JURNAL ILMIAH KOMPUTER GRAFIS, Vol.17, No.1, Juli 2024, pp. 179-185
p-ISSN : 1979-0414(print), e-ISSN : 2621-6256 (online)
DOI : https://doi.org/10.51903/pixel.v17i1.2010

Emotion Detection Using Contextual Embeddings for Indonesian Product Review Texts on E-commerce Platform

Amelia Devi Putri Ariyanto*[1], Fari Katul Fikriah[2], Arif Fitra Setyawan[3]
ameliadev26@gmail.com[*1], farichatulfikriyah45@gmail.com[2], arif.fitra.setyawan@uwhs.ac.id[3]
Widya Husada University
Information Technology and Systems Study Program
Subali Raya Street Number 12, Semarang City

ARTICLE INFO

Received 23 April 2024
Accepted 27 Mei 2024
Published 24 Juli 2024

ABSTRACT

The advancement of e-commerce has changed the way people shop. However, there is a mismatch between the actual quality of a product and the seller's description. Product reviews are an important source of information for making purchasing decisions. However, processing large numbers of reviews manually is difficult. This research aims to detect emotions in Indonesian language product review texts using contextual embeddings. The public dataset used was PRDECT-ID, which comprises five emotion labels. The methods used include data preprocessing, feature extraction using contextual embeddings such as Bidirectional Encoder Representations from Transformers (BERT), and classification using Decision Tree, Naïve Bayes, and k-Nearest Neighbors (KNN). Among the compared models, the KNN model demonstrated the highest improvement, achieving a 15.09% enhancement over the decision tree results. This research provides insights into the effectiveness of contextual embeddings in detecting emotions in Indonesian language product review texts.

Keywords: BERT, Contextual Embeddings, E-commerce Platform, Emotion Detection, Product Review Texts

1. Introduction

The rapid progress in e-commerce has changed people's way of shopping to become completely online because it can provide convenience and the prices offered are more affordable for consumers. Tokopedia, which was founded in 2009, has grown into one of the largest e-commerce giants in Indonesia. However, there is a mismatch between the actual quality of a product and the description provided by the seller on the e-commerce platform, thus making many consumers look for product information through e-commerce reviews that cover various aspects, such as price, service, and logistics [1].

Reviews on e-commerce platforms are crucial because they contain a lot of product information that makes it easier for new consumers to make correct purchasing decisions. Sellers also have the opportunity to understand consumer needs, identify product weaknesses, and encourage product innovation based on consumer feedback [2]. However, for laypeople, processing many online reviews and extracting consumer opinions from them is a difficult task. Therefore, it is necessary to use automatic review processing with emotion extraction techniques [3].

Emotions are an inseparable element in human life that significantly impact decision-making. Emotion detection is the process of recognizing various types of individual feelings or emotions, such as happiness, sadness, and anger. Detecting emotions from text data originating from e-commerce platforms is a challenging task because texts are typically short and brief and sometimes contain incomplete sentences [4]. The existence of various ambiguities due to the use of daily conversations that contain a lot of slang also becomes a challenge when detecting emotions in short texts from e-commerce platforms, making emotion detection even more difficult [5].

Text representation techniques are used to transform text into numerical form for easier processing for emotion detection [6]. Previous studies on emotion detection in Indonesia have predominantly employed statistical text representation methods such as TF-IDF [7], [8], [9]. The TF-IDF is a word weighting method used to assess the importance of a word in a document [10]. The TF-IDF method disregards word order and contextual meaning and relies solely on word frequency to determine concept relationships.

Text representation techniques using contextual embeddings can overcome the problem of statistical-based text representation techniques, which take into account the words around the word to capture richer contextual meaning and nuances. One algorithm for contextual embeddings is Bidirectional Encoder Representations from Transformers (BERT) ([11]. BERT can read text in two directions, from left to right and from right to left, to understand the full context of the word by utilizing the self-attention mechanism of the Transformer model.

This study aims to detect emotions by representing text as a numerical vector using BERT contextual embeddings. The self-attention mechanism in BERT considers the importance of each word in the overall context of the given sentence as a whole. After the text is represented into a vector with contextual embeddings, a classification process is performed to determine the emotions of a review text using several distance-based (k-Nearest Neighbors), tree-based (Decision Tree) and probability-based (Naïve Bayes) classifier algorithms.

