Real-time Classification and Span-based Detection of Comments on Facebook Sales Posts

Tam Huu Minh Nguyen[1,2], Tinh Pham Phuc Do[1,2], Ngoc Dinh Duy Cao[1,2], The Tran Gia Nguyen[1,2], and Trong-Hop Do[1,3]

[1] University of Information Technology, VNU-HCM, Vietnam
[2] {20521871,20522020,20521661,20521940}@gm.uit.edu.vn
[3] hopdt@uit.edu.vn

Abstract. The exchange of goods, buying, and selling on social networks is now prevalent; in a sales post, many practical and useless comments exist for businesses and sellers. Sellers often filter out valuable comments manually. However, manual filtering takes a lot of effort and human resources while comments constantly increase. Therefore, in this project, we build a data set ViCCSP for the comment classification problem based on user intent, along with the ViSDC dataset for named entity recognition task. The ViCCSP dataset includes five labels (question, order, judge, spam, other), collected from sales posts in social networking groups on Facebook. The ViSDC dataset consists of four entities (phone number, quantity, product, and address). We build real-time systems for streaming data collection, training, and visualization. In this paper, we use Kafka language models for system implementation. PhoBERT-base gives the highest results in task 1 (comment classification), reaching 91.15% for measuring accuracy and 87.24% for measuring F1 scores. For task 2 (NER), XLM-RoBERTa-large gives the highest results, reaching 96.03% - accuracy, 94.37% - F1 at the word level, 94.77% - accuracy, 89.39% - F1 at the span level, 82.88%-strict. In addition, we have implemented the system and satisfied the users.

Keywords: Text classification · Facebook · Sales posts · Kafka.

1 INTRODUCTION

E-commerce is developing strongly, and the need to buy and sell online is gradually becoming popular, bringing high revenue and promoting products more widely than traditional ones. In a sales post, many people comment with different purposes, such as questions about products and services, comments to the order, and buying products; there are also judging comments, product reviews, service, seller, or even spam comments that are disruptive, meaningless or have nothing to do with the product. Therefore, for sellers to filter out valuable comments for their business from a large number of comments is time-consuming. Having order comments provide detailed information such as phone number, address, product quantity, and other requirements not only helps sellers grasp information quickly but also saves time, helping the transaction go smoothly but also creating a professional and trustworthy trading environment. Moreover, social media users comment continuously, so manual, fast, and continuous filtering requires much effort.

The automatic comment classification problem is one of the famous and essential tasks in natural language processing. The input of this task is a text to be classified, and the output is the label of that text. This problem allows us to categorize texts for different purposes. For example, in the "Review spam detection", this task helps classify spam and non-spam, allowing the removal of spam reviews and keeping useful ones. As a result, businesses can rely on such reviews to improve their efficiency.

Named Entity Recognition (NER) is a vital task in natural language processing, involving identifying and classifying specific entities like names of people, organizations, and locations within text. It plays a crucial role in various applications such as information extraction and sentiment analysis. For instance, in financial news analysis, NER helps extract key entities enabling investors to make informed decisions.

Facebook is one of the largest social media platforms in the world, making it a bustling marketplace for buying and selling goods. As a result, there are numerous comments on a single sales post with various purposes, making it difficult for sellers to track and identify comments used for purchasing, inquiring about products or services, or spam comments. Therefore, technology is needed to assist sellers in dealing with this issue along with comment classification tasks. Several frameworks, such as Spark Streaming, Apache Flink, and Apache Kafka, are available to handle continuous and large-scale comment processing.

In this project, we use comment classification for Facebook sales posts and extract information of order comments. Besides, build a real-time classification system to help businesses on Facebook filter out valuable comments quickly and continuously over time. Our contributions are as follows:

- We have built a dataset named ViCCSP of 14,352 data points for the classification problem of 5 labels: Question, Order, Judge, Spam, and Other, with strict rules and high quality.
- We have built a dataset named ViSDC of 4,017 data points for the named entity recognition problem of 5 labels: Phone Number, Quantity, Product and Address.
- With the classification task on the ViCCSP dataset, we have implemented various models and identified the model with the highest performance, achieving an accuracy of 91.15% and an F1-score of 87.24%. Additionally, we have conducted a detailed analysis of the results.
- We've completed the task of named entity recognition on the ViSDC dataset by employing various models and identifying the one with the highest performance. Our top-performing model achieved an accuracy of 94.8% and an F1-score of 89.55%. Additionally, we conducted a detailed analysis of the results for further insights.

- We developed a system that directly collects Facebook data, classifies comments, outputs comment order information, and visualizes posts and comments in real-time using Kafka and Streamlit.

Our dataset ViCCSP is the first in Vietnam to perform this task. Our system can be applied in practice, supporting businesses on the Facebook social network to achieve better sales, faster customer engagement, and reduced effort.

The remaining sections of the paper are as follows. Section 2 presents related datasets and methods for the comment classification task. Section 3 provides a detailed description of the data collection process and preliminary statistics about the dataset. Section 4 discusses the language models that we experimented with. Section 5 presents the experimental pipeline and the results obtained. Section 6 discusses the data streaming part using Kafka in this report. Section 7 is the conclusion and our future directions for development.

2 RELATED WORK

2.1 Datasets

The available foreign language datasets are plentiful in the field of research and analysis of social media comments. However, the Vietnamese dataset is still small, especially in sales via the social network Facebook. Relevant datasets:

A Large-scale Dataset for Hate Speech Detection on Vietnamese Social Media Texts [8]. The Vietnamese dataset for detecting hate language on social networks includes more than 30,000 comments, with each comment assigned one of three labels: CLEAN, OFFENSIVE, and HATE.

ToxLex_bn [12]: A curated dataset of bangla toxic language derived from Facebook comment. The dataset is collected from more than 2 million comments on Facebook. After processing, the dataset consists of 1959 lines with nine attributes for toxic detection problems on social networks.

UIT-ViSD4SA [10] was released in 2021 as a dataset comprising 35,396 human-annotated spans from 11,112 feedback comments on phones for evaluating span detection for aspect-based sentiment analysis. All these feedbacks were collected from e-commerce platforms. These 4 aspects that users pay the most attention to and provide feedback on are performance, battery, features, and camera, with three emotional sentiment levels: positive, negative, and neutral.

PhoNER_COVID19 [14], released in 2021, is a dataset for the NER task, comprising essential information related to COVID-19, extracted from reputable Vietnamese news sources such as VnExpress, ZingNews, BaoMoi, and ThanhNien. The dataset includes 35,000 entities with 10 entity types across a total of 10,000 sentences. This dataset contains the largest number of entities compared to existing Vietnamese NER datasets.

ViHOS [3], a social media dataset, consists of 11,056 comments with each comment assigned one of three labels: CLEAN, OFFENSIVE, and HATE, derived from the ViHSD dataset. ViHOS has 5,360 comments with hate and offensive spans and 5,696 clean comments.

Currently, datasets for comments in the context of online selling on Facebook are limited. Building and publishing new datasets in this field would be crucial to support research and evaluation of interactions between sellers and buyers, as well as customer opinions and feedback about products and services. This new dataset will expand the research and analysis capabilities in online selling through social media and assist businesses in gaining a better understanding of customer attitudes and opinions, enabling them to make accurate and effective business decisions.

2.2 Method

The classification problem has many approaches, such as traditional machine learning and deep learning models. Some works using traditional machine learning methods can be mentioned as [5], The authors used Random Forest, MultiNormial Naive Bayes, Logistic Regression, Decision Tree, Support Vector Classifier, and XGBoost models to perform the emotional classification of tweets on Twitter. The authors also used Kafka and Apache Spark to build a system that directly collects and classifies tweets in real-time.

For deep learning, we can mention the work of [16], the authors used combined models such as CNN-LSTM, LSTM-CNN, BiLSTM-CNN, and BiLSTM-TE to classify the topics of reviews, achieving quite good results. In addition, language models have also been used for classification tasks and have achieved state-of-the-art performance, including models like RoBERTa, XLM-RoBERTa, and PhoBERT.

