Jurnal Teknik Informatika (JUTIF)
Vol. 5, No. 4, October 2024, pp. 1385-1391
DOI: https://doi.org/10.52436/1.jutif.2024.5.5.2675
p-ISSN: 2723-3863
e-ISSN: 2723-3871

SENTIMENT ANALYSIS FOR E-COMMERCE PRODUCT REVIEWS BASED ON FEATURE FUSION AND BIDIRECTIONAL LONG SHORT-TERM MEMORY

Habibullah Akbar[*1], Diah Aryani[2], Marwan Kadhim Mohammed Al-shammari[3], M. Bahrul Ulum[4]

[1]Master of Computer Science, Faculty of Computer Science, Universitas Esa Unggul, Indonesia
[2,4]Informatics Engineering, Faculty of Computer Science, Universitas Esa Unggul, Indonesia
[3]University of Baghdad, Computer Center, Baghdad, Iraq
Email: [1]habibullah.akbar@esaunggul.ac.id, [2]diah.aryani@esaunggul.ac.id,
[3]marwan.kazem@cc.uobaghdad.edu.iq, [4]m.bahrul_ulum@esaunggul.ac.id

(Article received: August 17, 2024; Revision: September 9, 2024; published: October 25, 2024)

Abstract

E-commerce platforms would benefit from performing sentiment analysis of their customer's feedback. However, the vast amount of transaction data makes manual sentiment analysis of product reviews impractical. This research proposes an approach to automatically classify the sentiment of a given product review based on three major steps: data preprocessing, text representation, and classification model development. First, review data is cleaned to remove ambiguity and non-meaningful elements. Second, Word2Vec and GloVe features are combined to represent the words in a more unified vector space. Lastly, these combined features are classified to determine sentiment polarity using the Bidirectional Long Short-Term Memory Network (BiLSTM) model. The test results demonstrate that the proposed BiLSTM model achieves 91% uniform performance for all four metrics (accuracy, precision, recall, and F1-score), which is 3% higher than the results achieved by the standard LSTM model. Moreover, the BiLSTM model requires 9.91 seconds less training computation time than the LSTM.

Keywords: BiLSTM, e-commerce, GloVe, sentiment analysis, Tokopedia, Word2Vec

1. INTRODUCTION

Sentiment can be used to assess attitudes, opinions, or emotions expressed in sentences or gestures [1]. Sentiment can be used to see customer perceptions on social media, product reviews, and online forums. E-commerce platforms would benefit from doing sentiment analysis to enhance their customer experience. Sentiment analysis has gained widespread popularity in both academic research and the business world, largely due to the big data generated by Internet users [2].

One example of a giant e-commerce platform in Indonesia is Tokopedia (born in 2009). According to [3], in September 2023 the Tokopedia web page received 88.9 million visits, a decrease of 31% compared to the beginning of 2023. Regardless of declining market share, The Business Times reported that in March 2024 Tokopedia was acquired by an e-commerce platform from China (TikTok) [4]. The acquisition value reached 75.01% of Tokopedia's shares, about US$1.8 billion.

Why did such a significant drop occur? Besides various factors such as competition, technical problems, and national economic conditions, the user experience and sentiment may be the culprit due to the fierce competition. Unfortunately, the vast amount of transaction data makes manual sentiment analysis of product reviews impractical. Alternatively, sentiment analysis can be performed automatically [5] using Natural Language Processing (NLP) [6].

Challenges in developing sentiment analysis models include application domain dependencies, the large number of vocabularies in a natural language, abbreviations of words or abbreviations in terminology, figures of speech, and the number of predicted classes. [7]. Apart from that, sentiment also correlates with other psychological variables such as emotions [8], type of language (e.g. Arabic, Chinese, French, and Italian) [9], the type of data used (audio, video, image, or physiological signals of the body) [10], [11], and [12].