2. Method

The stages of emotion detection using contextual embeddings in product review text data include data preparation, data preprocessing, feature extraction, detection, and evaluation. Figure 1 shows the overall flow of the proposed method. A detailed explanation of each process will be explained in the sub-chapter below.

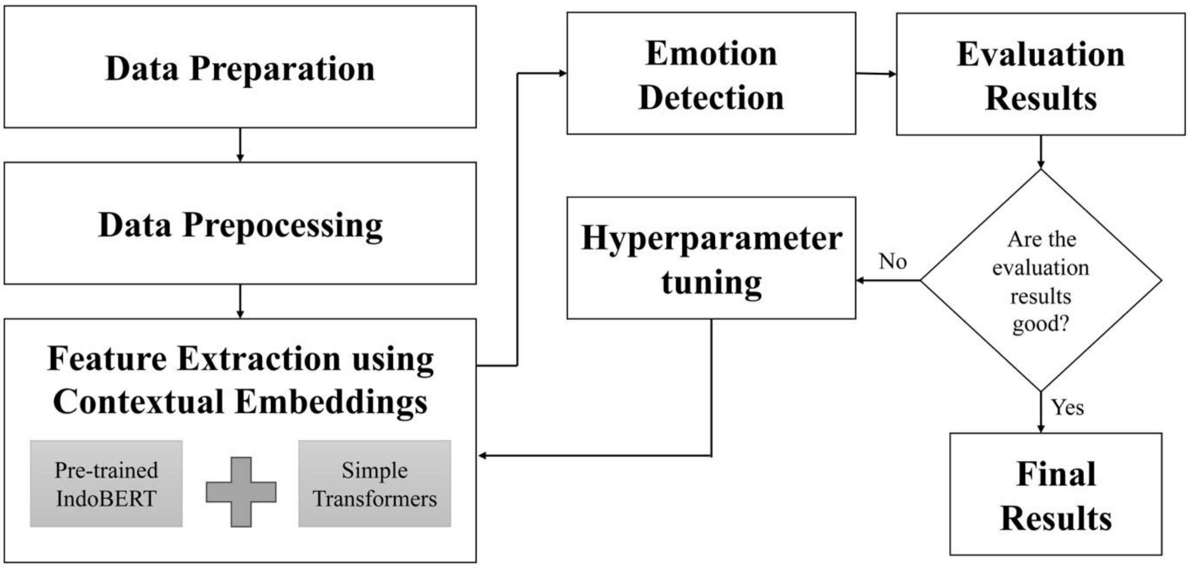

Figure 1. Flowchart of Emotion Detection Using Contextual Embeddings in a Product Review Text

2.1. Data Preparation

This research uses the public dataset PRDECT-ID [12], which is a dataset for classifying emotions in Indonesian language product reviews taken from the Tokopedia e-commerce platform. The PRDECT-ID contains 5400 product reviews from 29 different product categories. Each product review was annotated with one emotion using Shaver's emotion model because it is simple and powerful enough to build emotion models. Each emotion has unique sentence characteristics. For example, anger generally contains curse words and expressions of dislike, whereas the emotion of fear contains warnings and doubts about the quality of a product or seller. The emotion distribution in the PRDECT-ID dataset is shown in Figure 2. This research did not consider imbalance data; thus, 600 emotion labels were created for each emotion label.

