

Article

Not peer-reviewed version

Application of Deep Learning in Aspect-Based Sentiment Analysis from E-Commerce User Reviews

[Quỳnh Phạm](#) *

Posted Date: 15 May 2025

doi: 10.20944/preprints202505.1235.v1

Keywords: sentiment analysis; deep learning; PhoBERT; ABSA; E-Commerce



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Application of Deep Learning in Aspect-Based Sentiment Analysis from E-Commerce User Reviews

Quynh Thi Thuy Pham

Ho Chi Minh City University of Transport; thuyquynh.phamthi2410@gmail.com

Abstract: In the context of the rapid growth of e-commerce, understanding customer emotions and needs through online feedback has become a key factor in strategies to improve product and service quality. On platforms such as Shopee, users often leave reviews after experiencing products - these comments not only reflect customer satisfaction but also provide detailed information regarding various aspects such as price, quality, shipping, and customer service. This study leverages over 11,000 customer reviews on many shoe-related products from the e-commerce platform - Shopee, annotated across 8 different aspects and classified into 4 sentiment categories: positive, negative, neutral, and undefined. The main objective is to perform Aspect-Based Sentiment Analysis (ABSA) to uncover insights into user experiences and propose improvements to products and services. By combining advanced text preprocessing techniques, data visualization, and modern machine learning models, this research not only evaluates overall sentiment but also analyzes detailed factors influencing consumer purchasing decisions and satisfaction. This study contributes to the development of Vietnamese natural language processing and opens up future research directions for low-resource languages.

Keywords: sentiment analysis; deep learning; PhoBERT; ABSA; E-commerce

1. Introduction

In recent years, the strong development of e-commerce in Vietnam has led to an explosion in the number of online product reviews. E-commerce platforms such as Shopee¹, Lazada², Shein³ hay Tiktok Shop⁴ record millions of reviews annually, with the fashion sector, especially footwear, accounting for a significant proportion, thereby creating a rich source of user experience data. According to statistics, the proportion of Vietnamese consumers participating in online shopping and leaving feedback has been steadily increasing, particularly after the COVID-19 pandemic, and is expected to continue to grow in the coming years. The explosion in the volume of reviews has led to an urgent need for automatic analysis, mining of sentiments, and extraction of insights from these feedbacks to enhance product quality and service.

Aspect-Based Sentiment Analysis (Do et al., 2019) - a subfield of natural language processing (NLP) - plays a crucial role in extracting subjective information from text to determine consumer attitudes. It is an essential tool for businesses to improve product quality, enhance services, and optimize customer experiences. However, Vietnamese sentiment analysis still faces many limitations compared to other languages due to the complexity of its syntax, tonal system, and the lack of large-scale annotated datasets (Van Thin et al., 2023).

To contribute to addressing this challenge, this study focuses on developing a sentiment analysis model based on a dataset of over 11,000 user comments on footwear products from the Shopee

¹ <https://shopee.vn>

² <https://www.lazada.com>

³ <https://www.shein.com>

⁴ <https://shop-vn.tiktok.com>

platform. The data is labeled across 8 aspects (Price, Shipping, Outlook, Quality, Size, Shop_Service, General, and Others) and classified into 4 sentiment levels (Positive, Negative, Neutral), with an emphasis on Aspect-Based Sentiment Analysis. We propose a novel sentiment analysis framework specifically designed for Vietnamese shoes reviews on Shopee, integrating PhoBERT⁵ with various preprocessing techniques to optimize the model's ability to learn Vietnamese text representations. To demonstrate the effectiveness of the proposed approach, we also train and compare SVM (Xue et al., 2009) and Naive Bayes (Webb et al., 2010) classifiers on the same dataset. These models are selected due to their proven effectiveness in text classification tasks, serving as reference baselines to evaluate the extent of performance improvement when applying Transformer-based models such as PhoBERT. The objective of this study is not only to accurately classify the overall sentiment but also to detect detailed aspect-based sentiments, thereby extracting valuable insights into consumer experiences to support businesses in product improvement and customer satisfaction enhancement.