In Vietnamese, some works also use language models for classification tasks. In the paper [15], PhoBERT achieved the highest accuracy of 86.89% for the task of classifying reviews as spam or not and 72.17% for the task of classifying reviews into different spam categories. The paper [11] used Spark Streaming and various machine learning and deep learning models to classify comments on social media in real-time for the hate speech task. For offline experiments, the PhoBERT-CNN model achieved the highest results on both datasets, reaching 64.43% F1-score and 87.17% accuracy on the ViHSD dataset and 90.89% F1-score and 98.26% accuracy on the HSD-VLSP dataset.

3 DATASET

In this project, we introduce two datasets ViCCSP and ViSDC, collected from sales posts in Facebook groups. The dataset contains information about sales posts, including the content of the posts and the comments within those posts. We use this dataset in the real-time comment classification task and name entity recognition using Kafka to help sellers quickly and accurately classify comments and extract information from order comments.

Our dataset will improve the quality and effectiveness of research related to the online marketplace on Facebook while meeting the increasing demand for Vietnamese data in online business.

3.1 Data Collection

Selenium[4] is a popular automation tool widely used for software testing and performing automated actions on web browsers. Specifically, Selenium supports popular browsers like Chrome, Firefox, and Safari, allowing interaction and control of the web browser using Python programming code, automating tasks, and efficiently collecting data from web pages.

Beautiful Soup [13] is a powerful Python library for parsing and extracting data from HTML and XML. With Beautiful Soup, we can easily filter and extract elements, classes, and attributes in the HTML document of a web page. This allows us to extract the necessary data from posts and comments in Facebook groups quickly and efficiently.

In this study, we use Selenium and XPath to find the element containing the post content and BeautifulSoup to find the elements containing the comments of the corresponding posts. Additionally, we use Selenium to automate the data collection process, starting with opening the browser and logging into Facebook, then accessing the sales posts and displaying the comments.

3.2 Annotation Guidelines

3.2.1 ViCCSP In this part, we construct a new dataset based on the post's purpose for the comment classification task. Therefore, we have built a set of labeling rules to create this dataset. Our dataset consists of five labels: 1-Question, 2-Order, 3-Judge, 4-Spam, and 5-Other (The numbers represent the labels). Below are the definitions of each label:

- **Question:** Questioning comments are questions or opinions about the product that users want answered or clarified. These comments indicate a need for more understanding or further information before purchasing. Questioning comments often relate to technical specifications, features, usage instructions, quality, warranty, or other inquiries related to the product..

- **Order:** Ordering comments are inquiries related to ordering products. These are comments on the nature of the order.

- **Judge:** Judging comments are opinions, feedback, and product descriptions provided by buyers and sellers (product introductions), whether they have purchased the product (initial evaluations based on the images posted for sale).

- **Spam:** Spam comments are comments that are not meaningful or contain irrelevant content, such as links to unrelated websites, advertisements for loans, online games, gambling, etc.
- **Other:** These comments do not fall into the categories mentioned above and include promotional comments, advertising other products, or tagging other users.

[4] https://www.selenium.dev/

Some comments are difficult to categorize in the data set due to various challenges, such as non-standard language, spelling mistakes, sentence structure, and abbreviations. Through the labeling process and discussions among team members, we encountered some challenging cases, which include the following:

– In cases where a comment could belong to multiple labels, we assigned the label based on the priority of each category. The priority was determined according to the prevalence of each label in reality. The priority ranking from high to low is as follows: "Order", "Question", "Judge", "Spam", and "Other".

Table 1: An example of multiple labels in one comment.

Comment	Labels and explanations
Mình mua 2 cái, bạn gởi ship xuống thị trấn cần thạnh huyện cần giờ được ko tính ship luôn là bao nhiêu cho mình biết nha (***I bought 2 pieces****, can you send it to Can Thanh town, Can Gio district? If it includes the shipping, how much is the total? Let me know)*	Ở bình luận có 1 ý có nội dung về Đặt hàng (***Mình mua 2 cái***) do vậy bình luận này được gán là nhãn 1 (Thắc mắc). (*In the comment, there is a comment with content about Order (****I bought 2 pieces****), so this comment is labeled as 1 (Questions)*.)

– Comments containing URLs, if the URLs are from e-commerce platforms such as Shopee, Lazada, Tiki, etc., will be labeled as "Other", while all other URLs will be labeled as "Spam".
– Comments containing only the phrase "**ib**" (inbox) or comments with similar meaning will be labeled "Questions".

Table 2: Some examples in the dataset.

Example 1 **Input:** **Post:**"xả kho bát sứ trắng giá #20k 1chục bấm tham.gia săn đồ giá rẻ để mua mua 3 chục miễn phí ship bát loại 1 đẹp lắm" (Discharge stock of white porcelain bowls price #20k 10 bowls. Join the "săn đồ giá rẻ" group to buy 30 beautiful bowls of type 1 for freeship.) **Comment:** "có cho kiểm tra không sóp ơi" (Can I check, shop?) **Output: 1** (**Explanation:** Because the comment content intends to ask the seller about the service that comes with the product, the label is "Question", encoded as number 1.)
Example 2 **Input:**

Post:“xả lỗ đi đẻ thanh lý hết #39k/1 can dầu 5 lít chú ý: mỗi người chỉ đc đặt 1 can ạ ktra hàng trc khi thanh toán tặng thêm 1 chai nước mắm áp dụng cho khách tham gia săn đồ giá rẻ”
(Discharge the hole to give birth to liquidate all #39k/1 5 liter can of oil. Attention: each person can only order 1 can, check the goods before paying, get an extra bottle of fish sauce, applied to customers participating in "hunting cheap goods")
Comment: “cho chị 1 can dầu 5lit” (Please give me a 5-liter can of oil.)

Output: 2
(**Explanation:** The comment means that she wants to buy a 5-liter can of oil, so this comment belongs to label 2 (Order).

Example 3
Input:
Post:“ms 2146 chả cá bao ngon mọi người ơi 1kg : #120k (2 bịch được 4 miếng tròn vừa) chả được làm từ cá hố và cá mú nên dai và rất ngọt thịt mọi người yên tâm về chất lượng chả được ướpxem thêm đậm vị dùng để ăn cơm , bánh mì làm chả sốt cà chả chiên nước mắm hoặc có thể đem nấu canh chua , nấu bánh canh hủ tiếu đều được ship đồng giá 10k mua từ 1kg e freeship”
(ms 2146 delicious grilled chopped fish, everyone 1kg: #120k (2 bags get 4 medium pieces). They are made from halibut and grouper and are chewy and very sweet. Everyone can be assured of the quality of the marinated grilled chopped fish, see more flavor used to eat rice, bread to make fried fish sauce, or you can cook sour soup or cook noodles soup. Ship for the same price 10k, buy from 1kg I freeship)
Comment: “chả ngon lắm chị” (The grilled chopped fish is so delicious)

Output: 3
(**Explanation:** The content of the Post is to sell grilled chopped fish. The comment has content praising the quality of "delicious" fish cakes, so this comment belongs to label 3 (Judge).

Example 4
Input:
Post:“thanh lý hết nghỉ bán kệ tủ lạnh, máy giặt sale còn #59k siêu thị đang bán 250k ạ kệ máy lọc nước #49k kệ chống rung, giảm ồn, chống ẩm ướt. bảo vệ máy giặt, tủ lạnh máy nào cũng vừa các bác nhé. nó điều chỉnh đc độ dài để vừa với máy ạ còn vài cái dọn kho chứ ko lời lãi gì nữa aj”
(Liquidate all the shelves of refrigerators and washing machines on sale, and #59k supermarkets are selling 250k shelves of water purifiers #49k shelves for anti-vibration, noise reduction, and moisture resistance. Protect your washing machine, refrigerator, or any machine that will fit you. It can be adjusted in length to fit the machine. I still have a few to clean the warehouse but not make any profit.)
Comment: “xinchosdftothon de”

Output: 4
(**Explanation:** The above comment has no meaning, cannot be translated, and does not belong to acronyms, teen code, or used on social networks, so this comment belongs to label 4 (Spam)..

