Utilizing Transformer Models to detect Vietnamese fake news on Social media flatforms

Supervisor: Dr. Tran Thanh Phuoc
Natural Language Processing and Knowledge Discovery Laboratory,
Faculty of Information Technology,
Ton Duc Thang University
Ho Chi Minh City, Vietnam
Student: Huynh Anh Tuan
Computer science student,
Ton Duc Thang University
Ho Chi Minh City, Vietnam

Abstract

The proliferation of fake news on social media has become a serious issue, leading to misinformation and societal harm. This project aims to develop a system for detecting Vietnamese fake news using transformer models, particularly PhoBERT—a version of BERT optimized for Vietnamese. To address this issue, I collected a dataset consisting of Vietnamese posts from the social media platform Facebook and several official Vietnamese news articles, covering topics such as lifestyle, news, and politics. However, there were challenges due to the imbalance between the number of real and fake news. The posts were manually labeled as real or fake, then preprocessed and trained using transformer models and PhoBERT for Vietnamese, followed by evaluating their performance using metrics like accuracy, recall, and F1-score.

Our results indicate that PhoBERT outperforms other transformer models in detecting Vietnamese fake news, achieving high accuracy and reliability. This report outlines the context, objectives, methods, and future research directions, providing a comprehensive overview of the project and its contributions to the field of fake news detection.

Keywords: Vietnamese fake news, Fake News Detection, Transformer Models, PhoBERT, Social Media Analysis

1. Introduction

In the context of global modernization, social media platforms are becoming increasingly popular, accompanied by both positive and negative impacts. Among these, the rapid spread of fake news on social media has emerged as a serious societal issue, where the dissemination of false information has led to misunderstandings and even conflicts worldwide.

For the case of Vietnam, fake news has frequently caused public uproar, such as misinformation about epidemics, traffic accidents, incorrect knowledge in daily life, and even politically subversive content. These types of fake news often spread quickly within the community, leading to public confusion and affecting people's lives. Therefore, researching and detecting fake news is a necessary task to support and maintain social stability. This is why we chose this topic for our research.

In the past few years, deep learning has been recognized as a powerful tool in artificial intelligence, particularly in natural language processing (NLP). However, traditional deep learning models often rely on sequential data processing, which can be limiting when handling complex language tasks [1]. Transformers, a novel architecture, have revolutionized NLP by utilizing attention mechanisms that allow for more effective context and relationship processing within text [2]. This makes Transformers especially valuable for tasks such as fake news detection, where understanding nuanced language and context is crucial.

Throughout this research, we focus on leveraging transformer models to detect fake news. Specifically, we use PhoBERT, an advanced variant of the BERT model designed specifically for the Vietnamese language [3]. Our goal is to develop an effective system for identifying fake news on social media platforms, such as Facebook—the most widely used social media platform in Vietnam.

By harnessing the strengths of PhoBERT, we aim to improve the accuracy and effectiveness of Vietnamese fake news detection. However, we face significant challenges due to the lack of large-scale datasets containing both real and fake Vietnamese news. We have conducted data crawling from legitimate Facebook pages of official Vietnamese news outlets and fake data sources from impersonation pages, anti-establishment sources, and misinformation sites covering various fields from social life to politics. To

achieve this, we utilized several tools, such as Selenium for data crawling, followed by data processing through cleaning and encoding.

The structure of the remainder of this paper is as follows: In Part 2, we review related works on transformer models and fake news detection, focusing on methods and models applicable to our study in Vietnam. Next, in Part 3, we detail the proposed methodology, including the overall model and specific steps taken to develop the system. Part 4 covers the experimental setup, describes the dataset, results, and discussion. Finally, Part 5 concludes the paper by summarizing our findings and outlining future research directions.

2. Background

2.1 Related work

The detection of fake news has become a significant research area due to the increasing prevalence of misinformation globally. Several studies have explored various approaches to address this challenge.

The journey of fake news detection has advanced significantly with the development of Transformer models. Vaswani et al. (2017) [4] introduced the Transformer architecture, utilizing self-attention mechanisms to process sequential data efficiently, thus laying the foundation for modern NLP.

