International Journal of Advances in Intelligent Informatics
Vol. 9, No. 2, July 2023, pp. 273-285


# A hybrid model for aspect-based sentiment analysis on customer feedback: research on the mobile commerce sector in Vietnam

Thanh Ho [a,b,1,*], Hien Minh Bui [c,2], Thai Kim Phung [c,3]

[a] University of Economics and Law, Ho Chi Minh City, Vietnam
[b] Viet Nam National University Ho Chi Minh City, Vietnam
[c] UEH College of Technology and Design (UEH-CTD), University of Economics Ho Chi Minh City (UEH), Vietnam
[1] thanhht@uel.edu.vn; [2] hienbui.192118003@st.ueh.edu.vn; [3] phungthk@ueh.edu.vn
* corresponding author

ARTICLE INFO

**Article history**
Received December 28, 2022
Revised April 8, 2023
Accepted April 23, 2023
Available online June 2, 2023

**Keywords**
Customer feedback
Sentiment analysis
Mobile commerce
Machine and deep learning
POS tagging

ABSTRACT

Feedback and comments on mobile commerce applications are extremely useful and valuable information sources that reflect the quality of products or services to determine whether data is positive or negative and help businesses monitor brand and product sentiment in customers' feedback and understand customers' needs. However, the increasing number of comments makes it increasingly difficult to understand customers using manual methods. To solve this problem, this study builds a hybrid research model based on aspect mining and comment classification for aspect-based sentiment analysis (ABSA) to deeply comprehend the customer and their experiences. Based on previous classification results, we first construct a dictionary of positive and negative words in the e-commerce field. Then, the POS tagging technique is applied for word classification in Vietnamese to extract aspects of model commerce related to positive or negative words. The model is implemented with machine and deep learning methods on a corpus comprising more than 1,000,000 customer opinions collected from Vietnam's four largest mobile commerce applications. Experimental results show that the Bi-LSTM method has the highest accuracy with 92.01%; it is selected for the proposed model to analyze the viewpoint of words on real data. The findings are that the proposed hybrid model can be applied to monitor online customer experience in real time, enable administrators to make timely and accurate decisions, and improve the quality of products and services to take a competitive advantage.



## 1. Introduction

With today's Internet and e-commerce platforms explosion, online shopping has become easier and more convenient than ever. Mobile commerce applications have been rapidly developing. In Vietnam, four mobile commerce apps (i.e., Shopee, Lazada, Sendo, and Tiki) have the most visits on Google Play Store, with total monthly traffic of 143 million in Q4, 2020 [1]. In addition, a large amount of data from users represents online responses in the form of daily texts on mobile commerce applications. These reviews are a valuable resource for businesses to understand the users' experiences and opinions about products and services, which is helpful for both users and manufacturers [2]. However, it is becoming more difficult to identify the main patterns with the increasing number of comments every day. Therefore, an automated approach to extracting and summarizing the main patterns of online commentary is essential, with opinion mining through sentiment analysis (SA), specifically aspect-based sentiment analysis (ABSA), posing a significant challenge. In particular, the Vietnamese language is a complex language that consists of a 29-character alphabet including Latin characters, using additional tones, such as accents (´), hypotenuses (`), question marks (ˀ), tildes (~), and heavy accents (.). Additionally, many borrowed words are derived from Chinese, French, and English [3].

Various studies with machine learning approaches in sentiment analysis have been published. Yanuar Nurdiansyah *et al.* [4] suggested a system that utilizes the Naïve Bayes Classifier method to classify sentiment in Bahasa Indonesia movie reviews into two categories (positive and negative), with an average classification accuracy of 88.37%. Al Amrani *et al.* [5] proposed a hybrid approach to identify product reviews offered by Amazon using Random Forest (RF) and Support Vector Machine (SVM). Bolbol & Maghari *et al.* [6] conducted an experiment with various machine learning classifiers and found that Logistic Regression (LR) achieved the highest accuracy of 93% on the Arabic Tweets dataset. Other comparative studies used machine learning techniques (Naïve Bayes, SVM, Decision Tree, K-Nearest Neighbor, and Artificial Neural Network) to classify customer opinions [7]–[11]. Recently, deep learning approaches produced better performance than traditional machine learning. Peng *et al.* [12] employed Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM) [13], [14], and Convolutional Neural Networks (CNN) and the results have shown that CNN has reported the accuracy of 88.22%, RNN and LSTM have reported accuracy of 68.64% and 85.32% respectively. Li *et al.* [15] combined BERT+BiLSTM+CNN to classify sentiment on the Weibo text dataset and achieved the accuracy of 92.4%.

In Vietnamese, Nguyen *et al.* [16] used the PhoBERT model to classify sentiment-based stock article titles, with an accuracy of 93%. In another research, Nguyen *et al.* [17] proposed fine-tuning BERT method for sentiment analysis of Vietnamese Reviews, the results show that the BERT-RCNN model achieves the best result with the F1-score is 91.15%. In addition, Truong *et al.* [18] used the pre-trained model PhoBERT, with other fine-tuning techniques was achieved with an accuracy of 94.28% on the UIT-VSFC dataset.

ABSA is an interesting research topic in the opinion mining field [19]. People tend to talk about many aspects in their comments about a product or service and provide many words with positive or negative connotations. Specifically, each comment will include the following four categories: (1) an aspect of a word that has a positive or negative meaning, (2) an aspect of many words with a positive or negative meaning, (3) many aspects of one word that have positive or negative connotations, and (4) many aspects of many words that have positive or negative connotations [20]. For example, "Lazada has a lot of discount codes, but app still many bugs to fix". Here, users gave positive reviews of the "discount" aspect but negative reviews of the "app" aspect. Dealing with the comments that have many facets in sentences is an extremely challenging task.

One of the early studies in the field of user opinion aspect mining was introduced by Hu & Liu [21], which is based on the frequency of occurrence of nouns and noun phrases to exploit aspects. Subsequently, various approaches have been studied, focusing on the aspect mining problem, with the most widely used approach being the rule-based approach [22], which extracts aspects based on the grammatical relationships of words in a sentence. An aspect extraction system [23] was introduced from products that considered nouns to be aspect terms and extracted them based on POS tagging and term frequency-inverse document frequency (TF-IDF) methods. Classification algorithms, such as Naïve Bayes and SVMs which have also been applied for aspect extraction. Mai & Le [24] proposed a sequence-labeling approach that combines BiRNN and Conditional Random Fields (CRF) to concurrently extract opinion targets and detect their associated sentiments in smartphone-related datasets. Luc Phan *et al.* [25] give a suggested method for the Vietnamese aspect-based sentiment task is based on the Bi-LSTM architecture, using fastText word embeddings, their experiments demonstrate that this approach achieves the highest F1-scores of 84.48% for the aspect task and 63.06% for the sentiment task in Vietnamese Smartphone Feedback Dataset (UIT-ViSFD).

