VIETNAMESE SENTIMENT ANALYSIS ON SOCIAL MEDIA BASED ON BERT ARCHITECTURE

Thi Dung Tran[1], Duc Anh Tran[1], Thi Ngoc Oanh Pham[1], Nhat Tung Le[2]*

[1] Campus in Ho Chi Minh City, University of Transport and Communications
[2] Faculty of Information Technology - Communication, Dong Nai Technology University

dungtt_ph@utc.edu.vn, anhtd_ph@utc.edu.vn, opham414@gmail.com, lenhattung@dntu.edu.vn

ABSTRACT — This study conducts a comparative evaluation of machine learning models for Vietnamese sentiment analysis, including ViBERT4News, PhoBERT, XLM-RoBERTa, ViBERT, SVM, and LSTM. The dataset comprises 10,169 text samples collected from Facebook related to students at the Ho Chi Minh City Campus of the University of Transport and Communications (UTC2), labeled into three classes: positive, negative, and neutral. Results show that BERT-based models achieve superior performance compared to traditional methods, with ViBERT4News achieving the highest accuracy (89.19%), followed by PhoBERT (89.06%). The study not only demonstrates the effectiveness of pre-trained language models in processing emotions on Vietnamese social media data but also suggests potential applications in platforms for analyzing and visualizing student feedback, supporting academic monitoring, and improving educational quality.

Keywords — Sentiment analysis, PhoBERT, ViBERT, ViBERT4News, XLM-RoBERTa.

I. INTRODUCTION

Sentiment analysis, also known as opinion mining, is an important research field in natural language processing, focusing on analyzing opinions, emotions, evaluations, assessments, and attitudes of users toward entities such as products, services, issues, and topics expressed in text [1]. In the context of ongoing digital transformation, extracting unstructured data from social media has become increasingly important, helping organizations, businesses, and educational institutions capture timely feedback from stakeholders [2].

In the field of education, traditional survey methods such as evaluation forms, end-of-term surveys, or emails often reveal many limitations: low response rates, formal data nature, and prolonged collection time. According to practical research, the rate of student participation in online surveys only ranges around 30-50%, leading to insufficient basis for schools to adjust policies promptly [3]. In contrast, social media - especially Facebook - is emerging as a channel that truthfully and vividly reflects student life, with dozens to hundreds of posts and comments each week containing valuable information about students' thoughts and aspirations.

For Vietnamese, sentiment analysis faces many specific challenges: flexible grammatical structure, rich word class system, high polysemy, and diverse emotional expression styles [4]. Particularly, social media texts contain many non-standard elements such as abbreviations, emoticons, slang, and writing that does not follow spelling rules.

Recent years have witnessed the superiority of deep learning models, especially the BERT architecture [5] and its variants, in capturing context and complex semantic representations of language.

This study focuses on systematically collecting and analyzing 10,169 data samples from forums and discussion groups on Facebook related to the Ho Chi Minh City Campus of the University of Transport and Communications (UTC2). The main objective is to evaluate the performance of Vietnamese sentiment analysis models: ViBERT4News, PhoBERT, XLM-RoBERTa, ViBERT, SVM, and LSTM in classifying student feedback into three groups: positive, negative, and neutral. In addition to comparing model effectiveness, the study also develops a result visualization system, supporting schools in grasping opinion trends and student concerns, thereby providing a scientific basis for policy planning and improving training quality in the digital age.

II. RELATED WORK

Sentiment analysis has undergone significant development, transitioning from rule-based semantic methods designed through inductive experience and expert knowledge to data-driven learning models. In the early stages, systems mainly exploited emotion dictionaries along with semantic rule sets built based on expert understanding, clearly showing characteristics of knowledge-based approaches [6,7]. The emergence of classical machine learning algorithms such as Naive Bayes, Maximum Entropy, and especially SVM [8] marked an important transition toward supervised learning, significantly improving classification effectiveness [9]. Experiments on movie review datasets showed that machine learning models achieved superior results compared to knowledge-based methods [10]. However, these models still depend on static feature representations, independent of context, which lack the capacity to represent complex semantic relationships and semantic combination phenomena, leading to reduced effectiveness when applied to diverse, flexible, and non-standardized natural corpora in real-world environments.

