VIETNAM NATIONAL UNIVERSITY OF HOCHIMINH CITY

THE INTERNATIONAL UNIVERSITY

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

FEEDBACK ANALYTICS

MINING USER INSIGHTS FOR TECH ADVANCEMENTS

IN E-WALLET APPS

By

Dang Quoc Anh Duy

A thesis submitted to the School of Computer Science and Engineering

in partial fulfillment of the requirements for the degree of

Bachelor of Data Science

Ho Chi Minh City, Vietnam

Year 2024

FEEDBACK ANALYTICS: MINING USER INSIGHTS FOR TECH ADVANCEMENTS IN E-WALLET APPS

APPROVED BY: Advisor

_________________________________,

Assoc. Prof. Nguyen Thi Thuy Loan

APPROVED BY: Committee

_________________________________,

Assoc. Prof. Nguyen Van Sinh

_________________________________,

Assoc. Prof. Nguyen Thi Thuy Loan

_________________________________,

Dr. Mai Hoang Bao An

_________________________________,

Dr. Le Duy Tan

_________________________________,

Thesis Committee

ACKNOWLEDGMENTS

First and foremost, I would like to express my deepest gratitude to my advisor, Assoc. Prof. Nguyen Thi Thuy Loan, for her invaluable guidance, unwavering support, and insightful feedback throughout the research process. Her expertise, patience, and encouragement have been instrumental in shaping this work.

I am profoundly thankful to the School of Computer Science and Engineering for providing the necessary resources, facilities, and conducive research environment essential for conducting this study. Special thanks to the thesis team of Associate Professor Nguyen Thi Thuy Loan, whose constructive criticisms, collaborative efforts have significantly enriched this research endeavor.

My gratitude goes to my family for their unconditional love, support, and encouragement throughout this academic journey. Their belief in my abilities has been a constant source of motivation. Lastly, I am thankful to my friends for their understanding, and encouragement, and for being a sounding board for my ideas. This thesis would not have been possible without the contributions and support of all these individuals, and for that, I am profoundly grateful.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

This thesis delves into the intricate domain of feedback analytics, pivotal for propelling technological advancements within E-Wallet applications. Focusing on the Vietnamese market, the study explores the application of sentiment analysis and aspect extraction methodologies to user reviews sourced from digital distribution platforms, notably the App Store, about E-Wallet apps. Leveraging sophisticated computational techniques, including advanced sentiment analysis algorithms and aspect extraction methodologies, this research aims to uncover subtle insights hidden in user feedback.

The author proposes a comprehensive analysis of how to enhance digital wallet services in Vietnam. The sentiment analysis methodology employed involves a sophisticated fusion of Bidirectional Long Short-Term Memory (Bi-LSTM) and Bidirectional Gated Recurrent Unit (Bi-GRU) neural network architectures. Additionally, the study explores an alternative approach utilizing PhoBERT, a pre-trained language model based on the transformer architecture specifically tailored for Vietnamese. These deep learning techniques are used to effectively capture and analyze the temporal dependencies in textual data extracted from user reviews.

For aspect extraction, Latent Dirichlet Allocation (LDA) is applied, enabling the identification of latent topics and key aspects discussed within the reviews. Subsequently, multi-label models are constructed for aspect detection, facilitating the simultaneous categorization of reviews into multiple predefined aspects or topics. By integrating these advanced techniques, this research aims to achieve comprehensive sentiment analysis and aspect detection, thereby providing invaluable insights for refining and enhancing E-Wallet applications.

CHAPTER 1: INTRODUCTION

1.1. Background

With the expansive growth of the internet and technological advancements, traditional activities have undergone significant digital transformation across various aspects of life. This shift is particularly noticeable in financial transactions, where E-Wallet applications have become essential tools in modern digital commerce. E-Wallets facilitate convenient and secure transactions, enabling users to make purchases and conduct transactions seamlessly through their smartphones with just a touch. As a result, the necessity for physical cash is diminishing, giving way to the convenience of digital payments via E-Wallets.

The adoption of mobile payment solutions is accelerating globally. This adoption not only simplifies the transaction process but also contributes to the exponential growth of user metadata and behaviors on online platforms. The data generated from these transactions provides valuable insights that businesses and organizations can leverage to inform strategic decisions and enhance user experiences, highlighting the critical role of E-Wallet applications in shaping the future of digital finance and commerce.

According to FiinGroup –Vietnam’s leading integrated financial data service provider, the number of active digital wallets in Vietnam is projected to increase by 40% to 50 million by the end of 2024 [1]. This underscores the importance of sentiment analysis as a key research area within Natural Language Processing (NLP)—where numerous businesses actively engage in discerning users’ attitudes towards their applications. In the context of digital wallet, valuable insights can be gleaned from analyzing user reviews on digital distribution service platforms like the App Store, Google Play, and Microsoft Store. Generally, sentiments are categorized as positive, neutral, or negative.

Moreover, enhancing services significantly depends on effectively extracting aspects from textual customer reviews. Businesses seek to understand these aspects—especially those associated with positive feedback—to maintain and improve services. Similarly, they aim to identify aspects mentioned in negative feedback to refine and enhance the application. By actively listening to customer insights, businesses strive to develop mobile wallet apps offering superior user experiences, fostering customer satisfaction and loyalty.

Motivated by the rapid growth of digital wallets in Vietnam, the author aims to perform sentiment analysis and aspect extraction from Vietnamese feedback shared by users through their experiences with these platforms.

1.2. Problem Statement

In the midst of digital transformation, cash transactions are becoming increasingly rare, particularly among Generation Z. The convenience offered by digital wallets has led to a surge in their usage—a trend substantiated by significant growth statistics in this sector [2]. According to Vietnamese Government News, the central bank reported a 68.54 percent increase in mobile banking transactions, with QR code payments experiencing an astounding 1,062.01 percent rise in value in 2023 [3]. Notably, nearly 80% of Gen Z consumers utilize digital wallets, marking a significant shift in payment habits for the new generation. Consequently, many businesses are investing heavily in analyzing customer reviews of E-Wallet applications to gain deeper insights.

The insights derived from such analyses serve various purposes: they can identify technical issues requiring attention, offer comparative analyses against competitors, and aid in personalizing user experiences by customizing the app's features and interface to align with user preferences. Additionally, marketing campaigns can harness positive reviews to attract new users. Reviews also provide critical feedback on the app's features, indicating which are favored and should be retained or enhanced, and which are less popular and may need reconsideration.

However, understanding the sentiments and specific aspects mentioned in user reviews of E-Wallet applications remains challenging. The challenges include dealing with the complexities of Vietnamese grammar, tone nuances, and homophones [4]. The occurrence of sarcasm and irony in Vietnamese discourse often leads to interpretative challenges for sentiment analysis tools, resulting in potential inaccuracies. Moreover, the diversity of idiomatic expressions and regional dialects adds complexity to accurately extracting sentiment from textual data.

Without systematic methods to analyze and extract insights from these reviews, this wealth of information remains largely untapped. This paper addresses this gap by developing methodologies for sentiment analysis and aspect extraction from user reviews of E-Wallet apps. In doing so, it aims to improve the functionality, competitiveness, and user satisfaction of E-Wallet apps, thereby contributing to the advancement of digital wallet technology.

1.3. Scope and Objectives

The research scope is mainly focused on the top digital wallet app reviews in Vietnam and the ratings of customers after using services. The selected applications to be explored are Momo, ShoppePay, ZaloPay, VNPay, and ViettelPay. The main objectives are to gain insights from user reviews to make decisions on various applications like app improvements, marketing, user churn detection, etc. Here is the list of purposes that the author wants to achieve:

- Performing comprehensive analysis of review data with the comparison in trends, distribution, and popularity of the app
- Designing and implementing feature extraction techniques to capture relevant information from text patterns.
- Investigating the effectiveness of machine learning and deep learning approaches for sentiment analysis and aspect extraction

- Developing a robust classification model for sentiment analysis to determine whether the feedback is Positive, Neutral, or Negative (Multi-class Classification problem)
- Extracting aspects from reviews with 10 different aspects generated by Topic Modeling (Multi-label Classification problem)
- Evaluating the performance of the proposed system using benchmark datasets.
- Comparing performances with other existing papers.

The author aims to perform the text classification and aspect extraction for the Vietnamese e-wallet app reviews, applying both machine learning and deep learning techniques. This thesis addresses the problem by exploring the application of Bidirectional Long-Short Term Memory (Bi-LSTM) combined with Bidirectional Gated Recurrent Units (Bi-GRU) and PhoBERT (the state-of-the-art pre-trained model for Vietnamese text based on RoBERTa) for sentiment analysis and aspect extraction from customers' reviews.

Furthermore, this thesis also covers experiments with other supervised learning techniques to implement the multi-label classification for aspect extraction, such as Binary Relevance, Classifier Chain, and Label Powerset, to compare performance against the deep learning approach.

1.4. Tools and Resources

This paper employs a comprehensive suite of tools and resources to analyze user reviews of digital wallet applications. The data collection is facilitated through the use of the google-play-scraper, which enables efficient extraction of user reviews from the Google Play Store. This tool ensures a rich dataset for subsequent analysis.

For data preprocessing, a combination of Pandas and Numpy is utilized to handle and manipulate the data effectively. vncorenlp is used for natural language processing tasks, particularly for Vietnamese text, ensuring accurate tokenization and sentiment analysis.

Additionally, regex is employed for pattern matching and text cleaning, which is essential for preparing the data for analysis.

Data visualization is a critical component of this thesis, aiding in the interpretation and presentation of findings. Tools such as matplotlib and seaborn are used to create detailed and informative visualizations, providing clear insights into the data trends and patterns.

The classification models are developed using powerful machine learning frameworks. sklearn is used for traditional machine learning techniques, while Keras and PyTorch are employed for building and training deep learning models. These tools collectively enable a robust analysis, offering a comprehensive understanding of user sentiments and preferences regarding digital wallet applications.

1.5. Assumption and Solution

The demand for digital payments is escalating in today's digital landscape, fueled by technological advancements, widespread smartphone usage, and a growing preference for cashless transactions. This thesis is contextualized within a dynamic digital economy where consumers, especially those from Generation Z, are progressively embracing digital wallets as their preferred payment method. This trend is observable in both developed and emerging markets, with urban regions exhibiting higher adoption rates due to enhanced internet connectivity and access to contemporary financial services.

The fundamental premise of this thesis is that user reviews of digital wallet app, sourced from platforms such as the Google Play Store, offer invaluable insights into user satisfaction, encountered challenges, and future expectations. These reviews are indicative of authentic usage and experiences, thus serving as a fertile dataset for discerning consumer behavior and inclinations. The assumption here is that the collected data encapsulates a wide spectrum of user perspectives, encompassing varied opinions and experiences across multiple demographics.

Additionally, it is presumed that the analysis of this data can uncover actionable insights that can drive technical improvements, enhance competitive strategies, personalize user experiences, and inform marketing and promotional efforts. The underlying assumption is that companies are willing to invest in analyzing user feedback to remain competitive and meet their users' evolving demands.

To meet the increasing demand for digital payment solutions and augment the user experience of digital wallet applications, this thesis advocates an approach predicated on systematically examining user reviews. This approach employs advanced data processing techniques alongside machine learning and deep learning methodologies to extract recommendations for technical refinements—such as improving app functionality, introducing novel features, or rectifying glitches. It also facilitates competitive analysis to gauge performance against other digital wallet applications, pinpointing opportunities for differentiation.

1.6. Structure of thesis

This Thesis is comprised of 5 chapters: Introduction, Literature Review, Methodology, Implementation & Results, and Conclusion & Future work.

Chapter 1 - Introduction

This chapter provides an overview of the research topic, presents the problem statement, defines the scope and objectives of the thesis, outlines the assumptions, and describes the overall structure of the thesis.