In this study, a hybrid research model that combines aspect mining and comment classification is developed. We propose an experimental method for the model that consists of five main tasks: first, we provide a Vietnamese customer reviews dataset in the e-commerce industry; Second, we offer dictionaries and techniques to address issues in pre-processing Vietnamese text; Third, we propose the Bi-SLTM model for text sentiment classification; Fourth, we identify aspects of the product and service that users comment on, such as positive or negative comments, using POS tagging based on the result of the proposed model's output; Finally, we build a the dashboard to help managers gain deeper understanding of customers' needs and preferences, thereby improving the level of customer satisfaction and expectations about products and services. Fig. 1 provides an overview of the hybrid research model.

The remainder of this paper is structured as follows: Section 2 presents the methodology used in this study. Section 3 outlines the results and associated discussion. In this section, we detail our experiments and provide corresponding discussions. Finally, in Section 4, we summarize our work and discuss future directions for research.

## 2. Method

To build the hybrid model and research method, the secondary data from previous studies which are related to the research objectives of the article on topics of aspects extraction and sentiment analysis in the field of mobile commerce and others are surveyed and studied. Specifically, we surveyed the theoretical basis, analyzed business models, and proposed appropriate research models. And then, experimental research method will be applied to implement and evaluate the proposed model through the following stages: identifying business problems and collecting online comments from customers, cleaning data, integrating, and applying statistical methods, and employing machine and deep learning techniques for natural language processing to uncover hidden knowledge and insights. The implicit knowledge in the data is represented by the nuances in customer comments, specific aspects, and issues related to the nuances of words. Based on the results, recommendations related to business and management will be proposed. The proposed hybrid model for aspect-based sentiment analysis in Fig. 1 includes 5 stages.

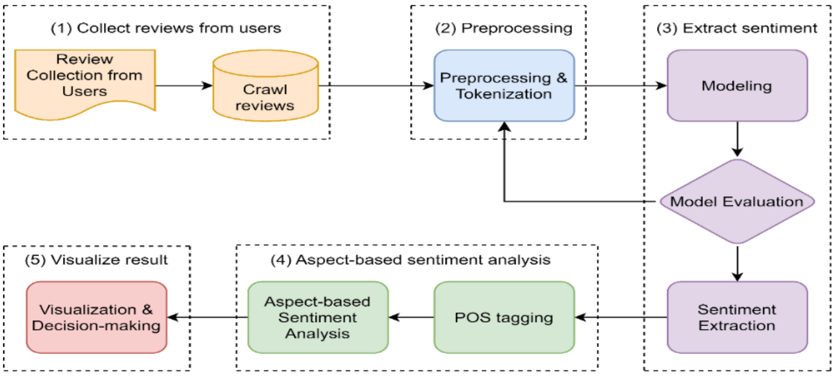


**Fig. 1.** A hybrid model for aspect-based sentiment analysis (Source: Authors)

### 2.1. Data Collection

Python language with google_play_scraper library is used to collect data from Google Play Store, which includes over 1,2 million customer feedback on four mobile commerce apps (Shopee, Lazada, Sendo, and Tiki) from 2013 to 2022. The dataset is named VN_E-commerce_Review and is published at https://www.kaggle.com/datasets/hienbm/vietnamese-ecommerce-review.

### 2.2. Data Pre-processing

After observing the original data set, there are so much noisy data such as other words not in the Vietnamese alphabet, emoji, emoticons, teen code, missing accents, and abbreviations. Therefore, the study highly focuses on pre-processing. As a result, the accuracy of proposed model has improved efficiently. The results after pre-processing will be shown in Table 1, including: (1) Converting all the words to lowercase; (2) Removing words with hashtag (#), html (< >), url (http://), mention (@); (3) Removing characters number, punctuation, and strange characters that are not in Vietnamese alphabet;

(4) Removing elongated characters, duplicate and continuous words or phrases (e.g., “ok ok okkkkkk” -> “ok”); (5) Removing duplicate emojis; (6) Adding Vietnamese accents with dictionary mapping utilizing Vietnamese dictionary words with "key" as the original word and "value" as the word without the accents. (7) Mapping teen code and abbreviation words with a dictionary (e.g. “thjk” -> “thích $_{like}$”); (8) Removing extra white space; (9) Applying word tokenizer with pyvi library [26] (the library was published by Viet -Trung Tran) (eg: “tẩy chay” -> “tẩy chay $_{boycott}$”). Table 1 shows the examples of the dataset before and after pre-processing.

**Table 1.** Dataset examples after pre-processing (Source: Authors)

| Original text (in Vietnamese) | After pre-processing (in Vietnamese and English) |
|---|---|
| dịch vụ tốt và nhiệt tình 😜 😜 | dịch_vụ tốt $_{good\ service}$ và $_{and}$ nhiệt_tình $_{enthusiastic}$ |
| tuyet voi :))) | tuyệt_vời $_{wonderful}$ |
| rất tiện và chất lượng lắm... | rất $_{very}$ tiện $_{convenient}$ và $_{and}$ chất_lượng $_{quality}$ lắm $_{much}$ |
| đề nghị xử lý các shop lừa đảo | đề_nghị $_{suggest}$ xử_lý $_{dealing\ with}$ các shop $_{shops}$ lừa_đảo $_{scam}$ |

### 2.3. Modeling

This study not only intends to build a novel technique to improve sentiment classification on comments using the Bi-LSTM model, but it also aims to conduct a sentiment dictionary and extract aspects based on sentiment. The structure of the suggested model is explored in detail in this section. Fig. 2 depicts the general structure of the Bi-LSTM model.

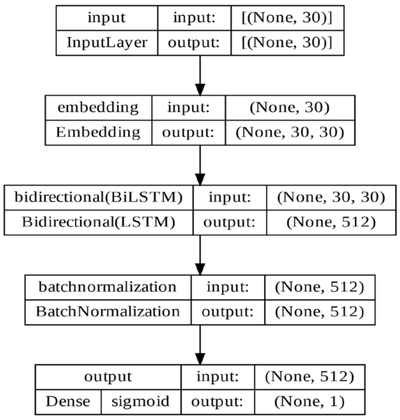


**Fig. 2.** The architecture of Bi-LSTM model (Source: Authors)

After pre-processing, the dataset is divided into 3 sets as input data: a training set with 1,039,240 reviews (~80% of random data), a validation test set with 129,905 reviews (~10% of random data remaining), and a test set with 129,905 reviews (~10% of random data remaining). The label results have scores from 1 to 5, and then we divide them into 3 groups corresponding to scores 1-3 as Negative and scores 4-5 as Positive [26]. Fig. 3 shows the number of text reviews for each class for the training, validation, and test sets.

