Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 7 No. 3 (2023) 586 - 591 ISSN Media Electronic: 2580-0760

Aspect Based Sentiment Analysis Marketplace Product Reviews Using BERT, LSTM, and CNN

Syaiful Imron[1], Esther Irawati Setiawan[2*], Joan Santoso[3], Mauridhi Hery Purnomo[4]
[1,2,3]Informatics Engineering Department, Faculty of Science and Technology, Institut Sains dan Teknologi Terpadu Surabaya
[4]Department of Electrical Engineering, Department of Computer Engineering, University Center of Excellence on Artificial Intelligence for Healthcare and Society (UCE AIHeS), Institut Teknologi Sepuluh Nopember
[1]syaiful_i20@mhs.stts.edu, [2]esther@stts.edu, [3]joan@stts.edu, [4]hery@ee.its.ac.id

Abstract

Bukalapak is one of the largest marketplaces in Indonesia. Reviews on Bukalapak are only in the form of text, images, videos, and stars without any special filters. Reading and analyzing manually makes it difficult for potential buyers. To help with this, we can extract this review by using aspect-based sentiment analysis because an entity cannot be represented by just one sentiment. Several previous research stated that using LSTM-CNN got better results than using LSTM or CNN. In addition, using BERT as word embedding gets better results than using word2vec or glove. For this reason, this study aims to classify aspect-based sentiment analysis from the Bukalapak marketplace with BERT as word embedding and using the LSTM-CNN method, where LSTM is for aspect extraction and CNN for sentiment extraction. Based on testing the LSTM-CNN method, it gets better results than LSTM or CNN. The LSTM-CNN model gets an accuracy of 93.91%. Unbalanced dataset distribution can affect model performance. With the increasing number of datasets used, the accuracy of a model will increase. Classification without using stemming on datasets can increase accuracy by 2.04%.

Keywords: *aspect-based sentiment analysis; BERT; LSTM; CNN*

1. Introduction

In Indonesia, marketplaces are growing rapidly, as evidenced by the many online buying and selling sites that have sprung up in Indonesia, such as Shopee, Tokopedia, Bukalapak and many others. Based on iPrice data, the Bukalapak marketplace in Q1 2022 had 23 million visitors. It can be concluded that Bukalapak still has many users. In addition, based on research [1] that the results of the analysis of marketplace trends on social media Twitter based on the highest percentage of positive reviews were achieved by Bukalapak with a score of 49.71%, Shopee second place with a value of 38.24%, while Tokopedia is in third place with a percentage of 28.84%.

Purchasing goods through the Bukalapak, buyers can provide reviews regarding the items purchased. Reviews can be text, images, videos, and stars. These reviews are quite useful for potential buyers and marketplace parties. With the information contained in the review, customers can obtain further information about products or services in the marketplace [2]. As for the marketplace, the review is an input to improve the quality of the marketplace [3].

Currently, most reviews on the marketplace are just text, images, videos, and stars without a specific filter from the marketplace. Reading and analyzing it manually makes it difficult for potential buyers. It is hoped that the marketplace will provide a filter that specifically addresses existing reviews. To help with this, a method is needed, one method that can be used is sentiment analysis [4].

Sentiment analysis is one of the topics in NLP. Sentiment analysis is also known as opinion mining, which is a process to determine whether the expression is positive, negative, or neutral [5]. There are several levels of granularity in sentiment analysis, but among all, aspect-based sentiment analysis (ABSA) is the most appropriate in this study. The reason is that the valuation of an entity cannot be represented by one sentiment. In addition, reviews can evaluate various aspect entities differently [6].

Aspect-based sentiment analysis performs the extraction of aspects and sentiments in a sentence so that the number of aspects in the sentence can be determined and the polarity of the sentiments of each aspect can be determined. For example, the review

Accepted: 08-12-2022 | Received in revised: 15-04-2023 | Published: 02-06-2023

“barang bagus, tapi pengiriman lama (good item, but long delivery)”, in the review the sentence “barang bagus (good item)” shows the context of “quality” which has a “positive”. While the sentence "pengiriman lama (long delivery)" shows the context of "delivery" which has a "negative".