Chapter 2 - Literature Review

This chapter reviews the existing literature on the growth of digital wallets, and text classification, especially the techniques used to perform sentiment analysis and aspect extraction using supervised learning approaches. It also explores relevant studies, and techniques employed in the field, highlighting their strengths and limitations.

Chapter 3 - Methodology

This chapter describes the methodology used in this research, including data collection, data preprocessing, feature extraction techniques, modeling, how to prepare standardized data for model training with Machine Learning and Deep Learning Models and evaluate procedures, and performance metrics.

Chapter 4 - Implementation and Results

This chapter presents the experimental results obtained from the proposed system. It includes the performance evaluation of the combined Bi-LSTM and Bi-Gru model and pre-trained model PhoBERT for Vietnamese text using benchmark datasets, along with a comparative analysis of their performance with previous papers. This chapter also discusses the findings of the research and provides insights into the strengths and limitations of the proposed approach.

Chapter 5 - Conclusion and Future Work

The final chapter concludes the thesis, summarizes the key findings, and discusses the contributions and significance of the research. It also outlines potential directions for future work in the field of sentiment analysis and aspect extraction using advanced algorithms.

CHAPTER 2: LITERATURE REVIEW

2.1 Important Factors Affecting Digital Wallet User Experience

The adoption of digital wallets in Vietnam has seen significant growth, influenced by various socio-economic and technological factors. This literature review explores the key factors affecting the user experience of digital wallets in Vietnam, drawing on existing research and empirical studies.

2.1.1 Convenience and Speed of Transactions

The primary appeal of digital wallets lies in their convenience and speed. As reported in performance expectancy [5], it highlights the convenience and speed benefits of digital payment systems, particularly digital wallets, in various contexts such as online utility bill payments, money transfers, online shopping, and ticket booking. Digital wallets enable users to perform these tasks swiftly and efficiently, consolidating multiple financial accounts and payment methods into one platform. This integration saves users considerable time, allowing for quick transactions with just a single click. Moreover, compared to traditional methods, digital wallets facilitate faster money transfers, contributing to their appeal and adoption among users seeking convenience and efficiency in financial transactions.

2.1.2 User Interface and Design

The design and user interface (UI) of digital wallet applications are crucial for enhancing user satisfaction. A user-friendly interface can improve the user experience by streamlining navigation and transaction processes [6]. The study also emphasizes the continuous efforts of digital wallet companies to improve user experience to transform business practices in Vietnam through the broad adoption of cashless payment systems. Applications featuring simple, attractive, and easily navigable designs are more likely to attract new users as they meet the preference for simplicity and visual appeal.

2.1.3 Security and Privacy Concerns

Security remains a paramount concern for digital wallet users. Recent research highlights the substantial worries people have about making financial transactions online. The threat of data leaks and scams is real and can deter people from fully embracing digital wallets [7]. These fears are why both businesses and consumers approach e-commerce with caution. Moreover, safeguarding personal privacy during digital transactions is another critical concern [8]. The potential for personal data to be misused or accessed without permission can erode confidence in digital wallet platforms. So, the need to analyze review data is very crucial to address these security and privacy challenges effectively.

2.1.4 Promotion and Incentives

Promotional activities and incentives are critical in attracting and retaining users. Hoang and Le's research model in 2020 demonstrated that Promotional Benefits exerted the most impact on users' intention to use e-wallets within the Vietnamese context [9]**.** Cashback offers, discounts, and loyalty rewards are effective in encouraging users to adopt and continue using digital wallets. These incentives not only attract new users but also increase the frequency of transactions among existing users, as they feel they are getting added value from the service.

2.1.5 Customer Support and Service Quality

Quality customer support is essential to ensuring users have a good experience. The level of customer service significantly impacts how people perceive their interactions with e-wallet applications. Research by Chang, Lan, and Zhu in 2017 highlights that customer service quality greatly influences the decision to keep using mobile payment platforms [10]. As Chu, Lee, and Chao noted in 2012, in the context of e-banking, strong customer support boosts both satisfaction and trust, promoting user loyalty [11]. The study emphasizes that dependable and attentive customer service is vital for a favorable user experience and for motivating ongoing engagement with e-wallet apps.

2.2 Challenges And Opportunities to Enhance Mobile Wallet Services

2.2.1 Challenges: Security and Fraud Prevention

Security and fraud prevention are critical challenges in adopting and enhancing mobile wallet services. Users are concerned about their financial information's security **[12].** Data breaches, hacking, and identity theft are prominent threats that can deter users from adopting mobile wallets. Advanced security measures such as encryption, tokenization, and biometric authentication are essential to address these concerns. These technologies can protect sensitive information and prevent unauthorized access, thereby enhancing user trust and confidence in mobile wallet services.

Another aspect of security involves fraud detection and prevention. Real-time monitoring systems powered by artificial intelligence (AI) and machine learning can identify suspicious activities and prevent fraudulent transactions. The implementation of these technologies can significantly reduce the risk of fraud and enhance the overall security of mobile wallet services.

2.2.2 Challenges: User Adoption and Trust

Building trust and encouraging user adoption are persistent challenges in the mobile wallet industry. Trust is a critical factor influencing the adoption of mobile payments since users need to feel confident in the reliability and security of mobile wallet services to transition from traditional payment methods [13].

Resistance to change is particularly evident among older demographics, who may be more comfortable with cash or card payments. Overcoming this reluctance requires targeted education and awareness campaigns to inform users about the benefits and security of mobile wallets. Demonstrating the convenience and safety of digital payments can help build trust and encourage broader adoption.

2.2.3 Opportunities: Advancements in Technology

The advancement in technology is a significant opportunity to improve digital wallet services since the information collected can assist in gaining a better insight into what clients are saying. They are further explained as follows: sentiment analysis is the task of interpreting the tone of the expressed emotions, and aspect extraction is the task of finding out components or features mentioned in customer feedback. These technologies can help employees of a company monitor specifics regarding clients' opinions and perceptions so that these can be addressed and the satisfaction of the clients improved. Thus, having the opportunity to extract useful bits from all the reviews, the providers can improve their services, focusing on the aspects that improve or worsen the dynamics of the sphere pursued.

Furthermore, the extension's enriched sentiment analysis enables targeted interactions and offers based on user sentiment, as well as feature customization. Incorporating the identification of negative sentiments in advance contributes to enhancing consumer loyalty and ensuring customer satisfaction, which consequently reduces customer churn age.

2.2.4 Opportunities: Financial Inclusion

Digital wallets present a significant opportunity for promoting financial inclusion, particularly in underserved and unbanked populations. These technologies can bridge the gap for individuals who lack access to traditional banking services, offering a secure and convenient means of participating in the financial ecosystem. By leveraging mobile technology, digital wallets can reach remote and rural areas where brick-and-mortar banking infrastructure is limited or non-existent.

Moreover, digital wallets often come with lower fees and fewer barriers to entry compared to traditional banks, making financial services more accessible to a broader audience. This inclusivity can empower individuals by providing them with tools for savings, payments, and transfers, which are essential for personal and economic development. The ability to conduct financial transactions digitally also reduces the risks associated with carrying cash and enhances security for users.

2.3 Sentiment Analysis Approaches on Vietnamese Text

Sentiment analysis is one of the developing fields in natural language processing (NLP) with huge applications in real-life such as customer feedback analysis, public opinion and social media comments. The approaches of sentiment analysis can be divided into three main categories: lexicon-based, machine-learning, and hybrid approach [14]. In recent years, many researchers have focused on the adaption of sentiment analysis techniques to apply it for Vietnamese text. Here is the evolution of the SA techniques until the research finding:

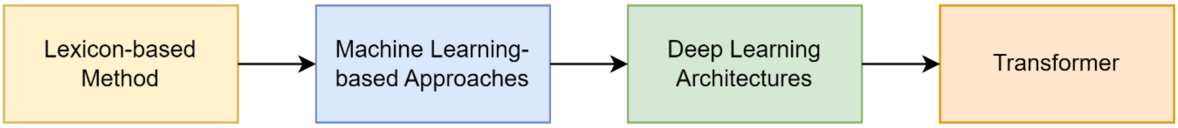

Figure 2.1: The Evolution of Sentiment Analysis on Text Classification

2.3.1 Lexicon-based Method

The computational study of opinions and emotions in text has undergone significant evolution. Initially, lexicon-based approaches dominated the field. These methods relied on precompiled dictionaries of words, each assigned a sentiment value (positive, negative, or neutral) [15]. The process involved cleaning and tokenizing text data, then using these lexicons to score the sentiment based on the presence and frequency of sentiment-laden words. Although straightforward and interpretable, lexicon-based approaches often struggled with context sensitivity and the nuances of language, limiting their effectiveness.

2.3.2 Machine Learning-based Approaches

To address the drawback of context sensitivity, machine learning techniques had been introduced, bringing a new level of sophistication to sentiment analysis. These methods involve collecting labeled datasets where text is tagged with sentiment labels, preprocessing the text data to remove noise, and transforming it into numerical features through techniques such as

Bag of Words (BoW) or Term Frequency-Inverse Document Frequency (TF-IDF). Support Vector Machines (SVM), Naive Bayes, and other classifiers are then trained on these features to learn patterns associated with different sentiments. Machine learning models significantly improved accuracy and adaptability compared to lexicon-based methods when it works on handling large and diverse datasets.

In the context of Twitter dataset, machine learning has been proven as the successful approach for reaching good performance of sentiment analysis and outperforming the lexicon-based method [16].

2.3.3 Deep Learning Architectures

The exploration of deep learning techniques built a great milestone in sentiment analysis application. Recurrent Neural Networks (RNNs), particularly those with Long Short-Term Memory (LSTM) units, became popular for their ability to capture sequential dependencies and contextual information in text data [17]. The process involves similar initial steps of data collection and preprocessing, followed by the use of word embeddings like Word2Vec or GloVe to represent text in dense, continuous vector spaces. LSTM networks are then trained to understand long-term dependencies, making them adept at handling the complexities of human language. Consequently, this architecture has proven to be more effective than the ML-based model since the ability to analyze sentiment reaches higher accuracy and better generalization of complex contexts.

2.3.4 Transformer

The most recent advanced approach in SA is the use of transformer models, which have revolutionized natural language processing (NLP). Transformers, such as BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer), utilize self-attention mechanisms to process text [18]. This new advanced mechanism, allows the model to capture context from both directions, providing a deep understanding of language contexts. The process involves fine-tuning pre-trained transformer models on sentiment-labeled data, which greatly reduces the need for large and labeled datasets. Transformers excel at understanding context, managing long-range dependencies, and generalizing across different domains, setting a new benchmark for accuracy and effectiveness in sentiment analysis.

2.4 Aspect Extraction Approaches on Vietnamese Text

Aspect extraction in natural language processing (NLP) involves the identification of specific aspects or features within text that users are expressing opinions or sentiments about. The use of the aspect extraction model is particularly necessary for the analysis of customer reviews, where understanding key aspects of a product or service that is being criticized within the review can provide valuable insights for businesses to improve their services.

2.4.1 Rule-based Approaches

One of the foundational methods for aspect extraction in Vietnamese text is rule-based approaches. These techniques rely on predefined linguistic rules and patterns to identify aspects. For example, identifying nouns or noun phrases that are frequently associated with opinion words (such as adjectives or verbs indicating sentiment) can reveal aspects that users are discussing. Rule-based systems are intuitive and transparent, making them easy to interpret and modify based on specific domain requirements. However, these approaches can be labor-intensive as they require manual creation and refinement of rules, which may not capture all nuances of language use in diverse contexts [19].