Since then, numerous studies have employed Transformer models for fake news detection in English. Devlin et al. (2018) [5] introduced BERT (Bidirectional Encoder Representations from Transformers), which uses bidirectional attention to better understand word context. Building on this foundation, Liu et al. (2019) [6] enhanced BERT with RoBERTa, improving training efficiency and performance on NLP benchmarks. Additionally, Sanh et al. (2019) [7] proposed DistilBERT, a smaller and faster version of BERT, suitable for real-time applications.

In the context of fake news detection, Agarwal et al. (2021) [8] integrated a Bi-LSTM layer with attention into contextual embeddings for classifying trustworthy news in English. Monti et al. (2019) [9] studied graph neural networks, employing a four-layer Graph CNN network combining user activity and article information to predict news. Meanwhile, Qi et al. (2019) [10] emphasized the role of visual content, introducing a multidomain visual

neural network that uses CNN and RNN models to analyze image features, aiding in distinguishing between fake and real news.

In the Vietnamese context, Dat Quoc Nguyen et al. (2020) [3] developed PhoBERT, a Transformer-based model pre-trained on a large corpus of Vietnamese text, setting a significant benchmark for Vietnamese NLP tasks. Their results indicate that PhoBERT consistently outperforms the current leading XLM-R pre-trained multilingual model, setting new benchmarks in several Vietnamese-specific NLP tasks such as part-of-speech tagging, dependency parsing, named-entity recognition, and natural language inference.

Recent studies have focused on utilizing PhoBERT and other deep learning techniques for Vietnamese fake news detection. For instance, Cao Nguyen Minh Hieu et al. [11] proposed a tool during the ReINTEL 2020 Challenge that combined PhoBERT embeddings with temporal and community interaction metrics (shares, likes, comments). Their StackNet model achieved an AUC score of 0.9521, ranking first on the ReINTEL leaderboard.

Ngoc-Dong Pham et al. (2021) [12] proposed a hybrid method combining PhoBERT with TF-IDF for word embeddings and CNN for feature extraction, achieving a notable AUC score of 0.9538, though the reliance on the ReINTEL dataset may limit diversity. Cam-Van Nguyen Thi et al. (2022) [13] introduced v3MFND, a deep multimodal fake news detection model integrating text, images, and videos to enhance accuracy; however, the model's complexity may affect its real-time applicability. Khoa Dang Pham et al. (2023) [14] studied the vELECTRA model [15] with handcrafted features, achieving an AUC score of 0.9575 on the ReINTEL dataset; nevertheless, the dependence on handcrafted features may limit adaptability. Vo Trung Hung et al. (2023) [16] used CNN and RNN models to classify news into four groups, achieving an 85% accuracy rate. The dataset size may limit the generalizability of their results.

These studies highlight the effectiveness of Transformer models, particularly PhoBERT, in detecting Vietnamese fake news. They also underscore the importance of combining textual data with multimodal and meta-data to improve performance, while pointing out challenges related to dataset size, diversity, and computational complexity that need to be addressed in future research.

2.2 Theoretical basis

To effectively implement the project on detecting Vietnamese fake news on social media platforms using Transformer models, particularly PhoBERT and other BERT variants, it is essential to have a solid understanding of the following foundational knowledge:

2.2.1 Natural Language Processing (NLP)

Natural Language Processing (NLP) is a machine learning technology that enables computers to interpret, interact with, and understand human language [1] . NLP encompasses various tasks such as syntactic parsing, semantic analysis, entity recognition, and text classification.

In the context of text classification, NLP extracts information from text, processes semantics, and represents the text in feature forms suitable for input into machine learning or deep learning models. Techniques like Bag of Words, TF-IDF (Term Frequency-Inverse Document Frequency), and word embeddings help convert text into numerical forms. Subsequently, machine learning models like Naive Bayes, SVM (Support Vector Machine), and others can be trained to classify text into categories such as positive or negative sentiment, spam or non-spam, and real or fake news.

2.2.2 Transformer Model

The Transformer model represents a breakthrough in NLP, introduced by Vaswani et al. (2017) in the paper *"Attention Is All You Need"* [4] . The highlight of the Transformer lies in its self-attention architecture, which allows the model to learn relationships between words in a sentence without adhering to the sequential order used in previous models like RNN or LSTM.

