SENTIMENT ANALYSIS OF LABUBU POPULARITY TRENDS: A STUDY OF POPMART COMMENTS USING NATURAL LANGUAGE PROCESSING TECHNIQUES

KITTINUN POONSAWAT

AN INDEPENDENT STUDY SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIRMENT FOR THE DEGREE OF MASTER OF SCIENCE
IN DATA SCIENCE AND ANALYTICS
KMITL DIGITAL ANALYTICS AND INTELLIGENCE CENTER SCHOOL OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
2025

KMITL-2025-SC-M-017-003

Independent Study	Sentiment Analysis of LABUBU Popularity Trends: A Study of POPMART Comments Using Natural Language Processing Techniques
Student Name	Mr. Kittinun Poonsawat
Student ID	66056009
Degree	Master of Science (Data Science and Analytics) KMITL Digital Analytics and Intelligence Center
Year	2025
Independent Study	Asst. Prof. Dr. Pattama Charoenporn

ABSTRACT

In today's digital era, where consumer sentiment significantly influences brand perception, understanding customer emotions has become a strategic priority for companies like POP MART. Among its iconic collectibles, Labubu stands out as a symbol of the brand's global appeal, gaining tremendous popularity through the viral influence of Lisa from BLACKPINK, who showcased her Labubu collection on social media. This moment sparked a surge in interest, especially in Thailand, where Labubu achieved over 365,000 mentions in a single month, driven by its appeal to young, trend-conscious consumers.

This study, Sentiment Analysis of LABUBU Popularity Trends: A Study of POPMART Comments Using NLP Techniques, investigates customer perceptions and emotions toward Labubu by employing advanced transformer-based NLP models—WangchanBERTa, RoBERTa, and XLM-R. By analyzing Thai-specific and multilingual datasets collected from social media platforms, this research evaluates the models' performance in sentiment classification.

The analysis highlights WangchanBERTa's superior performance on Thai-specific datasets, achieving the highest F1-score of 87.51%. XLM-R demonstrated robust multilingual capabilities with an F1-score of 84.65%, while RoBERTa excelled in English sentiment analysis with an F1-score of 82.83%. These findings emphasize the importance of leveraging both language-specific and multilingual models to address cultural and linguistic nuances effectively.

Sentiment patterns reveal positive comments highlighting emotional connections to Labubu, neutral sentiments about logistical inquiries, and negative feedback on technical and availability issues. These insights guide strategies to enhance marketing, boost customer engagement, and address operational challenges.

This study advances NLP applications in sentiment analysis, offering POP MART data-driven insights to refine global and regional strategies while underscoring the value of multilingual transformer models in analyzing diverse consumer feedback.

Keywords: Sentiment Analysis, Labubu, POP MART, NLP Techniques, Transformer Models, F1-Score

ACKNOWLEDGEMENTS

This independent study has been successfully completed with the generous guidance and invaluable support of my advisor, Assistant Professor Dr. Pattama Charoenporn, whose dedication and expertise have greatly contributed to the success of this research. I am profoundly grateful for her time, advice, and encouragement throughout this journey, and I extend my heartfelt thanks to her.

I would also like to express my sincere gratitude to all the faculty members and guest lecturers who have shared their knowledge and insights over the past two years. Their teachings have been instrumental in enabling me to apply what I have learned to this research effectively. Additionally, I deeply appreciate the collaboration and assistance from many individuals who supported me from the initial stages to the completion of this study. Special thanks go to my peers for their unwavering support in academic matters, group projects, exam preparation, programming techniques, and sourcing relevant information that contributed to the success of this work.

Lastly, I am deeply indebted to my family, especially my parents, for their unconditional love, support, and encouragement. Their steadfast belief in my abilities and their emphasis on the importance of education have been my greatest inspiration throughout my academic journey.

Kittinun Poonsawat

TABLE OF CONTENT

CONTENT (CONT.)

CONTENT (CONT.)

CONTENT (CONT.)

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Background and Significance of the Problem

In today's digital era, consumer interaction with brands has transformed significantly, driven by the rapid growth of social media platforms. Customers now express their opinions, emotions, and preferences online, shaping purchasing decisions and brand perceptions. For brands like POP MART, this digital shift presents both opportunities and challenges: access to real-time consumer feedback is unprecedented, but interpreting this vast data requires advanced analytical tools.