2.4.2 Supervised Machine Learning-based

Supervised machine learning approaches have also been employed for aspect extraction in Vietnamese text, leveraging annotated datasets where text segments (such as sentences or phrases) are labeled with aspect categories. Techniques like Support Vector Machines (SVM), Naive Bayes, and neural networks are utilized with features including word embeddings, syntactic features, or specific linguistic patterns relevant to Vietnamese. These methods can achieve high accuracy when trained on large along with high-quality labeled datasets. However, their performance heavily relies on the availability of such datasets, which may be limited for Vietnamese compared to more widely studied languages like English.

2.4.3 Deep Learning Architectures

The recent ontological models have radically transformed the natural language processing approaches. Deep learning utilizes layers of computational nodes that include neurons in replicating the human brain, hence allowing the algorithm to distinguish the small differences between languages. This helps in correctly identifying aspects from Vietnamese text and analyzing customers' opinions and sentiments towards the company or product. Such architectures are specifically useful in managing the tone and the context effects of Vietnamese, which are inherent in the language and hence make such architectures invaluable in analyzing user generated content [20].

The architecture of this method is the same since it also belongs to the natural language processing field. The most applied methods are Recurrent Neural Networks (RNNs) which have many advantages of understanding the context and structure of text data. That is why they are very popular and applicable to many aspect extraction approaches.

To sum up, aspect extraction in Vietnamese text has seen advancements across many methods, including rule-based, supervised, unsupervised, and deep learning approaches. Each approach has its strengths and challenges, influenced by the availability of labeled data, linguistic characteristics of Vietnamese, and computational resources. So, in future research on aspect extraction, new solutions will focus on adapting and optimizing existing strong models in deep learning and machine learning to further improve he application of E-Wallet app reviews, especially working on the Vietnamese language context.

CHAPTER 3: METHODOLOGY

In this part, the author presents a detailed step-by-step explanation of the methodologies to implement the classification goals. The procedure started with the data collection, encompassing data processing, feature extraction, and classification approaches to perform the tasks of sentiment analysis and aspect extraction on Vietnamese customer reviews for E-Wallet applications. The research workflow is visually represented in Figure 3.1, providing a comprehensive overview of the sequential steps taken to conduct the study. The graphical design represents the roadmap to guide the reader through the classification procedure, and outline the initial setup to the final outcomes.

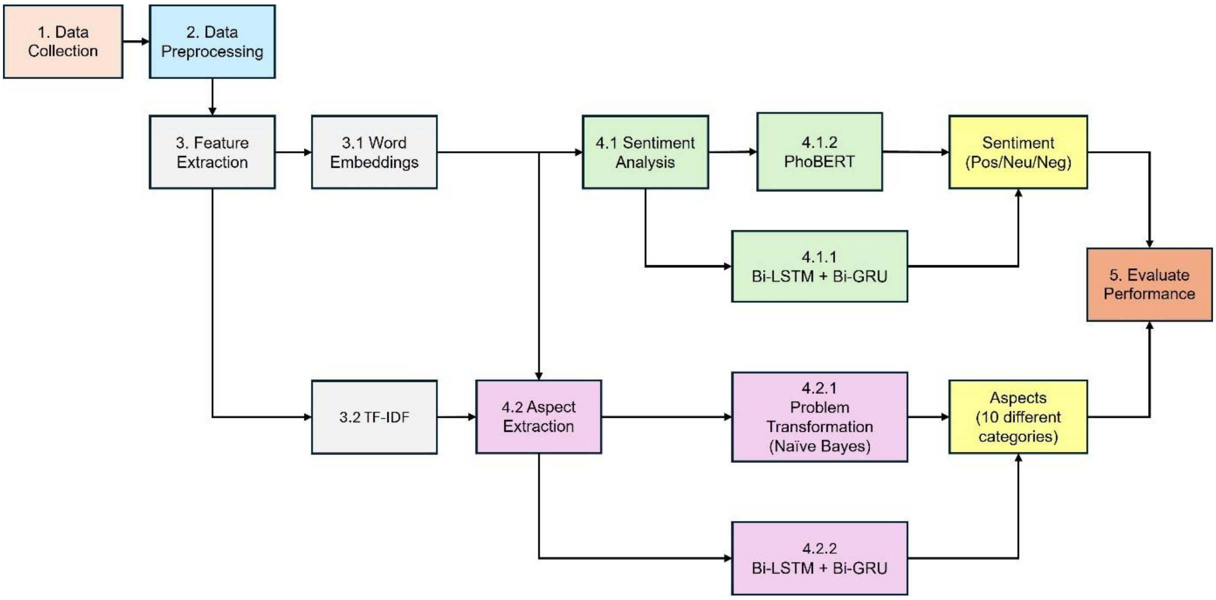

Figure 3.1: Project Workflow

3.1 Data Collection

With the aim to analyze data from the top digital wallets in Vietnam and given the absence of a standard dataset in this field, the author decided to scrape data from the Google Play Store - a digital distribution service developed by Google, which hosts a vast array of user reviews for various applications. To collect the necessary data, the author utilized the Google-play-scraper API in Python. This API was modified to retrieve customer reviews that met specified requirements in terms of quantity and quality, effectively breaking the API's limitations on the number of samples. The digital wallet applications selected for this study were Momo, ShopeePay, ZaloPay, VNPay, and ViettelPay.

The raw dataset includes over 50,000 samples with 12 different attributes related to the review metadata. Particular emphasis is placed on the Vietnamese review texts and associated product ratings.

3.2 Data Preprocessing

3.2.1 Data Cleaning

In the preprocessing phase, several essential techniques are employed to refine the textual data, ensuring its suitability for supervised learning model application. Here is the full process of data processing to achieve the cleaned data:

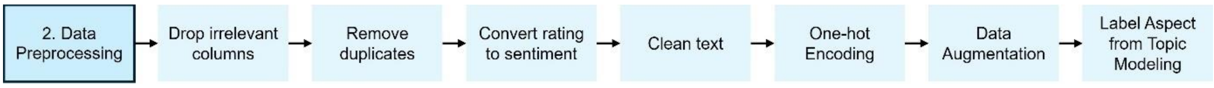

Figure 3.2: Data Processing Procedure

All irrelevant columns are omitted before the removal of duplicated data points. Next, to categorize sentiments, customer ratings are systematically distributed across three distinct categories: 'Positive,' 'Neutral,' and 'Negative'. Reviews with ratings exceeding 4 are categorized as 'Positive,' while those with a rating of 3 are designated as 'Neutral'. The remaining reviews are labeled as 'Negative' sentiment. After that, the sequences of text cleaning methods are involved before transforming the target label with a one-hot encoding approach.

However, in real-world applications, many Vietnamese reviews are written in non-accent format, making it difficult for the classification model to perform as accurately as possible. So, to address this problem, the author applies data augmentation in the customer reviews by removing accents and extending it to the original dataset. The augmented data are checked for duplicates in the reviews subset to maintain the quality of the data. Finally, the label aspects are figured out using a topic modeling method called LDA (Latent Dirichlet

Allocation). The author used the generated words from the Topic Modeling algorithm to define categories of the E-Wallet dataset. In order to label these aspects for the original dataset, the author defined a function to perform the keyword matching from customer reviews. Even though there is another approach using pre-trained model PhoBert to label aspects, it seems to be time-consuming and has incorrect outcomes. Here is the full process of data processing that the author mentioned above:

Text Cleaning Process

Within the cleaning text phase, several techniques are applied to ensure the trustworthiness of the review text. Initially, the list of emojis, punctuation, and digits are removed from the text to reduce noise since these stuff do not contribute any meaning to the classification task. After that, all reviews are transformed to lowercase to reduce the size of the final vocabulary before applying the word segmentation method. This task was achieved by using the VNCoreNLP toolkit, which is a fast and accurate Natural Language Processing annotation pipeline for Vietnamese. Next, the reviews after using word segmentation are joined into string before performing tokenization into single token with the use of built-in split function in Python. However, customer reviews at this stage were still not standardized since it may contain many sensitive words and teen code that needed to convert into the right format. So, the author decided to define the normalize acronyms function to handle this situation. This function iterates over each review text to find the sensitive words. If the review contain the bad words from the defined list, it will be removed or converted to the correct format of the teencode words. At the end, the final review texts become semantic coherence, reducing vocabulary sparsity and fostering robust, reliable, and actionable to perform modeling in sentiment analysis and aspect extraction within the dynamic, complex, and evolving landscape of textual data analytics.

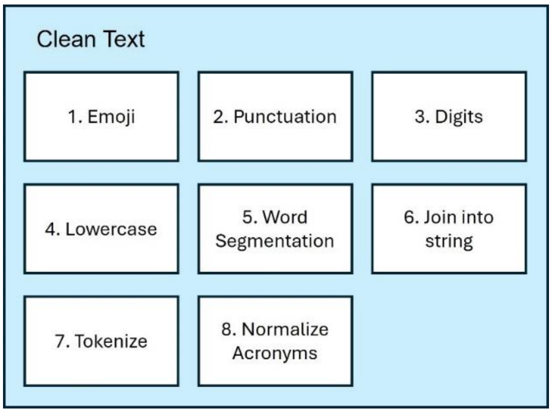

Figure 3.3: Cleaning Text Process

3.2.2 Exploratory Data Analysis

In the exploratory data analysis of customer reviews, a striking observation emerges concerning the distribution of overall ratings provided by consumers. Specifically, a pie chart in Figure 3.4 shows the predominance of 1-star ratings, constituting a substantial majority within the dataset with accounting for 61.9%. This graphical depiction highlights the overwhelmingly negative sentiment conveyed by customers, indicating low satisfaction levels, product quality, and favorable experiences associated with the products or services under review. As for the proportion of the 5-star ratings, nearly 26% of the reviews belong to this group without any duplications in reviews. In contrast to the previous distribution, the remaining rating group only appropriate small percentage of the distribution with below 5% in average.

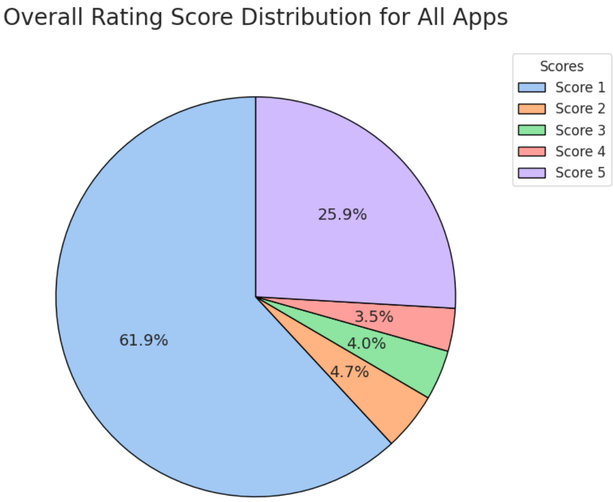

Figure 3.4: Distribution of Customers' Reviews over Ratings

Conversely, lower rating categories exhibit comparatively lesser frequencies, underscoring the prevalent positive sentiment and exemplary performance within the analyzed dataset. The distribution of sentiment classes underscores the significance of positive customer feedback, favorable brand perceptions, which can empower stakeholders with actionable insights, strategic foresight, and data-driven decision-making capabilities across diverse applications, industries, and research domains within the dynamic and complex ecosystem of customer reviews and sentiment analytics.