Example 5
Input:
Post:"#sg vĩnh lộc a bình chánh 13 pro vn/a máy nguyên zin hình thức 97% và có rịa tí lưng như ảnh pin 88% 13.800.000đ (máy và ốp) trách nhiệm xài 1 tuần lỗi em hoàn tiền 0931279599"
(#sg Vinh Loc a Binh Chinh 13 pro vn/a telephone is original 97% form and has a small scratch on the back like the picture battery 88% 13,800,000 VND (machine and case) responsible for using 1-week error, I refund 0931279599)
Comment: "trường sơn" (Truong Son)

Output: 5
(**Explanation:** The above comment is just mentioning the name of one person, not in any of the 4 labels: Question, Order, Judge, Spam, so this comment belongs to label 5 (Other).

3.2.2 ViSDC In this part, we constructed a dataset called ViSDC for extracting information from comments classified as orders in the previous task. After reviewing the dataset, we defined four main aspects in the order labels including: 1-Phone number, 2-Quantity, 3-Product, 4-Address.

E 1 chiếc$_{\text{Quantity}}$ màu xanh$_{\text{Product}}$ 0939074xxx$_{\text{Phone_Number}}$ ấp Long Thuận xã Long hậu huyện lai vung tinh đong thap$_{\text{Address}}$

Fig. 1: An example for 4 aspects in ViSDC dataset.

- **Phone number**: Text containing information about the buyer's phone number is mentioned. It may include special characters such as (090), 090.902.123, 094,1312,123.
- **Quantity**: Contains information about quantity and unit, often accompanied by the product. It can be written in numbers (10 lon, 5 thùng *(10 cans, 5 crates)*), written in words (hai cái, một thùng *(two pieces, one crate)*), weight (1kg, 5 cân, hai ký *(1kg, 5 pounds, two kilograms)*), and length (một mét, 1 mét *(one meter, 1 meter)*.
- **Product**: The name of the product that the buyer refers to. It can be the way the buyer refers to help the seller understand (1 set 99k).
- **Address**: The text containing information about the address, including: place name; house number, street number; hamlet, village; ward, commune; district, town; province, city.

Furthermore, we encountered many cases where characters were stuck together with no spaces, causing the labels to be stuck together, resulting in confusion. Therefore, we decided to ignore such cases.

3.3 Data preprocessing

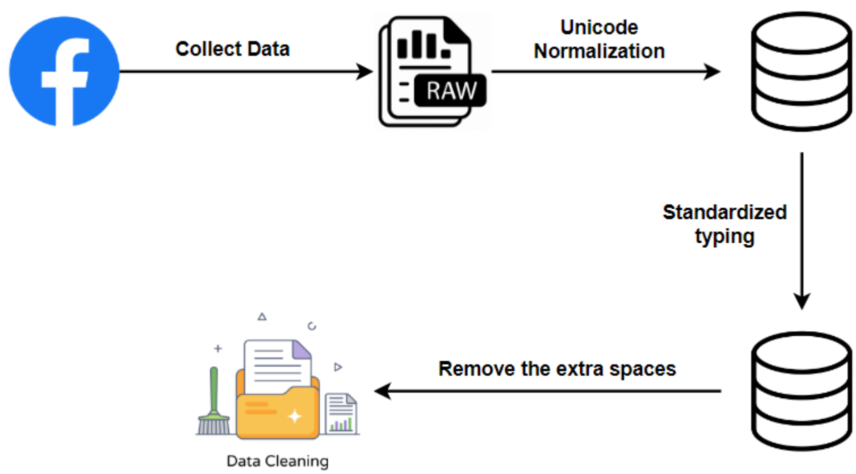

Fig. 2: Data preprocessing process.

The wording of the social media comment data domain is informal, standard Vietnamese style, contains many abbreviations, many bear typing methods, contains many emoticons, as well as sentences containing many residuals, this affects the learning of the model, can reduce performance, so we process comments and posts but in the same order as shown in the image 2.

Unicode Normalization: there are many Unicode Conversion Formats (UTFs), such as UTF-8, UTF-16, and UTF-32 are widely used, but our choice is Normalization to UTF-8.

Standardized typing: comments and posts on social networks will have many different ways of typing accents, but are one word: "hóa" and "hoá" These two words, if put into the computer, will be recognized as two different words but essentially the same. This problem increases the size of the lexicon as well as the wrong encoding, leading to confusion and reducing efficiency. That's why we process them before we label the data. For words with only one vowel, the accent will be on that vowel; for words with two vowels, the accent will be on the first vowel. For example, "hoá" will be converted to "hóa."

Finally, we remove the extra spaces at the beginning, the end of the paragraph, and the places that contain a lot of space into one space.

3.4 Annotation Process

3.4.1 ViCCSP Once the labeling rule is in place, we disseminate it to 4 annotators (CTVs) then each assigns 150 data points to test their understanding of the rule. After this round, four annotators try to assign 600 data points to calculate the consensus among annotators, and if the goal of good agreement among the annotators is achieved, the dataset will be officially built; if not, we will have to re-populate the rule and reassign until the good type is performed, in this round we also let the four annotators cross-check, discuss challenging cases to update the set of solutions, The results of the consensus among annotators calculated by the Krippendorff method reached more than 85%, According to the ranking table for annotation agreement in categorical data [4], Table 3 shows that the average agreement among pairs was 0.85, which is considered almost perfect as it falls between the range of 0.81 and 1.00. The results calculated by Cohen's Kappa index are presented in the table below table 3.

Table 3: Inter-annotator agreement between four annotators.

	CTV 1	CTV 2	CTV 3	CTV 4
CTV 1	1	0,87	0,82	0,84
CTV 2	0,87	1	0,85	0,86
CTV 3	0,82	0,85	1	0,86
CTV 4	0,84	0,86	0,86	1
Average [5]	0,84	0,86	0,84	0,85
Overall Average	0.85			

3.4.2 ViSDC For the information extraction task, we also disseminated the labeling rules to 4 annotators, ensuring consistency and reliability in the annotation process. Following this, we rigorously evaluated the inter-annotator agreement among the 4 annotators on a subset of 200 data points.

Furthermore, we iteratively refine the annotation process until the average agreement among annotators reaches the desired threshold of at least 80% on the F1-score metric. This iterative approach ensures that any discrepancies or ambiguities in the labeling rules are addressed, thereby enhancing the quality and consistency of the annotated dataset.

Additionally, we investigate the impact of excluding the O label during the evaluation process. The O label, representing entities that do not belong to any specific category, often constitutes a significant portion of the text and may introduce noise or ambiguity in the annotation. By evaluating the agreement both with and without the O label, we gain insights into the annotators' ability to accurately identify and categorize entities of interest.

Based on the findings presented in Table 4 and 5, The agreement among the 4 annotators reached a level of 93%, which demonstrate high levels of agreement among annotators, we can confidently conclude that the annotators have a strong grasp of the labeling rules and can effectively annotate the dataset.

Table 4: The agreement among the 4 annotators on the F1-score scale with the "O" label.

	CTV 1	CTV 2	CTV 3	CTV 4
CTV 1	1	0.93	0,94	0,94
CTV 2	0,93	1	0,94	0,92
CTV 3	0,94	0,94	1	0,92
CTV 4	0,94	0,92	0,92	1
Average [6]	0,94	0,93	0,93	0,93
Overall Average	0.93			

Table 5: The agreement among the 4 annotators on the F1-score scale without the "O" label.

	CTV 1	CTV 2	CTV 3	CTV 4
CTV 1	1	0.92	0,94	0,93
CTV 2	0,92	1	0,95	0,91
CTV 3	0,94	0,95	1	0,90
CTV 4	0,93	0,91	0,90	1
Average [7]	0,93	0,93	0,93	0,92
Overall Average	0.93			

3.5 Dataset Overview

After the above steps of collecting, pre-processing, and labeling, the data set ViCCSP is a data set of social network users' comments on Facebook sales articles, with size 14,352 data points, with three attributes "Post", "Comment", and "Label". Four annotators manually label the dataset with an overall average inter-annotator agreement 85%. We then divide the data into three sets of train, validation and test with the ratio of 8:1:1 to train the language models for solving the problem.