There are quite a lot of uses of deep learning in aspect-based sentiment analysis research, especially using LSTM and CNN. In research [7] used LSTM for aspect extraction and CNN for sentiment extraction. The results of this study found that using LSTM and CNN obtained better results than using only LSTM, CNN, or SVM. The results of the study [8], [9] also state that using LSTM and CNN produces better results than using only LSTM, CNN, or KNN.

LSTM is not good at taking local features from a text, this affects the sentiment classification. In addition, LSTM feature extraction has a different effect on sentiment polarity [7]. For this reason, it is necessary to take advantage of the advantages of CNN in taking local features from a text that have an impact on sentiment polarity [10].

Word embeddings are mappings between words and vectors that can capture similarities between words [9]. Bidirectional Encoder Representations from Transformers (BERT), a neural network-based pre-training technique for Natural Language Processing (NLP) [11]. BERT has 2 frameworks with different functions, pre-training, and fine tuning. BERT can be used as word embeddings. Compared to Word2Vec or GloVe based which only provide a single context independent representation for each token, BERT takes a sentence as input and computes a token level representation using information from the entire sentence [12]. In research [13], [14] that the use of BERT as word embedding gets a better accuracy value than using word2vec or glove as word embedding.

Based on some of the previous research above, there is a lot of research on aspect-based sentiment analysis that uses the LSTM method for aspect extraction and CNN for sentiment extraction. However, there has been no research on aspect-based sentiment analysis using BERT as word embedding with the LSTM method for aspect extraction and CNN for sentiment extraction simultaneously. Therefore, this study aims to determine the results of combining BERT as word embedding, the LSTM method for aspect extraction and CNN for sentiment extraction.

2. Research Methods

This section describes architectures and methods for classifying aspect-based sentiments. At this stage there are 5 processes. Data collection, data labeling, preprocessing, ABSA model, testing and evaluation. A description of the process steps performed in this research is shown in Figure 1.

First, the process of collecting review data from the marketplace Bukalapak. Scrapping technique is used to retrieve review data. After the data collection process, it is followed by the process of labeling aspects and sentiment data. For labeling done manually one by one. The next process is preprocessing, which in this process consists of case folding, punctuation removal, word normalization, and stemming. The output in the preprocessing process consists of data that uses stemming and data that does not use stemming. This is because this research also wants to know the impact of stemming on the result. After preprocessing, the word embedding process is carried out. The purpose of word embedding is for data to be processed into machine learning algorithms. In this process, use BERT embedding. The next process is aspect classification using LSTM and sentiment classification using CNN. The last process is testing and evaluation. For evaluation using confusion matrix.

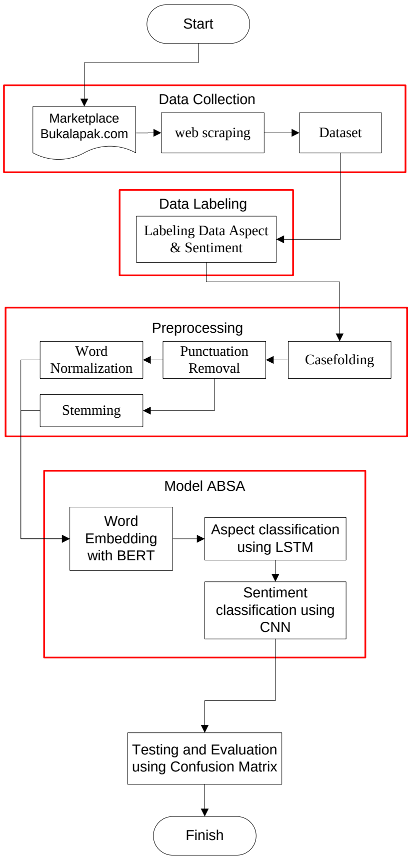
Start
Data Collection
Marketplace Bukalapak.com
web scraping
Dataset
Data Labeling
Labeling Data Aspect & Sentiment
Preprocessing
Word Normalization
Punctuation Removal
Casefolding
Stemming
Model ABSA
Word Embedding with BERT
Aspect classification using LSTM
Sentiment classification using CNN
Testing and Evaluation using Confusion Matrix
Finish

Figure 1. Research Flow

2.1 Dataset

The dataset was obtained from the Bukalapak review. The data captured is only in the form of review data in the form of descriptive text in Indonesian. The dataset is obtained by scrapping using google colab and will be saved in csv format. The results of scrapping which are still in the form of raw data, are then labeled.