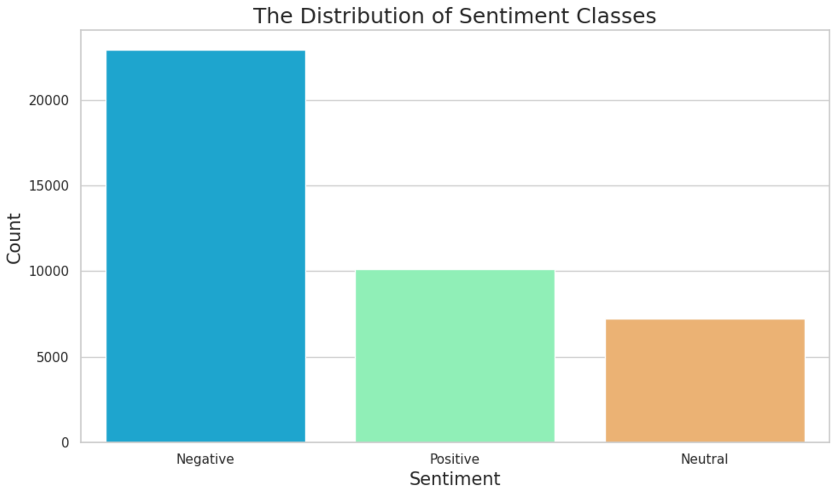

Figure 3.5: The distribution of Sentiment Classes

The pie chart in Figure 3.5 offers a compelling visual representation of the sentiment distribution within the customer reviews with distinct categories based on sentiment orientations: Positive, Negative, and Neutral. Notably, the chart vividly illustrates a dominant presence of positive sentiments, encompassing a staggering 87.9% of the total sentiment distribution. This dominant value is the proof of the prevalent satisfaction, favorable experiences, and positive perceptions conveyed by customers within the e-wallet reviews dataset. Conversely, the chart portrays a relatively diminutive representation of negative sentiments, accounting for a mere 4.55%, and neutral sentiments, constituting 7.53%, within the dataset.

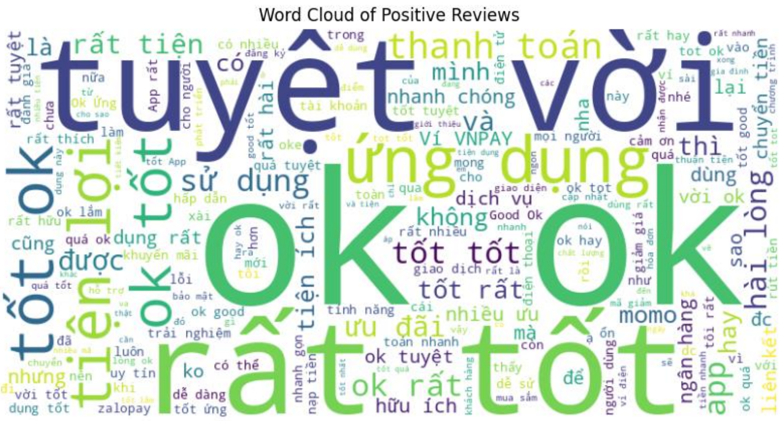

Figure 3.6: Distribution of positive words in Word Cloud

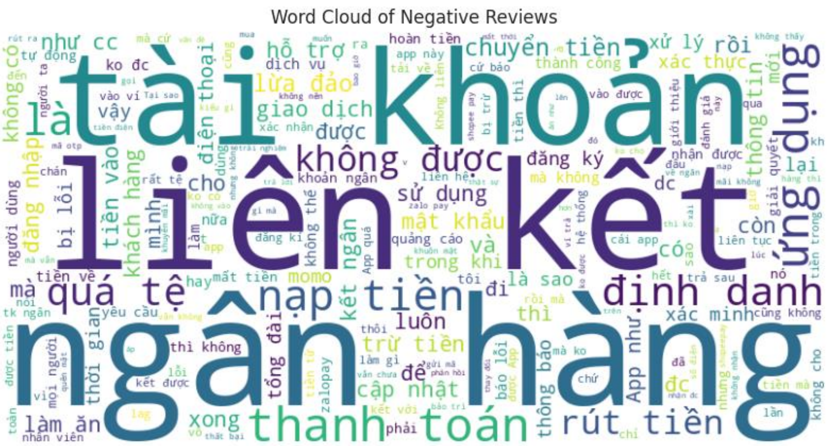

Figure 3.7: Distribution of negative words in Word Cloud

The above World Clouds provide a comprehensive visualization of the positive and negative words extracted from the reviews within the e-wallet field. Notably, terms related to transactions occurred frequently, indicated the most feature that user are interested in when using the e-wallet apps. Some of the highlighted terms in negative reviews are "tài khoản", “liên kết", and "ngân hàng". For the distribution of positive reviews, the most frequent word are quite general such as "ok, "rất tốt", and "tuyệt vời", which are not contributed much to the evaluation of aspect extraction task.

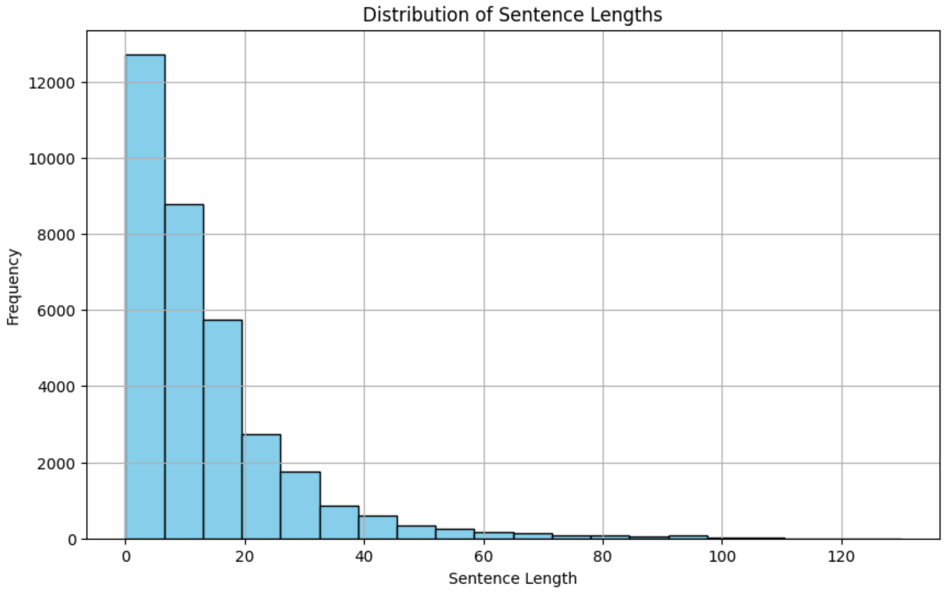

Figure 3.8: Histogram of Sentence Lengths

In order to make a decision on the max length of the author's model implementation, the histogram of sentence lengths is plotted to achieve this need. As you can see, the sentence lengths are distributed around 10-20 words, and the maximum value of sentence length is 130 words in the working dataset.

3.3 Feature Extraction

In the methodology for feature extraction, the conversion of human language into a readable and understandable format for computers is applied using various existing natural language processing techniques. The initial step involves transforming textual data into a numerical format conducive for integration into machine learning and deep learning models. These techniques include Word Embeddings and Term Frequency-Inverse Document Frequency (TF-IDF). The subsequent sub-sections comprehensively elaborate on the application and implementation of these techniques in the context of the project.

3.3.1 Word Embeddings

Word embeddings are dense vector representations of words that capture their meanings, syntactic properties, and semantic relationships by placing similar words closer together in the vector space. These embeddings are foundational in natural language processing (NLP) tasks, enabling models to understand and process text efficiently.

Figure 3.9: Word Embeddings Process

In deep learning models, such as those built with the Keras library, word embeddings can be generated using the Keras Tokenizer, which converts text into sequences of integers, where each integer corresponds to a word in the vocabulary. Once the vocabulary is formed, we can create word embeddings using Vietnamese pre-trained Word2Vec, which involves training a neural network to learn each word's context based on its surrounding words. We can construct an embedding matrix by utilizing the tokenizer's vocabulary. Each row in this matrix corresponds to a word in the tokenizer's vocabulary and contains the Word2Vec vector for that word. This embedding matrix can be used in various natural language processing tasks, enabling the model to capture semantic meanings and relationships between words effectively.

Moreover, another approach leverages PhoBERT, a pre-trained transformer model specifically designed for the Vietnamese language, to perform text classification by utilizing contextualized word embeddings. Initially, the input sentences are tokenized into subwords using PhoBERT's tokenizer, converting these subwords into their respective IDs, thus transforming the raw text into a format compatible with PhoBERT. Input tensors are then prepared, and attention masks are created, which assist the model in distinguishing between actual tokens and padding tokens. By loading the pre-trained PhoBERT model configured for sequence classification, this method harnesses its capability to generate rich, contextualized embeddings for each token in the input sequences.

During the training loop, token IDs and attention masks are passed through PhoBERT. The model processes these inputs to generate contextual embeddings, subsequently used by the classification head to predict labels. This approach applies advanced word embeddings dynamically, capturing the nuanced meanings of words based on their context within sentences. The pre-trained nature of PhoBERT on extensive Vietnamese text ensures a robust understanding of the language's structure and semantics, enhancing its performance on classification tasks.

Compared with the word embeddings using Keras, PhoBERT approach is different in the way of performing embeddings since it applies the powerful of the pre-trained model specified for Vietnamese text. Instead of using static word embeddings like Word2Vec or GloVe, PhoBERT generates embeddings dynamically based on the context of each word in a sentence. This is a more advanced and powerful approach, as it allows the model to capture the nuanced meanings of words depending on their usage.

3.3.2 TF-IDF (Term Frequency – Inverse Document Frequency)

TF-IDF (Term Frequency-Inverse Document Frequency) is a numerical statistic used to evaluate the importance of a word in a document relative to a collection of documents (corpus). It combines two metrics: term frequency (TF), which measures how frequently a term appears in a document, and inverse document frequency (IDF), which assesses how important a term is across the entire corpus.

Several approaches can be employed to calculate TF-IDF, including using libraries like Scikit-learn in Python, which provides straightforward functions to transform a text corpus into TF-IDF features. For the aspect extraction task in a machine learning context, TF-IDF is a popular feature selection method due to its effectiveness in highlighting significant terms that can be indicative of specific aspects. TF and IDF are calculated with the following formulas:

$$TF(t,d) = \frac{number\ of\ times\ t\ appears\ in\ d}{total\ number\ of\ terms\ in\ d}$$

$$IDF(t) = log\frac{N}{1+df}$$

$$TF - IDF(t,d) = TF(t,d) * IDF(t)$$

Equation 1: TF-IDF Formula

The TF-IDF selection method is crucial because it helps identify and weight terms that are most relevant to the aspects being analyzed. By focusing on terms with high TF-IDF scores, the model can better distinguish between different aspects based on their contextual importance.

In the aspect extraction task using a machine learning approach, TF-IDF serves as a robust feature extraction technique. By transforming the text data into numerical format, TF-IDF enhances the model's ability to identify and extract relevant aspects. The techique reflects the importance of terms within and across documents and weight these words more essential with higher score. This preprocessing step is critical for preparing the data for various machine learning algorithms to make sure that all informative features are used to train the model.

3.4 Sentiment Analysis

3.4.1 Combined Bi-LSTM and Bi-GRU

Long Short-Term Memory (LSTM) networks are a type of recurrent neural network (RNN) designed to address the vanishing gradient problem, which hampers traditional RNNs in learning long-range dependencies. The architecture of an LSTM consists of memory cells and three main gates: input, forget, and output gates. These gates regulate the flow of information, allowing the network to retain or forget information as needed.