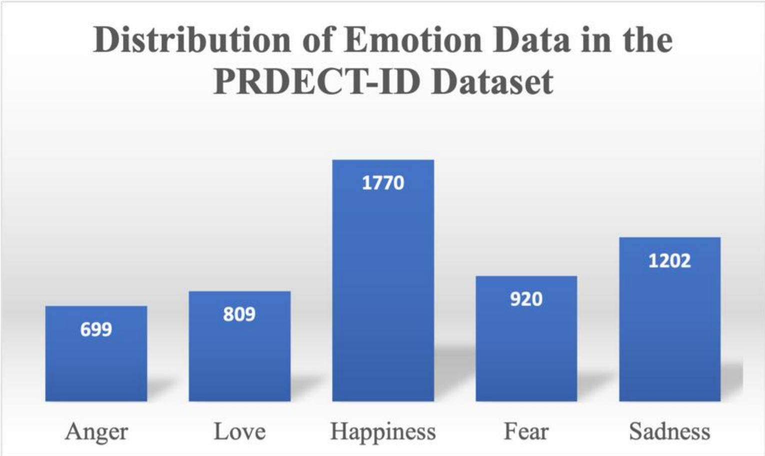

Figure 2. Distribution of emotions in PRDECT-ID Dataset

2.2. Data Preprocessing

Preprocessing converts unstructured data into processed forms that can be customized to meet the requirements of subsequent processing stages. Preprocessing comprises several stages, namely tokenization, removing unclear symbols, and normalization. Tokenization is the process of converting a document word into a word by removing spaces. Normalization is the process of changing nonstandard words into standard forms [10]. Emotion labels are encoded numerically before contextual embeddings for feature extraction. The emotion label Happiness was transformed into label 0, Sadness into label 1, Fear into label 2, Love into label 3, and Anger into label 4.

2.3 Feature Extraction Using Contextual Embeddings

The BERT model can be used for Natural Language Processing (NLP) tasks such as emotion detection, using two main approaches: fine-tuning and feature-based. The fine-tuning approach involves retraining the model with parameter adjustments, whereas the feature-based approach uses feature representations extracted from a pretrained BERT model as input to the classification model without the need for fine-tuning [13]. Simple Transformers is a Python library that makes it easy to use advanced Transformer models, such as BERT, GPT-2, and RoBERTa, for emotion detection. The library is built on Hugging Face's Transformers library, which provides implementations of well-known Transformer models, such as pre-trained IndoBERT.

IndoBERT is a variant of BERT that has been trained specifically on Indonesian language texts. Pre-trained IndoBERT indicates that the BERT model has undergone a previous training process using a large text corpus in Indonesia. Thus, this model has a good understanding of the Indonesian language structure and vocabulary. By using Simple Transformers and pretrained IndoBERT, researchers can quickly and easily apply powerful NLP models to specific tasks in Indonesia.

2.4 Emotion Detection

After obtaining features using contextual embeddings, the next step is the classification process to determine emotions from product review texts on e-commerce platforms. Several classifier algorithms are used to compare the performance, namely tree-based classification (Decision Tree), probability-based classification (Naïve Bayes), and distance-based classification (k-Nearest Neighbors).

The way the Decision Tree classifier works is to divide the dataset into smaller subsets while developing an associated decision tree. At each node in the tree, the algorithm selects the most effective features to classify the data into different classes. This process is repeated for each sub-branch until all data in that node have the same class or feature are shared. To make predictions, the Decision Tree follows a path from root to leaf according to the features of the incoming data until it reaches the predicted class [14].

The KNN algorithm stores the entire training dataset and makes predictions based on the proximity of new instances to the training data. The use of a distance metric (generally Euclidean Distance) finds the K nearest neighbors of a new instance. The class of a new instance is determined by majority voting from its K closest neighbors [15].

Naive Bayes works by calculating the probability that a data instance falls into a certain class based on its features. There is an assumption of independence from the Naïve Bayes algorithm, where each feature in the dataset is considered independent. For example, if features X_1 and X_2, the value of X_1 does not affect that of X_2, and vice versa. Bayes' theorem is also used to calculate the posterior probability of a class based on the feature distribution in the given class [16].

2.5 Evaluation Results

The metrics used to measure the model’s performance in detecting emotions were precision, recall, and the F1 score metrics. Precision measures the accuracy of a model's positive predictions, meaning how many of its positive predictions are actually positive. Recall measures how well the model finds all true positive cases. The F1 Score is the harmonic average of precision and recall metrics, which provides a balance between these two metrics [17].

3. Result and Discussion

This research uses Google Collaboratory with the Python programing language for emotion detection in product review text from e-commerce platforms. The implementation process is in accordance with the previous section, where the product review text is transformed into numerical form using contextual embeddings (BERT). Table 1 lists the evaluation results obtained by several classifiers using contextual embeddings for emotion detection. The lowest performance was achieved using the decision tree classifier with an F1 score of 0.53, while the k-Nearest Neighbors classifier obtained the highest performance score of 0.61.