The grouping of aspect categories follows the research [4]. Where there are six aspects, accuracy, quality, service, price, packaging, and delivery. Meanwhile, there are only 2 sentiments, positive and negative. The dataset obtained was 3,114 reviews where the distribution of the dataset can be seen in Table 1. The distribution of the reviews is uneven between aspects. The quality aspect has the most reviews. While the service aspect has the fewest reviews.

Table 1. Dataset Distribution

Aspect	Positive	Negative	Total
Quality	305	819	1,124
Delivery	452	140	592
Accuracy	313	191	504
Packaging	200	112	312
Price	272	30	302
Service	194	86	280
Total	1,736	1,378	3,114

This research is a multi-class classification, that is each review can only define one aspect and one sentiment. The labeling of review aspects and sentiments is done manually. The first review is labeled aspect. Furthermore, reviews that already have an aspect label are labeled sentiment. Examples of reviews that have been labeled aspects and sentiments can be seen in Table 2.

Table 2. Reviews Labeled

Reviews	Aspect	Sentiment
Namun akan lebih baik apabila pengiriman dipercepat	Delivery	Negative
Barang sesuai pictnya	Accuracy	Positive
Packingnya rapi dan bagus	Packaging	Positive
Cukup mahal dengan harga segitu	Price	Negative

2.2 Preprocessing

Before being processed further, the dataset is preprocessed. preprocessing is needed when creating models because preprocessing has quite an effect on model performance [15]. In this preprocessing stage, there are 4 processes, this is case folding, punctuation removal, word normalization, and stemming. Case folding is the process of changing all letters into lowercase or lowercase [16]. Punctuation removal is the process of removing characters other than letters, including punctuation marks. Word normalization aims to change abbreviated, misspelled, and non-raw words into corresponding words in the great Dictionary of Bahasa Indonesia (KBBI) [15]. This process is carried out by first creating a manually created slang word dictionary based on observations of the dataset used. Where there are 3,252 words in the vocabulary. Stemming is the process of removing affixes, prefixes, and endings from a word. in the end every word in the document contains only the base word (17). This research is in Indonesian, so the stemming in this research uses the Sastrawi library.

2.3 Word Embedding

Word embedding is a method used to create a vector by changing the word representation which allows words with the same meaning to have a similar representation. In this research for word embedding using BERT embedding. The pre-trained BERT model used is the indoBERT model (indobert-base-p2) obtained from the research [15], [17]. The indoBERT model is trained from the indo4B dataset.

In the BERT embedding stage, the first input sentence is tokenized with a tokenizer. Each input sentence is tokenized into a word or sub word and matched in the vocabulary. If there are words that are not in the vocabulary, then the word or sub-word is added with the ## symbol [16], [18]. Each sentence is assigned a special token, [CLS] at the beginning of the sentence and [SEP] at the end of the sentence. The [SEP] token is also used to separate one sentence and another. The input sentence is adjusted to a predetermined maximum length by padding it with a special token [PAD].

The next process is the input sentence is matched with a unique number or ID. The unique number or ID is obtained based on the word index in the vocabulary, because the vocabulary is arranged based on the level of occurrence. As in the indoBERT model (indobert-base-p2), the [CLS] token is at index 101 and is followed by the [SEP] token which is at index 102. Substitution of token to id is fixed. So, if there is a word like the word "barang", the index or id of the word is 1,1162.

In the indoBERT model (indobert-base-p2) there is an attention mask. Attention mask is a variable used to prevent the model from processing the padding as part of the input. The mask variable consists of values 0 and 1. Where a value of 0 means that the token does not need to be included in the model calculation. While the value of 1 means token that need to be calculated by the model. For details, see Figure 2.