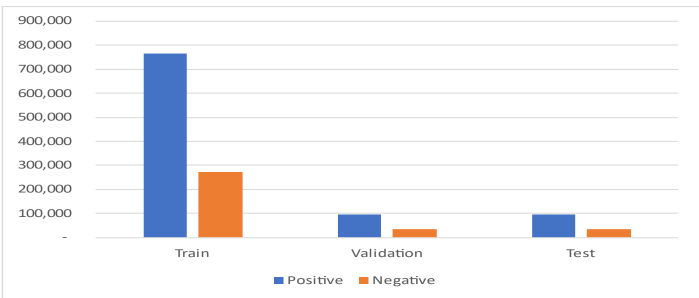


**Fig. 3.** The number of text reviews for training, validation, and test sets

**Embedding** layer takes the integer-encoded vocabulary and looks up the embedding vector for each word index. The embedding layer using training samples from the VN_E-commerce_Review dataset is trained rather than a pre-trained embedding word model such as word2vec [27] or GloVe [28]. In this study, we used an embedding layer given by Keras with a set of 32,526 unique vocabularies, with each word is embedded in a 30-dimension vector space.

**Bi-LSTM** layer is an extension of the LSTM models that combine Bi-RNN models and LSTM units to capture the context information. In the first round, an LSTM is applied on the input sequence (i.e., forward layer). In the second round (i.e., backward layer) of the LSTM model, the reverse form of the input sequence is fed into the LSTM model [29]. Using the LSTM twice improves the learning of long-term dependencies, which allows it to learn the context more efficiently. The architecture of Bi-LSTM is illustrated in Fig. 4.

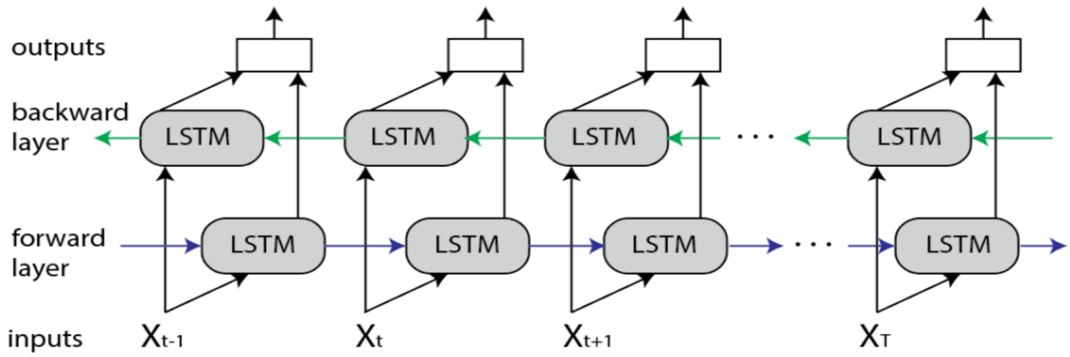


**Fig. 4.** The architecture of Bi-LSTM [30]

**Sigmoid** function is equivalent to a 2-element Softmax, with the second element being assumed to be zero. The output of the sigmoid function is always within a range between 0 to 1:

$$sigmoid\ (x) = \frac{1}{1+e^{-x}} \tag{1}$$

### 2.4. Hyper-parameter setting

Hyperparameter tuning is a crucial step in improving model performance. This method involves modeling the relationship between hyperparameters and the model's performance using a probabilistic model. The model is iteratively updated based on the performance of the evaluated hyperparameters, resulting in a more efficient search of the hyperparameter space. It can prevent overfitting or underfitting and selects the best model as the final model. In this study, we tune four different hyperparameters below and optimize the accuracy. After that, the Bayesian optimization function from Keras Tuner [30] for hyperparameter tuning is used. The objective of parameter is set to "val_accuracy" and the "max_trials" parameter is set to 50, which means that the optimization function will attempt to optimize the validation accuracy by trying a maximum of 50 different combinations of hyperparameters. Table 2 describes the best hyperparameter values in the proposed model.

### 2.5. Metrics

Confusion matrix is applied to evaluate the model. It is a matrix with the number of True Positives (TP), True Negatives (TN), False Negatives (FN), and False Positives (FP). The metrics that are most widely used for evaluation are described as follows:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{2}$$

$$precision = \frac{TP}{TP+FP} \tag{3}$$

$$recall = \frac{TP}{TP+FN} \tag{4}$$

$$F1 - score = \frac{2*precision*recall}{precision+recall} \quad (5)$$

**Table 2.** Hyper-parameters setting

| Parameters | Values | Search space |
|---|---|---|
| Embedding dimension | 30 | - |
| Filter | 256 | Values between 32 and 512 (inclusive) in steps of 32. These represent the number of units in a neural network layer |
| (LSTM) dropout | 0.5 | Values between 0.1 and 0.5 (inclusive) in steps of 0. These represent the fraction of input units to drop in a neural network layer. |
| Bi-LSTM output size | 512 | - |
| Activation | tanh | One of two values "tanh" or "relu". It represents the activation function to use in a neural network layer. |
| Batch size | 256 | - |
| Number of epoch | 100 | - |
| Batch normalization | Yes | - |
| Loss function | binary_crossentropy | - |
| Optimizer | Adam | - |
| Learning rate | 0.002 | Values between 1e-5 and 1e-1 (inclusive). We use logarithmic sampling to explore the search space that represents the learning rate of a neural network optimizer. |

### 2.6. Building sentiment dictionary and extracting aspects by using POS tagging method

POS tagging stands for "Part-of-Speech tagging". It is a process in natural language processing that involves assigning a part of speech (such as noun, verb, adjective, etc.) to each word in a sentence. The process involves analyzing the grammatical structure of the sentence and identifying the role of each word in the sentence. In this study, to apply the technique of labeling from type, we used the pyvi library [31], which is considered one of the libraries with the best results for natural language processing in the Vietnamese language, with a F1 - score of 0.925.

### 2.7. Building sentiment dictionary

Based on using the POS tagging method and the results of model classification, we built a dictionary to classify positive and negative nuances. The dictionary was built by filtering adjective words in the dataset and determining their frequency of occurrence with respect to the rating score to assign them to either the positive or negative group. If a word's frequency of occurrence in one group is more than the other, it will be classified into that respective group. Table 3 shows some examples of words classified as positive or negative.