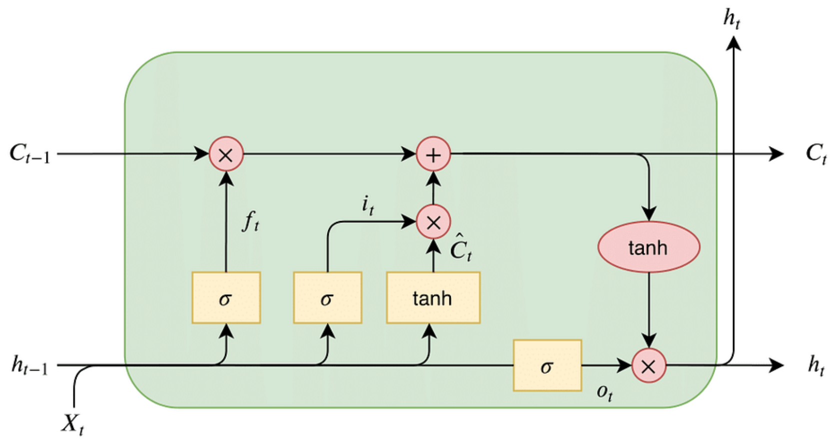

Figure 3.10: Long Short-Term Memory Architecture [21]

Bidirectional LSTM (Bi-LSTM) enhances this architecture by incorporating two LSTMs: one processes the sequence forward, and the other processes it backward. This dual approach captures contextual information from both past and future states, providing a more comprehensive understanding of the sequence.

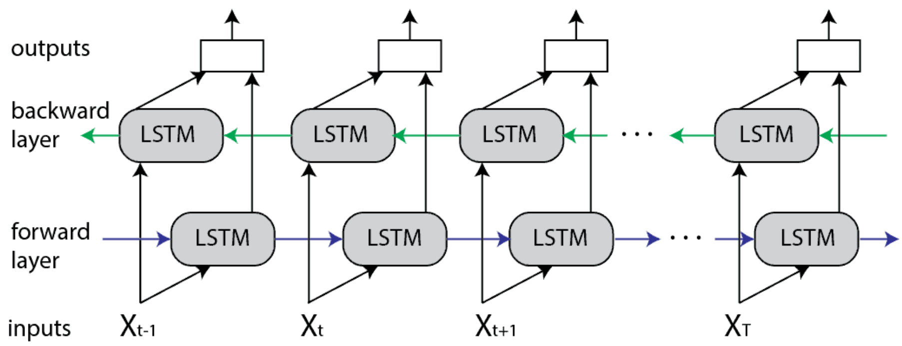

Figure 3.11: Bidirectional LSTM Architecture [22]

Gated Recurrent Unit (GRU) networks, another form of RNNs, simplify the LSTM's architecture by combining the cell and hidden states and using only two gates: the reset and update gates. This makes GRUs more efficient while maintaining performance comparable to LSTMs. The architecture of a GRU allows it to effectively manage the flow of information through sequences. A Bidirectional GRU (Bi-GRU) extends this by employing two GRUs, one processing the sequence in the forward direction and the other in the backward direction. This bidirectional processing ensures that both past and future context are taken into account, enhancing the model’s ability to capture the nuances in the data, whereas a regular GRU only has access to information from the past.

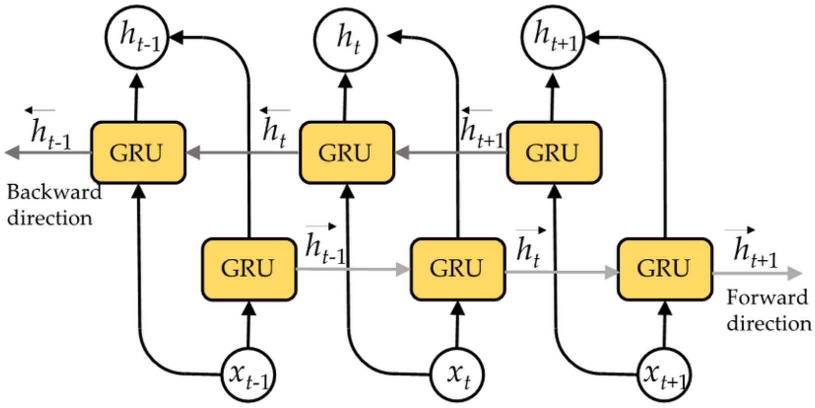

Figure 3.12: Bi-GRU Working Architecture [23]

The author decided to combine Bi-LSTM and Bi-GRU models to leverage the strengths of both architectures for sentiment analysis of Vietnamese customer reviews. The Bi-LSTM’s robust mechanism for handling long-term dependencies and comprehensive bidirectional context complements the Bi-GRU’s efficiency and effective sequential data management. This hybrid architecture benefits from Bi-LSTM's thorough contextual understanding and Bi-GRU's streamlined processing capabilities, making it particularly suited for capturing the complex and nuanced sentiment expressed in Vietnamese language reviews. By integrating these models, the combined network can achieve more accurate and nuanced sentiment analysis, improving the ability to understand and categorize customer sentiments.

The diagram in figure 3.13 illustrates the structure of a designed model for tasks such as sentiment analysis and aspect extraction. The model begins with an input sentence, which is processed through an embedding layer to convert words into dense vector representations. These embeddings are then fed into two parallel branches: one using a Bidirectional Long Short-Term Memory (Bi-LSTM) network and the other using a Bidirectional Gated Recurrent Unit (Bi-GRU) network. These layers are designed to capture the sequential dependencies in the text data, processing the input sentence in both forward and backward directions to understand the context from both ends.

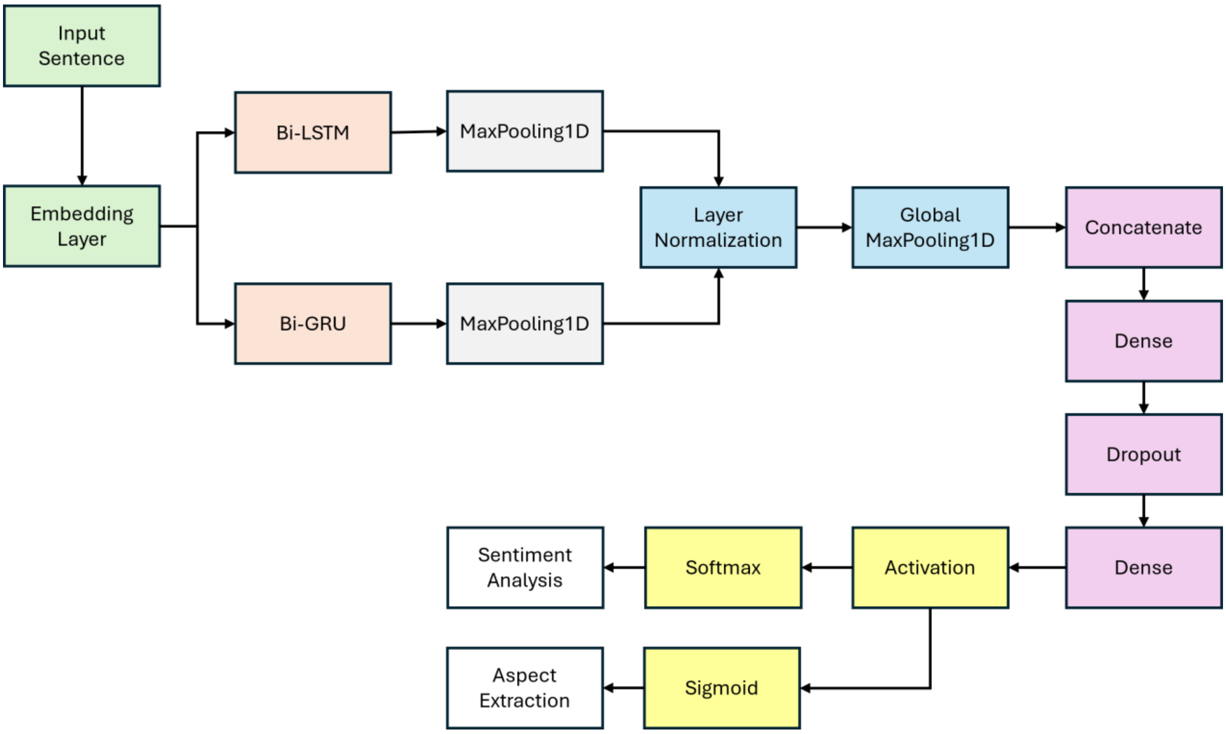

Figure 3.13: Bi-LSTM + Bi-GRU Proposed Architecture

Following each recurrent layer, MaxPooling1D is applied to reduce the dimensionality of the input, retaining only the most significant features by selecting the maximum value over a pooling window. This reduction helps in minimizing overfitting and computational complexity while preserving important features. After pooling, layer normalization is applied to stabilize and accelerate the training process by ensuring that the inputs to each layer have a stable distribution. Subsequently, Global MaxPooling1D further reduces the dimensionality by taking the maximum value of each feature map across the entire sequence length, resulting in a fixed-length vector and making the model invariant to the input length while focusing on the most prominent features.

The outputs from the two branches (Bi-LSTM and Bi-GRU) are then concatenated, combining the features learned by both types of recurrent layers. This fusion of features enhances the model’s ability to capture diverse patterns in the data. Following concatenation, a fully connected (dense) layer helps in learning complex representations by combining the features from the previous layers. A dropout layer is applied next as a regularization technique to prevent overfitting by randomly setting a fraction of input units to zero during training, thereby making the model more robust and generalizable.

The model has two separate output layers tailored for different tasks: for sentiment analysis, a softmax activation function is used to output probabilities for each sentiment class, while for aspect extraction, a sigmoid activation function is applied to handle multi-label classification, where each aspect can be independently identified. This structured approach of combining Bi-LSTM and Bi-GRU with pooling, normalization, and concatenation layers enables the model to effectively capture and utilize the rich contextual information from the input sentences. The use of dropout and specific activation functions further refines the model's performance, making it well-suited for complex NLP tasks such as sentiment analysis and aspect extraction.

3.4.2 Pre-trained model PhoBERT

The designed model, with its combination of Bidirectional Long Short-Term Memory (Bi-LSTM) and Bidirectional Gated Recurrent Unit (Bi-GRU) layers, inherently has a high time complexity due to the sequential nature of these layers. Each Bi-LSTM and Bi-GRU layer processes input tokens in both forward and backward directions, which increases computational demands. Additionally, the inclusion of multiple pooling layers, normalization, and concatenation steps further adds to the computational load. To mitigate time complexity, efficient implementation practices such as parallel processing, model pruning, and optimizing the batch size can be employed. However, the most significant impact can be achieved through the selection of a pre-trained model, such as PhoBERT, which is already optimized for such tasks.

The use of pre-trained models is advantageous due to their ability to leverage knowledge gained from vast amounts of data during the pre-training phase. This reduces the need for extensive computational resources and time for training from scratch. Pre-trained models have already learned useful features and representations, which can be fine-tuned for specific tasks, significantly accelerating the training process and improving performance.

PhoBERT is specifically designed for the Vietnamese language, making it particularly suitable for tasks involving Vietnamese text. It is built upon the robust architecture of RoBERTa, which has demonstrated superior performance in various natural language processing (NLP) tasks. PhoBERT's pre-training on a large Vietnamese corpus ensures that it captures the nuances and structure of the language, providing a strong foundation for downstream tasks such as sentiment analysis and aspect extraction.

PhoBERT is based on RoBERTa (A Robustly Optimized BERT Pretraining Approach), which itself is an optimized version of BERT (Bidirectional Encoder Representations from Transformers). RoBERTa improves upon BERT by training with more data, removing the next sentence prediction task, and using longer sequences and larger batch sizes. The architecture of PhoBERT, like RoBERTa, consists of multiple layers of transformers. Each transformer layer has a self-attention mechanism and feed-forward neural networks, enabling the model to capture complex patterns and dependencies in the text.