To address remaining issues in traditional methods, deep learning architectures have been widely applied and become the foundation for the most recent advances in natural language processing. Transformer [11] is a breakthrough architecture that replaces sequential processing mechanisms with self-attention mechanisms, allowing models to efficiently learn dependency relationships between positions in input sequences without being limited by distance, while utilizing parallel computing capabilities to build global semantic representations according to context, and thereby becoming the basis for the development of large-scale pre-trained language models. Based on this, BERT [5] was developed and plays the role of a pre-trained language model with the ability to learn bidirectional contextual representations, thereby significantly improving effectiveness in text classification tasks. Additionally, other neural network architectures such as RvNN with the ability to model hierarchical structures of sentence semantics [12,13], LSTM with long-short term memory mechanisms [14,15], and CNN with the ability to detect local semantic features like emotionally charged phrases [16,17], have also been widely applied in learning semantic representations from unstructured language data. The automatic representation learning capability of deep learning models becomes particularly essential in the context of social media data, where non-standard elements such as slang, emoticons, abbreviations, and syntactic structures that do not comply with grammar frequently appear.

For Vietnamese, this problem poses many specific challenges due to the characteristics of a tonal language, flexible syntactic structure, and the prevalence of non-standard elements in online communication. Initial efforts, typically the hybrid model combining Hierarchical Dirichlet Process (HDP) and SVM, achieved 87% accuracy, showing potential for application in Vietnamese natural language processing [18]. Following this success, many specialized pre-trained language models have been developed, including PhoBERT [19] built on RoBERTa architecture [20] and integrating phonological information, ViBERT4News trained on large-scale journalistic datasets, along with ViBERT [21,22] and XLM-RoBERTa [23]. These models show different levels of effectiveness in processing Vietnamese text.

BERT has demonstrated superior effectiveness in sentiment analysis across multicontextual social media platforms. Hondor Saragih et al., 2024 [24] developed a comprehensive sentiment analysis framework using BERT to process multimedia content, combining both text and multimedia features. This research achieved a high probability in accurate sentiment classification with significantly improved accuracy and low cross-entropy loss.

Vietnamese sentiment analysis using BERT architecture shows promising results, with state-of-the-art performance achieved through innovative pre-trained language model approaches. Hong-Viet Tran et al., 2025 specifically developed a novel sentiment analysis system for Vietnamese using PhoBERT-V2, a RoBERTa-based optimization of the BERT model, combined with SentiWordNet [25]. Their experimental validation on the VLSP 2016 and AIVIVN 2019 datasets demonstrated excellent performance in classifying sentiment across different contexts. Recent studies have also explored multi-class emotion recognition beyond binary sentiment classification. Tran et al. (2024) [26] proposed an emotion recognition system that classifies Vietnamese learner feedback into seven emotion categories (enjoyment, trust, hope, sadness, surprise, fear, and others) using PhoBERT combined with emoji sentiment analysis, achieving 74.1% accuracy on a dataset of 13,590 student opinions. Their work demonstrates the importance of incorporating emojis in emotion recognition, as modern communication on social media platforms frequently uses emojis to express emotions more vividly than text alone.

Despite achieving many advances, most current research still approaches binary classification (positive-negative). However, natural communication, especially on social media, often includes many statements that are neutral or do not clearly express emotions. Moreover, current research mainly relies on general datasets such as product reviews or movie comments, while specific target groups like university students have distinct linguistic characteristics, topics of interest, and ways of expression. Student feedback on digital platforms is often personal, unstructured, and contains many nuances such as satisfaction, disappointment, constructive criticism, or neutrality. Therefore, expanding to three-label classification models (positive, negative, neutral) on specific datasets becomes necessary to more fully reflect the actual emotional spectrum in specific communication contexts.