The Transformer model consists of two main components: the Encoder and the Decoder:

- **Encoder:** The Encoder receives a sequence of words as input and represents them as semantic vectors. Each Encoder comprises multiple sequential layers, with two main components in each layer: the self-attention mechanism and a feed-forward neural network. The self-attention mechanism enables the model to learn the semantic relationships between relevant words in a sequence while ignoring unrelated words. The feed-forward neural network processes these attended vectors to produce deeper semantic representations.

- **Decoder:** The Decoder has a similar structure to the Encoder, using self-attention for the target input. Additionally, it employs cross-attention to connect with the Encoder's output. This enables the Decoder to generate semantic representations based on both the initial input sequence and the previously generated output sequence.

The collaboration between the Encoder and Decoder allows the Transformer to process language tasks such as machine translation, text summarization, text generation, and text classification with flexibility and efficiency.

2.2.2.1 BERT (Bidirectional Encoder Representations from Transformers):

BERT is a pre-trained language model designed to understand word context in both directions (left-to-right and right-to-left) within a sentence [5]. BERT is trained on two primary tasks: Masked Language Modeling (MLM) and Next Sentence Prediction (NSP).

Masked Language Modeling (MLM): In this task, some words in a sentence are replaced with the [MASK] token, and the model is required to predict the masked words based on the surrounding context.

Next Sentence Prediction (NSP): This task asks the model to predict whether a given sentence is the next sentence of a preceding sentence, thereby improving the model's ability to understand relationships between sentences.

BERT has achieved outstanding results in various NLP tasks such as text classification, entity recognition, and question answering.

2.2.2.2 RoBERTa (A Robustly Optimized BERT Pretraining Approach):

RoBERTa is a variant of BERT that has been optimized to improve performance by removing the Next Sentence Prediction (NSP) task and using a larger training dataset [6]. RoBERTa applies the masked language modeling (MLM) approach with enhancements in training and data. RoBERTa has demonstrated superior performance in NLP tasks such as text classification and entity recognition, owing to optimized hyperparameters and training data.

2.2.2.3 PhoBERT

PhoBERT is a variant of BERT that has been trained entirely on Vietnamese text data [3], allowing the model to better capture the semantic and syntactic characteristics specific to this language.

PhoBERT incorporates improvements from RoBERTa, such as eliminating the Next Sentence Prediction (NSP) task and using only Masked Language Modeling (MLM), while being trained on a large-scale dataset.

This approach enables PhoBERT to perform more effectively compared to BERT or RoBERTa models trained on other languages.

2.2.3 TF-IDF (Term Frequency-Inverse Document Frequency)

TF-IDF (Term Frequency-Inverse Document Frequency) is a technique commonly used in Natural Language Processing (NLP) and text mining [17], [18]. It is a statistical measure used to evaluate the importance of a word within a document in a collection of documents or a text corpus.

In this study, we used TF-IDF as a preprocessing step to transform text into feature vectors. These vectors can then be combined with Transformer models to enhance the ability to classify real and fake news. TF-IDF helps the model focus on important keywords and minimizes the impact of common words that carry less information during the model training process.

3. Proposed methods

3.1 The designed system

Our system can be divided into four main stages , which are generally shown in **Figure 1**: (1) Data Collection, (2) Data Processing, (3) Model Training, and (4) Model Evaluation.

- **Data Collection:** In the first stage, we collect data from Facebook posts on official news pages as well as pages that spread fake, misleading, or disruptive content across topics like current affairs, lifestyle, and politics. We gather details such as the author, content, post link, and comments. This stage is crucial as the dataset will significantly influence the research outcomes.
- **Data Processing:** The collected data will undergo a series of preprocessing steps, including cleaning, text normalization, and the crucial step of manually labeling the posts as true or fake. After preprocessing, the data will be divided into training and testing sets and prepared for model training.
- **Model Training:** We use the processed data to train Transformer models: BERT, RoBERTa, and PhoBERT. We apply various training techniques to each model to optimize performance, including hyperparameter tuning and

cross-validation techniques. After training, we compare the results of the three models to evaluate their effectiveness in detecting fake news.