Process of performing PhoBERT:

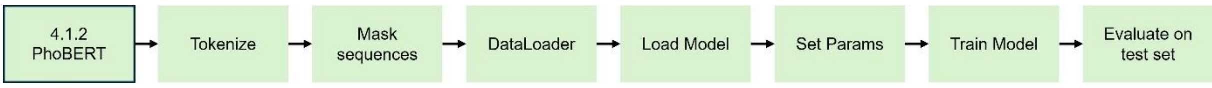

Figure 3.14: PhoBERT Implementation Flow

3.4.3 Evaluation Methods

The use of performance evaluation parameters was introduced to evaluate the results after applying each model. The assessment of classification methods relies on key metrics such as Accuracy, F-Score, Cross-entropy, Recall, and Precision. These metrics play a crucial role in evaluating the effectiveness of supervised machine learning algorithms, and they are derived from the confusion matrix or contingency table. The confusion matrix is a visual tool that helps assess algorithm performance. Terms like 'True Positive (TP),' 'False Positive (FP),' 'True Negative (TN),' and 'False Negative (FP)' are employed to compare class labels in this matrix, as illustrated in the Table below [16].

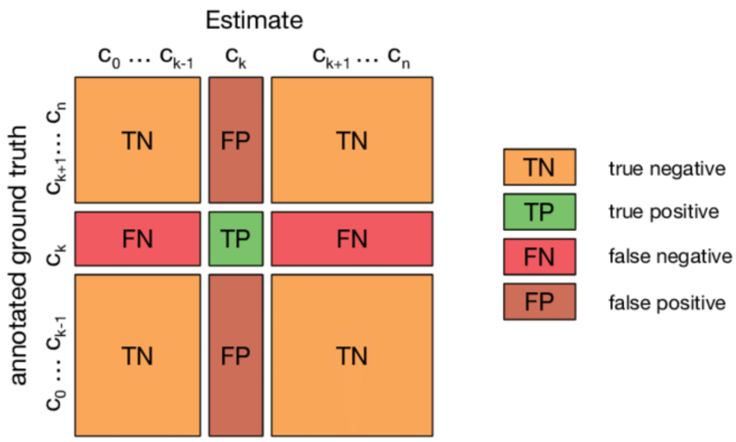

Figure 3.15: Confusion Matrix for multi-class Classification [24]

True Positive signifies positive reviews correctly identified as positive, while False Positive represents instances predicted as negative but was positive. Conversely, True Negative denotes negative reviews correctly classified as negative, while False Negative indicates instances predicted as positive but was negative. Precision, recall, f-measure, and accuracy, derived from the confusion matrix data, serve as vital indicators for evaluating classifier performance.

- **Accuracy**

The ratio of the customers' reviews that are correctly classified to the total number of reviews. While accuracy is a widely used metric, it may be misleading in imbalanced datasets, where one class dominates. It is suitable for scenarios where class distribution is relatively even.

$$Accuracy = \frac{TP + TF}{TP + FP + TN + FN} \ (3)$$

- **Cross-entropy**

Cross-Entropy Loss quantifies the difference between predicted probabilities and true class labels. It is a common loss function used during the training of classification models. The output of log loss is a probability value between 0 and 1.

- **F1-score**

The F1-score is the harmonic means of precision and recall, providing a balanced measure of a model's overall performance. It aims to find a compromise between precision and recall.

$$F1 - score = 2 * \left(\frac{Precision * Recall}{Precision + Recall}\right) \ (4)$$

3.5 Aspect Extraction

3.5.1 Problem Transformation (Naive Bayes)

In multi-label classification, methods like Binary Relevance and Classifier Chains employ Naive Bayes as a base classifier, transforming the problem into multiple binary tasks or sequential predictions. These approaches are often used to benchmark their performance against deep learning algorithms, which directly model intricate label dependencies using neural networks, potentially achieving superior accuracy albeit demanding more data and computational resources.

- **Binary Relevance**

Binary Relevance is a straightforward and commonly used technique in multi-label classification where each label is treated as a separate single-label classification problem. Each binary classifier predicts the presence or absence of its corresponding label for each instance, simplifying the problem by assuming label independence. While straightforward to implement and interpret, it may overlook label correlations and can lead to imbalanced datasets across classifiers, affecting predictive performance in scenarios where label dependencies are significant.

X	Y_1	Y_2	Y_3
$x^{(1)}$	1	0	1
$x^{(2)}$	1	1	0
$x^{(3)}$	1	1	1
$x^{(4)}$	1	1	1
$x^{(5)}$	0	0	1

https://huytranvan2010.github.io/

X	Y_1	Y_2	Y_3
$x^{(1)}$	1	0	1
$x^{(2)}$	1	1	0
$x^{(3)}$	1	1	1
$x^{(4)}$	1	1	1
$x^{(5)}$	0	0	1

Figure 3.16: Binary Relevance Method [25]

- **Classifier Chains**

For the implementation of classifier chains, the first classifier is trained just on the input data and then each next classifier is trained on the input space and all the previous classifiers in the chain. Each label in the dataset is predicted in sequence, considering the predictions of previous labels as additional features. It starts by training a base classifier on the input features and the first label. For subsequent labels, the input features are augmented with the predictions of the previously predicted labels. This chaining of classifiers continues until all labels have been predicted.

X	Y_1
$x^{(1)}$	1
$x^{(2)}$	1
$x^{(3)}$	1
$x^{(4)}$	1
$x^{(5)}$	0

X	Y_1	Y_2
$x^{(1)}$	1	0
$x^{(2)}$	1	1
$x^{(3)}$	1	1
$x^{(4)}$	1	1
$x^{(5)}$	0	0

X	Y_1	Y_2	Y_3
$x^{(1)}$	1	0	1
$x^{(2)}$	1	1	0
$x^{(3)}$	1	1	1
$x^{(4)}$	1	1	1
$x^{(5)}$	0	0	1

Figure 3.17: Classifier Chains Rule [25]

- **Label Powerset**

Label Powerset is a technique in multi-label classification where each unique combination of labels in the training set is treated as a separate class. This method works by transforming a multi-label classification problem into a multi-class classification problem where each unique combination of labels in the training data becomes a distinct class. This approach allows classifiers to predict sets of labels directly, accommodating label dependencies but potentially increasing computational complexity and requiring sufficient data to handle all possible label combinations effectively.

X	Y_1	Y_2	Y_3
$x^{(1)}$	1	0	1
$x^{(2)}$	1	1	0
$x^{(3)}$	1	1	0
$x^{(4)}$	1	1	1
$x^{(5)}$	0	0	1

X	Y_1	Y_2	Y_3
$x^{(1)}$	1	0	1
$x^{(2)}$	1	1	0
$x^{(3)}$	1	1	0
$x^{(4)}$	1	1	1
$x^{(5)}$	0	0	1

Figure 3.18: Label Powerset Approach [25]

3.5.2 Combined Bi-LSTM and Bi-GRU

In the domain of natural language processing, particularly in aspect extraction, the combined architecture of Bidirectional Long Short-Term Memory (Bi-LSTM) and Bidirectional Gated Recurrent Unit (Bi-GRU) has proven to be effective. This architecture mirrors that of sentiment analysis models, leveraging the sequential data processing capabilities of both Bi-LSTM and Bi-GRU to capture context from both directions in the text data. The choice of sigmoid as an activation function is strategic for its ability to map predictions to a probability distribution between 0 and 1, which is particularly useful when the model is designed to generate two distinct outputs. This dual-output approach allows for a more nuanced understanding of the text, distinguishing between different aspects and their associated sentiments within a single model framework.

3.5.3 Evaluation Methods

- **Macro Average**

Macro average calculates the metric independently for each class and then takes the average of those metrics. The formulas to compute these metrics are:

$$Precision_{macro} = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i + FP_i}$$

$$Recall_{macro} = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i + FN_i}$$

$$F1 - Score_{macro} = \frac{1}{N}\sum_{i=1}^{N}(\frac{2 * Precision_i * Recall_i}{Precision_i + Recall_i})$$

where:

- N is the total number of classes
- TP_i is the number of true positives for class i
- FP_i is the number of false positives for class i
- FN_i is the number of false negatives for class i

- **Hamming Score**

For multilabel classification problems, accuracy is calculated as an average across labels. Technically, this is known as a ‘Hamming score'.

$$H = \frac{1}{N}\sum_{i=1}^{N} 1 - \frac{d(Y_i, \hat{Y}_i)}{n}$$

where:

- N is the total number of instances.
- $d(Y_i, \hat{Y}_i)$ is the Hamming distance between the true labels Y_i and the predicted labels $\hat{Y}_i$
- n is the number of labels.

CHAPTER 4: IMPLEMENTATION AND RESULTS

4.1 Implementation

The “clean_text” function preprocesses textual data by performing several key cleaning operations. Initially, it handles punctuation by iterating through each character in the string.punctuation list, replacing each punctuation mark in the text with the mark surrounded by spaces to facilitate later removal. Emojis are removed using a regular expression that matches characters outside the Unicode Basic Multilingual Plane (BMP).

Next, a regular expression removes all remaining punctuation, and another pattern removes digits. The text is then converted to lowercase to ensure uniformity. Word segmentation is performed using the rdrsegmenter.word_segment method, which is particularly useful for languages without spaces between words. Finally, the segmented words are joined into a single string, separated by commas and spaces. The raw text after mapping over the clean_text function will be cleaned and structured, making it suitable for various natural language processing tasks.

```
def clean_text(text):
    for char in string.punctuation: # Add spaces before and after punctuation
        text = text.replace(char, ' ' + char + ' ')

    emoji_pattern = re.compile(pattern="[^\u0000-\uFFFF]",flags=re.UNICODE)
    text = emoji_pattern.sub(r"", text) # Remove emojis
    text = re.sub(r'[^\w\s]', '', text) # Remove punctuation
    text = re.sub(r'\d+', '', text)     # Remove digits
    text = text.lower() # Convert to lowercase
    text = rdrsegmenter.word_segment(text) # Word Segmentation
    # text = ViTokenizer.tokenize(text) # Word Segmentation
    text = ', '.join(text)     # Join the text into a string
    return text
```

Figure 4.1: Cleaning Text Process

The “normalize_acronyms” function is essential for ensuring coherence and accuracy within a list of words, primarily by replacing acronyms with their corresponding full forms as defined in the “replace_list” dictionary. It incorporates a safeguard mechanism to exclude any substituted words that appear in the Vietnamese “bad_words” list, thus enhancing data quality by filtering out potentially inappropriate or undesired terms. This function is particularly valuable in text preprocessing tasks where consistency and appropriateness of language are paramount, ensuring that the processed data remains suitable for further analysis or application-specific tasks.

```
def normalize_acronyms(word_list):
    normalized_words = []
    for word in word_list:
        # Replace word if it exists in the replace_list dictionary
        replaced_word = replace_list.get(word, word)
        # Check if the replaced word is not a bad word
        if replaced_word.lower() not in bad_words:
            normalized_words.append(replaced_word)
    return normalized_words

reviews['tokenized'] = reviews['tokenized_text'].apply(normalize_acronyms)
```

Figure 4.2: Normalize Acronyms Function

The “convert_sents_ids” function plays a crucial role in preparing textual data for the application of PhoBERT, a specialized language model tailored for Vietnamese text processing. This function processes a list of sentences by first adding start and end tokens to each sentence. It then encodes these modified sentences into sequences of token IDs using a vocabulary (vocab). This encoding step ensures that each word or subword in the sentences is represented numerically, which is essential for feeding the data into PhoBERT or similar models that require fixed-length input sequences of token IDs.

```
def convert_sents_ids(sents):
    ids = []
    for sent in sents:
        subwords = '<s> ' + bpe.encode(sent) + ' </s>'
        encoded_sent = vocab.encode_line(subwords, append_eos=True, add_if_not_exist=False).long().tolist()
        ids.append(encoded_sent)
    ids = pad_sequences(ids, maxlen=max_length, dtype="long", value=0, truncating="post", padding="post")
    return torch.tensor(ids)
```