Research conducted by [13] represent a product review as a TF-IDF feature and then the Naive Bayes model to predict sentiment from Tokopedia app review data collected from the Google Play Store. Even though we have used the 10-fold cross-validation method, the accuracy obtained is less than satisfactory, namely 76%. However, the TF-IDF feature only calculates word frequency and the presence of words in the text without considering the semantic meaning of words or the relationships between words [14]. Although TF-IDF is easier to implement, the order of words contained in a sentence is not supposed to be ignored [15].

Similar studies were carried out by [16], but using Random Forest to predict the sentiment orientation with much higher accuracy (97.05%) based on data collected directly from Tokopedia's backend. However, there is a major concern due to the labeling. The sentiment labels given to the dataset are determined based on the lexicon and not carried out by experts. Indeed, the presence of positive words does not necessarily mean that the user's review is positive because human language styles can include implied meanings such as figures of speech [17]. The problem of the dataset also occurs in [18]. They use the K-Nearest Neighbor (KNN) to automatically assign labels to Tokopedia user review data collected from the Google Play and App Store (Apple).

Recently, [19] has used deep learning to perform sentiment analysis on consumer product review data on Tokopedia. Review data labeling is determined based on the rating value that users give to the product. There are two models studied (CNN and LSTM). The CNN model achieved 91% accuracy and 85% F-score, while the LSTM model achieved 94% accuracy and 89% F-score. The hyperparameters used in their experiment were learning rate = 0.01, epoch = 25, and dropout = 0.2. A comparison to Naive Bayes, Support Vector Machine (SVM), and Logistic Regression (LR) models produced accuracy and F-scores of 87% and 88%, 90% and 89%, and 87% and 86%, respectively.

Throughout 2023, the LSTM method has also been used for fake news classification [20], political election [21], rules and regulation [22], universities [23], stock prediction [24], and classification of public opinion about electric cars in Indonesia obtained from YouTube video comments [25].

A review of the use of deep learning methods for sentiment analysis can be read at [26]. There were 105 articles indexed by Scopus that were reviewed and it was found that the popular deep learning method (13 articles) used for sentiment analysis was LSTM. Based on the review, the average accuracy of the LSTM method in this article reached 89%. Apart from LSTM, other models that were also used in previous research were Convolutional Neural Network (CNN), Bidirectional Encoder Representations from Transformers (BERT), and Gated Recurrent Unit (GRU). Unfortunately, the review does not show the application domain of sentiment analysis.

From several previous studies, we addressed three gaps. First, the TF-IDF feature has become very popular. It is unlikely to find research such as in [27] that experiment with the GloVe feature for sentiment analysis. Second, sentiment analysis research has focused on traditional machine learning such as Naive Bayes, SVM, KNN, Logistic Regression, and Random Forest. Third, performance evaluation in previous studies ignored the computational time and instead weighted more on evaluating classification effectiveness such as accuracy and F-score [19].

To answer the gap above, this research highlighted three main contributions. First, we propose a unified vector representation by combining Word2Vec and GloVe. Second, we use deep learning i.e. BiLSTM network model to classify sentiment polarity. Third, we evaluate the proposed approach based on effectiveness (accuracy, precision, recall, F1-score) and efficiency (computation time). This research uses a dataset labeled by a clinical psychologists [28]. The proposed approach achieved better performance compared to the standard LSTM, and also better precision than the benchmarked method [19].

2. RESEARCH METHOD

The block diagram of the proposed method can be seen in Figure 1. The three main stages of the sentiment analysis approach are data preprocessing, text representation (word embedding), and model development (training and evaluation).

4.1. Dataset

This research uses the PRDECT-ID dataset consisting of product reviews collected from Tokopedia, one of the largest online platforms in Indonesia. This dataset can be accessed at [28].