The remainder of this paper is organized as follows: **Section 2** presents the theoretical background related to sentiment analysis and aspect-based sentiment analysis in the context of the Vietnamese language. **Section 3** describes the proposed method, focusing on integrating preprocessing techniques with the PhoBERT model and setting up the training process with the Focal Loss function to handle data imbalance. **Section 4** presents the experimental setup, the results obtained from both traditional and deep learning models, and a detailed analysis of model performance based on multiple evaluation metrics. Finally, **Section 5** concludes the paper, summarizing the key findings and proposing potential future research directions for Vietnamese sentiment analysis.

2. Theoretical Background

2.1. Sentiment Analysis

Sentiment analysis is a research field that intersects natural language processing, text mining, and computational linguistics (Mejova, 2009). The goal of sentiment analysis is to identify and classify linguistic expressions that convey private states such as emotions, opinions, or speculations (Dang et al., 2023). Sentiment in text can appear explicitly or implicitly, with polarity typically categorized into two groups: positive and negative. But sometimes expressed as a continuous range of intensity. A major challenge of sentiment analysis lies in accurately identifying the target towards which the sentiment is directed, as well as distinguishing the author's true sentiment, particularly in complex quotation structures. In practice, much of the research has focused on product reviews, where feature extraction - such as the design, material, durability, or size of footwear - plays an important role in enabling detailed aspect-based sentiment analysis.

2.2. Aspect-Based Sentiment Analysis

Aspect-Based Sentiment Analysis (ABSA) is an advanced technique in natural language processing (NLP) that enables deeper exploration of consumer reviews by identifying sentiments toward specific aspects of a product rather than analyzing only the overall sentiment (Tran et al., 2022). In the field of e-commerce, particularly on platforms such as Shopee, a user review often refers to multiple factors such as price, quality, delivery, and customer service, rather than providing a general evaluation. For example, in the sentence "The shoes are beautiful but the delivery is slow" an ABSA model would detect positive sentiment towards the "design" aspect and negative sentiment towards the "delivery" aspect. ABSA typically involves two main steps: (1) aspect extraction and (2) sentiment classification corresponding to each extracted aspect..

⁵ <https://github.com/VinAIRResearch/PhoBERT>

2.3. PhoBERT – Pre-Trained Language Models for Vietnamese

PhoBERT is a state-of-the-art language model developed specifically for Vietnamese, in which "Pho" (Phở) refers to a famous traditional dish of Vietnam (Nguyen & Tuan Nguyen, 2020). PhoBERT was pre-trained based on the RoBERTa method (Liu et al., 2019), an improvement over BERT (Devlin et al., 2019) aimed at optimizing pre-training performance by increasing the amount of input data and modifying certain training techniques. PhoBERT has two versions, including PhoBERT-base and PhoBERT-large, which adopt the same architecture as BERT-base and BERT-large, respectively. A key highlight of PhoBERT is that it was trained on the largest monolingual Vietnamese corpus to date (comprising Vietnamese Wikipedia data and a large news dataset after deduplication). The raw data were preprocessed by segmenting words and sentences using the VnCoreNLP tool, and then encoded into subwords using fastBPE with a vocabulary of 64,000 subwords. The success of PhoBERT demonstrates the importance of developing specialized pre-trained language models for each language, especially for low-resource languages such as Vietnamese.

3. Data preparation and text preprocessing

3.1. Dataset

In this study, we use an open-source dataset from Kaggle⁶, which was collected from user reviews of various footwear products on the e-commerce platform Shopee, comprising approximately 11,700 user feedback entries. The data are labeled with four different sentiment categories and eight specific aspects. Table 1 defines the aspects included in the dataset.

Table 1. Definition of Aspects in the Dataset.

Khía cạnh	Định nghĩa
Price	User feedback is related to price (high/low/....)
Shipping	User feedback related to shipping service (fast/slow/....)
Outlook	User feedback related to the shoe's appearance (beautiful/ugly/dirty/....)
Quality	User feedback regarding the quality of the product (good/bad/damaged/....)
Size	User feedback regarding shoe size (tight/medium/small/wide/....)
Shop_Service	User feedback related to the quality of the seller's customer care (satisfied/unsatisfied/not good/....)
General	User feedback regarding the overall condition of the shoe (ok/ no problem/....)
Others	User feedback is not related to the product or aspects mentioned above

As shown in Figure 1, the dataset captures multiple aspects simultaneously within a single review. Each user comment may express sentiments toward several product attributes, such as price, shipping, service, Outlook, quality, size, and service. ...