- **Model Evaluation:** The final stage involves assessing the performance of the trained model. We use a separate test dataset to evaluate the model's accuracy, class precision, recall, and F1 score. Based on the evaluation results, we may further fine-tune the model or adjust preprocessing techniques to enhance performance.

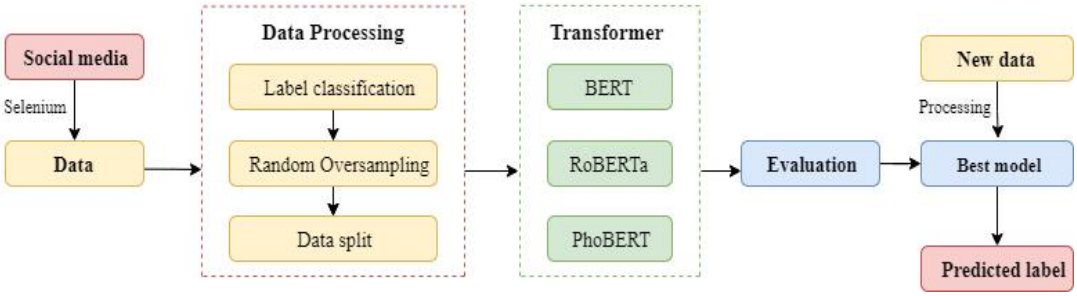

Figure 1. General model of the system.

3.2 Data Collection

Due to the limitations of existing datasets on Vietnamese news and social media posts, and the outdated nature of available information, we decided to collect our own data to contribute to future research resources.

We manually selected and filtered posts. For authentic news, we identified reliable sources on Facebook, such as major Vietnamese news outlets and government pages. For fake news, we targeted sensationalist tabloids and Facebook groups that frequently share misleading information.

After selecting the necessary sources, we used Selenium to automate data collection, simulating user actions such as navigating websites and extracting data. Ultimately, we collected two datasets: one for authentic news and one for fake news, as shown in **Figure 2.**

However, the data collection process faced several challenges, including limited collection time and difficulties in finding fake news sources due to some articles being removed after being reported. As a result, there is a discrepancy in the number of authentic and fake news items in our data, as illustrated in **Figure 4**.

date	author_id	content	label	link	comment_list
29/07/2024 13:45	https://www.facebook	Vụ xe bán tải cố vượt rào chắn, bị tàu hỏa tông ở Đ	0	https://www.fật" }, { "comment_id": "c36", "author": "Trần Phúc Hậu",	
30/07/2024 23:58	https://www.facebook	TPHCM: Hơn 4.600 ca mắc sốt xuất huyết, nhiều 'đ	0	https://www.fuyễn", "content": "Ánh Tây coi chừng bối nhen" }, { "comr	
30/07/2024 22:59	https://www.facebook	Nóng: Ngộ độc hàng loạt tại trụ sở công ty mẹ TikT	0	https://www.for": "Lê Nguyễn Bảo Thư", "content": "Nhii Mai Kim Ngân Hồng	
31/07/2024 14:50	https://www.facebook	Ngày mai: Giá xăng trong nước có thể giảm lần thứ	0	https://www.făng nào" }, { "comment_id": "c6", "author": "Lắm Chuyện	
31/07/2024 12:30	https://www.facebook	Pin dự phòng của hành khách bốc cháy tại nhà ga s	0	https://www.fẽm quá" }, { "comment_id": "c20", "author": "Thang Vo",	
31/07/2024 10:50	https://www.facebook	Thương tâm quá: Trong lúc chờ nhau trên xe máy v	0	https://www.fl": "c14", "author": "Vũ Hà", "content": "Nam mô a di đà phật	
30/07/2024 22:50	https://www.facebook	THƯƠNG TÂM HÀ GIANG: ĐẤT ĐÁ LĂN TỪ TALUY D	0	https://www.fiuy", "content": "A di đà Phật" }, { "comment_id": "c9",	
30/07/2024 22:30	https://www.facebook	NỮ TÀI XẾ Ô TÔ ĐẠP NHẦM CHÂN GA GÃY TNGT LI	0	https://www.fạy mẹ trẻ ..may k chết ng ..." }, { "comment_id": "c12", "au	

Figure 2. An examples of the structure and content of some data.