**Table 3.** Example of nuance words (Source: Authors)

| word (in Vietnamese and English) | pos_tag | negative | positive |
|---|---|---|---|
| tốt $_{good}$ | A | 4.68% | 95.32% |
| nhiều $_{multiple}$ | A | 41.37% | 58.63% |
| tuyệt_vời $_{excellent}$ | A | 1.41% | 98.59% |
| nhanh $_{fast}$ | A | 12.43% | 87.57% |
| rẻ $_{cheap}$ | A | 16.69% | 83.31% |
| ok $_{okay}$ | A | 6.43% | 93.57% |
| tuyệt $_{great}$ | A | 1.87% | 98.13% |
| cao $_{high}$ | A | 58.91% | 41.09% |
| đúng $_{true}$ | A | 46.85% | 53.15% |
| tệ $_{bad}$ | A | 96.08% | 3.92% |
| tiện_lợi $_{convenient}$ | A | 2.94% | 97.06% |
| hài_lòng $_{happy}$ | A | 5.96% | 94.04% |

#### 2.7.1. Aspects Extraction

The aspects by filtering the most frequently occurring noun words in the dataset were extracted that are related to segments such as goods, price, customer care, and customer experience UX/UI, etc. The aspect-based sentiment extraction results are presented in Table 4.

**Table 4.** Example of aspect based on nuance words (Source: Authors)

| Word (in Vietnamese and English) | pos_tag |
|---|---|
| hàng $_{goods}$ | N |
| người $_{people}$ | N |
| sản_phẩm $_{product}$ | N |
| cửa_hàng $_{shop}$ | N |
| tiền $_{money}$ | N |
| ứng_dụng $_{app}$ | N |
| chất_lượng $_{quality}$ | N |
| hàng $_{order}$ | N |
| phí $_{cost}$ | N |
| giá $_{price}$ | N |

## 3. Results and Discussion

### 3.1.Experimental setup

Colaboratory Pro which is released by Google for all data pre-processing, training, and testing: 24GB of Reading Access Memory (RAM), GPU Nvidia Tesla P100-PCIe-16GB, and Intel(R) Xeon(R) CPU @ 2.30GHz. To train the model, the binary cross-entropy [32] for the loss function is applied:

$$\Lambda o\sigma\sigma = -\sum_{\chi=1}^{M} y_{o,c}\ \lambda o\gamma(p_{o,c})\ , \tag{6}$$

In the given equation, M represents the total number of classes, log refers to the natural logarithm, $y_{o,c}$ is a binary indicator taking values of either 0 or 1, representing whether class label c is the correct classification for observation , and $p_{o,c}$ is the predicted probability that observation o belongs to class c. For the optimizers, the optimal Adam method [33] with learning rate equal 0.002 is applied. The number of epochs is set to 100 with EarlyStopping to monitor the metric val_accuracy, and patience equals 15. This means stopping training when the monitored metric has stopped improving after 15 epochs in a row. Fig. 5 displays the accuracy and loss training histories.

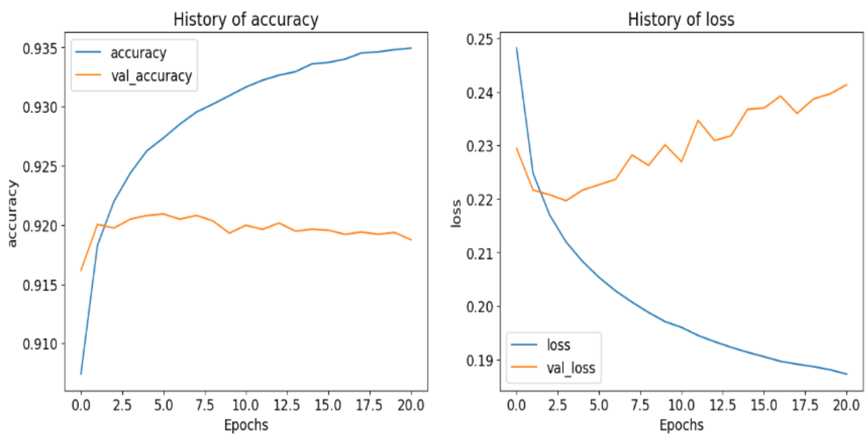


**Fig. 5.** The training history of the accuracy and the loss (Source: Authors)

### 3.2. Experimental results

After training, the test data set is used to assess the performance of the proposed model. Fig. 6 shows the result in confusion matrix.

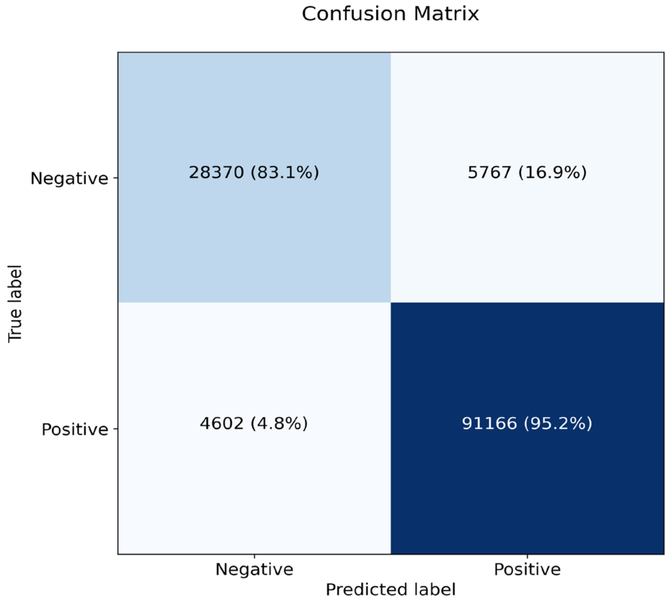


**Fig. 6.** The confusion matrix (Source: Authors)

Table 5 illustrates the accuracy, precision, recall, and F1-score values for the proposed technique in comparison to conventional machine learning, as well as the assessment findings for both before and after pre-processing.

**Table 5.** Performance Comparison of Bi-LSTM Model (Source: Authors)

| Model | Original | | | | After pre-processing | | | |
|---|---|---|---|---|---|---|---|---|
| | *Accuracy* | *Precision* | *Recall* | *F1-score* | *Accuracy* | *Precision* | *Recall* | *F1-score* |
| Naïve Bayes | 88.78 | 92.59 | 92.15 | 92.37 | 89.15 | 92.65 | 92.63 | 92.64 |
| SVM | 90.28 | 92.3 | 94.72 | 93.49 | 90.53 | 92.55 | 94.78 | 93.65 |
| Logistic regression | 90.49 | 92.37 | 94.95 | 93.64 | 90.75 | 92.61 | 95.04 | 93.81 |
| LSTM | 91.81 | 93.69 | 95.31 | 94.49 | 91.95 | 93.94 | 95.22 | 94.58 |
| **Bi-LSTM** | **91.73** | **93.99** | **94.85** | **94.42** | **92.01** | **94.07** | **95.17** | **94.61** |

After obtaining the results, the Bi-LSTM method is the most suitable for the dataset used in this study, with an accuracy of 92.01% and F-score of 94.61% after pre-processing step, which are higher than the remaining methods and before pre-processing.