⁶ <https://www.kaggle.com/>

# Review	# Price	# Shipping	# Outlook	# Quality	# Size	# Shop_Serv...	# General	# Others
Giày hơi có mùi nồng, lưu ý đôi LA không phải để xam nên mng cần nhắc kĩ nhê a nhma với giá tiền này...	-1	-1	-1	2	-1	-1	2	-1
Hàng về đẹp lắm nha ship thần thiên đi giày vừa in❤	-1	1	1	-1	-1	-1	-1	-1
Hàng ok nên mua đây rất đẹp	-1	-1	1	-1	-1	-1	2	-1
Bun . OTI gửi Oke số 6 đi sisdđ. Bôi đi được đạn sisdđ ke à đi dính sưng.	-1	-1	-1	-1	-1	-1	-1	2
Màu đẹp giống trong hình mọi người nên mua nha Minh thấy cực kỳ đẹp luôn ý	-1	-1	1	-1	-1	-1	-1	-1

Figure 1. Examples of the dataset.

3.2. Data Preprocessing

Data preprocessing is a critical stage aimed at enhancing the quality of the input dataset for machine learning models. The performed steps include:

- **Converting all text to lowercase:** This eliminates unnecessary distinctions between uppercase and lowercase letters.
- **Removing null values and invalid comments:** Empty comments or those containing inappropriate content are removed from the dataset.
- **Standardizing vocabulary and language:** The text is normalized by replacing abbreviations, slang, and spelling errors with standard Vietnamese expressions. We use a customized dictionary specifically designed for e-commerce data related to footwear. For example, "giày đẹp vl" is standardized to "giày rất đẹp".
- **Converting emojis and punctuation:** Symbols such as ":", "<3", and "^^" are converted into explicit emotional expressions ("vui vẻ", "yêu thích", "hài lòng").
- **Word segmentation:** Texts are segmented using the VNCoreNLP tool to ensure that Vietnamese compound words such as "giày thể thao" are correctly recognized rather than split into separate words like "giày", "thể", "thao".
- **Removing short comments:** Comments with fewer than five words are considered insufficiently informative and are removed to improve training accuracy. Additionally, generic or vague comments containing keywords like "ok", "ổn", "rồi", "bình thường", "tạm", "được", "giao" are also filtered out to ensure specificity in sentiment analysis.

After preprocessing, the dataset achieves consistency and cleanliness, providing a solid foundation for the machine learning phase.



Figure 2. Word cloud of footwear product data on the Shopee platform.

Before training the model, we performed data balancing using the *Oversampling* technique. For each aspect, the data were grouped according to different sentiment labels. Then, the group with fewer samples was randomly duplicated until it matched the number of samples in the majority class within the same aspect. This process ensures that each sentiment class within every aspect is evenly distributed, preventing the model from being biased during training. This preprocessing approach enhances the model’s ability to learn from minority classes while maintaining the diversity of aspects and user feedback in the training dataset.

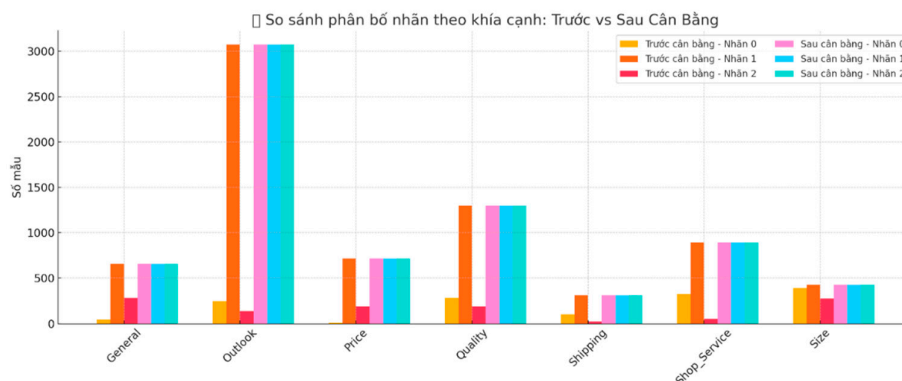


Figure 3. Distribution of sentiment classes: before *vs.* after oversampling.