2.1. Data Preprocessing

This data cleansing involves two processes. First, characters in the text are converted to lowercase using the lower() function. This step aims to avoid differences between upper and lower case letters which could affect the analysis results. For example, the words "Goods" and "goods" will be considered the same after conversion to lowercase. Second, non-alphabetic characters such as numbers and punctuation marks are removed from the text. Removal of these characters helps in reducing noise in product reviews.

2.2. Text Representation

Cleaned data is then converted into a special form of representation. Instead of TF-IDF, we use Word2Vec and GloVe representations. Word2Vec is a vector space model developed by Google converting words into vectors with fixed dimensions [29]. The authors created Word2vec as a word embedding algorithm to encode every word in text into a vector. This algorithm has also been widely used in NLP research.

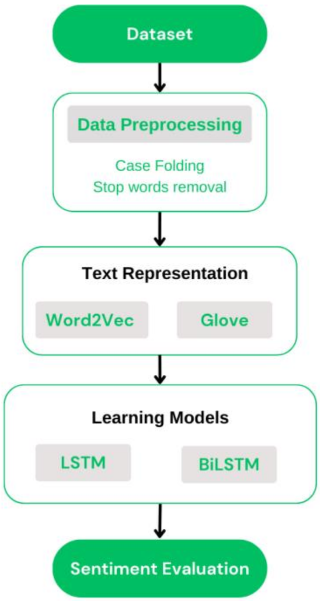

Figure 1. Stages of Proposed Sentiment Analysis Approach

Word2vec uses a context-based approach, generated based on the surrounding words. There are two main variants of Word2Vec: Continuous Bag of Words (CBOW) and Skip-gram. CBOW predicts target words based on context words, while Skip-gram predicts context words based on target words. In this research, Word2Vec (CBOW) captures the semantic meaning of words in reviews, providing vectors that allow the model to understand the relationship between words and meaning in the context of product reviews.

On the other hand, GloVe (Global Vectors for Word Representation) is a matrix-based word learning method developed by Stanford [30]. GloVe leverages global statistics from a text corpus that calculates the frequency of occurrence of words in their context across documents. The word vectors generated by GloVe reflect the probability of those words appearing together in a broader context. This method allows the creation of text representations based on global information and provides vectors that can be used to compare semantic similarities between words. These two representations will be combined into a single representation which will then be used for classification models.

2.3. Model Development

The models used for sentiment analysis of Tokopedia product reviews are the LSTM and BiLSTM models. These models can capture temporal dynamics and context in text data. These models are designed to handle data sequences and retain important information from long contexts, which is especially relevant in sentiment analysis.

More specifically, LSTM has a structure that allows the model to store information over long periods using memory units called state cells. This unit consists of three main gates i.e. input gate, forget gate, and output gate. These gates control the flow of incoming, deleting, and expelling information from the cell state, allowing LSTM to capture long-term relationships in text data.

Meanwhile, BiLSTM is an extension of LSTM which processes text input data in two directions: from left to right and vice versa.

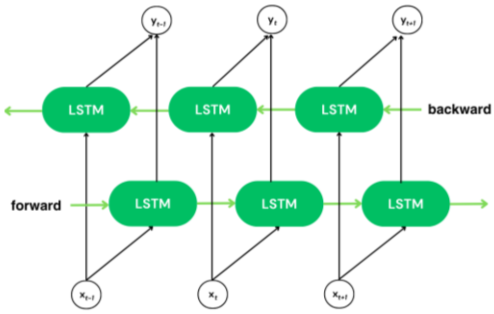

Figure 2. BiLSTM architecture

By using two LSTM layers functioning in parallel, BiLSTM can capture information from both directions in a text sequence, providing a better understanding of the context and meaning of words in the text. The BiLSTM architecture used in the research is illustrated in Figure 2.

2.4. Metric Evaluation

Evaluation is performed by splitting the dataset into training data and test data. The hyperparameters used to train the LSTM model can be seen in Table 1. Experiments were performed using the Python programming language and the Google Colab editor running on the T4 Graphics Processing Unit (GPU) runtime. We use several values for learning rate, epoch, and dropout to find more optimal results.