3.3 Data Processing

We performed data cleaning through the following steps: removing empty, invalid, or duplicate entries, converting all text to lowercase, and eliminating special characters and URLs. Then we select the data fields that will be used and the remaining data will be shown as shown in **Figure 3**.

	content	label
0	vụ xe bán tải cố vượt rào chắn bị tàu hỏa tông...	0
1	tphcm hơn 4600 ca mắc sốt xuất huyết nhiều điể...	0
2	nóng ngộ độc hàng loạt tại trụ sở công ty mẹ t...	0
3	ngày mai giá xăng trong nước có thể giảm lần t...	0
4	pin dự phòng của hành khách bốc cháy tại nhà g...	0

Figure 3. Data after being cleaned and using information fields selected.

At this stage, we are focusing solely on the content of the posts and their classification labels, but we plan to extend our research to include analysis of comments in the future.

It is evident that the number of fake news samples is significantly lower compared to true news (shown in the chart in **Figure 4**), which could lead to bias in model training and inaccurate results. To address this issue, we have implemented two solutions:

- Incorporating Additional Data from VFND: We integrated posts from the VFND dataset, as described in the thesis by Ho Quang Thanh, “VNFD Vietnamese Fake News Datasets: Tập hợp các bài báo tiếng Việt và các bài post Facebook phân loại 2 nhãn Thật & Giả.” [19] . However, since this dataset was collected in 2019, we selectively included only news that is not affected by time, such as scientifically debunked information, superstitious news, or distorted lifestyles. The supplemental data comprises no more than 20% of our current collection of fake news.
- Using Random Oversampling Technique: We applied the Random Oversampling technique from the ‘imbalanced-learn’ library. This method effectively balances the dataset by increasing the number of samples from

the minority class. It works by randomly duplicating existing samples in the minority class until the class distribution is balanced.

Figure 5 illustrates the proportion of the two labels after balancing the dataset. Balancing the labels helps prevent the model from being biased toward the majority class, improving accuracy for both labels, and ensuring that evaluation metrics such as precision, recall, and F1-score accurately reflect the model's true performance. This also enhances the model's ability to detect significant patterns, leading to improved AUC values.

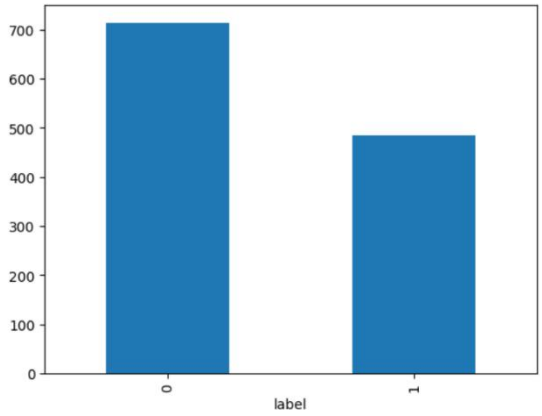

Figure 4. The size of 2 data sets after collection.

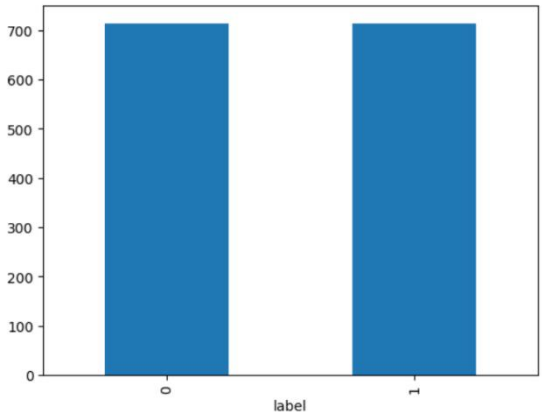

Figure 5.The size of 2 data sets after processing.