III. VIETNAMESE SENTIMENT ANALYSIS METHODOLOGY

A. DATASET CONSTRUCTION

1. DATA COLLECTION

Data in this study were collected from Facebook social media platforms associated with student life at UTC2. Main sources include the "UTC2 Confessions" community page, "Listen to Students Speak" forum, along with official and unofficial Facebook groups of Faculties, Departments, and Student Clubs. The collection process was conducted from October 2024 to April 2025 through the Apify automation platform, allowing data extraction in

JSON format with information fields such as post/comment content, interaction numbers (emotional reactions, shares, comments), and posting time. A total of over 10,000 data items (including posts and comments) were collected, including content posted from 2013 to the time of research. Table 1 below presents the detailed distribution of exploited data sources.

Table 1. *Distribution of data collection sources*

Data Source	Number of Samples	Percentage (%)
Listen to Students Speak Forum	4,278	42.07
UTC2 Confessions	1,827	17.97
Faculty/Department and Student Club Groups	4,064	39.96
Total	10,169	100

All collected data was processed through a procedure to remove personally identifiable information to ensure no elements that could trace user identity were retained. The study only used publicly available data sources or data with transparent consent from participants, while strictly complying with ethical standards in social science research and current legal regulations related to privacy rights and personal information security.

2. DATA STANDARDIZATION

The data standardization process is designed to normalize input corpora and minimize information noise, thereby ensuring consistency and improving machine learning model performance. Details of the process are shown in Table 2.

Table 2. *Summary of data standardization process*

Processing Step	Description	Tools/Resources
Standardize abbreviations	Replace slang and abbreviations with standard forms	Self-built dictionary (520 pairs)
Clean special characters	Remove special characters, URLs, formats, non-standard symbols	Regular expressions
Process emoticons	Map 63 emojis to 3 basic emotion groups	Emoji dictionary (63 emojis)
Remove stop words	Remove words without semantic value	Stop word list (1,000 words)
Word segmentation & tokenization	Segmentation and tokenization according to model architecture	VnCoreNLP, WordPiece, SentencePiece
Format input	Padding/truncation and label encoding	Custom tokenizer

Specifically, the first stage performs standardization of abbreviations and slang through lookup in a self-built dictionary of 520 mapping pairs ("mik" → "mình", "k" → "không"). Next, the text cleaning process is conducted to remove noise elements, including special characters, URL links, unnecessary text formatting, rare emoticons, and numbers.

For emoticon processing, the study maps 63 common emojis to three basic emotion groups: positive, negative, and neutral, to support the emotion classification process (😍 → positive, 😭 → negative, '🙄': 'Neutral'). Simultaneously, a list of about 1,000 Vietnamese stop words is applied to remove words without significant semantic value.

After the cleaning stage, text is segmented using VnCoreNLP tools and continues to be tokenized according to each model's architecture: WordPiece for BERT variants and SentencePiece for PhoBERT. Padding and truncation techniques are applied to ensure uniform input sequence length. Finally, emotion labels are converted from text to numerical format for supervised learning algorithms.

This process contributes to improving input corpus quality through noise reduction and data standardization. For illustration, the sentence "mik k thik môn học này 😠 😠 " after going through the entire preprocessing procedure will be converted to: "mình không thích môn học này cảm_xúc_tiêu_cực".

3. DATA LABELING

The three-class emotion data labeling process (positive, negative, neutral) is systematically designed to ensure consistency and reliability of the dataset. The research team built a detailed guideline with clear definitions for each emotion type.

The labeling process includes three stages:

- Each data sample is independently evaluated by two assessors
- Cases of disagreement (~11%) are discussed for consensus
- A random sample set (~10%) is re-evaluated to check consistency level

Labeling quality is evaluated using Cohen's Kappa coefficient [27], with a value of 0.84 showing very high agreement between independent evaluators. The final dataset includes 10,169 samples, with relatively balanced label distribution: 32.3% positive, 35% negative, and 32.7% neutral. The distribution of sample numbers by emotion label is presented in Figure 1.