Table 1. Values used to Train the Classifiers

Hyperparameters	LSTM
input_size	100
embedding_size	128
dropout_rate	[0.1, 0.2, 0.3, 0.4, 0.5]
learning_rate (lr)	[0.001, 0.01, 0.1]
epoch	[5. 10, 15, 20]
optimizer	Adam

Model performance is based on metrics derived from the confusion matrix (see Table 2).

Table 2. Confusion Matrix

Clinical Expert Sentiment Labels	Sentiment Prediction	
	Positive	Negative
Positive	TP	FN
Negative	FP	TN

True Positive (TP) shows that the prediction result is positive according to the label given by the expert and True *Negative* (TN) shows that the predicted sentiment is negative according to the label. Meanwhile, FP and FN respectively show positive prediction error (where the label is negative) and negative prediction error (where the label is positive). From this matrix, the metrics of accuracy, precision, recall, and F1-Score. The metric values will be averaged from the prediction results in each positive and negative class (weighted average).

3. RESULTS

The PRDECT-ID dataset consists of 5400 rows from 29 different product categories. Each row of data has a product review column and has been given a sentiment and emotion label by a clinical psychologist. Of the 5400 reviews, there was a positive sentiment of 47.8% (2579 lines) and a negative sentiment of 52.2% (2821 lines). In addition to reviews, this dataset also contains columns for location, price, overall rating, number of items sold, total reviews, and customer ratings. The results of the data-cleaning process are shown in Table 3.

Table 3. Data Cleaning Results

Products	Review Before Cleaning	Review After Cleaning
Wireless Keyboard i8 Mini TouchPad Mouse 2.4G ...	Alhamdulillah berfungsi dengan baik. Packaging...	alhamdulillah berfungsi dengan baik packaging...
LG brand LG LCD LED TV Monitor Charger Adapter...	Barang Bagus, pengemasan Aman, dapat Berfungsi...	barang bagus pengemasan aman dapat berfungsi...
Ultrasonic Aroma Diffuser Humidifier Colorful LED 300ml+remote	Mungil tapi bekerja dng baik. Dan murahh terjangkau.. pas dng kebutuhan	mungil tapi bekerja dng baik dan murahh terjangkau pas dng kebutuhan
TDS Meter 3 Aquarium Water Hydroponic Measuring Instrument ppm Pool Nutrients TDS-3 - TDS 3 Meter	Produk sesuai deskripsi, packing aman terlindung, pengiriman cepat, toko/ penjual responsif, alat sudah dicoba dan berfungsi, sangat direkomendasikan.	produk sesuai deskripsi packing aman terlindung pengiriman cepat toko penjual responsif alat sudah dicoba dan berfungsi sangat direkomendasikan
AFL Bidirectional HDMI Switcher 1-In 2-Out & 2...	Works fine. Respon seller cepat, barang...	works fine respon seller cepat barang...

Conversion to lowercase and removal of non-alphabetic characters reduces noise in the data, while removal of stop words filters out elements that are not relevant for sentiment analysis. With cleaned and processed text, sentiment analysis models can operate with more consistent and representative data, increasing accuracy and effectiveness in understanding customer sentiment.

The results of the learning rate experiment are given in Table 4. In the first experiment, the epoch value used was 5 and the dropout value used was 0.5. The learning rate results of 0.001 and 0.01 are not significantly different. However, a learning rate of 0.001 requires a slightly shorter training time than a value of 0.01. The value 0.1 gives the lowest accuracy, precision, recall, and F1-score results even though the training time is more efficient. Note that the precision, recall, and F1-score results are calculated based on the macro average that takes the arithmetic mean across all classes.