In addition, the study applies a class weighting technique based on the frequency of labels in the training set, which is integrated into the Focal Loss (Lin et al., 2018) function to optimize the classification performance for minority classes. The input data are tokenized using the PhoBERT-base tokenizer, with a maximum sequence length limited to 256 tokens to ensure input size consistency. Furthermore, Gradient Clipping (Chen et al., 2020) is employed to prevent the gradient explosion phenomenon during the training process.

4. Experimental Results

4.1. Experimental Setup

In this study, I develop a training pipeline for the PhoBERT-base model based on aspect-based sentiment analysis. The main steps include data preprocessing, model training using the Focal Loss technique and performance evaluation based on comprehensive metrics. The workflow is divided into three stages: data preprocessing, model training, and result evaluation. All experiments were

conducted in the Google Colab Pro environment with Tesla T4 GPU hardware. The PhoBERT-base-v2 model was loaded from the Huggingface Transformers library (Wolf et al., 2020) and fine-tuned for the aspect-based sentiment analysis task.

The model training was performed using the following optimized hyper parameters: the number of epochs is set to 8, the batch size is 16, the learning rate is set at 2^{-5} , and epsilon was configured to 1^{-8} to enhance the stability during weight updates. The AdamW optimizer (Loshchilov & Hutter, 2019) was used in combination with a linear learning rate scheduler with warmup steps to mitigate early overfitting.

Evaluation Metrics

To evaluate the effectiveness of the models, we use commonly adopted metrics in classification tasks, including:

- **Accuracy:** The ratio of the total number of correct predictions to the total number of instances in the test set. This metric reflects the overall correctness of the model.
- **Precision:** The average of precision scores computed separately for each class. Precision measures the proportion of correctly predicted positive instances over all instances predicted as positive, helping assess the model's ability to avoid false positives.
- **Recall:** The average of recall scores computed separately for each class. Recall evaluates the model's ability to correctly identify positive instances within the actual dataset.
- **F1-score:** The average of F1-scores computed separately for each class. The F1-score is the harmonic mean of precision and recall, balancing both metrics, especially important in imbalanced datasets.

All metrics are macro-averaged (unweighted average across classes), ensuring that classes with different numbers of samples are fairly evaluated.

4.2. Experimental Results

The experimental results demonstrate that the PhoBERT model exhibits superior performance due to its ability to capture deep semantic meaning and contextual information in Vietnamese. Notably, the model maintains high accuracy even when processing reviews with informal or non-standard language structures, which are commonly found in product evaluations on e-commerce platforms.