This study focuses on Labubu, one of POP MART's most popular collectibles, which has achieved significant success globally and particularly in Thailand. The brand's marketing strategy, centered on the blind box concept, leverages consumer psychology by emphasizing scarcity, surprise, and exclusivity. Labubu's popularity in Thailand surged notably after Lisa from BLACKPINK showcased her collection on social media. Lisa's influence, combined with her global fanbase, generated a massive wave of attention for Labubu, sparking over 365,000 mentions in just one month according to Nielsen Social Listening. This viral moment highlights the power of celebrity endorsements and the role of social media in driving consumer trends, especially among young women aged 18-24, a key demographic for POP MART.

As POP MART expands internationally, it faces the challenge of navigating diverse cultural and emotional consumer preferences. Strategies that resonate in one market, such as Thailand, may not yield the same results elsewhere. Understanding these nuances is essential for sustained global success.

This research seeks to bridge this gap by analyzing consumer sentiment toward Labubu using pretrained Natural Language Processing (NLP) models, including WangchanBERTa, RoBERTa, and XLM-R, for sentiment classification. By examining comments from social media platforms, specifically POP MART GLOBAL and POP MART THAILAND Facebook pages, the study aims to uncover emotional drivers that influence purchasing decisions. Insights from this analysis will enable POP MART to refine its marketing strategies, enhance customer engagement, and adapt to varying cultural contexts.

This study highlights the broader value of sentiment analysis in deciphering unstructured data to inform decision-making. The findings will provide actionable insights for POP MART, ensuring its continued relevance and competitive edge in the evolving global marketplace.

1.2 Objectives of the Study

1.2.1 To compare consumer sentiment between Thai and global markets based on social media comments about Labubu.

1.2.2 To evaluate the effectiveness of WangchanBERTa for Thai sentiment analysis compared to the multilingual XLM-R model.

1.2.3 To assess the performance of RoBERTa and XLM-R for analyzing English and multilingual comments.

1.2.4 To study social media user behavior and its influence on purchasing decisions for POP MART products, focusing on Labubu.

1.3 Scope of the Study

This research examines consumer sentiment towards Labubu by analyzing comments from the POP MART GLOBAL and POP MART THAILAND Facebook pages. Comments were collected from two primary sources: the Popmart-th page for Thai-specific data to evaluate the performance of a Thai language model and the Popmart-global page for multilingual data to facilitate model comparison in a multilingual context. These datasets were utilized to test the effectiveness of different models in sentiment analysis

1.3.1 Data Collection

The research data will be obtained by scraping comments from Facebook posts specifically related to Labubu. This includes posts mentioning Labubu or featuring the hashtag #Labubu. Comments will be sourced from the official POP MART GLOBAL and POP MART THAILAND fan pages, focusing on user interactions with Labubu-related content. This ensures relevance for sentiment analysis. The comments will be categorized into 3 sentiment classes positive, neutral and negative class

1.3.2 Models Used

This research employs three pretrained NLP models for sentiment classification:

1) WangchanBERTa: A Thai-specific model fine-tuned for capturing nuances in Thai text, applied to comments from the POP MART THAILAND fan page.

2) RoBERTa: An optimized English-language model used for analyzing comments on the POP MART GLOBAL fan page.

3) XLM-R: A multilingual model capable of handling over 100 languages, used for analyzing both Thai and multilingual comments on the POP MART GLOBAL fan page.

1.4 Expected Benefits

1.4.1 Enhanced Product Development: Sentiment analysis will help POP MART identify consumer preferences and emotional drivers, guiding product improvements and trend predictions to ensure Labubu's continued appeal across markets.

1.4.2 Cultural Sensitivity in Marketing: The study will reveal cultural differences in sentiment, enabling POP MART to create region-specific campaigns that align with diverse consumer values and preferences.

1.4.3 Comparison of NLP Models: The research will evaluate the performance of WangchanBERTa, RoBERTa, and XLM-R in sentiment analysis, providing insights into their strengths and limitations for analyzing unstructured social media data.