For the ViSDC dataset, we have 3654 data points with 4 extraction aspects: "Phone number", "Quantity", "Product", and "Address". With an inter-annotator agreement of over 93% among 4 annotators. We also divided the ViSDC dataset into three sets: train, validation, and test with a ratio of 8:1:1 for model training.

3.5.1 ViCCSP The distribution of labels in the datasets (Fig 3) is not uniform. Labels "Question" and "Order" account for a large number of samples in both train, validation, and test sets. Meanwhile, the "Judge" and "Spam" labels account for fewer samples. The "Other" label has an average sample count.

The main reason for this is that because comments on sales articles often focus on asking questions about the product or service provided, as well as ordering and asking for detailed information about the product, users often need to understand product quality, delivery method better, or verify the information before making a purchase decision. On the other hand, with sales articles on Facebook, the number of comments leaving phone numbers or closing orders online also accounts for a large number.

With spam comments, Facebook has a monitoring policy. Spam comments can be deleted or hidden automatically. Therefore, the number of spam comments remaining on the article will be less, resulting in a small number of spam labels in the dataset.

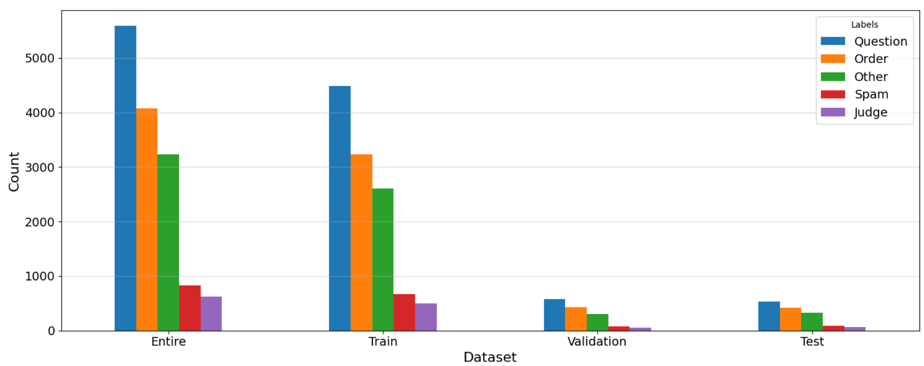

Fig. 3: Number of each label in the dataset.

3.5.2 ViSDC According to Figure 4, we can observe certain characteristics regarding the distribution of aspects within the ViSDC dataset. Aspects such as "Quantity" and "Phone Number" constitute the highest proportion, whereas "Address" and "Product" have a lower prevalence within the sentences. This might suggest that when placing orders, buyers typically provide order quantities to the seller or furnish their phone numbers for direct consultation.

In addition to the frequency of each aspect within a sentence, analyzing the distribution of aspect quantities within each sentence is an important method to gain deeper insights into how aspects are distributed throughout the text. According to Figure 5, we can observe that the distribution peaks from 1 to 2 aspects per sentence, with very few sentences having 5 or more aspects. This indicates that the texts in the dataset tend to focus on a few key aspects predominantly. Combining this analysis with the previous findings, we can infer that users generally leave comments with fewer aspects while placing orders.

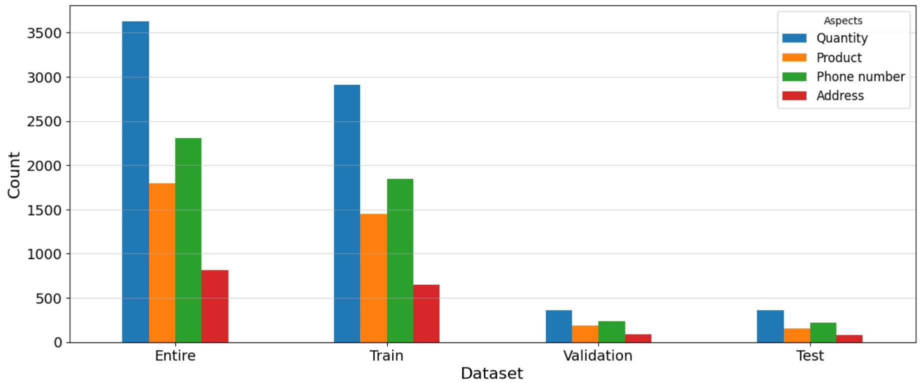

Fig. 4: Number of comments with aspects in dataset.

Furthermore, each sentence in the dataset always has at least one label because we selectively chose sentences containing aspects for labeling purposes.

4 MODEL

4.1 BERT

BERT [2] is a natural language processing (NLP) model based on the BERT (Bidirectional Encoder Representations from Transformers) architecture. One notable feature of BERT is its ability to understand the context and meaning of each word in a sentence by considering both the preceding and succeeding parts of that word. This allows BERT to capture subtle nuances and semantic relationships between words accurately. BERT has achieved impressive results in various language tasks, surpassing many previous traditional models. With its strong contextual understanding and rich language representation, BERT plays a crucial role in applications such as text classification, machine translation, information extraction, and many other fields in natural language processing. In task 1 (comment classification), we use BERT and multilingual BERT (mBERT) and in task 2 (NER), we use mBERT.

4.2 XLM-RoBERTa

XLM-RoBERTa [1] is a deep learning model based on the Transformer architecture developed by Facebook AI Research (FAIR). Introduced in 2019, XLM-RoBERTa is an extended and improved version of the RoBERTa [7] model (Robustly Optimized BERT Pretraining Approach), and it has achieved impressive results in various natural language processing tasks. XLM-RoBERTa is trained on data from multiple languages (2.5TB of data), including Vietnamese and other languages worldwide. This enables XLM-RoBERTa to learn the syntax and

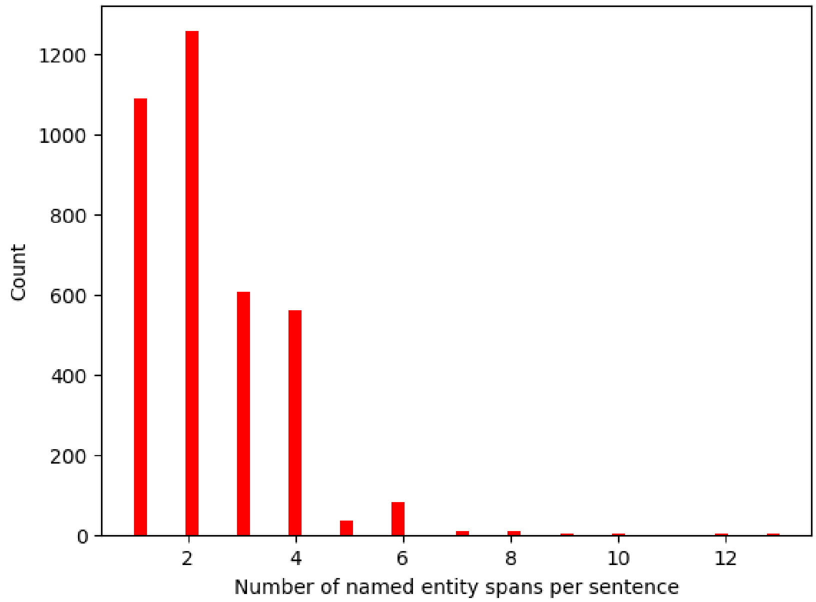

Fig. 5: Number of spans per comment in the dataset.

semantics of multiple languages, enhancing its ability to comprehend and analyze multilingual text.