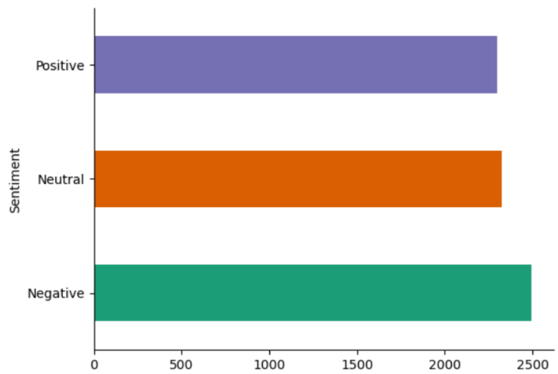

Figure 1. *Distribution of samples by emotion label*

B. EXPERIMENTAL MODELS

SVM: Uses Bag-of-Words representation combined with TF-IDF with 10,000 features, linear kernel, parameter C=1.0, and balanced class weight mode to handle data imbalance.

LSTM: Architecture includes 100-dimensional embedding layer, 2 LSTM layers with 256 hidden units, dropout 0.5, and fully-connected layer with softmax. The model is trained with Adam optimizer, learning rate 0.001, batch size 64.

PhoBERT: Uses PhoBERT-base-v2 with 12 Transformer layers, 768 hidden dimensions, 135 million parameters. Input data requires pre-segmentation using VnCoreNLP and uses SentencePiece tokenizer.

ViBERT: Architecture similar to BERT-base but pre-trained on 20GB Vietnamese corpus from Wikipedia, electronic newspapers, and forums, using WordPiece tokenizer with 30,000 tokens.

ViBERT4News: ViBERT version specially pre-trained on Vietnamese news data, suitable for sentiment analysis in journalistic texts and social media.

XLM-RoBERTa: Multilingual model based on RoBERTa, trained on 100 languages (including Vietnamese) with about 270 million parameters.

IV. EXPERIMENTS

A. EXPERIMENTAL PARAMETERS

Data is divided into three sets: training (70%), validation (15%), and testing (15%) using stratified method, to ensure balanced label distribution among emotion classes in each set. Experiments are deployed on Google Colab and Kaggle platforms, using NVIDIA Tesla T4 GPU. The programming environment includes libraries: PyTorch 2.6.0, Transformers 4.49.0, VnCoreNLP 1.1.1, along with other data processing libraries.

The hyperparameters we configured:

- PhoBERT Model
  - Batch Size: 16 (for both training and evaluation)
  - Learning Rate: 2e−5 with a decaying schedule (learning rate scheduler)
  - Number of Epochs: 3 (with early stopping based on validation set accuracy)
  - Optimizer: AdamW with weight decay 0.01

    - Max Sequence Length: 100 tokens
    - Dropout Rate: 0.1
- Support Vector Machine (SVM)
    - TF-IDF Max Features: 10,000
    - TF-IDF N-gram Range: (1,2)
    - SVM Kernel: linear
    - Regularization (C): 1.0
    - Class Weight: balanced
    - Cross-Validation Folds: 5
    - Probability Estimation: enabled (True)
- LSTM
    - Hidden Size: 256
    - Number of Layers: 2
    - Dropout: 0.5
    - Embedding Dimension: 100
    - Batch Size: 64
- ViBERT and ViBERT4news
    - Batch Size: 8 (for both training and evaluation)
    - Learning Rate: 3e−5 with a decaying schedule (learning rate scheduler)
    - Number of Epochs: 3 (with early stopping based on validation set accuracy)
    - Optimizer: AdamW with weight decay 0.01
    - Max Sequence Length: 100 tokens
    - Dropout Rate: 0.1
- XLM-RoBERTa
    - Batch Size: 8 (for both training and evaluation)
    - Learning Rate: 2e−5 with a decaying schedule (learning rate scheduler)
    - Number of Epochs: 3 (with early stopping based on validation set accuracy)
    - Optimizer: AdamW with weight decay 0.01
    - Max Sequence Length: 100 tokens
    - Dropout Rate: 0.1