Figure 4.3: Padding sequences in PyTorch

The “make_mask” function generates binary masks for batches of token IDs, essential for models like BERT and PhoBERT in natural language processing. These masks distinguish between actual tokens and padding tokens (0), ensuring models handle variable-length sequences efficiently. By converting masks into PyTorch tensors, the function supports seamless integration with deep learning frameworks, optimizing computational resources and enhancing model accuracy in tasks requiring fixed-length input sequences.

```
def make_mask(batch_ids):
    batch_mask = []
    for ids in batch_ids:
        mask = [int(token_id > 0) for token_id in ids]
        batch_mask.append(mask)
    return torch.tensor(batch_mask)
```

Figure 4.4: Make mask function

The design of Bi-LSTM and Bi-GRU architecture with detailed information on each output is presented in the below graph:

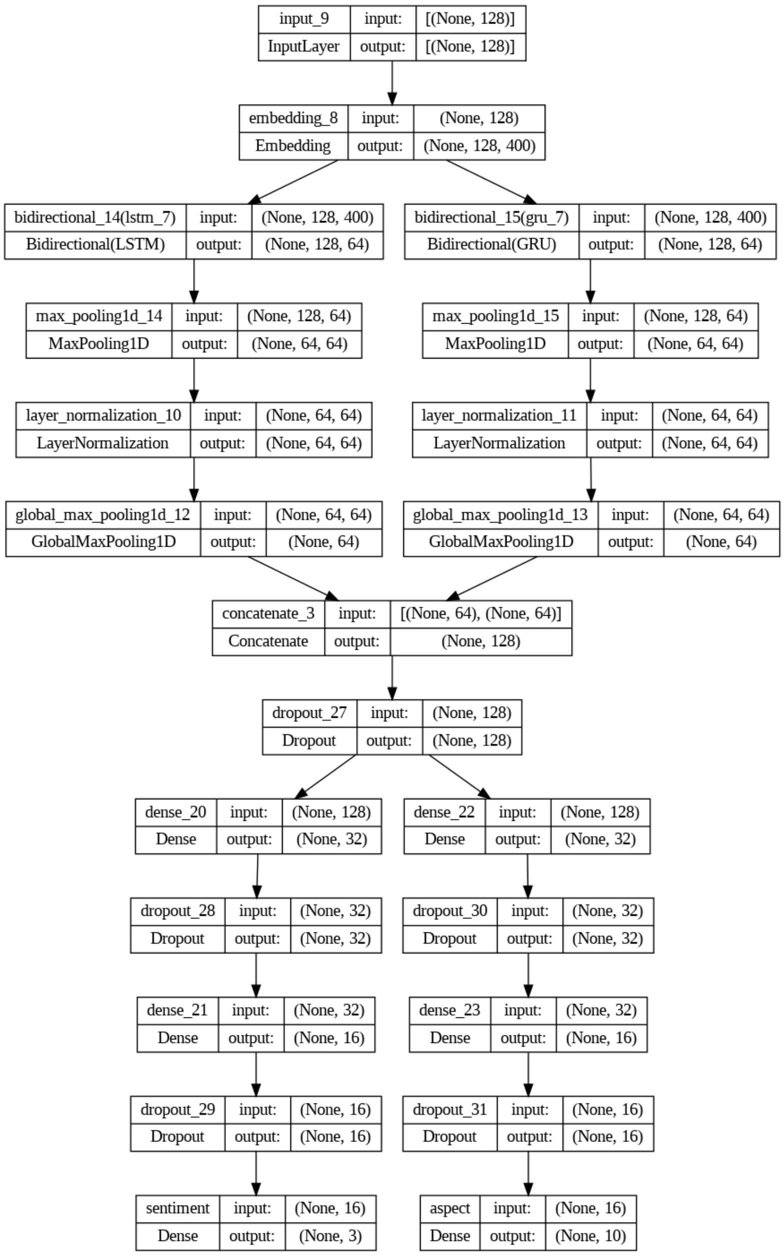

Figure 4.5: Details of the Bi-LSTM and Bi-GRU architecture

4.2 Results

4.2.1 Results of Sentiment Analysis

The results presented in Table 4.1 highlight the performance of two sentiment analysis models: the combined Bi-LSTM and Bi-GRU model, and the PhoBERT model. The combined model of Bi-LSTM and Bi-GRU exhibited a cross-entropy loss of 0.607, an accuracy of 0.739, and an F1-score of 0.753. On the other hand, the PhoBERT model showed a lower cross-entropy loss of 0.447, a higher accuracy of 0.773, and an F1-score of 0.702. In the comparison in the running time, PhoBERT model outperforms the Bi-LSTM and Bi-GRU architecture with 9784 second training time.

PhoBERT model's superior accuracy and lower cross-entropy loss, indicating better predictive accuracy and less error compared to the Bi-LSTM + Bi-GRU model. Despite this, the Bi-LSTM + Bi-GRU model achieved a higher F1-score, suggesting a more balanced performance across precision and recall. Additionally, PhoBERT demonstrated greater efficiency with a shorter runtime. Future work should focus on optimizing these models further. Enhancements for the Bi-LSTM + Bi-GRU model could include incorporating additional layers or attention mechanisms, while fine-tuning PhoBERT with larger datasets might improve its F1-score.

Table 4.1: Results of Sentiment Analysis models

	Cross Loss	*Accuracy*	*F1-score*	*Runtime*
Bi-LSTM + Bi-GRU	0.607	0.739	0.753	12350s
PhoBERT	0.447	0.773	0.702	9784s

When the Bi-LSTM and Bi-GRU models' training and validation losses are compared, the training loss reduces dramatically over 18 epochs from nearly 1 to below 0.5 at the end. In contrast, the validation loss reduces faster at the first three epochs before remaining stable at 0.68 until the end of the training phase.

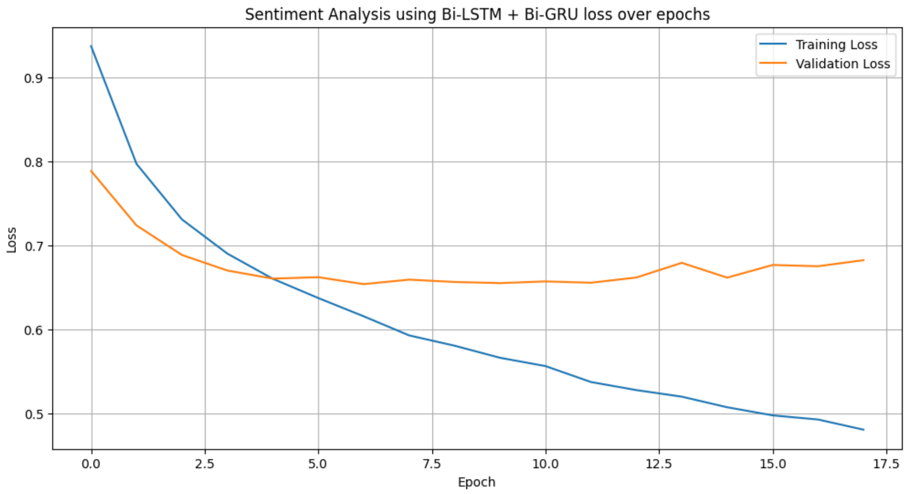

Figure 4.6: Comparision in loss of Bi-LSTM and Bi-GRU model for sentiment analysis

4.2.2 Results of Aspect Extraction

The results from Table 4.2 provide a comparative analysis of different aspect extraction models, highlighting the performance of Binary Relevance, Classifier Chains, Label Powerset, and the combined Bi-LSTM and Bi-GRU model. The Bi-LSTM + Bi-GRU model significantly outperforms the others, achieving the highest accuracy, macro precision, macro recall, macro F1-score, and Hamming score. However, this superior performance comes with the longest runtime, indicating higher computational complexity.

In contrast, Binary Relevance and Classifier Chains exhibit similar performance, with moderate accuracy and macro precision, but they fall short in macro recall and macro F1-score compared to the Bi-LSTM + Bi-GRU model. These models are more efficient with relatively short runtimes, making them quicker but less effective.

The Label Powerset model shows the lowest accuracy and macro recall, resulting in a low macro F1-score, despite having the shortest runtime. While the Bi-LSTM + Bi-GRU model provides the most comprehensive and accurate aspect extraction, it requires more computational resources and time. Simpler models like Binary Relevance and Classifier Chains offer quicker results but with less precision and recall.

Table 4.2: Results of Aspect Extraction over different models

	Accuracy	*Macro Precision*	*Macro Recall*	*Macro F1-Score*	*Hamming Score*	*Runtime*
Binary Relevance	0.5933	0.9049	0.5572	0.6404	0.1743	27s
Classifier Chains	0.5934	0.8831	0.5521	0.6316	0.1760	23s
Label Powerset	0.5398	0.9377	0.3038	0.3405	0.1775	16s
Bi-LSTM + Bi-GRU	0.8455	0.9298	0.9150	0.9219	0.9125	12350s

4.3 Model Evaluation

By comparsion between models along the thesis research, the key findings of the research are PhoBERT for the sentiment analysis task and Bi-LSTM + Bi-GRU for the aspect extraction. Some of important metrics of the best model will be presented in the below table as the summary for the findings.

Table 4.3: Best model for Sentiment Analysis

	Cross Loss	*Accuracy*	*F1-score*	*Runtime*
PhoBERT	0.447	0.773	0.702	9784s

Table 4.4: Best Model for Aspect Extraction

	Accuracy	*Macro Precision*	*Macro Recall*	*Macro F1-Score*	*Hamming Score*	*Runtime*
Bi-LSTM + Bi-GRU	0.8455	0.9298	0.9150	0.9219	0.9125	12350s

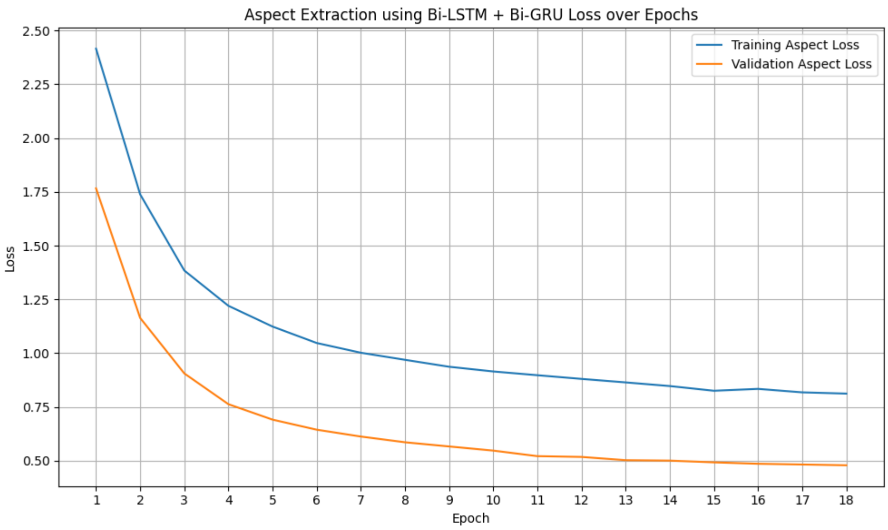

Figure 4.7: Loss of Bi-LSTM +Bi-GRU model for Aspect Extraction

The classification report of the Bi-LSTM and Bi-GRU models displays a good classification performance with very high metrics values. This model's drawback occurs only when predicting security and privacy in the e-wallet reviews app.