4. Experimental results

4.1 Corpus

After completing the data processing steps, including cleaning and balancing the data as discussed in sections 3.2 and 3.3, we obtained a dataset comprising social media posts and news from Vietnamese sources. The dataset contains over 1,400 samples, including both real and fake news across various domains. We then randomly split the dataset into a training set and a test set with an 80/20 ratio, resulting in 1,124 samples for training and 282 samples for testing. This approach allows the model to learn patterns effectively, increasing the likelihood of better performance on new, unseen data.

4.2 Evaluation tool

The classification results will be evaluated using Accuracy, Precision, Recall, F1 Score, and AUC.

- Accuracy: The ratio of correct predictions to total predictions, reflecting overall model performance, though it may not account for class imbalances.

- Precision: The ratio of true positive predictions to total positive predictions, indicating the correctness of positive predictions.
- Recall: The ratio of true positive predictions to actual positives, measuring the model's ability to identify all positive instances.
- F1 Score: The harmonic mean of Precision and Recall, balancing these metrics, particularly useful for imbalanced data.
- AUC (Area Under the Curve): This represents the area under the ROC (Receiver Operating Characteristic) curve, a graph that shows the relationship between the True Positive Rate and the False Positive Rate across different classification thresholds. AUC measures the model's ability to distinguish between classes; a higher AUC indicates better model performance in differentiating between positive and negative classes.

4.3 Results

Table 1. The results of model evaluation.

Model	Accuracy	Precision	Recall	F1-Score	AUC
PhoBERT + TF-IDF	**0.888112**	**0.863014**	**0.913043**	**0.887324**	0.922689
PhoBERT	0.872340	0.850649	0.909722	0.879195	**0.947665**
BERT	0.787234	0.850000	0.708333	0.772727	0.858343
RoBERTa	0.741135	0.844660	0.604167	0.704453	0.834541

Table 1 presents the evaluation results based on various metrics from the models when tested on the same training and testing datasets.

- Roberta: Roberta's classification performance is relatively poor, with an Accuracy of 0.741 and an AUC of 0.835. While Precision is reasonably high at 0.845, the Recall is only 0.604, indicating that the model misses many instances of fake news. The F1 Score of 0.704 shows that while the model performs reasonably, it is less effective compared to other models.
- BERT: Performs better than Roberta in classification, with an Accuracy of 0.787 and an AUC of 0.858. Precision is 0.850 and Recall is 0.708, demonstrating a balanced performance between detecting fake news and legitimate news. The F1 Score of 0.773 indicates that BERT is a strong model but still not the best among those tested.

- PhoBERT: Achieves the highest performance with an Accuracy of 0.872 and an AUC of 0.948, indicating excellent ability to distinguish between real and fake news. The F1 Score of 0.879 shows that this model is well-balanced between Precision and Recall, although it is slightly lower compared to PhoBERT + TF-IDF.
- PhoBERT TF-IDF: Achieves the highest Accuracy of 0.888 among the models, with excellent Precision (0.863), Recall (0.913), and AUC (0.923). This model is well-balanced between accuracy and detection capability, making it the most effective for the task.

4.4 Discussion

The experimental results indicate that PhoBERT and PhoBERT + TF-IDF are both highly effective models for analyzing and classifying fake news. PhoBERT + TF-IDF achieved the best performance, with the highest Accuracy, Precision, and F1 scores, demonstrating its exceptional classification capabilities. PhoBERT also performed very well, with high Accuracy, Precision, Recall, and AUC scores.

Table 2. Comparison of the predictive performance of the models on the training dataset.

	Posts/News	**True label**	**Model's predicted label**			
			RoBERTa	BERT	PhoBERT	PhoBERT + TF-IDF
1	Tuyến Metro Nhổn Ga Hà Nội vận hành thương mại vào ngày 09/08/2024.	**Real**	Real	Real	Real	Real
2	Hà Nội gặp khó khăn di dời người dân ra khỏi vùng lũ.	**Real**	Fake	Fake	Real	Real
3	Ban tổ chức Olympic hủy buổi tập 3 môn phối hợp lần thứ hai vì chất lượng nước sông Seine.	**Real**	Fake	Real	Real	Real
4	Hiện trường kinh hoàng xe tải cố vượt đường ray khiến tàu hỏa trật bánh, ít nhất 100 người thương vong, hành khách hoảng loạn.	**Fake**	Fake	Fake	Fake	Fake
5	Tai nạn sập hầm lò đặc biệt nghiêm trọng ở Quảng Ninh khiến 5 công nhân tử vong,... đảng bộ và công đoàn bù nhìn chưa bao giờ lo cho điều kiện lao động của người dân.	**Fake**	Real	Real	Real	Fake