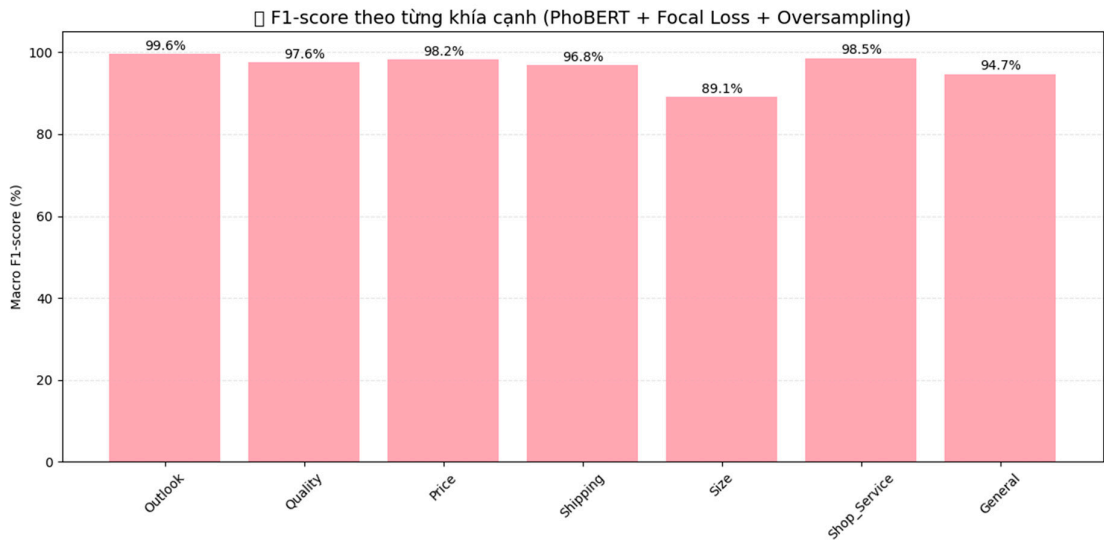


Figure 4. F1-scores by aspect for the proposed model.

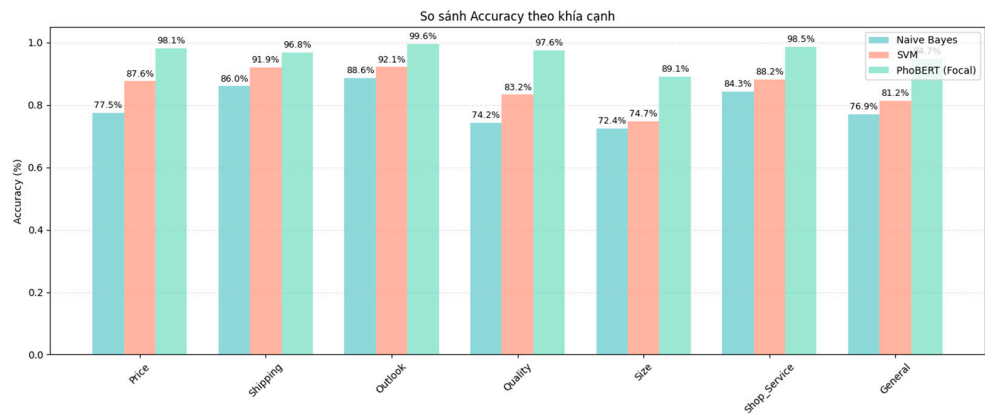


Figure 5. Accuracy values by aspect for PhoBERT in comparison with two traditional models, SVM and Naive Bayes.

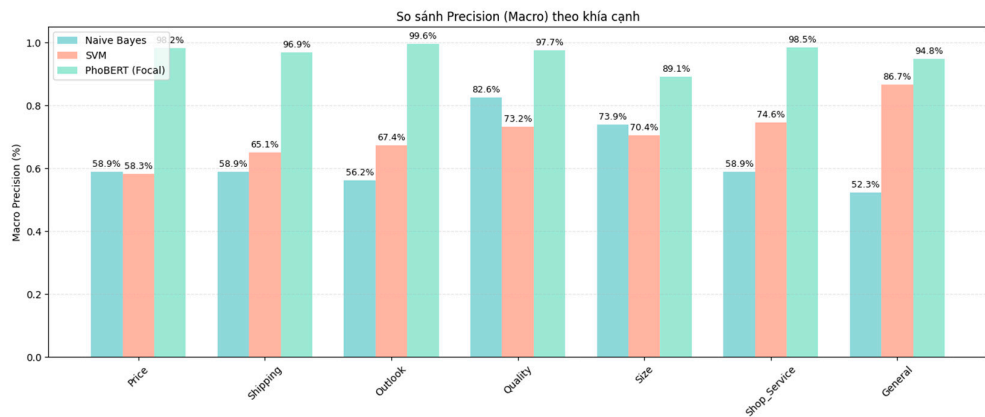


Figure 6. Comparison of macro-averaged Precision values by aspect among PhoBERT, SVM, and Naive Bayes.