### 3.3. Bussiness application

After implementing the labeling model from categories using the POS tagging method, Fig. 7 show the analysis of positive and negative sentiment trends in detail, including the total number of comments in the nine years from 2013 to 2022 covered by the corpus, with the Lazada application having the highest total number of comments (530,055), of which there were 410,021 (accounting for 77.35%) positive comments (sentiment = 1.0) and 120,034 (22.65%) negative comments (sentiment = 0.0). The Shopee application ranked second with a total of 518,716 comments, with 352,094 positive (67.88%) and 166,622 negative comments (32.12%). In third place, the Sendo application had a total of 179,321 comments, with 143,170 positive (79.74%) and 36,368 (20.26%) negative comments. Lastly, Tiki application had a total of 70,602 comments, with 52,310 positive (74.09%) and 18,292 negative comments (25.91%).

Fig. 7 shows that most applications have a decreasing trend of a positive proportion and have an increasing negative proportion of comments from 2019 to 08/2022.

| Appid | Target | 2013 | 2014 | 2015 | 2016 | 2017 | At<br>2018 | 2019 | 2020 | 2021 | 2022 | Grand Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| lazada | 0 | 23<br>29.87% | 187<br>19.50% | 1,706<br>18.64% | 3,404<br>24.97% | 5,227<br>24.70% | 12,297<br>27.88% | 17,585<br>12.39% | 18,811<br>17.02% | 42,110<br>31.85% | 18,684<br>33.17% | 120,034<br>22.65% |
| | 1 | 54<br>70.13% | 772<br>80.50% | 7,446<br>81.36% | 10,231<br>75.03% | 15,933<br>75.30% | 31,802<br>72.12% | 124,374<br>87.61% | 91,683<br>82.98% | 90,083<br>68.15% | 37,643<br>66.83% | 410,021<br>77.35% |
| shopee | 0 | | | 339<br>18.72% | 1,577<br>39.60% | 4,478<br>40.73% | 17,037<br>51.60% | 26,394<br>30.33% | 25,387<br>20.92% | 48,081<br>31.17% | 43,329<br>40.77% | 166,622<br>32.12% |
| | 1 | | | 1,472<br>81.28% | 2,405<br>60.40% | 6,516<br>59.27% | 15,982<br>48.40% | 60,617<br>69.67% | 95,980<br>79.08% | 106,170<br>68.83% | 62,953<br>59.23% | 352,094<br>67.88% |
| sendo | 0 | | | 405<br>15.79% | 1,742<br>26.12% | 2,862<br>23.05% | 6,650<br>22.03% | 14,416<br>17.63% | 5,779<br>21.81% | 3,520<br>22.85% | 994<br>24.76% | 36,368<br>20.26% |
| | 1 | | | 2,160<br>84.21% | 4,926<br>73.88% | 9,555<br>76.95% | 23,537<br>77.97% | 67,372<br>82.37% | 20,714<br>78.19% | 11,885<br>77.15% | 3,021<br>75.24% | 143,170<br>79.74% |
| tiki | 0 | | | 105<br>38.46% | 445<br>27.28% | 767<br>50.26% | 2,074<br>20.18% | 5,318<br>17.47% | 4,454<br>31.50% | 3,807<br>38.71% | 1,322<br>53.18% | 18,292<br>25.91% |
| | 1 | | | 168<br>61.54% | 1,186<br>72.72% | 759<br>49.74% | 8,203<br>79.82% | 25,115<br>82.53% | 9,687<br>68.50% | 6,028<br>61.29% | 1,164<br>46.82% | 52,310<br>74.09% |

**Fig. 7.** The total numbers of positive and negative comments on the four applications from 06/2013 to 08/2022

Dive into more detail, Fig. 8 presents the top 10 words with positive nuances in the period (01/2018 - 08/2022). "Good" was the word most appreciated by users on all four applications, with Lazada having the most comments from 2018-2021. Followed by words such as "multiple", "excellent", "fast", "cheap", "okay", "great", "true", "better", and "happy".

| At / Appid | 2018 | | | | 2019 | | | | 2020 | | | | 2021 | | | | 2022 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Word | lazada | sendo | shopee | tiki | lazada | sendo | shopee | tiki | lazada | sendo | shopee | tiki | lazada | sendo | shopee | tiki | lazada | sendo | shopee | tiki |
| tốt (good) | 8,543 | 7,261 | 3,516 | 2,115 | 35,475 | 21,172 | 13,899 | 6,792 | 24,432 | 6,272 | 22,009 | 2,731 | 23,391 | 3,587 | 23,268 | 1,681 | 9,771 | 928 | 14,099 | 275 |
| nhiều (multiple) | 1,939 | 1,337 | 2,100 | 445 | 4,301 | 3,146 | 5,820 | 1,155 | 4,091 | 1,222 | 8,217 | 734 | 7,417 | 693 | 10,322 | 395 | 3,636 | 171 | 6,927 | 148 |
| tuyệt_vời (excellent) | 909 | 630 | 556 | 283 | 5,022 | 2,989 | 3,958 | 1,192 | 4,262 | 945 | 6,395 | 335 | 3,397 | 557 | 6,670 | 166 | 1,502 | 165 | 4,113 | 39 |
| nhanh (fast) | 1,376 | 1,118 | 970 | 682 | 4,464 | 2,428 | 2,591 | 2,109 | 3,057 | 968 | 4,626 | 1,114 | 3,521 | 410 | 4,547 | 739 | 2,457 | 115 | 2,858 | 169 |
| rẻ (cheap) | 1,260 | 873 | 1,125 | 202 | 2,729 | 1,642 | 2,217 | 526 | 1,904 | 529 | 3,137 | 267 | 2,661 | 283 | 3,482 | 167 | 1,033 | 63 | 2,088 | 49 |
| ok (okay) | 661 | 495 | 340 | 109 | 2,597 | 1,456 | 1,415 | 486 | 2,294 | 472 | 2,950 | 269 | 2,623 | 331 | 3,391 | 206 | 1,495 | 82 | 1,708 | 24 |
| tuyệt (great) | 626 | 476 | 343 | 175 | 3,031 | 1,572 | 1,868 | 667 | 2,228 | 392 | 3,023 | 157 | 1,534 | 168 | 3,124 | 66 | 491 | 47 | 1,897 | 19 |
| đúng (true) | 755 | 656 | 722 | 178 | 2,028 | 1,385 | 1,147 | 551 | 1,896 | 536 | 1,778 | 397 | 2,329 | 313 | 2,056 | 288 | 1,463 | 64 | 1,333 | 68 |
| hơn (better) | 978 | 535 | 1,073 | 184 | 1,147 | 934 | 1,547 | 533 | 1,014 | 319 | 1,930 | 343 | 1,525 | 235 | 2,436 | 313 | 771 | 66 | 1,468 | 57 |
| hài_lòng (happy) | 766 | 447 | 341 | 260 | 1,942 | 1,079 | 962 | 1,054 | 1,497 | 409 | 1,758 | 334 | 2,072 | 211 | 1,711 | 204 | 1,141 | 62 | 802 | 15 |