Decision Trees tend to overfit the training data, particularly when handling varied and informal text data. This causes model performance to decrease when applied to previously unknown data. Informal text data such as product reviews on e-commerce platforms have very high variations in terms of language style, slang use, and sentence structure, making it difficult for Decision Trees to handle this uncertainty and variation. Decision Tree's rigid structure also makes it not sufficiently flexible to capture the complexity and nuance of informal texts.

The proposed KNN method employs a simple but effective approach that measures the distance between text samples to classify emotions. This makes the KNN better able to handle variations and uncertainty in informal text than the decision tree. The proposed KNN makes no assumptions about data distribution; thus, it is more adaptive to variations in informal text data. The use of nearest-neighbor data can provide more contextual and relevant information for determining emotions, which is useful in informal texts that are often ambiguous. There was an increase in F1 score of 15.09% from using Decision Tree to k-Nearest Neighbors.

Table 1. Evaluation Results with Contextual Embeddings for Emotion Detection

Classifier	**Evaluation Results**			
	Precision	**Recall**	**F1-Score**	**Accuracy**
Decision Tree	0.53	0.53	0.53	0.53
Naïve Bayes	0.58	0.56	0.56	0.56
k-Nearest Neighbors	**0.61**	**0.62**	**0.61**	**0.62**

Table 2. Confusion Matrix from the k-Nearest Neighbors Classifier (Highest Performing) for Emotion Detection

		Predicted Labels				
		Happy	**Sadness**	**Fear**	**Love**	**Anger**
Actual Labels	**Happy**	31	1	1	14	0
	Sadness	1	45	15	0	8
	Fear	1	18	32	0	12
	Love	13	0	0	54	0
	Anger	0	15	15	1	23

Table 2 shows the confusion matrix of the k-Nearest Neighbors classifier, which obtained the highest emotion detection performance. A confusion matrix is a tool for evaluating the performance of classification models by comparing model predictions with actual values. There was a significant prediction error between the emotions “Happy” and “Love”. For example, 14 cases of "Happy" were predicted as "Love". This may occur because the words used to express happiness are often similar to those used to express love. The prediction error between “Sadness” and “Fear” was quite high, with 15 cases of “Sadness” being predicted as “Fear” and vice versa. The model found it challenging to differentiate between sadness and fear in textual contexts. The emotion “Anger” showed relatively fewer prediction errors than the other emotions. However, there are still some significant errors, for example, 15 cases of “Anger” being predicted as “Fear”.

4. Conclusion

This study successfully detected emotions in product review texts on e-commerce platforms using contextual embedding as a technique for representing text, which was then classified using several classification algorithms. Based on our research results, we found that the decision tree-based classifier algorithm (Decision Tree) is less effective for informal text data because it is overfitting and inflexible in handling text variations. The probability-based classifier algorithm (Naïve Bayes) outperforms the decision tree algorithm but faces challenges in handling informal text variations.

Future work can be evaluated on larger and more diverse datasets to ensure the generalizability of the model. Data augmentation techniques to increase the amount and variety of training data can be used in future work. In addition, classifier algorithms with ensemble methods can be used to improve performance.

References

[1] Y. Liu, J. Lu, J. Yang, and F. Mao, "Sentiment analysis for e-commerce product reviews by deep learning model of Bert-BiGRU-Softmax," *Mathematical Biosciences and Engineering*, vol. 17, no. 6, pp. 7819–7837, Nov. 2020, doi: 10.3934/MBE.2020398.

[2] Y. S. Mao, L. Y. Zhang, and Y. R. Li, "Finding product problems from online reviews based on BERT-CRF model," 2019.

[3] P. Nandwani and R. Verma, "A review on sentiment analysis and emotion detection from text," Dec. 01, 2021, *Springer*. doi: 10.1007/s13278-021-00776-6.

[4] A. D. P. Ariyanto, C. Fatichah, and D. Purwitasari, "Semantic Role Labeling for Information Extraction on Indonesian Texts: A Literature Review," in *2023 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, IEEE, Jul. 2023, pp. 119–124. doi: 10.1109/ISITIA59021.2023.10221008.