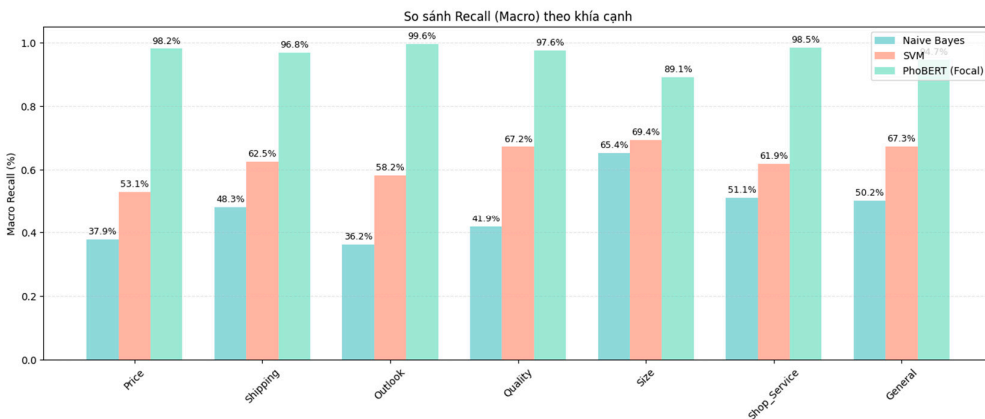


Figure 7. Comparison of macro-averaged Recall values by aspect among PhoBERT, SVM, and Naive Bayes.

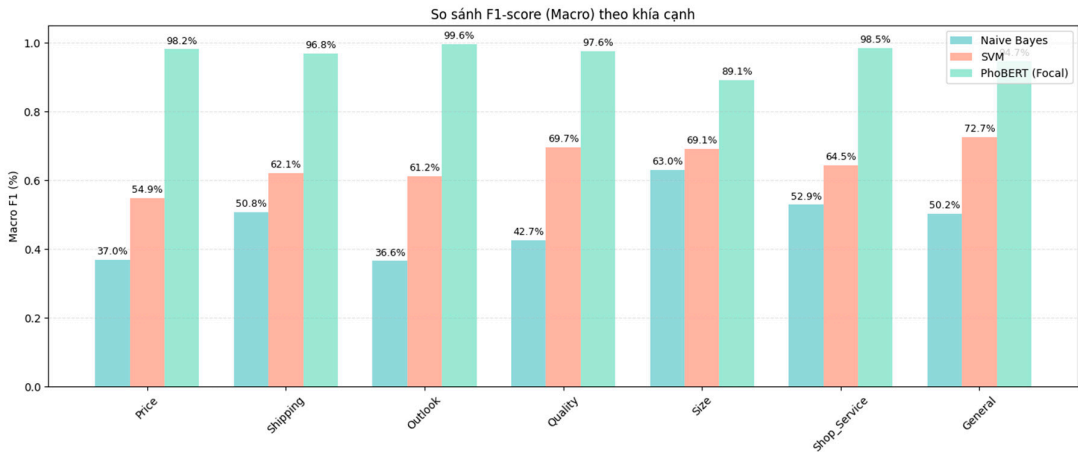


Figure 8. Macro-averaged F1-score by aspect for PhoBERT, SVM, and Naive Bayes.

The comparative evaluation across multiple metrics, including Macro F1-score, Macro Recall, Macro Precision, and Accuracy, highlights the superior effectiveness of the proposed PhoBERT (Focal) model over traditional machine learning approaches such as Naive Bayes and SVM. In terms of Macro F1-score and Recall, PhoBERT consistently achieved outstanding results, with values approaching or exceeding 98% across most aspects. This demonstrates the model’s strong ability to accurately identify and comprehensively cover different sentiment groups. The results on Precision further reinforce that PhoBERT not only correctly detects most relevant cases but also minimizes false predictions, with Macro Precision values exceeding 96% across all aspects.

The experimental results indicate that the PhoBERT-base model, combined with preprocessing techniques and the Focal Loss function, achieved high performance in the task of aspect-based Vietnamese sentiment analysis. The model effectively captures the complex linguistic features of Vietnamese and handles label imbalance well within the dataset. Specifically, it outperforms traditional machine learning approaches such as Naive Bayes and SVM across key evaluation metrics, including Accuracy, Macro Precision, Macro Recall, and Macro F1-score. These findings confirm the strength of modern deep learning methods when applied to Vietnamese natural language processing tasks. The combination of a pre-trained language model (PhoBERT) with Vietnamese-specific preprocessing techniques has significantly improved prediction quality and opened up broad potential for practical applications in customer feedback analysis systems.