**Fig. 8.** Top 10 words with positive nuances in the period (01/2018 - 08/2022)

The template is designed so that author affiliations are not repeated each time for multiple authors of the same affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization). This template was designed for two affiliations.

Fig. 9 shows the top 10 negative words in the period (01/2018 - 08/2022). In 2022, for the Shopee app, the negative words to care about were "bad", "long", "different" and "slow" with more than 1000 occurrences more than the rest of the top 10 words. For the Lazada, words with an occurrence number above 1000 were "bad". The negative words that Tiki needs to pay attention to include "bad", "long".

| At / Appid | 2018 | | | | 2019 | | | | 2020 | | | | 2021 | | | | 2022 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Word | lazada | sendo | shopee | tiki | lazada | sendo | shopee | tiki | lazada | sendo | shopee | tiki | lazada | sendo | shopee | tiki | lazada | sendo | shopee | tiki |
| cao (high) | 3,144 | 696 | 2,017 | 59 | 2,321 | 2,039 | 1,739 | 196 | 1,015 | 403 | 1,936 | 203 | 945 | 285 | 1,959 | 156 | 573 | 56 | 926 | 35 |
| tệ (bad) | 460 | 229 | 904 | 103 | 622 | 620 | 1,633 | 421 | 1,386 | 329 | 1,697 | 361 | 2,505 | 247 | 3,721 | 490 | 1,150 | 74 | 3,039 | 153 |
| khác (different) | 574 | 362 | 729 | 154 | 862 | 730 | 1,375 | 525 | 923 | 332 | 1,718 | 362 | 1,832 | 205 | 2,723 | 276 | 933 | 66 | 1,671 | 78 |
| lâu (long) | 469 | 373 | 761 | 84 | 878 | 632 | 1,082 | 457 | 824 | 224 | 1,301 | 261 | 1,420 | 207 | 3,376 | 309 | 883 | 75 | 2,443 | 110 |
| chậm (slow) | 347 | 418 | 1,325 | 124 | 643 | 585 | 1,915 | 360 | 612 | 207 | 1,973 | 237 | 819 | 145 | 2,608 | 305 | 496 | 75 | 2,380 | 58 |
| kém (less) | 363 | 219 | 420 | 36 | 746 | 516 | 722 | 185 | 987 | 216 | 831 | 178 | 1,250 | 139 | 1,201 | 150 | 406 | 25 | 678 | 42 |
| đắt (expensive) | 1,100 | 200 | 1,174 | 57 | 663 | 451 | 902 | 155 | 329 | 92 | 727 | 137 | 380 | 75 | 1,141 | 124 | 255 | 28 | 596 | 56 |
| ít (little) | 186 | 150 | 184 | 94 | 301 | 280 | 392 | 187 | 283 | 102 | 518 | 99 | 609 | 77 | 892 | 64 | 377 | 30 | 624 | 43 |
| gần (near) | 260 | 79 | 362 | 45 | 279 | 174 | 476 | 171 | 259 | 63 | 555 | 106 | 451 | 41 | 971 | 137 | 226 | 13 | 580 | 29 |
| khó (difficult) | 167 | 155 | 195 | 31 | 285 | 233 | 420 | 85 | 392 | 141 | 569 | 73 | 744 | 105 | 588 | 65 | 350 | 21 | 315 | 32 |

**Fig. 9.** Top 10 negative words in the period (01/2018 - 08/2022)

In terms of words with a nuance-related aspect, Fig. 10 shows the top 10 words in the period (01/2018 - 08/2022) for the four applications including "app", "order", "cost", "customer", "error", "voucher", "advertising", "phone", "time", and "pr*ck". Based on the 2022 data, the positive and negative related aspects that all four applications need to care about are as follows "app", which includes feedback regarding the use of the application, ease of use, access speed, and the response on applications. This aspect group had the highest occurrence frequency in the top 10 aspect-related words.

| | At / Appid | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2018 | | | | 2019 | | | | 2020 | | | | 2021 | | | | 2022 | | | |
| Word | lazada | sendo | shopee | tiki | lazada | sendo | shopee | tiki | lazada | sendo | shopee | tiki | lazada | sendo | shopee | tiki | lazada | sendo | shopee | tiki |
| ứng_dụng (app) | 883 | 754 | 1,433 | 381 | 1,927 | 1,856 | 3,576 | 831 | 2,319 | 745 | 4,271 | 595 | 4,304 | 505 | 5,737 | 483 | 2,021 | 129 | 3,895 | 136 |
| đơn (order) | 924 | 700 | 1,690 | 245 | 1,165 | 1,235 | 1,856 | 846 | 1,373 | 644 | 2,300 | 492 | 2,862 | 498 | 5,220 | 755 | 1,733 | 142 | 3,987 | 198 |
| phí (cost) | 3,670 | 808 | 2,967 | 83 | 2,618 | 1,853 | 2,345 | 249 | 1,210 | 360 | 2,469 | 211 | 1,154 | 223 | 2,245 | 152 | 666 | 45 | 1,064 | 29 |
| khách_hàng (customer) | 651 | 492 | 623 | 121 | 1,266 | 1,052 | 1,385 | 458 | 1,468 | 583 | 2,221 | 379 | 2,572 | 371 | 3,178 | 485 | 1,117 | 100 | 1,732 | 132 |
| lỗi (error) | 450 | 185 | 901 | 225 | 447 | 367 | 1,712 | 329 | 502 | 207 | 1,908 | 264 | 1,184 | 94 | 3,409 | 191 | 669 | 26 | 5,976 | 70 |
| mã (voucher) | 190 | 108 | 630 | 86 | 296 | 529 | 1,382 | 140 | 397 | 224 | 2,003 | 153 | 1,025 | 153 | 2,887 | 125 | 648 | 29 | 3,137 | 54 |
| quảng_cáo (advertising) | 202 | 182 | 437 | 51 | 495 | 450 | 780 | 130 | 742 | 178 | 919 | 151 | 3,969 | 111 | 2,193 | 75 | 1,690 | 35 | 1,352 | 20 |
| điện_thoại (phone) | 271 | 357 | 1,319 | 104 | 401 | 699 | 1,449 | 223 | 475 | 267 | 1,624 | 186 | 986 | 161 | 2,379 | 138 | 469 | 26 | 1,897 | 46 |
| giờ (time) | 682 | 188 | 575 | 64 | 512 | 610 | 930 | 220 | 516 | 209 | 1,042 | 197 | 1,012 | 97 | 1,713 | 205 | 524 | 25 | 1,171 | 53 |
| c*c (pr*ck) | 246 | 100 | 335 | 49 | 521 | 282 | 738 | 115 | 637 | 109 | 651 | 111 | 1,556 | 64 | 1,774 | 80 | 977 | 21 | 1,909 | 34 |