4.3 PhoBERT

PhoBERT [9] is a specialized natural language processing (NLP) model designed for the Vietnamese language, developed by the research team at VinAI Research. It is built upon the RoBERTa model and trained on a large amount of Vietnamese data (20GB of data from Wikipedia and news websites). This enables the model to effectively understand and work with the Vietnamese language, with the ability to comprehend the contexts and meanings of words in Vietnamese. We only use PhoBERT for task 1.

4.4 BART

BART [6] (Bidirectional and Auto-Regressive Transformers) is a language model based on the transformer architecture, introduced by researchers at Facebook AI Research (FAIR) in 2019. The BART model combines elements of both bidirectional transformer models (like BERT) and autoregressive models (like GPT). It is pretrained on large-scale unsupervised tasks and can be fine-tuned for supervised tasks such as classification, summarization, machine translation, and various other natural language tasks. We only use BART for task 1.

4.5 ViSoBERT

ViSoBERT is a pre-trained language model for Vietnamese that was released in 2023. It shares the same architecture as XLM-RoBERTa and has approximately 100M parameters. It was trained on nearly 1GB of social media data. We only use ViSoBERT for task 1.

4.6 ViBERT

ViBERT is a BERT model but pre-trained in Vietnamese. The training data used for ViBERT is approximately 10GB, with a vocabulary size of over 38,000 tokens. We only use ViBERT for task 1.

4.7 ViDeBERTa

ViDeBERTa is a pre-trained language model for Vietnamese that was released in 2023. It shares the same architecture as XLM-RoBERTa and comes in three versions: ViDeBERTa_xsmall, ViDeBERTa_base, and ViDeBERTa_large, with 22M, 86M, and 304M parameters, respectively. In this paper, we use ViDeBERTa_base for task 2.

5 EXPERIMENTS AND RESULTS

5.1 Experiment settings

The experimental process is illustrated in Figure 6. After performing the preprocessing steps in Section 3, we proceeded with word segmentation to prepare the data for the PhoBERT model. We used the dataset without performing word segmentation for the BERT, XLM-RoBERTa, and BART models.

All four models were experimented with using the base version. The BERT, XLM-RoBERTa, and BART models were configured with five epochs, a max sequence length of 512 to ensure capturing the entire content of posts and comments, a batch size of 8, and a learning rate of 5e-06. As for the PhoBERT model, we conducted experiments with three epochs, a max sequence length of 256, and a learning rate of 5e-05.

After training the models, they were evaluated based on various metrics. The best-performing model would be selected and used for real-time comment classification. The data would be collected and processed online before being fed into the prediction model.

5.2 Evaluate metrics

In this study, we evaluate two tasks: classification and name entity recognition (NER).

Classification task

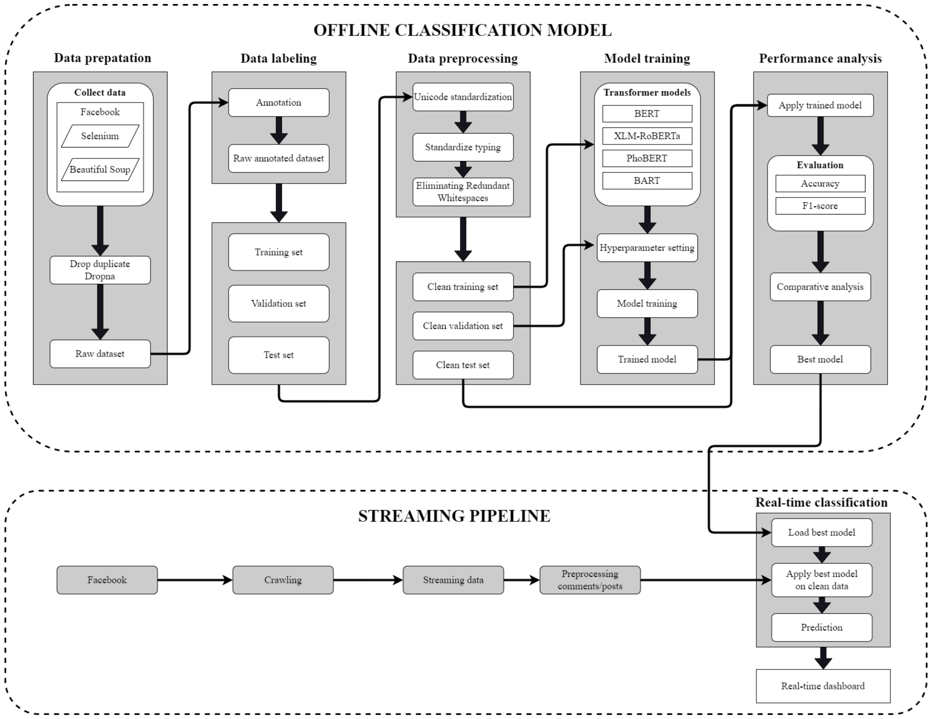

Fig. 6: Experimental procedure.

In the classification task, we focus on two main evaluation metrics: Accuracy and F1-macro.

$$Accuracy = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \tag{1}$$

$$F1\text{-macro} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{2}$$

By emphasizing F1-macro, we ensure that the model performs well across all classes, including the minority classes. This implies that the model has the ability to classify accurately and consistently across different instances, ensuring fairness and reliability in the evaluation process.

NER task

For the NER task, we evaluate accuracy, F1, and strict metrics. For F1 and accuracy, we assess at two levels: word level and span level. Strict evaluation is performed at the word level and is defined as follows: if two samples have corresponding labels with perfectly matching words, they are counted as 1 point; otherwise, no points are awarded. At the word level, we calculate scores for each comment and then compute the average.

5.3 Experiment results

TASK1

Model	F1-score (%)	Accuracy (%)
PhoBERT-base	**87,24**	**91,15**
XLM-R-base	84,11	88,92
mBERT-base-cased	83,96	88,43
BERT-base-uncased	80,91	87,53
BERT-base-cased	79,34	86,06
BART-base	75,31	84,11

Table 6: Experimental results of the models on the test set.

Table 6 presents the results of the different models on the test set. The results show that there is no significant difference between the models. However, the multilingual or single-language models specifically designed for Vietnamese perform better than others. The highest-performing model is PhoBERT, followed by XLM-RoBERTa, and then mBERT, which ranks third.

TASK2

Based on Tables 7, 8, and 9, we can observe that the highest results on the test set are achieved when training the model with ViSoBERT (96.03% - accuracy, 94.37% - F1 at the word level, 94.77% - accuracy, 89.39% - F1 at the span level, 82.88%-strict). ViSoBERT achieves the best performance because its pre-training data shares characteristics similar to our data (social media domain).

Furthermore, comparing Tables 7 and 8, we can see that the results based on word-level evaluation generally yield higher accuracy and F1-score than those based on span-level evaluation. This is because an aspect can contain multiple words, and incorrectly predicting even a single word can lead to an incorrect prediction of the number of aspects in a sentence. On the other hand, the number of spans labeled for each aspect in a sentence is smaller, which significantly increases the error rate and reduces the performance when evaluating at the span level.

5.4 Results analysis

Analysis with aspect

To gain a deeper understanding of the data, we analyze the performance of the models on the test set for each aspect. As can be seen in Figure 7, the Phone number aspect achieves the best result (99.10%), which is significantly higher than the other aspects for all six models. This is because the phone number aspect spans are simply a sequence of numbers, making it easy for the models to identify them.

According to Figure 5, the Address aspect has fewer instances than the Product but achieves a higher result. This is because the Address structure has a

Model	DEV		TEST	
	Accuracy (%)	F1 (%)	Accuracy(%)	F1 (%)
XLM-RoBERTa-large	96.13	94.34	95.18	93.26
XLM-RoBERTa-base	96.2	94.52	95.77	94.34
mBERT-base-uncased	96.42	94.7	95.75	94.09
ViSoBERT	96.36	94.93	**96.03**	**94.37**
ViBERT	94.82	92.42	94.84	92.7
ViDeBERTa-base	87.13	80.55	86.91	80.53

Table 7: Results of the models on the dev and test datasets. The results are computed at the word level for each comment and averaged across comments.