B. EVALUATION METRICS

To evaluate the model performance in sentiment analysis, the research team employed the Accuracy metric, assuming data is evenly distributed among classes, to reflect the overall proportion of correctly classified samples. Accuracy, along with other key metrics (Recall, Precision, and F1 Score), is calculated based on the Confusion Matrix, utilizing the following formal definitions:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$Precison = \frac{TP}{TP + FP} \quad (3)$$

$$F1 - score = 2.\frac{Precision. Recall}{Precision + Recall} \quad (4)$$

The metrics are defined as follows: FN (False Negative) - positive samples mistakenly predicted as negative, FP (False Positive) - negative samples mistakenly predicted as positive, TN (True Negative) - correctly predicted negative samples, TP (True Positive) - correctly predicted positive samples.

C. EXPERIMENTAL RESULTS

ViBERT4News is the highest performing model with 89.19% accuracy. PhoBERT ranks second with 89.06%, followed by XLM-RoBERTa (87.29%), ViBERT (84.93%), SVM (84.99%), and LSTM (80.93%). Detailed results are presented in Table 3.

Table 3. *Model results on test set*

Model	Accuracy	Precision	Recall	F1-Score
ViBERT4News	0.8919	0.8933	0.8925	0.8929
ViBERT	0.8493	0.8497	0.8497	0.8497
PhoBERT	0.8906	0.8912	0.8920	0.8916
XLM-RoBERTa	0.8729	0.8708	0.8750	0.8729
SVM	0.8499	0.8499	0.8499	0.8499
LSTM	0.8093	0.8093	0.8093	0.8093

V. DISCUSSION

Experimental results show clear advantages of Transformer-based models, especially ViBERT4News and PhoBERT, in Vietnamese sentiment analysis tasks. Compared to previous studies, the achieved accuracy (89.19%) is relatively high, especially in the context of data collected from social media - which is inherently non-standard and contains much noise. However, although ViBERT4News achieves the highest accuracy on the test set, this model shows unstable generalization capability when applied to out-of-domain data, raising questions about effectiveness when deployed in real-world environments. Meanwhile, PhoBERT shows better generalization capability and may be a more suitable choice for practical deployment applications. A major challenge that still exists is the ability to process Vietnamese texts with complex characteristics such as containing multiple conflicting emotions, being sarcastic, or using variant words and slang - which are common in social media language. These characteristics require models to have stronger semantic and contextual representation mechanisms. This study contributes two main points: (i) building a labeled Vietnamese student feedback emotion dataset, with potential to serve future research; and (ii) developing a complete sentiment analysis system to support educational management and improvement activities.

VI. CONCLUSION AND FUTURE DIRECTIONS

The study has successfully built a sentiment analysis system to extract UTC2 student attitudes on Facebook social media platform, using modern natural language processing models. Among the surveyed models, ViBERT4News achieved the highest accuracy (89.19%), however PhoBERT (89.06%) is evaluated as a more suitable choice for practical deployment due to its stability. Besides positive results, the study still has some limitations. Current models face difficulties in processing sarcastic texts, expressing multiple emotions, and have not yet exploited aspect-based sentiment analysis. Additionally, the data scope is still limited mainly to one target group and one platform. Future development directions include: (i) improving the ability to process complex text types and unstructured language, (ii) expanding data sources from many other platforms such as YouTube, TikTok, academic forums, or internal feedback systems, (iii) developing aspect-level sentiment analysis to provide more detailed information about specific components in the training environment, and (iv) integrating the system into the school's IT infrastructure to build early warning mechanisms, supporting decision-making in training management and student care. In terms of practical applications, the study proposes piloting the system at UTC2, integrating analysis results into existing management systems such as student information portals and opinion survey systems, thereby contributing to improving training quality, enhancing learning experiences, and providing effective student support.