5. Conclusions

In this study, I developed and evaluated an aspect-based sentiment analysis model for Vietnamese, based on the PhoBERT-base architecture combined with Focal Loss and advanced natural language preprocessing techniques. The model demonstrated strong sentiment classification capabilities on product review datasets from e-commerce platforms.

Despite achieving promising results, there remains significant potential for further improvement. Future research directions include: Exploring more advanced deep learning architectures such as transformers or attention-based networks; Experimenting with transfer learning techniques to adapt to new data domains; Applying ensemble methods by combining multiple strong models to enhance accuracy; Developing larger and more diverse annotated datasets across various domains beyond e-commerce; and Extending the scope of sentiment analysis to more complex tasks such as topic modeling or spam/invalid comment detection. I believe that pursuing these research directions will contribute to building more efficient and comprehensive Vietnamese natural language processing systems in the future. This work lays a strong foundation for future advancements in Vietnamese sentiment analysis and highlights the potential of deep learning methods in tackling complex language understanding tasks.

References

- Chen, X., Wu, S. Z., & Hong, M. (2020). Understanding Gradient Clipping in Private SGD: A Geometric Perspective. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, & H. Lin (Eds.), *Advances in Neural Information Processing Systems* (Vol. 33, pp. 13773–13782). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2020/file/9ecff5455677b38d19f49ce658ef0608-Paper.pdf
- Dang, C., Moreno-García, M. N., De la Prieta, F., Nguyen, K. V., & Ngo, V. M. (2023). Sentiment Analysis for Vietnamese – Based Hybrid Deep Learning Models. *Preprints*. <https://doi.org/10.20944/preprints202306.1318.v1>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. <https://arxiv.org/abs/1810.04805>
- Do, H. H., Prasad, P. W., Maag, A., & Alsadoon, A. (2019). Deep learning for aspect-based sentiment analysis: A comparative review. *Expert Systems with Applications*, 118, 272–299.
- Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2018). Focal Loss for Dense Object Detection. <https://arxiv.org/abs/1708.02002>
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. <https://arxiv.org/abs/1907.11692>
- Loshchilov, I., & Hutter, F. (2019). Decoupled Weight Decay Regularization. <https://arxiv.org/abs/1711.05101>
- Mejova, Y. (2009). Sentiment analysis: An overview. *University of Iowa, Computer Science Department*, 5, 1–34.
- Nguyen, D. Q., & Tuan Nguyen, A. (2020). PhoBERT: Pre-trained language models for Vietnamese. In T. Cohn, Y. He, & Y. Liu (Eds.), *Findings of the Association for Computational Linguistics: EMNLP 2020* (pp. 1037–1042). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.findings-emnlp.92>
- Tran, Q.-L., Le, P. T. D., & Do, T.-H. (2022). Aspect-based sentiment analysis for Vietnamese reviews about beauty product on E-commerce websites. *Proceedings of the 36th Pacific Asia Conference on Language, Information and Computation*, 767–776.
- Van Thin, D., Hao, D. N., & Nguyen, N. L.-T. (2023). Vietnamese Sentiment Analysis: An Overview and Comparative Study of Fine-tuning Pretrained Language Models. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, 22(6). <https://doi.org/10.1145/3589131>
- Webb, G. I., Keogh, E., & Mikkilainen, R. (2010). Naïve Bayes. *Encyclopedia of Machine Learning*, 15(1), 713–714.
- Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., Platen, P. von, Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. L., Gugger, S., ... Rush, A. M. (2020). HuggingFace's Transformers: State-of-the-art Natural Language Processing. <https://arxiv.org/abs/1910.03771>
- Xue, H., Yang, Q., & Chen, S. (2009). SVM: Support vector machines. In *The top ten algorithms in data mining* (pp. 51–74). Chapman and Hall/CRC.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.