Table 4. LSTM Performance Based on Learning Rate

Hyperparameters (average)	*lr(epoch=5+dropout=0.5)*		
	0.001	0.01	0.1
accuracy	0.88	0.88	0.63
precision	0.88	0.88	0.67
recalls	0.88	0.88	0.63
F1-score	0.88	0.88	0.61
time (second)	**21.05**	21.33	17.05

Table 5. LSTM Performance Based on Epoch

Hyperparameters (average)	*epoch (lr = 0.001 + dropout = 0.5)*		
	10	15	20
accuracy	0.88	0.88	0.88
precision	0.88	0.88	0.88
recalls	0.88	0.88	0.88
F1-score	0.88	0.88	0.88
time (second)	33.02	45.65	60.07

Table 6. LSTM Performance Based on Dropout

Hyperparameters (average)	*dropout (lr = 0.001 + epoch = 5)*			
	0.1	0.2	0.3	0.4
accuracy	0.85	0.86	0.87	0.87
precision	0.86	0.88	0.88	0.88
recalls	0.85	0.86	0.87	0.87
F1-score	0.85	0.86	0.87	0.87
time (second)	15.98	21.09	18.50	16.39

The best value from the learning rate (0.001) was then used in the epoch experiment. Table 5 shows no difference in performance, which means that the LSTM model is already optimal at epoch 5. Adding more epochs does not have a better effect and leads to longer training time.

Table 6 shows that in dropout there are no positive changes when the dropout value is reduced, which means the optimal value is 50%. Apart from that, we have also tried a dropout value of 60% and the results are the same as a 50% dropout with a slight increase in computing time. From these experiments, we found that the optimal hyperparameter values for learning rate is 0.001, 5 epoch, and 50% dropout.

After getting the optimal parameters for the learning rate, epoch, and dropout hyperparameters, we continue experiments on standard BiLSTM models and those using Word2Vec and GloVe representations. The embedding size used in Word2Vec and GloVe to train BiLSTM is 100. This value is equal to the training for the LSTM model in the previous experiment.

In the BiLSTM model experiment with feature fusion (Word2Vec and GloVe), we converted review data previously written in the Indonesian language into the English language using Python's deep translator library (Google Translator). This process is compulsory as the GloVe features were trained using an English corpus (see Table 7). The translation results are quite interesting. The sentence" alhamdulillah berfungsi dengan baik packaging..." is translated into " thank god works well safe packaging...". This shows that the translation is satisfactory even for Arab-Latin texts.

The main results are shown in Table 8. It shows that BiLSTM+Word2Vec+GloVe produced the best performance, while there are no differences between the performance of standard LSTM and BiLSTM. BiLSTM requires a longer training time than LSTM. This is caused by BiLSTM which has a more complex neural architecture because it processes sequence data in two directions: from left to right and from right to left.

Table 7. Review Data in English

Products	Review Before Translation (Indonesian)	Review After Translated (English)
Wireless Keyboard i8 Mini TouchPad Mouse 2.4G ...	alhamdulillah berfungsi dengan baik packaging...	thank god works well safe packaging...
LG brand LG LCD LED TV Monitor Charger Adapter...	barang bagus pengemasan aman dapat berfungsi...	good item safe packing works well..
Ultrasonic Aroma Diffuser Humidifier Colorful LED 300ml+remote	Mungil tapi bekerja dng baik. Dan murahh terjangkau.. pas dng kebutuhan...	small works well cheap affordable right needs...
TDS Meter 3 Aquarium Water Hydroponic Measuring Instrument ppm Pool Nutrients TDS-3 - TDS 3 Meter	Produk sesuai deskripsi, packing aman terlindung, pengiriman cepat, toko/ penjual responsif, alat sudah dicoba dan berfungsi, sangat direkomendasikan	product described safe secure packaging fast delivery responsive shopseller tool tested works highly recommended
AFL Bidirectional HDMI Switcher 1-In 2-Out & 2...	works fine respon seller cepat barang...	works fine seller responds quickly item...