[5] A. D. P. Ariyanto, D. Purwitasari, and C. Fatichah, "A Systematic Review on Semantic Role Labeling for Information Extraction in Low-Resource Data," *IEEE Access*, vol. 12, no. April, pp. 57917–57946, 2024, doi: 10.1109/ACCESS.2024.3392370.

[6] A. D. P. Ariyanto, Chastine Fatichah, and Agus Zainal Arifin, "Analisis Metode Representasi Teks Untuk Deteksi Interelasi Kitab Hadis: Systematic Literature Review," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 5, pp. 992–1000, 2021, doi: 10.29207/resti.v5i5.3499.

[7] K. S. Nugroho, F. A. Bachtiar, and W. F. Mahmudy, "Detecting Emotion in Indonesian Tweets: A Term-Weighting Scheme Study," *Journal of Information Systems Engineering and Business Intelligence*, vol. 8, no. 1, pp. 61–70, Apr. 2022, doi: 10.20473/jisebi.8.1.61-70.

[8] Muljono, A. S. Winarsih, and C. Supriyanto, "Evaluation of classification methods for Indonesian text emotion detection," *Proceedings - 2016 International Seminar on Application of Technology for Information and Communication, ISEMANTIC 2016*, pp. 130–133, 2016, doi: 10.1109/ISEMANTIC.2016.7873824.

[9] L. Efrizoni, S. Defit, M. Tajuddin, and A. Anggrawan, "Komparasi Ekstraksi Fitur dalam Klasifikasi Teks Multilabel Menggunakan Algoritma Machine Learning," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 3, pp. 653–666, 2022, doi: 10.30812/matrik.v21i3.1851.

[10] A. D. P. Ariyanto, L. A, A. Z. A, M. Maryamah, R. W. S, and R. I, "Metode Pembobotan Kata Berbasis Cluster Untuk Perangkingan Dokumen Berbahasa Arab," *Techno.Com*, vol. 20, no. 2, pp. 259–267, May 2021, doi: 10.33633/tc.v20i2.4357.

[11] Amelia Devi Putri Ariyanto, "Interrelation Detection Between Hadith Books Using Word Embedding and Ensemble Learning (Thesis Magister)," Institut Teknologi Sepuluh Nopember, 2022. [Online]. Available: http://repository.its.ac.id/id/eprint/92825

[12] R. Sutoyo, S. Achmad, A. Chowanda, E. W. Andangsari, and S. Isa, "PRDECT-ID: Indonesian product reviews dataset for emotions classification tasks," *Data Brief*, 2022, doi: 10.17632/574v66hf2v.1.

[13] F. zahra El-Alami, S. Ouatik El Alaoui, and N. En Nahnahi, "Contextual semantic embeddings based on fine-tuned AraBERT model for Arabic text multi-class categorization," *Journal of King Saud University - Computer and Information Sciences*, 2021, doi: 10.1016/j.jksuci.2021.02.005.

[14] Rizky Haqmanullah Pambudi, B. D. Setiawan, and Indriati, "Penerapan Algoritma C4 . 5 Untuk Memprediksi Nilai Kelulusan Siswa Sekolah Menengah Berdasarkan Faktor Eksternal," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 7, pp. 2637–2643, 2018, [Online]. Available: http://j-ptiik.ub.ac.id5

[15] S. Satriya, R. H. D., Santoso, E., & Sutrisno, “Implementasi Metode Ensemble K-Nearest Neighbor untuk Prediksi Nilai Tukar Rupiah Terhadap Dollar Amerika,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (JPTIIK) Universitas Brawijaya*, vol. 2, no. 4, pp. 1718–1725, 2018.

[16] I. W. Saputro and B. W. Sari, “Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa,” *Creative Information Technology Journal*, vol. 6, no. 1, p. 1, 2020, doi: 10.24076/citec.2019v6i1.178.

[17] M. F. Fibrianda and A. Bhawiyuga, “Analisis Perbandingan Akurasi Deteksi Serangan Pada Jaringan Komputer Dengan Metode Naïve Bayes Dan Support Vector Machine (SVM),” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 9, pp. 3112–3123, 2018.