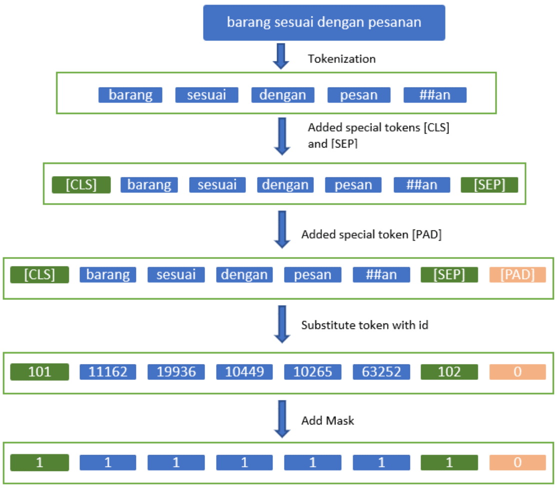

Figure 2. BERT Embedding

2.4 LSTM

The RNN method was developed into LSTM by adding LSTM cells in the RNN architecture [19], [17]. LSTM calculation takes longer than RNN, this is because LSTM stores weight. LSTM can work better than RNN in a long data sequence, because in LSTM cells there is a node that has self-recurrent [19], [17]. There are three important components in the LSTM process stage, forget gate, input gate, and output gate. Figure 3 is the LSTM architecture.

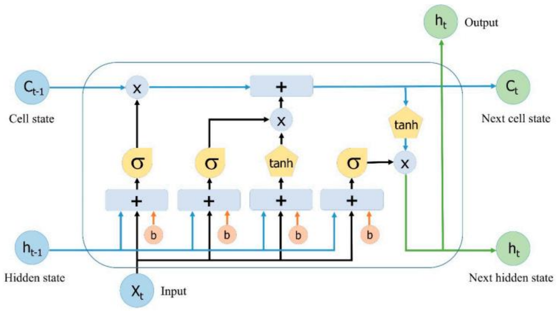

Figure 3. LSTM Architecture

Forget gate (ft) determines what information will be discarded and retained from Ct-1 using the sigmoid. Forget gate reads the values of ht-1 and Xt. Because it uses a sigmoid, the output t is “0” and “1”. If the value of t is "0" it will be forgotten, while the value of "1" will be preserved. The formula for the forget gate (ft) is written in formula 1.

$$f_t = \sigma\left(W_f . [h_{t-1}, x_t] + b_f\right). \quad (1)$$

The next process determines the new information that will be used in Ct. In this step there are 2 parts. First, the sigmoid gate or input gate (it) to determine which value to update, in formula 2. Second, tanh to generate a new candidate vector memory cell (C̃t), in formula 3. Furthermore, the two parts are combined to update the Ct.

$$i_t = \sigma(W_i . [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = tanh(W_c . [h_{t-1}, x_t] + b_c) \quad (3)$$

Next update the old cell state (Ct-1) to the new cell state (Ct). By multiplying the old cell state (Ct-1) by the forget gate (ft) to determine what information was forgotten. Then multiply the new candidate memory cell (C̃t) by the input gate (it), to determine how many Ct candidates are included. Then, the two are added together. For the formula to update the old cell state (Ct-1) to the new cell state (Ct) in formula 4.

$$c_t = (i_t \,. \tilde{C}_t + f_t \,. c_{t-1}) \quad (4)$$

Output based on cell state (C) value is passed to a filter. First, run the sigmoid gate or called the output gate (ot) in formula 5 to determine what parts of the cell state (C) are generated. Second, pass cell state (C) in formula 6 through tanh to produce a value between -1 and 1. Then multiply it by the sigmoid gate, resulting in the decided value.

$$o_t = \sigma(W_o . [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t \,. tanh(c_t) \quad (6)$$

2.5 CNN

Convolutional neural networks were initially used in the field of image processing or computer vision. In 2015, [20], [18] conducted research on natural language processing (NLP) using CNN to classify text. For text classification with CNN, use the convolution technique of sentences, paragraphs, or entire text documents. The Convolution technique works by dividing a matrix in the form of a text representation into a window or filter. The filters are then added together to become a new representation of the text which can be called feature maps. From each feature map, the largest value is taken using the maxpooling technique.

CNN architecture as in Figure 4 there are sentence representation, convolutional layer, max pooling, fully connected, dropout, softmax. Sentence representation is an input sentence in the form of a matrix, where this matrix is used as the input convolutional layer. To perform a convolutional layer using a filter that produces feature maps. The convolutional layer also determines the parameters used. The most important parameters are number of kernels and kernel size.
The results convolutional layer in the form of feature maps, then the pooling layer process is carried out. There are several methods for the pooling layer process, namely max pooling, average pooling, and others. In this research using max pooling, max pooling is taking the maximum value from a feature map. The result of max pooling is then carried out in a fully connected layer. To avoid overfitting, dropout is used. Optimization is used to reduce losses during training. There are many optimizers that can be used, Adam, Nadam, and SGD. The output from fully connected layer is then processed in the softmax layer. Softmax layer will provide output based on the greatest probability of a class.