4. DISCUSSION

An interesting result is provided by the performance of the BiLSTM+Word2Vec model. Empirically, it was found that the performance of the BiLSTM+Word2Vec model was 10% below the standard BiLSTM model. This indicates that solely Word2Vec representation was not able to improve LSTM while the GloVe feature produces similar performance as the standard BiLSTM model. This is quite surprising because the use of text representation did not succeed in improving model performance. However, combining the two text representations into BiLSTM, namely the BiLSTM+Word2Vec+GloVe model, has succeeded in increasing the average performance of the BiLSTM model by 3%.

Without the Word2Vec and GloVe fusion features, BiLSTM only achieved an average performance of 88%, whereas, with the fusion feature, it managed to achieve an average performance of 91%.

In detail, the precision, recall, and F1-score values in the negative class are 0.89, 0.94, and 0.92 respectively. For the positive class, the resulting values are 0.94%, 0.88%, and 0.90%. These results are comparable with the results of the research [19]. Our results have better precision but lower accuracy. The other limitation of [19] does not include the metrics of recall, F1-score, and computation time.

In theory, BiLSTM tends to require longer training time than the standard LSTM because its architecture has two directions so the number of BiLSTM parameters is also greater. The experimental results show that BiLSTM+Word2Vec+GloVe is more efficient (199.67 seconds) than BiLSTM (209.28 seconds). Additionally, we performed an epoch value of 10 and found that the average performance remained at 91% and the computing time increased by approximately two times. Combined features with 5 epochs and other experimental parameter values (learning rate 0.001 and dropout 0.5) achieve the best sentiment analysis prediction performance.

5. CONCLUSIONS

This research has explored the application of Word2Vec and GloVe into BiLSTM to perform sentiment classification on a public dataset (PRDECT-ID) containing product reviews on the Tokopedia platform. Experimental findings show that combining the two features can produce the highest precision, recall, and F1-score performance. The training computation time required is also better than the standard BiLSTM model which does not use the Word2Vec or GloVe features. This research also succeeded in testing the performance of LSTM-based models using more complete metrics. However, the accuracy performance of the resulting BiLSTM+Word2Vec+GloVe model is still below the LSTM+TF-IDF method. In future research, a more in-depth study needs to be carried out on the influence of various other features on LSTM-based models.

Table 8. Performance Comparison of Sentiment Analysis Approaches

Model	Accuracy	Precision	*Recall*	*F1-score*	*Time* (second)
Naive Bayes+TF-IDF[13]	0.76	-	-	-	-
CNN+TF-IDF[19]*	0.91	-	-	0.85	-
LSTM+TF-IDF[19]*	**0.94**	0.89	-	-	-
LSTM	0.88	0.88	0.88	0.88	21.05
BiLSTM	0.88	0.88	0.88	0.88	209.28
BiLSTM+Word2Vec	0.78	0.78	0.78	0.78	194.39
BiLSTM+GloVe	0.88	0.88	0.88	0.88	202.72
BiLSTM+Word2Vec+GloVe	0.91	**0.91**	**0.91**	**0.91**	**199.67**
BiLSTM+Word2Vec+GloVe (10 epochs)	0.91	0.91	0.91	0.91	406.81

*F-score used by [19] equivalent to the F1-score metric used in this study

DAFTAR PUSTAKA

[1] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, 2014, doi: 10.1016/j.asej.2014.04.011.

[2] L. J. Laki and Z. G. Yang, "Sentiment Analysis with Neural Models for Hungarian," *Acta Polytech. Hungarica*, vol. 20, no. 5, 2023, doi: 10.12700/APH.20.5.2023.5.8.