6	Sự cố nhánh cây dầu bị gãy, rơi từ độ cao khoảng 25 m ở công viên Tao Đàn, TPHCM. Vụ việc làm 2 người tử vong, 3 người bị thương.	**Real**	Real	Real	Real	Fake

However, it is important to note that while PhoBERT + TF-IDF has a slightly lower AUC compared to PhoBERT, it still maintains a high F1 Score, indicating a strong balance between Precision and Recall. This balance suggests that PhoBERT + TF-IDF may be more conservative, potentially missing some legitimate news but providing more accurate predictions overall.

Table 2 presents several representative cases extracted from the training set. For most straightforward instances of true and false news, such as notification-type case 1 and sensational case 4, all four models produced accurate results. However, in cases involving more complex news content, models like BERT and RoBERTa exhibited more classification errors, resulting in lower performance and reduced reliability.

Although PhoBERT and PhoBERT + TF-IDF demonstrated high accuracy in prediction, there were still some exceptions, particularly with cases where the news contained a mix of true and false information. For example, in case 5, "Tai nạn sập hầm lò đặc biệt nghiêm trọng ở Quảng Ninh khiến 5 công nhân tử vong,... đảng bộ và công đoàn bù nhìn chưa bao giờ lo cho điều kiện lao động của người dân" contains the true information "Tai nạn sập hầm lò đặc biệt nghiêm trọng ở Quảng Ninh khiến 5 công nhân tử vong" ("A particularly severe mining accident in Quảng Ninh killing 5 workers") but the additional part "…đảng bộ và công đoàn bù nhìn chưa bao giờ lo cho điều kiện lao động của người dân" ("the puppet party and trade unions have never cared about workers' conditions") is inaccurate and unverified. In this instance, PhoBERT + TF-IDF correctly classified it as fake news, whereas PhoBERT and other models were misled by the true part of the article. PhoBERT + TF-IDF's ability to accurately identify such cases is attributed to TF-IDF's emphasis on important keywords and its ability to minimize the influence of common but less informative words. This approach helps the model recognize that the additional information lacks authentic value and should not be considered true, thereby improving classification accuracy.

However, this cautious approach also led PhoBERT + TF-IDF to incorrectly classify some true news cases, such as case 6. To improve results, it is essential to refine the integration of PhoBERT with TF-IDF, ensuring that

the model can balance sensitivity to important keywords with a more nuanced understanding of context.

5. Conclusions and Future Work

In this study, we focused on leveraging Transformer models such as BERT, RoBERTa, and PhoBERT for fake news classification in Vietnam. We collected a dataset comprising Facebook posts from June to July 2024, covering topics such as lifestyle, society, and politics. Due to the limited number of fake news articles, we supplemented our dataset with additional fake news examples from the VFND dataset, as described in Ho Quang Thanh's thesis, "VNFD Vietnamese Fake News Datasets: Tập hợp các bài báo tiếng Việt và các bài post Facebook phân loại 2 nhãn Thật & Giả" [19] . We then applied Transformer models for classification, and the evaluation results showed that PhoBERT and PhoBERT combined with TF-IDF achieved the highest prediction performance for Vietnamese.

However, this model still has certain limitations, including the restricted dataset size and information loss due to the complexities of the Vietnamese language. Some posts use abbreviations, diverse grammar structures, euphemisms, or contain a mix of both true and false information, which can lead to incorrect predictions by the model. Additionally, our current research focuses solely on classifying news based on the content of the posts without considering other data sources like interaction metrics and comments, which are substantial and valuable data sources.

Moving forward, we plan to continue expanding our dataset and incorporate the analysis of comments on both authentic and fake posts. This will help us gain a deeper understanding of user sentiments and attitudes toward these two types of information, ultimately contributing to more accurate prediction results.