VII. REFERENCES

[1] B. Liu. (2012). Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies, vol. 5, no. 1, pp. 1-167. DOI: https://doi.org/10.2200/S00416ED1V01Y201204HLT016

[2] Giachanou, F. Crestani. (2016). Like It or Not: A Survey of Twitter Sentiment Analysis Methods. ACM Computing Surveys, 49(2), 1–41. DOI: https://doi.org/10.1145/2938640

[3] Wu, M.-J. (2022). Response rates of online surveys in published research: A meta-analysis. Computers in Human Behavior Reports, 7, 100206. DOI: https://doi.org/10.1016/j.chbr.2022.100206

[4] Nguyen, D., Ngo, X. B., & Tu, M. P. (2015). An empirical study on sentiment analysis for Vietnamese. 2014 International Conference on Advanced Technologies for Communications (ATC 2014), 7043403. DOI: https://doi.org/10.1109/ATC.2014.7043403

[5] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 4171-4186. DOI: https://doi.org/10.18653/v1/N19-1423

[6] Peter D. Turney. (2002). Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, 417-424. DOI: https://doi.org/10.3115/1073083.1073153

[7] Minqing Hu, Bing Liu. (2004). Mining and summarizing customer reviews. Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, 168-177. DOI: https://doi.org/10.1145/1014052.1014073

[8] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297. DOI: https://doi.org/10.1007/BF00994018

[9] Bo Pang, Lillian Lee. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1-2), 1-135. DOI: https://doi.org/10.1561/1500000011

[10] Bo Pang, Lillian Lee, Shivakumar Vaithyanathan. (2002). Thumbs up? Sentiment Classification using Machine Learning Techniques. Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002), 79-86. DOI: https://doi.org/10.3115/1118693.1118704

[11] Vaswani, A., et al. (2017). Attention Is All You Need. Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS 2017), 6000–6010.

[12] Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A., & Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, 1631–1642. DOI: https://doi.org/10.18653/v1/D13-1170

[13] Irsoy, O., & Cardie, C. (2014). Deep recursive neural networks for compositionality in language. Advances in Neural Information Processing Systems, 27.

[14] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. DOI: https://doi.org/10.1162/neco.1997.9.8.1735

[15] Tai, K. S., Socher, R., & Manning, C. D. (2015). Improved semantic representations from tree-structured long short-term memory networks. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics, 1, 1556–1566. DOI: https://doi.org/10.3115/v1/P15-1150

[16] Van, V. D., & Tran, T. T. (2018). Combining convolution and recursive neural networks for sentiment analysis. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1, 1038–1043.

[17] Kim, Y. (2014). Convolutional neural networks for sentence classification. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 1746–1751. DOI: https://doi.org/10.3115/v1/D14-1181

[18] Vo, H. T., Nguyen, T. T., Pham, H.-A., & Le, T. V. (2017). An Efficient Hybrid Model for Vietnamese Sentiment Analysis. Proceedings of the Asian Conference on Intelligent Information and Database Systems, 10192, 227-237. DOI: https://doi.org/10.1007/978-3-319-54472-4_22

[19] Nguyen, D. Q., Vu, T., & Nguyen, A. T. (2020). PhoBERT: Pre-trained language models for Vietnamese. Findings of EMNLP 2020, 1037-1042. DOI: https://doi.org/10.18653/v1/2020.findings-emnlp.92

[20] Liu, Y., et al. (2019). RoBERTa: A robustly optimized BERT pretraining approach. Computing Research Repository (CoRR), abs/1907.11692. DOI: https://arxiv.org/abs/1907.11692

[21] Bui, T. V., Tran, T. O., & Le-Hong, P. (2020). Improving Sequence Tagging for Vietnamese Text using Transformer-based Neural Models. Proceedings of the 34th Pacific Asia Conference on Language, Information and Computation, 13-20. DOI: https://aclanthology.org/2020.paclic-1.2/

[22] Pham, M. Q. N. (2020). An empirical study of using pre-trained BERT models for Vietnamese relation extraction task at VLSP 2020. Proceedings of the 7th International Workshop on Vietnamese Language and Speech Processing, 13–18.