1.4.4 Broader Industry Implications: Findings will benefit the toy and collectibles industry and beyond by improving understanding of multilingual sentiment, leading to better marketing strategies, product innovation, and customer engagement.

CHAPTER 2

THEORY AND LITERATURE REVIEWS

2.1 Natural Language Processing (NLP)

Natural Language Processing (NLP) is a subfield of artificial intelligence (AI) that focuses on the interaction between computers and human language. The primary goal of NLP is to enable machines to understand, interpret, and generate human language in a way that is meaningful and useful. NLP bridges the gap between human communication and machine understanding, allowing computers to process large amounts of natural language data (text or speech) to perform tasks like translation, sentiment analysis, question answering, and more.

2.1.1 NLP Techniques

2.1.1.1 Tokenization

Tokenization is the process of breaking down a string of text into smaller units called tokens. These tokens can be individual words, subwords, or even characters depending on the task. Tokenization is a foundational step in NLP because models work with numerical representations of text, and the first step toward that is breaking down the text into meaningful components.

1) Types of Tokenization

- Word-Level Tokenization: Each word in the text is treated as a token. For example, in the sentence:
  - Input: "Labubu is popular"
  - Tokens: [‘Labubu’, ‘is’, ‘popular’]

Word-level tokenization is straightforward but can struggle with languages that don't have clear word boundaries (e.g., Chinese or Thai) or when handling contractions (e.g., "don’t").

- Subword Tokenization: The words are broken down into subwords or character chunks. This is useful for handling rare words or morphologically rich languages. For example, Byte-Pair Encoding (BPE) is a popular subword tokenization technique used in models like BERT and RoBERTa. For the word "unhappiness":

- Input: "unhappiness"
- Tokens: [‘un’, ‘happiness’]

- Character-Level Tokenization: Each character in the text is treated as a token. This technique is useful in certain language models, especially for tasks involving low-resource languages or texts with noisy data

- Input: "Labubu"
- Tokens: [‘L’, ‘a’, ‘b’, ‘u’, ‘b’, ‘u’]

2) Important of Tokenization

Tokenization determines how the text is fed into models. Poor tokenization can lead to loss of meaning or context, especially in languages with complex morphology or compounding (like German or Finnish). In transformer models like BERT or RoBERTa, tokenization happens before text embeddings are created.

2.1.1.2 Named Entity Recognition (NER)

Named Entity Recognition (NER) is an NLP task that involves identifying and classifying entities in a text into predefined categories, such as people's names, organizations, locations, dates, and quantities.

1) NER Mechanism

- Entity Detection: The first step is identifying words or phrases in the text that correspond to entities.

- Entity Classification: The second step is classifying the identified entities into categories. For example, "Labubu" might be classified as a "PRODUCT," and "POP MART" as an "ORGANIZATION."

Example sentence:

- Input: "Labubu is a popular product sold by POP MART."
- NER Output: [‘Labubu’: PRODUCT, ‘POP MART’: ORGANIZATION]

2.1.1.3 Part-of-Speech (POS) Tagging

Part-of-Speech Tagging (POS) is the process of labeling each word in a sentence with its grammatical role. This helps NLP models understand the structure of a sentence and how different words relate to each other.

1) Common POS Tags

- Nouns (N): Words that refer to entities (people, places, things). For example, "Labubu" is a noun.

- Verbs (V): Words that describe actions or states. For example, "is" in the sentence "Labubu is popular" is a verb.

- Adjectives (ADJ): Words that describe nouns. In "Labubu is popular," "popular" is an adjective.

- Pronouns (PRON): Words that replace nouns (e.g., "he," "she," "it").

- Prepositions (PREP): Words that show the relationship between nouns (e.g., "in," "on," "at").

2) POS Tag Mechanism

- Rule-Based Approaches: Early methods of POS tagging relied on rule-based systems that used linguistic rules to assign tags.

- Statistical and Machine Learning Approaches: Modern POS taggers use machine learning models that are trained on large, annotated corpora. These models learn to predict the correct POS tag based on the word itself and its surrounding context. CRFs (Conditional Random Fields) and HMMs (Hidden Markov Models) were commonly used before deep learning became dominant.