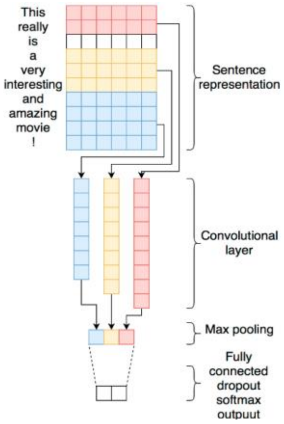

Figure 4. CNN Architecture

DOI: https://doi.org/10.29207/resti.v7i3.4751

2.5 Evaluation

Evaluation is a process to find out how well the resulting system performance. In this reasearch, the evaluation method used is the confusion matrix, where the confusion matrix will produce True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN), as shown in Figure 5. The results of confusion matrix used to calculate precision, recall, and F1 Score.

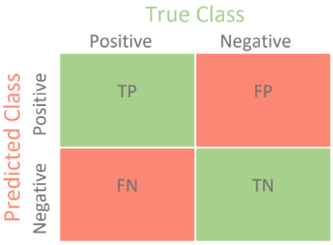

Figure 5. Confusion Matrix

Precision is how accurate the algorithm or model is to how many positive predictions are true positive (TP) from all positive predicted data (TP+FP). To calculate precision, use formula 7. Meanwhile, recall is how many positive predictions are true positive (TP) from all data that are originally positive (TP+FN) use formula 8. F1 Score is the balance value between precision and recall user formula 9.

$$Precision = \frac{TP}{TP+FP} \; x \; 100\% \qquad (7)$$

$$Recall = \frac{TP}{TP+FN} \; x \; 100\% \qquad (8)$$

$$F1\,Score = \frac{2 \; x \; Recall \; x \; Preceision}{Recall+Precision} \; x \; 100\% \qquad (9)$$

3. Results and Discussions

At this stage discusses the testing of the model created. This research uses a review dataset from the Bukalapak marketplace. The dataset used is 3,114 reviews, which are divided into 80% training data and 20% testing data. The first is to find the best model by testing several different parameters. Furthermore, trials with the division of the dataset. Lastly, testing the effect of stemming on the dataset on the final results.

In this study, 4 trials will be conducted with different parameter values. It's to find parameters that can optimize model performance. For parameter values can be seen in Table 3. For optimizers use adam.

Table 3. Parameter

Parameter	Value 1	Value 2	Value 3	Value 4
Filter Size	5	3	8	6
Feature Maps	300	200	500	90
Regularizer	0.1	0.01	0.1	0.001
Droup Out	0.5	0.5	0.5	0.5
Epoch	50	30	50	50
Batch Size	32	32	8	15
Learning Rate	1e-3	1e-3	1e-5	1e-5

3.1 Testing Parameter

After testing based on predetermined parameters, the first parameter gets the best results. Details of the results of the first test can be seen in Table 4. The results of the highest aspect of accuracy obtained aspects “accuracy” with an accuracy of 97,02%. While the lowest accuracy is obtained from the aspect “packaging” with an accuracy value of 87,3%. The aspect “packaging” for the F1-score is also the lowest with an F1-score of 86.55%. This is due to the unbalanced polarity distribution.

Table 4. Detail Test Result

Aspect	Accuracy	Precision	Recall	F1-Score
Quality	91,10%	90,45%	87,05%	88,55%
Delivery	96,72%	96,30%	93,45%	94,80%
Accuracy	97,02%	97,10%	96,70%	96,90%
Packaging	87,30%	87,80%	85,70%	86,55%
Price	96,72%	95,75%	97,66%	96,75%
Service	94,64%	93,25%	90,50%	91,50%
Average	93,91%	93,44%	91,84%	92,50%

3.2 Comparison with Other Models

At this stage, compares with several models LSTM, CNN, and LSTM-CNN for embedding use BERT. Based on the test, it was found that the LSTM-CNN model obtained the highest accuracy 93.91%. For comparison results can be seen in table 5.