Model	DEV		TEST	
	Accuracy (%)	F1 (%)	Accuracy(%)	F1 (%)
XLM-RoBERTa-large	95.48	88.81	94.04	87.09
XLM-RoBERTa-base	95.18	89.91	94.27	87.54
ViSoBERT	95.41	88.83	**94.77**	**89.39**
mBERT-base-uncased	95.48	88.08	94	86.92
ViBERT	93.65	85.3	93.46	85.39
ViDeBERTa-base	85.23	63.68	84.35	62.55

Table 8: Results of the models on the dev and test datasets. The results are computed at the span level.

clear pattern, including components such as street names, house numbers, cities, provinces, etc. This makes it easier for the model to identify spans belonging to the Address aspect. Identifying spans belonging to the Product aspect is more complex, as products can be described in many ways, such as product name, brand, description, etc. Therefore, identifying spans belonging to the Product aspect can be more challenging

5.5 Error Analysis

TASK1

After obtaining the results in Table 6, we analyzed the errors of the highest-performing model - PhoBERT. Figure 8 shows the confusion matrix of the model on the test set. We can observe that the two labels most frequently confused with each other are "Question" and "Other". Additionally, the label pair "Judge" and "Other" also show a considerable amount of confusion.

Table 10: Some mispredicted labels by the model.

Example 1

Post: “thủy sinh mềm xèo / cho anh em tập chơi, thả đâu cũng được, dễ sinh sản, không cần oxy bay lần anh em ưng em nha" *(Cheap seaweed / for beginners. You can place wherever; easy to breed, doesn't need oxygen, sell it gradually, please get in touch with me)*

Comment: "lao dai bạn kiểm tra tin nhắn ạ" *(lao dai please check your messages)*

Label: 5, **Predict: 1**

Example 2

Post: "GIƯỜNG PALLET GIÁ SIÊU YÊU GIƯỜNG PALLET SHIP TỈNH cực rẻ chiều cao linh động: 5cm, 7cm, 10cm, 15cm, 20cm, 30cm - size 1m2x1m9: #450k - size 1m4x2m: #520k - size 1m6x2m: #540k - size 1m8x2m: #590k - size 2mx2m: #790k _ size 2mx2m2: #850k khẳng định 1 lần nữa : bên em gỗ thông mới 100% không sử dụng gỗ pallet tái chế. có lỗi lầm cho đổi trả ngay lập tức, thợ đến tận nhà kiểm tra nếu hàng có vấn đề từ phía shop <PHONE NUMBER>" *(PALLET BED SUPER LOVED PRICE PROVINCE PALLET SHIP BED very cheap, flexible height: 5cm, 7cm, 10cm, 15cm, 20cm, 30cm - size 1m2x1m9: #450k - size 1m4x2m: #520k - size 1m6x2m: #540k - size 1m8x2m: #590k - size 2mx2m: #790k _ size 2mx2m2: #850k confirm once again: our pine wood is 100% new, not using recycled pallet wood. If there is a mistake, you can exchange it immediately. The mechanic will come to your home to check if there is a problem from the shop <PHONE NUMBER>)* "

Comment: "thúy m đã nt *(thúy i have messaged)*"

Label: 1, **Predict: 5**

Example 3

Post: "chi tiết hơn về mẫu ổ cắm sắp ra mắt" *(More details about the upcoming socket model)*

Comment: "tổng thể rồi hãy chi tiết. như này tôi thà mua cục sạc còn hơn." *(The big picture first, then the details. I'd buy a charger instead of this.)*

Label: 3, **Predict: 5**

Example 4

Post: "góc pass đồ dùng phú nhuận xin phép a/c, em cần dọn văn phòng nên cần pass nhanh các món đồ sau, ưu tiên đến tự lấy ạ. 2 kệ sách: 180k/ cái quạt treo: 200k bàn pha trà gỗ nguyên khối: 15 triệu tủ lạnh: 3tr 3 máy lạnh: 2.5tr/ cái 1,5 ngựa, 1.8 tr/ máy 1 ngựa bình ủ trà bình nấu nước sôi: 800k tủ chén: 500k bảng hiệu viết phấn: 600k xin phép a/c, em dọn nhà nên cần pass nhanh các món đồ sau, ưu tiên đến tự lấy ạ. 2 kệ sách: 180k/ cái quạt treo: 200k bàn pha trà gỗ nguyên khối: 15 triệu tủ lạnh: 3tr 3 máy lạnh: 2.5tr/ cái 1,5 ngựa, 1.8 tr/ máy 1 ngựa bình ủ trà bình nấu nước sôi: 800k tủ chén: 500k bảng hiệu viết phấn: 600k" *(I am passing on items quickly in Phu Nhuan District due to clearing my office, preferably trading in person. 2 bookshelves: 180k each/ wall fan: 200k/ wooden tea table: 15 million/ refrigerator: 3 million/ 3 air conditioners: 2.5 million for the 1.5 HP, 1.8 million for the 1 HP/ teapot: 800k/ cupboard: 500k/ chalkboard: 600k)*

Comment: "kathy mai c mua máy lạnh cũ k ạ" *(kathy mai Do you buy a second-hand air conditioner?)*

Label: 5, **Predict: 1**

Example 5

Post:“chuẩn bị đi nước ngoài nên pass vài món, tình trạng ghi rõ trên bài: 1 - loa kiểm âm edifier mr4 dùng được 2 tháng, còn rất đẹp và mới : 1100k 2 - bếp từ xiaomi dcl002cm fullbox còn bảo hành đến tháng 3 2024 : 380k 3 - quạt tích điện xiaomi 3life pin 8000mah gập fullbox còn bảo hành đến tháng 3 2024: 380k 4 - hút bụi giường deerma cm800 fullbox còn bảo hành tháng 3 2024: 380k 5 - tai nghe logitech g435 wireless nobox, đủ pk hết bảo hành : 550k 6 - bàn phím cơ ek1280s đen led rgb fullbox: 300k 7 - ram adata d50 8gb buss 3200 fullbox còn bảo hành : 350k xem trực tiếp tại nhà tp bắc giang,giá trên là đã giá chính xác, lấy trực tiếp hoặc bank ting ting có fix nhẹ xăng xe hoặc freeship, cod có cọc. không mặc cả quá nhiều cảm ơn mọi người đã đọc tin !” *(Preparing to go abroad, so I need to pass a few items. The condition is clearly stated on the post: 1 - edifier mr4 monitor used for two months, still very nice and new: 1100k, 2 - Xiaomi dcl002cm induction cooker full box with warranty until March 2024: 380k, 3 - Xiaomi 3life rechargeable fan 8000mah folding battery full box with warranty until March 2024: 380k, 4 - vacuuming derma bed cm800 full box with the warranty in March 2024: 380k, 5 - Logitech g435 wireless no box headset, full accessory out of warranty: 550k, 6 - mechanical keyboard ek1280s black led rgb fullbox: 300k, 7 - ram data d50 8gb buss 3200 full box with warranty: 350k watch directly at home in Bac Giang city, the above price is the exact price, have direct deposit, free shipping or light bank don't haggle too much, thank you everyone for reading!)*

Comment: “e lấy bàn phím nhé nhà bác chỗ nào ạ” **(I'll take the keyboard. Where is your house?)**

Label: 2, **Predict: 1**

The errors in the model mainly stem from the ambiguity in the data and guidelines. In examples #1 and #2, the guidelines state that responses from sellers to buyers should be labeled as "Other" if the context is clear and labeled as "Question" if it cannot be determined whether the comment is from the seller or the buyer. Accurately determining the context for each comment is a challenge for both humans and machines, which poses difficulties for the model in identifying the context in the sentences.

Another ambiguity in the guidelines is illustrated in example #3, where the label "Judge" is defined to include cases of both reviews from buyers and sellers (product introductions), whether they have made a purchase or not (initial reviews based on posted images), and reviews regarding the product and delivery process. With multiple cases being specified, some instances have too few reviews or lack clear indications, making them easily confused with the "Other" label.