2.1.1.4 Dependency Parsing

Dependency Parsing is an NLP task that involves analyzing the grammatical structure of a sentence and establishing relationships between "head" words and words that modify those heads. Essentially, it helps map out how words in a sentence depend on one another.

1) Dependency Parsing Mechanism:

In dependency parsing, every word in a sentence is connected to other words through syntactic relationships. Each relationship is represented as a head and a dependent. The root of the sentence (usually the main verb) is the central head, and all other words are either directly or indirectly connected to it.

Example sentence:

- Input: "Labubu is popular worldwide."

- Dependency Parse: "Labubu" (subject) depends on "is" (verb), and "popular" (predicate) also depends on "is."
- "Labubu" ⟶ "is" (subject-verb relationship)
- "is" ⟶ "popular" (verb-object relationship)

2.1.1.5 Text Embedding

Text Embedding is the process of converting words or tokens into numerical vectors that can be processed by machine learning models. These embeddings capture the semantic meaning of words and their relationships within a vector space.

1) Types of Embeddings:

1.1) Static Embeddings:

Word2Vec: One of the earliest popular embedding techniques, Word2Vec creates a fixed-length vector representation for each word based on its co-occurrence with other words in a large corpus. The key idea behind Word2Vec is the distributional hypothesis: words that occur in similar contexts tend to have similar meanings.

Limitation: Word2Vec provides a static representation, meaning a word always has the same embedding regardless of context. For example, "bank" (as in riverbank) and "bank" (as in financial institution) will have the same embedding, which can lead to confusion in contextual applications.

GloVe: Another static embedding technique, GloVe (Global Vectors for Word Representation) is similar to Word2Vec but focuses on the global word co-occurrence matrix rather than local context windows. It also produces fixed embeddings, which suffer from the same limitations as Word2Vec in terms of capturing polysemy (words with multiple meanings).

1.2) Contextual Embeddings:

BERT: Unlike static embeddings, models like BERT produce contextual embeddings, meaning the word's vector representation changes depending on the surrounding words. This enables the model to distinguish between different senses of a word based on context.

Example: In the sentence "The bank of the river" vs. "I deposited money in the bank," BERT will assign different embeddings to the word "bank" based on its context.

Transformer-Based Models: Modern transformer models like RoBERTa and XLM-R build on BERT's idea of contextual embeddings. These embeddings allow for better generalization across various NLP tasks (sentiment analysis, translation, etc.) since the model captures the full meaning of a word in the specific context it appears.

2.2 Sentiment Analysis

Sentiment Analysis is a subfield of Natural Language Processing (NLP) that focuses on determining the emotional tone or attitude conveyed in text. It is often used to analyze opinions expressed in social media, product reviews, blogs, or other written communication. The goal is to classify text into different sentiment categories, typically:

- Positive: Expressions that convey approval or a favorable opinion.
- Negative: Expressions of dissatisfaction or unfavorable opinions.
- Neutral: Text that does not express strong emotion either way, or is factual in nature.

2.2.1 Methods of Sentiment Analysis

There are different approaches to performing sentiment analysis:

- Lexicon-Based Approach: This is one of the earliest methods. It uses predefined dictionaries or lexicons of words that are associated with positive, negative, or neutral sentiments. For example, words like "happy" or "great" are tagged as positive, while words like "bad" or "terrible" are tagged as negative. However, lexicon-based approaches often struggle with context and sarcasm.

- Machine Learning-Based Approach: Instead of relying on predefined word lists, this approach uses statistical models to classify sentiment. The model is trained on labeled data (i.e., data where the sentiment is already known) and learns patterns in the text that correspond to positive, negative, or neutral sentiments. Traditional classifiers like Naive Bayes, SVM, and Logistic Regression were commonly used in earlier approaches.

- Deep Learning-Based Approach: With the advent of deep learning, models like RNNs (Recurrent Neural Networks) and LSTMs (Long Short-Term Memory Networks) became popular because they could learn long-term dependencies in the text, making them more adept at understanding the sentiment expressed across longer passages.
- Transformer-Based Approach: This is the most advanced and current method, using models like BERT, RoBERTa, and XLM-R. These models use the transformer architecture (discussed in detail below) and are pretrained on massive datasets, learning complex linguistic features. When fine-tuned on a sentiment analysis task, these models excel in classifying sentiments due to their bidirectional context understanding.