Aspect Extraction Classification Report:

	precision	recall	f1-score	support
convenience	0.88	0.82	0.85	1292
payment_integration	0.99	0.95	0.97	2882
accessibility	0.97	0.96	0.97	920
security_privacy	0.64	0.56	0.60	203
customer_support	0.97	0.94	0.96	272
technical_issues	0.97	0.98	0.98	1654
updates	0.97	0.98	0.97	186
fraud	1.00	0.98	0.99	155
promotion	0.97	1.00	0.98	177
functionality	0.94	0.97	0.96	276
micro avg	0.95	0.93	0.94	8017
macro avg	0.93	0.91	0.92	8017
weighted avg	0.95	0.93	0.94	8017
samples avg	0.94	0.94	0.93	8017

Figure 4.8: Classification Report of Aspect Extraction using Bi-LSTM and Bi-GRU

4.4 Discussion

The proposed methods outperformed the performance of machine learning based approaches in sentiment analysis of Vietnamese text where SVM approaches reach only 64.6% accuracy in total [26]. Moreover, the classification of each class is quite low compared to the approach of the author where the two positive and negative classes are well classified and acceptable prediction on the neutral class. The gap between the two approaches can be understood since the applications of that paper use traditional machine learning and the traning dataset is quite limited. Another comparison that is matched with the author’s topic is sentiment analysis using deep learning on Amazon reviews and ratings data [27]. This paper performance is slightly better performance metrics (81.82% accuracy) with the use of only basic GRU architecture. However, the dataset using for this approach are mainly in English, which has diverse resources and advanced approaches already. In contrast, the author worked on the dataset in Vietnamese with many difficulty in understanding context, the user behavior on non-latin reviews, complex characteristics of Vietnamese language. As the result, the proposed result of the author is quite acceptable in the context of Vietnames language where there are not many reasearch paper deep dived into this topic.

For the discussion on aspect extracion method, a research on multi-label classification of E-Commerce customer reviews is quite related with the same outcome goal. This paper hands on multiple models on aspect extraction to compare performance of each classification approach. The classifcation using BR-RF with embedding matrix of XGB achives the highest performance with hamming loss at 0.0589 and the micro along with macro average is quite good evaluation [28]. In comparison to the author's proposed model, the Bi-LSTM and Bi-GRU is greater than that paper in every metrics evaluation with the hamming loss only 0.0186 and all macro and micro averages are above 90%. In the context of Indonesian customer reviews using bidirectional encoder representations (BERT) with the use of transformer architecture, the fine-tuning of BERT reaches an accuracy of over 76% and a hamming loss of 0.0299 [29]. Again, the proposed model for aspect extraction using Bi-LSTM and Bi-GRU exceeds the performance of that paper with an accuracy of 84.55% with the least hamming loss 0.0186 against other related papers.

In overall, the performance of approaches on sentiment analysis and aspect extraction proposed by the author is quite acceptable in the context of digital wallet reviews in Vietnamese language. Therefore, these approaches can be applied in the real world applications to improve the technological feature of these e-wallet apps.

CHAPTER 5: CONCLUSION AND FUTURE WORK

5.1 Conclusion

Throughout the thesis research, the author has reached the stated objectives to create a robust and effective classification model for both sentiment analysis and aspect extraction. These models serve as essential tools for applying in the context of digital wallet industry where many businesses are trying to dominate the competitive market. The remarkable achievement is that the author could explore the new world applications of AI, especially for using the Vietnamese language. Another key takeaway after completing the thesis is the broadened knowledge of both s and y, which has improved and achieved the main goal of mining user insights for tech advancement in e-wallet app. Overall, the performance of deep learning approaches post-implementation is more accurate and reliable than traditional methods. If these stated models could be enhanced in computational time, they would become the highest quality choice in the field of NLP, particularly for sentiment analysis and aspect extraction.

However, the thesis encountered limitations regarding data quality, the author's domain knowledge in the digital wallet field, and the precise definition of neutral sentiment. The user review data scraped from app distribution platforms are not standardized and may not capture all significant features for generalization during the training phase. Furthermore, the reliance on keyword matching for labeling aspect data is a drawback, as it depends heavily on a human-defined dictionary. The lack of domain knowledge poses a challenge in defining aspect categories for the classification task. Consequently, the generated aspects are somewhat biased and may not encompass all existing categories in the real-world application of e-wallet apps. Furthermore, there is no precise definition of neutral sentiment, nor are there specific words to describe this state. As a result, the existence of a neutral class in the output can sometimes affect the effectiveness of the classification model by introducing additional noise and bias into the predictions of positive and negative sentiments.

5.2 Future work

Upon completing the sentiment analysis and aspect extraction from customer reviews of digital wallet apps, several avenues for future work can be explored to optimize further and enhance feedback analytics. These future works focus on improving advanced deep learning and machine learning methodologies to gain deeper insights and drive technological advancements in digital wallet services. Firstly, optimizing Bi-LSTM and Bi-GRU models with carefully selected hyperparameters to make the model more powerful and capture the complex context of the Vietnamese language. To increase the complexity and capability of the sentiment analysis models, future work can explore the integration of Convolutional Neural Networks (CNN) with Bi-LSTM and Bi-GRU architectures. This hybrid combination will take advantage of CNN to provide a more comprehensive analysis of user feedback, leading to more precise sentiment and aspect detection.

Another future direction is to develop real-time feedback analysis systems that can dynamically analyze user reviews as they are posted. This task will involve the need to create scalable and efficient deep learning models capable of processing large volumes of data in real-time. Additionally, predictive modeling can be integrated to forecast emerging sentiment trends and potential issues, allowing companies to proactively address user concerns and enhance their services.

REFERENCES

[1] Thong, V. (2024, February 27). Vietnam to have 50m Digital Wallets this year - vnexpress international. VnExpress International – Latest news, business, travel and analysis from Vietnam. https://e.vnexpress.net/news/business/economy/vietnam-to-have-50m-digital-wallets-this-year-4715807.html

[2] Brown, M., Hentschel, N., Mettler, H., & Stix, H. (2022). The convenience of electronic payments and consumer cash demand. Journal of Monetary Economics, 130, 86-102.

[3] Ha, H. (2024, March 9). Cashless payment on the rise in Viet Nam. en.baochinhphu.vn. https://en.baochinhphu.vn/cashless-payment-on-the-rise-in-viet-nam-111240309092453227.htm

[4] Thieu, H. T. T. Challenges In Classification Of Vietnamese Sentiment. ISSN: 2581 - 9941. Vol.6, Sept-Oct 2021.

[5] Wardana, A. A., Saputro, E. P., Wahyuddin, M., & Abas, N. I. (2022, June). The effect of convenience, perceived ease of use, and perceived usefulness on intention to use e-wallet. In International Conference on Economics and Business Studies (ICOEBS 2022) (pp. 386-395). Atlantis Press.

[6] Nguyen, C., Nguyen, T., & Tran, T. (2020). The Determinants of Consumer's Intention to Use E-wallet: The Case Study of MoMo in Vietnam. International Journal of Advanced Science and Technology, 29(3), 14284-14293.

[7] F. Nizam, H. J. Hwang, and N. Valaei, "Measuring the effectiveness of E-wallet in Malaysia," vol. 786. Springer International Publishing, 2019.

[8] Muhtasim, D. A., Tan, S. Y., Hassan, M. A., Pavel, M. I., & Susmit, S. (2022). Customer satisfaction with digital wallet services: an analysis of security factors. International Journal of Advanced Computer Science and Applications, 13(1), 195-206.

[9] Huynh, A. V. (2023). Determinants Affecting Intention to Use E-Wallets in Vietnam: An Empirical Case of Momo. KnE Social Sciences, 125-141.

5. Customer Service/ Service Quality

[10] Chang, Y.P., Lan, L.Y. and Zhu, D.H. (2017), "Understanding the intention to continue use a mobile payment", International Journal of Business and Information, Vol. 12 No. 4, pp. 363-390.

[11] Chu, P.-Y., Lee, G.-Y. and Chao, Y. (2012), "Service quality, customer SAT, customer trust, and loyalty in an e-banking context", Social Behavior and Personality: An International Journal, Vol. 40 No. 8, pp. 1271-1284.

[12] Linck, K., Pousttchi, K., & Wiedemann, D. G. (2006). Security issues in mobile payment from the customer viewpoint.

[13] GS, A. P., & Arivazhagan, R. (2021). A study on user perceived trust in mobile wallet. International Journal of Management (IJM), 12(3).

[14] Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. Knowledge-Based Systems, 226, 107134

[15] Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. Computational linguistics, 37(2), 267-307.

[16] Kolchyna, O., Souza, T. T., Treleaven, P., & Aste, T. (2015). Twitter sentiment analysis: Lexicon method, machine learning method and their combination. arXiv preprint arXiv:1507.00955.

[17] Yadav, A., & Vishwakarma, D. K. (2020). Sentiment analysis using deep learning architectures: a review. Artificial Intelligence Review, 53(6), 4335-4385.

[18] Naseem, U., Razzak, I., Musial, K., & Imran, M. (2020). Transformer based deep intelligent contextual embedding for twitter sentiment analysis. Future Generation Computer Systems, 113, 58-69.[19] Poria, S., Cambria, E., Ku, L. W., Gui, C., & Gelbukh, A. (2014, August). A rule-based approach to aspect extraction from product reviews. In Proceedings of the second workshop on natural language processing for social media (SocialNLP) (pp. 28-37).

[20] D. V. Thin, V. D. Nguye, K. V. Nguyen and N. L. -T. Nguyen, "Deep Learning for Aspect Detection on Vietnamese Reviews," 2018 5th NAFOSTED Conference on Information and Computer Science (NICS), Ho Chi Minh City, Vietnam, 2018, pp. 104-109, doi: 10.1109/NICS.2018.8606857.

[21] Hesaraki, S. (2023, October 27). Long Short-Term Memory (LSTM). Medium. https://medium.com/@saba99/long-short-term-memory-lstm-fffc5eaebfdc

[22] Zvornicanin, W. by: E. (2024, March 18). Differences between bidirectional and Unidirectional LSTM. Baeldung on Computer Science. https://www.baeldung.com/cs/bidirectional-vs-unidirectional-lstm

[23] Mekruksavanich, S., & Jitpattanakul, A. (n.d.). RNN-based deep learning for physical activity recognition using smartwatch sensors: A case study of simple and complex activity recognition. Mathematical Biosciences and Engineering. https://www.aimspress.com/article/doi/10.3934/mbe.2022265

[24] Krüger, F. (2016). Activity, context, and plan recognition with computational causal behaviour models. https://www.researchgate.net/publication/314116591_Activity_Context_and_Plan_Recognition_with_Computational_Causal_Behaviour_Models

[25] Huy, T. V. (2022, March 7). Multi-label classification. Multi-label classification – Tran Van Huy – Artificial Intellegence. https://huytranvan2010.github.io/Multi-label-classification/

[26] Bang, T. S., Haruechaiyasak, C., & Sornlertlamvanich, V. (2015, November). Vietnamese sentiment analysis based on term feature selection approach. In Proc. 10th International Conference on Knowledge Information and Creativity Support Systems (KICSS 2015) (pp. 196-204).

[27] Shrestha, N., & Nasoz, F. (2019). Deep learning sentiment analysis of amazon. com reviews and ratings. arXiv preprint arXiv:1904.04096.

[28] Deniz, E., Erbay, H., & Coşar, M. (2022). Multi-label classification of E-Commerce Customer Reviews via machine learning. Axioms, 11(9), 436. https://doi.org/10.3390/axioms11090436

[29] Nissa, N. K., & Yulianti, E. (2021). Multi-label text classification of Indonesian customer reviews using bidirectional encoder representations from transformers language model. International Journal of Electrical and Computer Engineering (IJECE). https://ijece.iaescore.com/index.php/IJECE/article/view/31006