Table 5. Comparation Model

Model	Accuracy	Precision	Recall	F1-Score
LSTM	88,60%	89,50%	89,40%	89,20%
CNN	90,04%	91,27%	90,80%	91,16%
LSTM-CNN	93,91%	93,44%	91,84%	92,50%

3.3 Testing Shared Dataset

At this stage a trial is carried out with variations in the amount of data used for training and testing data. The dataset is divided into 3 parts. The first trial used 1/3 random dataset, and the second trial used 2/3 dataset, and the third test used all data. The results of the comparison test of the three datasets are shown in Table 6. Based on these trials, it can be concluded that the more data used in the training and testing process, the better the performance of a model. This is due to the more datasets, the more sentence patterns in the dataset that can be learned by the model. So that it will add to the knowledge possessed by the model.

Table 6. Test Shared Dataset

Dataset	Accuracy	Precision	Recall	F1-Score
1/3 Dataset	82,50%	90,45%	87,05%	88,55%
2/3 Dataset	88,20%	96,30%	93,45%	94,80%
2/3 Dataset	93,91%	93,44%	91,84%	92,50%

3.4 Testing Stemming

At this stage, testing is carried out on the impact of stemming on the dataset. After testing, it was found that

DOI: https://doi.org/10.29207/resti.v7i3.4751

stemming has effect on accuracy. By doing stemming, the accuracy decreases by 2.04%. This is due to the loss of key words due to the stemming process. For comparison of accuracy results can be seen in Table 7.

Table 7. Test Stemming

Model	Accuracy	Precision	Recall	F1-Score
Model without stemming	93,91%	93,44%	91,84%	92,50%
Model with stemming	91,87%	89,84%	87,91%	90,82%

4. Conclusion

Based on the results of the analysis and several trials using several parameters, the best results were obtained with an accuracy of 93.91%, a precision of 93.44%, a recall of 91.84%, and an F1 score of 92.50%. The use of LSTM for aspect extraction and CNN for sentiment extraction obtains better accuracy than the LSTM or CNN models. The accuracy of a model can be affected by the number of datasets used. With more and more datasets used, the accuracy of a model will increase. The use of stemming affects the accuracy of a model, in this research the accuracy decreased by 2.04% when using stemming. This is due to the loss of keywords due to the stemming process.

This research concluded that the unbalanced distribution of datasets can affect the performance of a model. It is hoped that further research can use oversampling to overcome this. The suggestions that can be used for further research can add a method to handle negative sentences.

References

[1] D. L. Rianti, Y. Umaidah, and A. Voutama, “Tren Marketplace Berdasarkan Klasifikasi Ulasan Pelanggan Menggunakan Perbandingan Kernel Support Vector Machine,” STRING, vol. 6, no. 1, p. 98, Aug. 2021. https://doi: 10.30998/string.v6i1.9993.

[2] I. Ventre and D. Kolbe, “The Impact of Perceived Usefulness of Online Reviews, Trust and Perceived Risk on Online Purchase Intention in Emerging Markets: A Mexican Perspective,” Journal of International Consumer Marketing, vol. 32, no. 4, pp. 287–299, Aug. 2020. https://doi: 10.1080/08961530.2020.1712293.

[3] D. F. Nasiri and I. Budi, “Aspect Category Detection on Indonesian E-commerce Mobile Application Review,” in 2019 International Conference on Data and Software Engineering (ICoDSE), Pontianak, Indonesia, pp. 1–6, Nov. 2019. https://doi: 10.1109/ICoDSE48700.2019.9092619.

[4] M. T. Ari Bangsa, S. Priyanta, and Y. Suyanto, “Aspect-Based Sentiment Analysis of Online Marketplace Reviews Using Convolutional Neural Network,” Indonesian J. Comput. Cybern. Syst., vol. 14, no. 2, p. 123, Apr. 2020. https://doi: 10.22146/ijccs.51646.

[5] Md. A. Rahman and E. Kumar Dey, “Aspect Extraction from Bangla Reviews using Convolutional Neural Network,” in 2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR), Kitakyushu, Japan, pp. 262–267, Jun. 2018. htpps://doi: 10.1109/ICIEV.2018.8641050.