2.3 Transformer Models

The Transformer architecture, introduced in 2017 by Vaswani et al., revolutionized NLP by addressing the limitations of previous models like RNNs and LSTMs, which struggled with processing long sequences due to their sequential nature. Transformers, on the other hand, are highly parallelizable and capable of understanding long-term dependencies within text without requiring the sequence to be processed word by word.

Transformers form the foundation for state-of-the-art models like BERT, RoBERTa, and XLM-R. Their ability to handle complex dependencies, capture context bidirectionally, and scale to massive datasets makes them highly suitable for sentiment analysis and other NLP tasks.

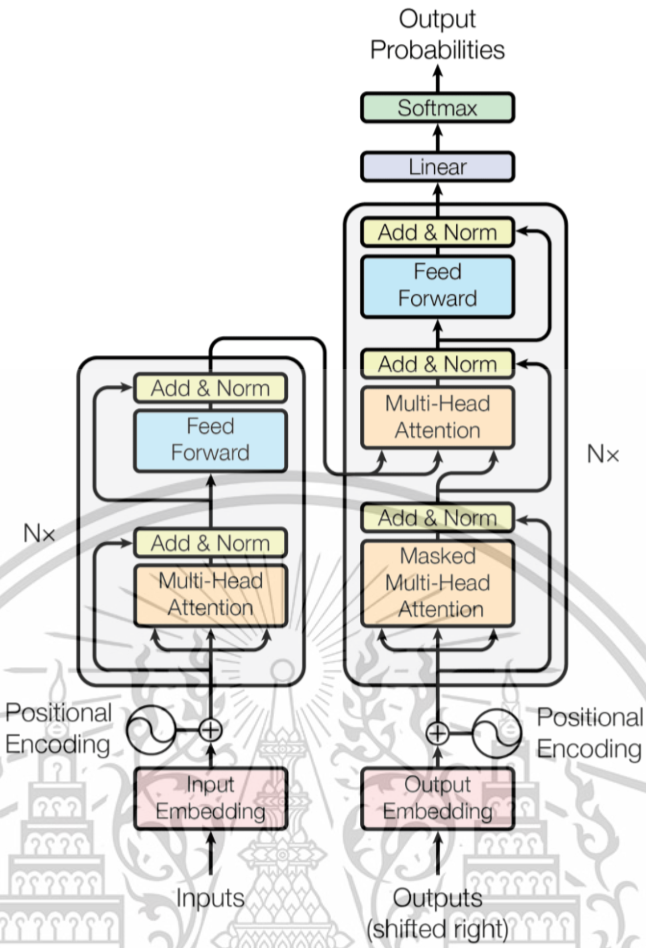

Figure 2.1 Tranformer Architecture

2.3.1 Key Components of the Transformer Architecture

From the Figure1 it shows the architecture of Transformer as following

2.3.1.1 Self-Attention Mechanism: The key innovation of transformers is the self-attention mechanism, which allows the model to weigh the importance of each word in a sentence relative to every other word. For example, in the sentence "Labubu is popular because it's unique," the self-attention mechanism allows the model to understand that "popular" is related to "Labubu," and "unique" explains why it is popular. Mathematically, self-attention is computed as follows:

$$Attention\ (Q, K, V) = softmax(\frac{QK^T}{\sqrt{D_k}})V \quad (1)$$

Where: Q : Query vector (projection of the input word)

K : Key vector (projection of the input word)

V : Value vector (representation of the input word)

D_K : Dimension of the key vector

The output of the attention mechanism is a weighted combination of the value vectors, allowing the model to focus on relevant parts of the sentence.

2.3.1.2 Positional Encoding: Unlike RNNs, transformers do not inherently understand the order of words in a sequence. To address this, positional encodings are added to the input embeddings to give the model information about the position of each word in the sentence. These encodings are added to ensure the model knows that "Labubu" comes before "popular" in a sentence, which affects how it interprets the relationship between the words.