**Fig. 10.** Top 10 aspects related to positive and negative nuances in the period (01/2018 – 08/2022)

### 3.4. Discussion

The experimental results have clarified the research model in Fig. 1 for the analysis of positive and negative aspects and nuances of customers toward mobile applications of commercial shops through feedback, comments, and reviews. In particular, when analyzing aspects related to negative nuances, the results show that the majority of opinions focus on six main issues related to the shops' products and services: (1) product quality is not as advertised; (2) the process of returning products is difficult and time consuming such that customers do not know when they will be refunded; (3) the mobile application is difficult to use; (4) when there is a response, the customer calls the switchboard and waits for a long time; (5) service attitude errors of staff, as there are promises that are not fulfilled; (6) the transportation of goods causes damage and long waiting times.

Based on customer feedback, the study also analyzed and found six solutions that can be recommended for shops to improve product and service quality based on the positive aspects and nuances deduced from the customers' opinions: (1) reminding about and taking proactive measures against stalls in stores; (2) updating customers on a regular basis about the product return process, what stage the return is in, and when to refund and through which channels; (3) providing specific instructions for customers via video, images, or directly on the mobile application when the customer makes a return or complaint; (4) preparing staffing plans during peak hours or promotional campaign days to always have sufficient numbers of customer support staff. In the future, stores should develop a system of call centers that automatically receive and answer calls or build AI chatbots to support customers in a timely manner; (5) provide continuous reviews of quality monitoring, listening to call recordings until the end, and viewing replies to emails and messages. Moreover, shops should identify common mistakes and provide training sessions for employees to follow the correct process, learn how to control emotions, and help them better understand customer wishes; (6) continuous checks of the shipping process, stowage, and the shipper's performance through customer reviews. In addition, there should be detailed instructions on how to pack items so that stores selling on mobile applications can properly pack products.

## 4. Conclusion

Based on the research and analysis results from our dataset of customer opinions about e-commerce products and services from the four largest mobile commerce applications in Vietnam (i.e., Lazada, Shopee, Sendo, and Tiki) on the Google Play Store from 01/2013 to 08/2022, the experimental results demonstrate that the proposed Bi-LSTM model achieved the best model among all existing models, with a high F1 score of 94.61%. The result of the Bi-LSTM model can be attributed to its ability to learn long-term dependencies and understand the context of the text, which is crucial in identifying sentiment. Additionally, the study obtained research results of scientific and practical significance with the following contributions: (1) providing a dataset of customer reviews specific to the Vietnamese e-commerce industry; (2) developing dictionaries and techniques to address pre-processing issues with Vietnamese text; (3) proposing a hybrid model for classifying sentiment and extracting aspects of products and services from reviews. These results have significant implications for both researchers and businesses. For instance, the model can be employed to build dashboards to monitor online customer experiences in real-time, assist administrators in making timely decisions, enhance the quality of products and services, and improve customer satisfaction.

The study has some limitations that could be addressed in future research. The model focuses on considering the positive and negative nuances and did not analyze neutral nuances. To expand the model's capabilities: (1) planning to incorporate more neutral nuances in the Vietnamese language; (2) improving the feature extraction for the network. An integrated word embedding approach may produce better results; (3) using pre-trained models and applying them to large datasets.

## Declarations

**Author contribution.** All authors contributed equally to the main contributor to this paper. All authors read and approved the final paper.
**Funding statement.** None of the authors have received any funding or grants from any institution or funding body for the research.
**Conflict of interest.** The authors declare no conflict of interest.
**Additional information.** No additional information is available for this paper.

## References

[1] "The Map of E-commerce in Vietnam," *iprice*, 2019. Accessed Apr. 19. 2021. [Online]. Available at: https://iprice.vn/insights/mapofecommerce/en/.

[2] B. Liu, M. Hu, and J. Cheng, "Opinion observer," in *Proceedings of the 14th international conference on World Wide Web - WWW '05*, 2005, p. 342, doi: 10.1145/1060745.1060797.

[3] "Vietnamese language," *Britannica*. [Online]. Available at: https://www.britannica.com/topic/Vietnamese-language.

[4] Y. Nurdiansyah, S. Bukhori, and R. Hidayat, "Sentiment analysis system for movie review in Bahasa Indonesia using naive bayes classifier method," *J. Phys. Conf. Ser.*, vol. 1008, no. 1, p. 012011, Apr. 2018, doi: 10.1088/1742-6596/1008/1/012011.

[5] Y. Al Amrani, M. Lazaar, and K. E. El Kadiri, "Random Forest and Support Vector Machine based Hybrid Approach to Sentiment Analysis," *Procedia Comput. Sci.*, vol. 127, pp. 511–520, Jan. 2018, doi: 10.1016/j.procs.2018.01.150.

[6] N. K. Bolbol and A. Y. Maghari, "Sentiment Analysis of Arabic Tweets Using Supervised Machine Learning," in *2020 International Conference on Promising Electronic Technologies (ICPET)*, Dec. 2020, pp. 89–93, doi: 10.1109/ICPET51420.2020.00025.

[7] B. Noori, "Classification of Customer Reviews Using Machine Learning Algorithms," *Appl. Artif. Intell.*, vol. 35, no. 8, pp. 567–588, Jul. 2021, doi: 10.1080/08839514.2021.1922843.

[8] M. Khan and K. Malik, "Sentiment Classification of Customer's Reviews About Automobiles in Roman Urdu," in *Advances in Intelligent Systems and Computing*, vol. 887, Springer Verlag, 2019, pp. 630–640, doi: 10.1007/978-3-030-03405-4_44.