[6] M. R. Yanuar and S. Shiramatsu, “Aspect Extraction for Tourist Spot Review in Indonesian Language using BERT,” in 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Fukuoka, Japan, pp. 298–302, Feb. 2020. https://doi: 10.1109/ICAIIC48513.2020.9065263.

[7] M. Jiang, W. Zhang, M. Zhang, J. Wu, and T. Wen, “An LSTM-CNN attention approach for aspect-level sentiment classification,” JCM, vol. 19, no. 4, pp. 859–868, Nov. 2019. https://doi: 10.3233/JCM-190022.

[8] I. Priyadarshini and C. Cotton, “A novel LSTM–CNN–grid search-based deep neural network for sentiment analysis,” J Supercomput, vol. 77, no. 12, pp. 13911–13932, Dec. 2021. https://doi: 10.1007/s11227-021-03838-w.

[9] C. Colón-Ruiz and I. Segura-Bedmar, “Comparing deep learning architectures for sentiment analysis on drug reviews,” Journal of Biomedical Informatics, vol. 110, p. 103539, Oct. 2020. https://doi: 10.1016/j.jbi.2020.103539.

[10] A. M. Alayba, V. Palade, M. England, and R. Iqbal, “A Combined CNN and LSTM Model for Arabic Sentiment Analysis,” in Machine Learning and Knowledge Extraction, vol. 11015, A. Holzinger, P. Kieseberg, A. M. Tjoa, and E. Weippl, Eds. Cham: Springer International Publishing, pp. 179–191, 2018. https://doi: 10.1007/978-3-319-99740-7_12.

[11] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” arXiv, May 2019. http://arxiv.org/abs/1810.04805

[12] W. Meng, Y. Wei, P. Liu, Z. Zhu, and H. Yin, “Aspect Based Sentiment Analysis With Feature Enhanced Attention CNN-BiLSTM,” IEEE Access, vol. 7, pp. 167240–167249, 2019. https://doi: 10.1109/ACCESS.2019.2952888.

[13] P. R. Amalia and E. Winarko, “Aspect-Based Sentiment Analysis on Indonesian Restaurant Review Using a Combination of Convolutional Neural Network and Contextualized Word Embedding,” Indonesian J. Comput. Cybern. Syst., vol. 15, no. 3, p. 285, Jul. 2021.https://doi: 10.22146/ijccs.67306.

[14] R. Man and K. Lin, “Sentiment Analysis Algorithm Based on BERT and Convolutional Neural Network,” in 2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), Dalian, China, pp. 769–772, Apr. 2021. https://doi: 10.1109/IPEC51340.2021.9421110.

[15] M. O. Ibrohim and I. Budi, “Multi-label Hate Speech and Abusive Language Detection in Indonesian Twitter,” in Proceedings of the Third Workshop on Abusive Language Online, Florence, Italy, pp. 46–57, 2019. https://doi: 10.18653/v1/W19-3506.

[16] N. A. Shafirra and I. Irhamah, “Klasifikasi Sentimen Ulasan Film Indonesia dengan Konversi Speech-to-Text (STT) Menggunakan Metode Convolutional Neural Network (CNN),” JSSITS, vol. 9, no. 1, pp. D95–D101, Jun. 2020. https://doi: 10.12962/j23373520.v9i1.51825.

[17] B. Wilie et al., “IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding,” p. 15. arXiv, Oct. 08, 2020. https://arxiv.org/abs/2009.05387

[18] M. M. Abdelgwad, “Arabic aspect-based sentiment classification using BERT.” arXiv, Nov. 27, 2021. https://arxiv.org/abs/2107.13290

[19] F. A. Prabowo, M. O. Ibrohim, and I. Budi, “Hierarchical Multi-label Classification to Identify Hate Speech and Abusive Language on Indonesian Twitter,” in 2019 6th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE), Semarang, Indonesia, pp. 1–5, Sep. 2019. https://doi: 10.1109/ICITACEE.2019.8904425.

[20] K. Sun, Y. Li, D. Deng, and Y. Li, “Multi-Channel CNN Based Inner-Attention for Compound Sentence Relation Classification,” IEEE Access, vol. 7, pp. 141801–141809, 2019. https://doi: 10.1109/ACCESS.2019.2943545.