2.3.1.3 Feed-Forward Networks: After self-attention, the output is passed through feed-forwardnetworks, which process each word's context and compute the final representation used for the downstream task (e.g., sentiment classification).

2.3.1.4 Multi-Head Attention: Transformers also use multi-head attention, where the attention mechanism is applied multiple times in parallel. This allows the model to capture different types of dependencies between words. For instance, one attention head might focus on syntactic relationships (e.g., subject-verb agreement), while other focuses on semantic relationships (e.g., meaning of words).

2.4 BERT (Bidirectional Encoder Representations from Transformers)

BERT (introduced by Google in 2018) was a breakthrough in NLP and is considered one of the foundational models in the transformer-based NLP revolution. Unlike previous models that processed text in a unidirectional manner, BERT is designed to understand language context by looking at both left and right words in a sentence simultaneously.

2.4.1 Architecture

BERT is based on the Transformer architecture and focuses entirely on the Encoder part of the Transformer. The Self-Attention Mechanism at the heart of BERT allows it to compute relationships between every word in a sentence and every other word, making it a bidirectional model. This is key to capturing the true context of a word within a sentence.

Key architectural components:

- Self-Attention Mechanism: Allows BERT to understand how words in a sentence relate to each other. For example, in the sentence "Labubu is popular," BERT can relate "Labubu" to "popular" even if the two words are not directly adjacent.

- Positional Encoding: Since BERT does not use recurrence or convolution (as RNNs and CNNs do), it uses positional encoding to capture the order of words. This ensures the model knows which words come first, second, and so on.

2.4.2 Pretraining Tasks

BERT is pretrained using 2 key tasks:

1) Masked Language Modeling (MLM): During training, BERT randomly masks some percentage of the input tokens (usually 15%) and tries to predict the masked words based on the context provided by the surrounding words. For example, in the sentence "Labubu is [MASK]," BERT must predict "popular" using the surrounding context.

2) Next Sentence Prediction (NSP): BERT is trained to predict whether two sentences appear consecutively in a text or not. This helps BERT understand the relationship between different parts of a document, making it more effective at tasks like question answering and natural language inference.

2.4.3 Fine-Tuning

Once BERT is pretrained, it can be fine-tuned on specific downstream tasks like sentiment analysis, question answering, or text classification. In fine-tuning, the model is trained on task-specific data with supervised learning, where labeled examples help the model adjust its weights for that particular task.

2.4.4 Strengths and Weaknesses

- Strengths: BERT’s bidirectional nature allows it to deeply understand the context, making it highly effective for tasks where understanding nuanced meaning is important (such as sentiment analysis, especially in longer sentences).

- Weaknesses: BERT’s pretraining objectives (MLM and NSP) may not be optimal for all tasks, and its training efficiency has been surpassed by newer models like RoBERTa.

2.5 RoBERTa (Robustly Optimized BERT Pretraining Approach)

RoBERTa (introduced by Facebook AI in 2019) is a direct improvement of BERT, designed to enhance its pretraining strategy. RoBERTa uses the same architecture as BERT but with some significant modifications to its pretraining methodology, making it more efficient and powerful for downstream tasks.

2.5.1 Key Improvements Over BERT

1) No Next Sentence Prediction (NSP): One of the key changes in RoBERTa is the removal of the Next Sentence Prediction (NSP) task, which BERT uses during pretraining. Research showed that removing this task improves the model's overall performance on many downstream tasks, as NSP was not particularly beneficial.

2) Larger and Longer Training: RoBERTa is trained on more data and for a longer period compared to BERT. RoBERTa uses 10x more data and trains for significantly more iterations. This allows the model to capture more linguistic patterns and dependencies that BERT might miss.

3) Dynamic Masking: In BERT, the masked tokens in MLM are fixed for every input sequence during training. RoBERTa introduces dynamic masking, where the masked tokens are selected randomly for each training step. This improves generalization by making the model less dependent on a specific mask configuration.

4) Batch Size and Learning Rate: RoBERTa uses much larger batch sizes and a more aggressive learning rate schedule, enabling it to converge faster and reach better results.