[3] A. Ahdiat, "Pengunjung Shopee Makin Banyak, Bagaimana E-Commerce Lain?," *databoks*, 2023. https://databoks.katadata.co.id/datapublish/2023/10/11/pengunjung-shopee-makin-banyak-bagaimana-e-commerce-lain

[4] J. Yordan, "Drop in Tokopedia's market share led to TikTok deal: GoTo chief," *The Business Times*, 2024. https://www.businesstimes.com.sg/startups-tech/startups/drop-tokopedia-s-market-share-led-tiktok-deal-goto-chief

[5] U. Singh, A. Saraswat, H. K. Azad, K. Abhishek, and S. Shitharth, "Towards improving e-commerce customer review analysis for sentiment detection," *Sci. Rep.*, vol. 12, no. 1, p. 21983, 2022, doi: 10.1038/s41598-022-26432-3.

[6] M. Hoang, O. A. Bihorac, and J. Rouces, "Aspect-based sentiment analysis using BERT," in *Proceedings of the 22nd nordic conference on computational linguistics*, 2019, pp. 187–196. [Online]. Available: https://hdl.handle.net/10062/93001

[7] D. M. E.-D. M. Hussein, "A survey on sentiment analysis challenges," *J. King Saud Univ. Sci.*, vol. 30, no. 4, pp. 330–338, 2018, doi: 10.1016/j.jksues.2016.04.002.

[8] A. Alslaity and R. Orji, "Machine learning techniques for emotion detection and sentiment analysis: current state, challenges, and future directions," *Behav. \& Inf. Technol.*, vol. 43, no. 1, pp. 139–164, 2024, doi: 10.1080/0144929X.2022.2156387.

[9] M. S. U. Miah, M. M. Kabir, T. Bin Sarwar, M. Safran, S. Alfarhood, and M. F. Mridha, "A multimodal approach to cross-lingual sentiment analysis with ensemble of transformer and LLM," *Sci. Rep.*, vol. 14, no. 1, p. 9603, 2024, doi: 10.1038/s41598-024-60210-7.

[10] H. Li, A. Guo, and Y. Li, "CCMA: CapsNet for audio--video sentiment analysis using cross-modal attention," *Vis. Comput.*, pp. 1–12, 2024, doi: 10.1007/s00371-024-03453-9.

[11] H. Yang *et al.*, "Large Language Models Meet Text-Centric Multimodal Sentiment Analysis: A Survey," *arXiv Prepr. arXiv2406.08068*, 2024, doi: 10.48550/arXiv.2406.08068.

[12] M. Bin Habib, M. F. Bin Hafiz, N. A. Khan, and S. Hossain, "Multimodal Sentiment Analysis using Deep Learning Fusion Techniques and Transformers.," *Int. J. Adv. Comput. Sci. \& Appl.*, vol. 15, no. 6, 2024, doi: 10.14569/IJACSA.2024.0150686.

[13] R. Meifitrah, I. Darmawan, and O. N. Pratiwi, "Sentiment analysis of tokopedia application review to service product recommender system using neural collaborative filtering for marketplace in Indonesia," in *IOP Conference Series: Materials Science and Engineering*, 2020, vol. 909, no. 1, p. 12071. doi: 10.1088/1757-899X/909/1/012071.

[14] S. Qaiser and R. Ali, "Text mining: use of TF-IDF to examine the relevance of words to documents," *Int. J. Comput. Appl.*, vol. 181, no. 1, pp. 25–29, 2018, doi: 10.5120/ijca2018917395.

[15] M. Taboada, "Sentiment analysis: An overview from linguistics," *Annu. Rev. Linguist.*, vol. 2, no. 1, pp. 325–347, 2016, doi: 10.1146/annurev-linguistics-011415-040518.

[16] B. Warsito, A. Prahutama, and others, "Sentiment analysis on tokopedia product online reviews using random forest method," in *E3S Web of Conferences*, 2020, vol. 202, p. 16006. doi: 10.1051/e3sconf/202020216006.