[9] M. A. Qureshi *et al.*, "Sentiment Analysis of Reviews in Natural Language: Roman Urdu as a Case Study," *IEEE Access*, vol. 10, pp. 24945–24954, 2022, doi: 10.1109/ACCESS.2022.3150172.

[10] S. Goyal, "Review Paper on Sentiment Analysis of Twitter Data Using Text Mining and Hybrid Classification Approach," *Internatio nal J. Eng. Dev. Res.*, vol. 5, no. 2, pp. 2321–9939, 2017, [Online]. Available: https://www.ijedr.org/papers/IJEDR1702032.pdf.

[11] A. Bayhaqy, S. Sfenrianto, K. Nainggolan, and E. R. Kaburuan, "Sentiment Analysis about E-Commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and Naïve Bayes," in *2018 International Conference on Orange Technologies (ICOT)*, Oct. 2018, pp. 1–6, doi: 10.1109/ICOT.2018.8705796.

[12] P. Cen, K. Zhang, and D. Zheng, "Sentiment Analysis Using Deep Learning Approach," *J. Artif. Intell.*, vol. 2, no. 1, pp. 17–27, Jul. 2020, doi: 10.32604/jai.2020.010132.

[13] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[14] Q. Lu, "Neural Network based Approaches for Aspect-Based Sentiment Analysis," *Highlights Sci. Eng. Technol.*, vol. 12, pp. 222–229, Aug. 2022, doi: 10.54097/hset.v12i.1457.

[15] H. Li, Y. Ma, Z. Ma, and H. Zhu, "Weibo Text Sentiment Analysis Based on BERT and Deep Learning," *Appl. Sci.*, vol. 11, no. 22, p. 10774, Nov. 2021, doi: 10.3390/app112210774.

[16] N. S. Tun, N. N. Long, T. Tran, N. T. Thao, D. T. Thu Phuong, and T. Nguyen, "Stock article title sentiment-based classification using PhoBERT," in *CEUR Workshop Proceedings*, 2021, vol. 3026, pp. 225–233.[Online]. Available at: https://ceur-ws.org/Vol-3026/paper25.pdf.

[17] Q. T. Nguyen, T. L. Nguyen, N. H. Luong, and Q. H. Ngo, "Fine-Tuning BERT for Sentiment Analysis of Vietnamese Reviews," in *2020 7th NAFOSTED Conference on Information and Computer Science (NICS)*, Nov. 2020, pp. 302–307, doi: 10.1109/NICS51282.2020.9335899.

[18] T.-L. Truong, H.-L. Le, and T.-P. Le-Dang, "Sentiment Analysis Implementing BERT-based Pre-trained Language Model for Vietnamese," in *2020 7th NAFOSTED Conference on Information and Computer Science (NICS)*, Nov. 2020, pp. 362–367, doi: 10.1109/NICS51282.2020.9335912.

[19] W. Zhang, X. Li, Y. Deng, L. Bing, and W. Lam, "A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges," in *IEEE Transactions on Knowledge and Data Engineering*, 2022, pp. 1–20, doi: 10.1109/TKDE.2022.3230975.

[20] A. S. Shafie, N. M. Sharef, M. A. Azmi Murad, and A. Azman, "Aspect Extraction Performance with POS Tag Pattern of Dependency Relation in Aspect-based Sentiment Analysis," in *2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP)*, Mar. 2018, pp. 1–6, doi: 10.1109/INFRKM.2018.8464692.

[21] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, Aug. 2004, pp. 168–177, doi: 10.1145/1014052.1014073.

[22] T. A. Rana and Y.-N. Cheah, "Exploiting sequential patterns to detect objective aspects from online reviews," in *2016 International Conference On Advanced Informatics: Concepts, Theory And Application (ICAICTA)*, Aug. 2016, pp. 1–5, doi: 10.1109/ICAICTA.2016.7803101.

[23] K. Srividya and A. M. S. Sowjanya, "Aspect Based Sentiment Analysis using POS Tagging and TFIDF," *Int. J. Eng. Adv. Technol.*, vol. 8, no. 6, pp. 1960–1963, Aug. 2019, doi: 10.35940/ijeat.F7935.088619.

[24] L. Mai and B. Le, "Aspect-Based Sentiment Analysis of Vietnamese Texts with Deep Learning," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10751 LNAI, Springer Verlag, 2018, pp. 149–158, doi: 10.1007/978-3-319-75417-8_14.

[25] L. Luc Phan *et al.*, "SA2SL: From Aspect-Based Sentiment Analysis to Social Listening System for Business Intelligence," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12816 LNAI, Springer Science and Business Media Deutschland GmbH, 2021, pp. 647–658, doi: 10.1007/978-3-030-82147-0_53.

[26] B. Liu, "Many Facets of Sentiment Analysis," in *A Partical Guide to Sentiment Analysis*, Springer, Cham, 2017, pp. 11–39, doi: 10.1007/978-3-319-55394-8_2.

[27] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," in *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings*, Jan. 2013, pp. 1–12, doi: 10.48550/arXiv.1301.3781.

[28] J. Pennington, R. Socher, and C. 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.

[29] M. M. Rahman, Y. Watanobe, and K. Nakamura, "A Bidirectional LSTM Language Model for Code Evaluation and Repair," *Symmetry (Basel).*, vol. 13, no. 2, p. 247, Feb. 2021, doi: 10.3390/sym13020247.

[30] T. O'Malley, E. Bursztein, J. Long, F. Chollet, H. Jin, and L. Invernizzi, "A Hyperparameter Tuning Library for Keras," *GitHub*, 2019. Accessed Apr. 30, 2020. [Online]. Available at: https://github.com/keras-team/keras-tuner.

[31] B.-T. Nguyen-Thi and H.-T. Duong, "A Vietnamese Sentiment Analysis System Based on Multiple Classifiers with Enhancing Lexicon Features," in *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST*, vol. 293, Springer Verlag, 2019, pp. 240–249, doi: 10.1007/978-3-030-30149-1_20.

[32] D. R. Cox, "The Regression Analysis of Binary Sequences," *J. R. Stat. Soc. Ser. B*, vol. 20, no. 2, pp. 215–232, Jul. 1958, doi: 10.1111/j.2517-6161.1958.tb00292.x.

[33] D. P. Kingma and J. L. Ba, "Adam: A Method for Stochastic Optimization," *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. International Conference on Learning Representations, ICLR, pp. 1–15, Dec. 22, 2014, doi: 10.48550/arXiv.1412.6980.