2.5.2 Pretraining and Fine-Tuning

RoBERTa still uses Masked Language Modeling (MLM) for pretraining, but without the NSP task. Like BERT, after pretraining, RoBERTa can be fine-tuned for specific tasks, including sentiment analysis, by adjusting the final layers of the network.

2.5.3 Strengths and Weaknesses

- Strengths: RoBERTa’s improvements in training (more data, dynamic masking, no NSP) make it more efficient and effective than BERT, especially in text classification tasks like sentiment analysis. RoBERTa often achieves better performance than BERT on benchmarks like GLUE.

- Weaknesses: Although RoBERTa improves on BERT in many areas, it still relies heavily on the Transformer encoder and has similar computational complexity, which can make it slower to fine-tune and deploy for real-time applications compared to more recent models like DistilBERT.

2.6 XLM-R (XLM-RoBERTa)

XLM-R (introduced by Facebook AI in 2019) is a multilingual extension of RoBERTa. It is designed to perform well across multiple languages, making it suitable for multilingual tasks such as cross-lingual sentiment analysis. XLM-R is pretrained on over 100 languages, including low-resource languages, making it highly versatile.

2.6.1 Key Features of XLM-R

1) Multilingual Pretraining: XLM-R is pretrained on text data from over 100 languages, making it one of the most powerful multilingual models available. It is designed to understand not just English but also languages like Thai, Bengali, Swahili, and many more, which is critical for projects that involve multilingual data

2) Cross-Lingual Transfer Learning: One of the most important innovations of XLM-R is its ability to perform cross-lingual tasks. This means that it can transfer knowledge learned from one language (e.g., English) to another (e.g., Thai). For example, XLM-R can be fine-tuned on an English sentiment analysis dataset and applied to analyze sentiment in Thai with minimal fine-tuning, making it very flexible for multilingual applications.

3) No Language-Specific Vocabulary: Unlike earlier multilingual models (e.g., mBERT), XLM-R does not use language-specific subword vocabularies. Instead, it uses a shared vocabulary across all languages, which helps it generalize better across languages, especially for low-resource languages where training data is scarce.

4) Training with Masked Language Modeling (MLM): Like BERT and RoBERTa, XLM-R uses Masked Language Modeling for pretraining, but on a much larger and more diverse dataset that includes multiple languages. This allows XLM-R to learn better cross-lingual representations.

2.6.2 Fine-Tuning for Multilingual Sentiment Analysis

Fine-tuning XLM-R for sentiment analysis involves training the model on labeled sentiment datasets from various languages. Given its pretraining on a multilingual corpus, XLM-R can generalize well across languages even when fine-tuned on a smaller dataset from a specific language, such as Thai. This makes it an ideal model for your project, where comments in multiple languages (Thai, English, and potentially others) need to be analyzed.

2.6.3 Strengths and Weaknesses

- Strengths: XLM-R's multilingual capabilities are its greatest strength. It excels at handling cross-lingual tasks and can transfer knowledge between languages. This makes it ideal for projects involving multiple languages, like analyzing comments on global social media pages.

- Weaknesses: While XLM-R is powerful, it is also resource-intensive. It requires significant computational power to fine-tune and deploy, especially when compared to monolingual models like RoBERTa. Additionally, XLM-R may not perform as well as monolingual models like WangchanBERTa for specific languages like Thai, which have their own specialized models.

2.7 WangchanBERTa

WangchanBERTa is a transformer-based model specifically developed and fine-tuned for the Thai language. It is derived from the BERT architecture and is tailored to handle the unique linguistic structure of Thai, which lacks spaces between words, making tokenization and language modeling more complex than in languages like English. Developed as part of a collaboration between Thai NLP researchers and the wider Hugging Face community, WangchanBERTa is designed to improve natural language processing (NLP) tasks in Thai, such as sentiment analysis, named entity recognition, and text classification.

2.7.1 Architecture

WangchanBERTa retains the same architecture as BERT (Bidirectional Encoder Representations from Transformers), which is based on a transformer encoder. It processes text bidirectionally, meaning it captures context from both the left and right of a word in a sentence, enabling it to understand nuanced meanings.