[17] R. Buyya, R. N. Calheiros, and A. V. Dastjerdi, *Big data: principles and*

paradigms. Morgan Kaufmann, 2016. doi: 10.1016/C2015-0-04136-3 Authors.

[18] R. A. E. V. T. Sapanji, D. Hamdani, P. Harahap, and others, "Sentiment Analysis of the Top 5 E-commerce Platforms in Indonesia using Text Mining and Natural Language Processing (NLP)," *J. Appl. Informatics Comput.*, vol. 7, no. 2, pp. 202–211, 2023, doi: 10.30871/jaic.v7i2.6517.

[19] D. Purnamasari, A. B. Aji, S. Madenda, I. M. Wiryana, and S. Harmanto, "Sentiment Analysis Methods for Customer Review of Indonesia E-Commerce," *Int. J. Innov. Comput. Inf. Control*, vol. 20, no. 1, pp. 47–60, 2024, doi: 10.24507/ijicic.20.01.47.

[20] A. Cardova and A. Hermawan, "Implementasi Metode LSTM Untuk Mengklasifikasi Berita Palsu Pada PolitiFact," *J. FASILKOM*, vol. 13, no. 3, pp. 471–479, 2023, doi: 10.37859/jf.v13i3.6175.

[21] R. D. Kurniawan and J. Muliawan, "Sentiment analysis of indonesian election 2024 using the k-nearest neighbor method," *J. Tek. Inform.*, vol. 5, no. 2, pp. 653–659, 2024.

[22] M. Wahiddianty, G. Firmasyah, H. Akbar, and others, "ANALISIS SENTIMEN OMNIBUS LAW PADA UMKM MENGGUNAKAN SUPPORT VECTOR MACHINE, NAIVE BAYES BERBASIS PARTICLE SWARM OPTIMIZATION," *J. Bisnis dan Manaj.*, vol. 3, no. 5, pp. 1086–1096, 2023.

[23] H. Akbar, "Analisis Sentimen Review Institusi Pendidikan Tinggi Dari Facebook Menggunakan Jaringan Saraf Tiruan," *J. Ilmu Komput.*, vol. 4, no. 1, pp. 19–30, 2019.

[24] M. F. Aslim, G. Firmansyah, B. Tjahjono, H. Akbar, and A. M. Widodo, "Utilization of LSTM (Long Short Term Memory) Based Sentiment Analysis for Stock Price Prediction," *Asian J. Soc. Humanit.*, vol. 1, no. 12, pp. 1241–1255, 2023.

[25] A. S. Widagdo, K. N. Qodri, F. E. N. Saputro, N. A. R. Putri, and others, "Analisis Sentimen Mobil Listrik di Indonesia Menggunakan Long-Short Term Memory (LSTM)," *J. FASILKOM*, vol. 13, no. 3, pp. 416–423, 2023, doi: 10.37859/jf.v13i3.6303.

[26] F. Fitroh and F. Hudaya, "Systematic Literature Review: Analisis Sentimen Berbasis Deep Learning," *J. Nas. Teknol. dan Sist. Inf.*, vol. 9, no. 2, pp. 132–140, 2023, doi: 10.25077/TEKNOSI.v9i2.2023.132-140.

[27] P. Demotte, K. Wijegunarathna, D. Meedeniya, and I. Perera, "Enhanced sentiment extraction architecture for social media content analysis using capsule networks," *Multimed. Tools Appl.*, vol. 82, no. 6, pp. 8665–8690, 2023, doi: 10.1007/s11042-021-11471-1.

[28] R. Sutoyo *et al.*, "Product Reviews Dataset for Emotions Classification Tasks - Indonesian (PRDECT-ID) Dataset," *Mendeley Data*, 2022. https://data.mendeley.com/datasets/574v66hf2v/1

[29] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv Prepr. arXiv1301.3781*, 2013, doi: 10.48550/arXiv.1301.3781.

[30] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543. doi: 10.3115/v1/D14-1162.