References

[1] K. Chowdhary and K. R. Chowdhary, "Natural language processing," *Fundamentals of artificial intelligence*, pp. 603–649, 2020.

[2] K. Han, A. Xiao, E. Wu, J. Guo, C. Xu, and Y. Wang, "Transformer in transformer," *Adv Neural Inf Process Syst*, vol. 34, pp. 15908–15919, 2021.

[3] D. Q. Nguyen and A. T. Nguyen, "PhoBERT: Pre-trained language models for Vietnamese," *arXiv preprint arXiv:2003.00744*, 2020.

[4] A. Vaswani, "Attention is all you need," *arXiv preprint arXiv:1706.03762*, 2017.

[5] J. Devlin, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[6] Y. Liu *et al.*, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

[7] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," *arXiv preprint arXiv:1910.01108*, 2019.

[8] A. Agarwal and P. Meel, "Stacked Bi-LSTM with attention and contextual BERT embeddings for fake news analysis," in *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, IEEE, 2021, pp. 233–237.

[9] F. Monti, F. Frasca, D. Eynard, D. Mannion, and M. M. Bronstein, "Fake news detection on social media using geometric deep learning," *arXiv preprint arXiv:1902.06673*, 2019.

[10] P. Qi, J. Cao, T. Yang, J. Guo, and J. Li, "Exploiting multi-domain visual information for fake news detection," in *2019 IEEE international conference on data mining (ICDM)*, IEEE, 2019, pp. 518–527.

[11] T. N. Hieu, H. C. N. Minh, H. T. Van, and B. V. Quoc, "ReINTEL Challenge 2020: Vietnamese Fake News Detection usingEnsemble Model with PhoBERT embeddings," in *Proceedings of the 7th international workshop on Vietnamese language and speech processing*, 2020, pp. 1–5.

[12] N.-D. Pham, T.-H. Le, T.-D. Do, T.-T. Vuong, T.-H. Vuong, and Q.-T. Ha, "Vietnamese fake news detection based on hybrid transfer learning model and TF-IDF," in *2021 13th International Conference on Knowledge and Systems Engineering (KSE)*, IEEE, 2021, pp. 1–6.

[13] C.-V. Nguyen Thi, T.-T. Vuong, D.-T. Le, and Q.-T. Ha, "v3mfnd: A deep multi-domain multimodal fake news detection model for Vietnamese," in *Asian Conference on Intelligent Information and Database Systems*, Springer, 2022, pp. 608–620.

[14] K. D. Pham, D. Van Thin, and N. L.-T. Nguyen, "Improving Vietnamese Fake News Detection based on Contextual Language Model and Handcrafted Features," *Science and Technology Development Journal*, vol. 26, no. 2, pp. 2705–2712, 2023.

[15] T. O. Tran and P. Le Hong, "Improving sequence tagging for Vietnamese text using transformer-based neural models," in *Proceedings of the 34th Pacific Asia conference on language, information and computation*, 2020, pp. 13–20.

[16] T. H. Vo, T. L. T. Phan, and K. C. Ninh, “DEVELOPMENT OF A FAKE NEWS DETECTION TOOL FOR VIETNAMESE BASED ON DEEP LEARNING TECHNIQUES.,” *Eastern-European Journal of Enterprise Technologies*, vol. 119, no. 2, 2022.
[17] B. Trstenjak, S. Mikac, and D. Donko, “KNN with TF-IDF based framework for text categorization,” *Procedia Eng*, vol. 69, pp. 1356–1364, 2014.
[18] A. R. Lubis, M. K. M. Nasution, O. S. Sitompul, and E. M. Zamzami, “The effect of the TF-IDF algorithm in times series in forecasting word on social media,” *Indones. J. Electr. Eng. Comput. Sci*, vol. 22, no. 2, p. 976, 2021.
[19] Ho Quang Thanh and ninh-pm-se, “thanhhocse96/vfnd-vietnamese-fake-news-datasets: Tập hợp các bài báo tiếng Việt và các bài post Facebook phân loại 2 nhãn Thật & Giả (228 bài),” Feb. 2019.