Furthermore, the model is prone to confusion between "tagged names" in comments, as shown in example #4. This comment consists of a question between two buyers and does not mention the seller. The guideline specifies that comments

Model	DEV	TEST
XLM-RoBERTa-large	80.5	81.79
XLM-RoBERTa-base	80.45	81.7
ViSoBERT	81.34	**82.88**
mBERT-base-uncased	80.34	81.58
ViBERT	75.77	79.07
ViDeBERTa-base	52.09	53.26

Table 9: Results of the models on the development and test datasets. The results are computed with the strict accuracy metric at the span level.

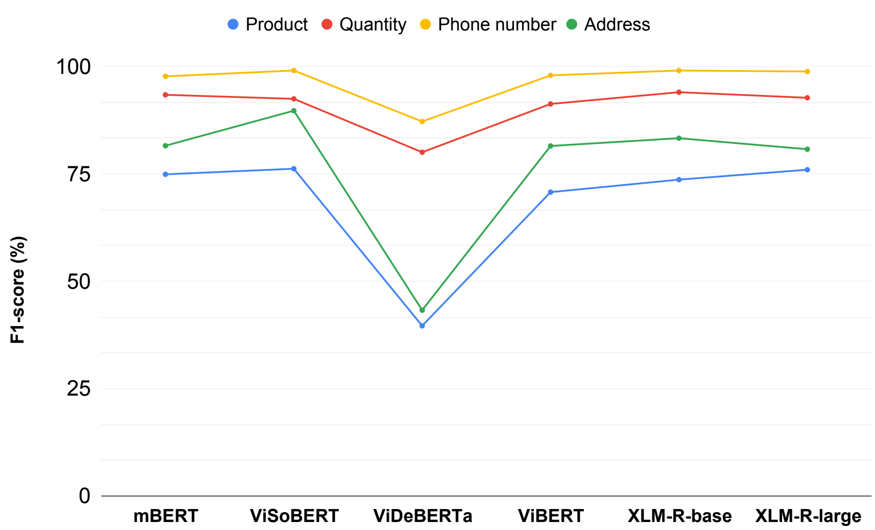

Fig. 7: Results of aspects for each model with F1-score at span level.

between users will be labeled as "Other". However, the context of this particular comment is ambiguous, leading the model to predict "Question".

The last case involves sentences that can have multiple labels. In example #5, the comment contains both "Question" and "Order" labels. The guideline prioritizes multiple labels with higher precedence ("Order" > "Question" > "Judge" > "Spam" > "Other"). However, occurrences of sentences with multiple labels are relatively rare, resulting in insufficient data for the model to learn these cases effectively.

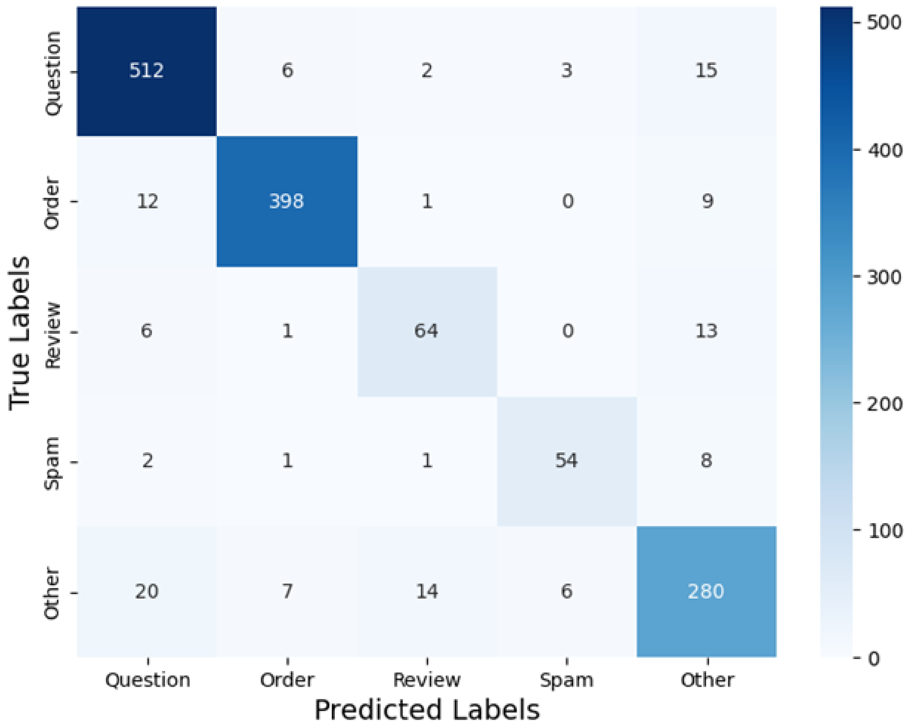

Fig. 8: Confusion matrix on the test set of the PhoBERT model.

6 STREAMING

6.1 Data streaming

Data Streaming is the process of transferring data continuously and without interruption at high speed. Instead of storing and processing data only once, as in the traditional way, data streaming processes the data immediately as it is created and continues while it is still in transit. This allows companies and organizations to react quickly and make decisions based on the latest data.

With the increasingly important role of data streaming, systems and technologies that support data transmission and real-time processing, such as Apache Kafka, Spark Streaming[8], Apache Flink[9],.. are becoming indispensable tools in modern IoT solutions and application development. In this report, we will implement processing data streams using Apache Kafka.

[8] https://spark.apache.org
[9] https://flink.apache.org

6.2 Kafka overview

Kafka[10] is a platform developed by LinkedIn, then transferred to the Apache Software Foundation and released as open source. Kafka is a real-time communication system that handles distributed data with high performance and reliability. With the simple and flexible 9 structure, Kafka has become one of the leading technologies in processing significant data streams, supporting building event-based applications, and integrating between different systems.

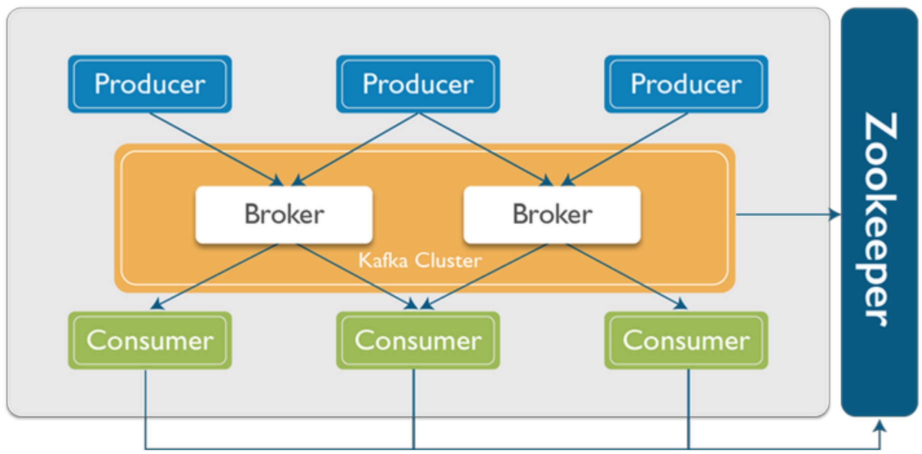

Fig. 9: Apache Kafka Architecture.

6.3 The architecture of the proposed application

The architecture of the proposed application is depicted in Figure 10. It consists of 2 parts as follows:

Part 1: Crawl Comments from Facebook and Submitted through Kafka Producer :

In this section, we combine Selenium and BeautifulSoup libraries to collect customer comments from sales articles on the social networking platform Facebook every 10 seconds. Along with that, we built a Kafka Producer that sends the messages including "Post", "Comment" and "User" collected above through a topic.

Before sending messages, we encode them with utf-8 encoding to ensure consistency in transmission. Messages are sent in separate lines for easy tracking and processing when received by the Consumer.

Part 2: Preprocessing and Classifying Comments using the XLM-R-Base Model:

[10] https://kafka.apache.org

Although the PhoBERT-base model gives good results, deploying it in a local environment is challenging. We have implemented using the XLM-R-Base model to classify comments in this application.

