit, "How fine can fine-tuning be? Learning efficient language models," *Proc. Mach. Learn. Res.*, vol. 108, pp. 2435–2443, 2020.
- [33] Z. Fu, H. Yang, A. M. C. So, W. Lam, L. Bing, and N. Collier, "On the Effectiveness of Parameter-Efficient Fine-Tuning," *Proc. 37th AAAI Conf. Artif. Intell. AAAI 2023*, vol. 37, pp. 12799–12807, 2023, doi: 10.1609/aaai.v37i11.26505.
- [34] Q. Wan, C. Wan, K. Xiao, H. Xiong, D. Liu, and X. Liu, "Token-Event-Role Structure-based Multi-Channel Document-Level Event Extraction," vol. 1, no. 1, 2023, [Online]. Available: <http://arxiv.org/abs/2306.17733>
- [35] A. S. Talaat, "Sentiment analysis classification system using hybrid BERT models," *J. Big Data*, vol. 10, no. 1, 2023, doi: 10.1186/s40537-023-00781-w.
- [36] H. Zhong et al., "Adam revisited: a weighted past gradients perspective," *Front. Comput. Sci.*, vol. 14, Oct. 2020, doi: 10.1007/s11704-019-8457-x.
- [37] P. Bianchi, "Convergence and Dynamical Behavior of the ADAM Algorithm for Nonconvex Stochastic Optimization," *SIAM J. Optim.*, vol. 31, pp. 244–274, Jan. 2021, doi: 10.1137/19M1263443.