[23] Conneau, A., et al. (2020). Unsupervised Cross-lingual Representation Learning at Scale. ACL 2020, 7440-7451. DOI: https://doi.org/10.18653/v1/2020.acl-main.747

[24] Saragih, H., Manurung, J. (2024). Leveraging the BERT Model for Enhanced Sentiment Analysis in Multicontextual Social Media Content. Jurnal Teknik Informatika C.I.T Medicom, 16(2), 82–89.

[25] Nguyen, T. A., Le, V. C., Tran, H. P. (2025). Expanding Vietnamese SentiWordNet to Improve Performance of Vietnamese Sentiment Analysis Models. Advances in Information and Communication Technology: Proceedings of the 4th International Conference ICTA 2025.

[26] Tran, T. D., Le, N. T., Bui, N. D., & Vu, H. (2024). Emotion recognition in learners with emoji sentiment accompaniment using the PhoBERT model. HNUE Journal of Science: Natural Sciences, 69(3), 46-56. DOI: https://doi.org/10.18173/2354-1059.2024-0034
[27] Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20, 37–46. DOI: https://doi.org/10.1177/001316446002000104

PHÂN TÍCH CẢM XÚC TIẾNG VIỆT TRÊN MẠNG XÃ HỘI DỰA TRÊN KIẾN TRÚC BERT

Trần Thị Dung, Trần Đức Anh, Phạm Thị Ngọc Oanh, Lê Nhật Tùng

TÓM TẮT — Nghiên cứu này thực hiện đánh giá so sánh hiệu suất của các mô hình học máy trong phân tích cảm xúc tiếng Việt, bao gồm ViBERT4News, PhoBERT, XLM-RoBERTa, ViBERT, SVM và LSTM. Bộ dữ liệu gồm 10.169 mẫu văn bản được thu thập từ các bài đăng và bình luận trên Facebook liên quan đến sinh viên Phân hiệu Tp. Hồ Chí Minh, Trường Đại học Giao thông vận tải, được gán nhãn thủ công thành ba lớp: tích cực (32,3%), tiêu cực (35,0%) và trung lập (32,7%). Kết quả thực nghiệm cho thấy các mô hình dựa trên kiến trúc BERT vượt trội hơn so với các phương pháp truyền thống, trong đó ViBERT4News đạt độ chính xác cao nhất (89,19%), tiếp theo là PhoBERT (89,06%). Tuy nhiên, PhoBERT thể hiện khả năng tổng quát hóa tốt hơn, phù hợp hơn cho việc triển khai ứng dụng thực tế. Nghiên cứu không chỉ chứng minh hiệu quả của các mô hình ngôn ngữ được tiền huấn luyện trong xử lý cảm xúc trên dữ liệu mạng xã hội tiếng Việt, mà còn đề xuất khả năng ứng dụng vào nền tảng phân tích và trực quan hóa phản hồi sinh viên, hỗ trợ giám sát học vụ và nâng cao chất lượng giáo dục.

Từ khóa: Phân tích cảm xúc, PhoBERT, ViBERT, ViBERT4News, XLM-RoBERTa.

Tran Thi Dung holds a Master's degree in Information Technology. She is currently a lecturer at the Ho Chi Minh City campus of the University of Transport and Communications. Her research interests include data mining, machine learning , and natural language processing.

Le Nhat Tung holds a Master's degree in Data Science. He is currently a lecturer at Dong Nai Technology University. His research interests include Data Science and Artificial Intelligence.

Tran Duc Anh is an IT Engineer and is currently a lecturer at the Ho Chi Minh City campus of the University of Transport and Communications. His main research interests include Artificial Intelligence, Machine Learning, and Natural Language Processing.

Pham Thi Ngoc Oanh is currently a student at the Ho Chi Minh City campus of the University of Transport and Communications. Her research interests include Data Mining, Machine Learning, and Natural Language Processing.