By creating a Kafka Consumer with the same topic, we can receive the messages sent by the Producer. After receiving the data, we decrypt the messages using the utf-8 encoder to restore the original and head data ready for preprocessing.

Next, we preprocess the user comment data, including cleaning, to match the input of the XLM-R-Base prediction model.

The results from the XLM-R-Base model will be stored in an Excel file and continuously updated. These data will then be made available for statistical analysis of the comments. We will display this analysis via a real-time dashboard. This result provides an overview of customer opinions on sales articles and helps businesses better understand customer feedback on their product or service.

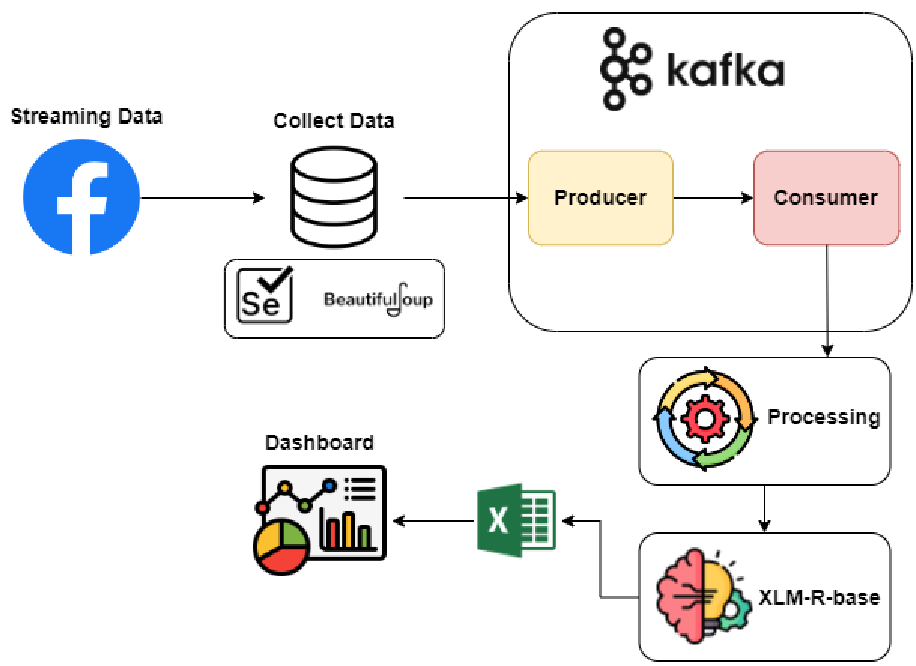

Fig. 10: The architecture of the proposed system.

7 CONCLUSION AND FUTURE WORK

In this article, we constructed two datasets: the first dataset, ViCCSP, contains comments and sales posts. ViCCSP is a dataset for comment classification with

5 labels: 1-Question, 2-Order, 3-Judge, 4-Spam, and 5-Other. For the ViCCSP dataset, we experimented, evaluated, and analyzed the dataset on PhoBERT, BERT, mBERT, XLM-RoBERTa, and BART models. PhoBERT-base achieved the highest results on ViCCSP, reaching 91.15% for accuracy and 87.24% for F1 score. We implemented and analyzed the performance of the models on these two datasets. The second dataset, ViSDC, contains comments regarding ordering content. ViSDC is a dataset for Name Entity Recognition with aspects such as b, c, d. For the entity recognition task, we evaluated on BCD models, and the best-performing model was ViSoBERT. The results with Accuracy, F1, Strict accuracy based on word evaluation method reached the highest at 96.03%, 94.37%, 82.88%, respectively, while for span level, the highest performance was 94.77% - Accuracy and 89.39% - F1 on the test set.

This study also presents a real-time comment classification application on Facebook using Apache Kafka, focusing on real-time capabilities and handling large volumes of data. We focused on training the model with a large amount of data and applying the trained model to process comments quickly and continuously.

In the future, for the entity recognition task, we aim to explore advanced NLP models capable of capturing detailed contextual information to improve entity identification in complex comments of the system. We will focus on refining the entity labeling process to ensure higher-quality datasets, enabling the model to effectively handle diverse and complex language expressions. Finally, we plan to build a system interface for practical deployment, making the system user-friendly and ready to serve.

References

1. Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., Stoyanov, V.: Unsupervised cross-lingual representation learning at scale. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 8440–8451. Association for Computational Linguistics, Online (Jul 2020). https://doi.org/10.18653/v1/2020.acl-main.747, https://aclanthology.org/2020.acl-main.747
2. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). pp. 4171–4186. Association for Computational Linguistics, Minneapolis, Minnesota (Jun 2019). https://doi.org/10.18653/v1/N19-1423, https://aclanthology.org/N19-1423
3. Hoang, P.G., Luu, C.D., Tran, K.Q., Van Nguyen, K., Nguyen, N.L.T.: Vihos: Hate speech spans detection for vietnamese. arXiv preprint arXiv:2301.10186 (2023)
4. Landis, J.R., Koch, G.G.: The measurement of observer agreement for categorical data. biometrics pp. 159–174 (1977)
5. Lasri, I., Riadsolh, A., Elbelkacemi, M.: Real-time twitter sentiment analysis for moroccan universities using machine learning and big data technologies. International Journal of Emerging Technologies in Learning **18**(5) (2023)

6. Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., Zettlemoyer, L.: BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 7871–7880. Association for Computational Linguistics, Online (Jul 2020). https://doi.org/10.18653/v1/2020.acl-main.703, https://aclanthology.org/2020.acl-main.703
7. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V.: Roberta: A robustly optimized bert pretraining approach (2019), http://arxiv.org/abs/1907.11692, cite arxiv:1907.11692
8. Luu, S.T., Nguyen, K.V., Nguyen, N.L.T.: A large-scale dataset for hate speech detection on vietnamese social media texts. In: Advances and Trends in Artificial Intelligence. Artificial Intelligence Practices: 34th International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE 2021, Kuala Lumpur, Malaysia, July 26–29, 2021, Proceedings, Part I 34. pp. 415–426. Springer (2021)
9. Nguyen, D.Q., Tuan Nguyen, A.: PhoBERT: Pre-trained language models for Vietnamese. In: Findings of the Association for Computational Linguistics: EMNLP 2020. pp. 1037–1042. Association for Computational Linguistics, Online (Nov 2020). https://doi.org/10.18653/v1/2020.findings-emnlp.92, https://aclanthology.org/2020.findings-emnlp.92
10. Nguyen, K.T.T., Huynh, S.K., Phan, L.L., Pham, P.H., Nguyen, D.V., Van Nguyen, K.: Span detection for aspect-based sentiment analysis in vietnamese. arXiv preprint arXiv:2110.07833 (2021)
11. Quoc Tran, K., Trong Nguyen, A., Hoang, P.G., Luu, C.D., Do, T.H., Van Nguyen, K.: Vietnamese hate and offensive detection using phobert-cnn and social media streaming data. Neural Computing and Applications **35**(1), 573–594 (2023), https://arxiv.org/abs/2206.00524
12. Rashid, M.M.O.: Toxlex_bn: A curated dataset of bangla toxic language derived from facebook comment. Data in Brief **43**, 108416 (2022)
13. Richardson, L.: Beautiful soup documentation. April (2007)
14. Truong, T.H., Dao, M.H., Nguyen, D.Q.: Covid-19 named entity recognition for vietnamese. arXiv preprint arXiv:2104.03879 (2021)
15. Van Dinh, C., Luu, S.T., Nguyen, A.G.T.: Detecting spam reviews on vietnamese e-commerce websites. In: Asian Conference on Intelligent Information and Database Systems. pp. 595–607. Springer (2022)
16. Zeng, Y., Zhang, R., Yang, L., Song, S.: Cross-domain text sentiment classification method based on the cnn-bilstm-te model. Journal of Information Processing Systems **17**(4), 818–833 (2021), http://jips-k.org/digital-library/2021/17/4/818