2.7.2 Pretraining and Fine-tuning:

WangchanBERTa was pretrained on a large corpus of Thai texts, which allows it to generate high-quality language representations specifically suited to the intricacies of Thai syntax, grammar, and vocabulary. Fine-tuning WangchanBERTa on specific tasks such as sentiment analysis helps it adapt to domain-specific datasets, improving its performance for Thai-based NLP tasks.

Tokenization: Thai does not use spaces between words, so traditional tokenization methods often struggle with this language. WangchanBERTa uses subword tokenization techniques like WordPiece to handle this challenge, ensuring that words are properly segmented and understood by the model.

2.7.3 Strengths and Weaknesses

Strengths: WangchanBERTa outperforms general-purpose models like multilingual BERT or XLM-R when applied to Thai-specific tasks. It provides more accurate tokenization, contextual embedding, and understanding of the Thai language, making it ideal for analyzing sentiment and emotion in Thai social media comments.

Weaknesses: While WangchanBERTa excels in Thai language tasks, it may not perform well in multilingual contexts or when exposed to mixed-language texts. This is where models like XLM-R, which are designed for cross-lingual tasks, are better suited.

2.8 Performance Evaluation Metrics

2.8.1 Overall Accuracy

The accuracy metric represents the proportion of correct predictions made by the model out of all predictions. Accuracy can be calculated using the following formula:

$$Accuracy = \frac{(TP+TN)}{(TP+FN+FP+TN)} \quad (2)$$

Where: TP (True Positive): The number of correctly predicted positive cases.

TN (True Negative): The number of correctly predicted negative cases.

FP (False Positive): The number of cases predicted as positive but are actually negative.

FN (False Negative): The number of cases predicted as negative but are actually positive.

In summary, the accuracy is the ratio of correctly predicted cases (both positive and negative) to the total number of predictions made. The values for TP, TN, FP, and FN are derived from the Confusion Matrix. An example of a Confusion Matrix is shown in Table 2.1.

Table 2.1 Confusion Matrix Table

		Actual Data	
		Positive	Negative
Predicted Data	Positive	TP	FP
	Negative	FN	TN

2.8.2 Precision (Positive Predictive Value)

Precision measures the accuracy of positive predictions by calculating the proportion of true positive predictions out of all predicted positive cases. It can be computed as:

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

Precision indicates how many of the predicted positive cases were correctly classified.

2.8.3 Recall (Sensitivity or True Positive Rate)

Recall measures the ability of the model to correctly identify all actual positive cases. It is defined as the proportion of true positive predictions out of all actual positive cases and is given by:

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

This metric helps determine how well the model identifies positive cases and is particularly useful in cases where missing a positive case has a high cost.

2.8.4 F1-Score

The F1-score is the harmonic mean of Precision and Recall. It provides a balance between precision and recall, particularly useful when the dataset is imbalanced. The formula is as follows:

$$F1 = 2 \times \left(\frac{Precision \times Recall}{Precision \times Recall}\right) \tag{5}$$

F1-score helps in evaluating models when both precision and recall are equally important.

2.9 Relevant Studies

2.9.1 Fine-tuned Sentiment Analysis of COVID-19 Vaccine-Related Social Media Data

- Objective: This paper focuses on fine-tuning BERT-based models for the sentiment analysis of social media posts related to COVID-19 vaccines.
- Key Findings: It highlights the importance of sentiment analysis in public health communication by identifying emotions tied to vaccine discussions. The results suggest that fine-tuning domain-specific models can significantly improve sentiment classification.
- Relevance to the project: This paper demonstrates how BERT-based models can be fine-tuned to extract public sentiment from social media platforms, similar to our use of pretrained models for extracting sentiments from POP MART-related comments.
- Key Point: Effective in demonstrating the utility of fine-tuned models for real-time sentiment analysis, which aligns well with our approach of using pretrained models.
- Results: The fine-tuned DistilRoBERTa model identified that Twitter had more negative sentiment compared to Reddit, and positive sentiment increased post-vaccine availability, highlighting the success of domain-specific fine-tuning in capturing sentiment trends.

2.9.2 BERT Base Model for Toxic Comment Analysis on Indonesian Social Media

- Objective: The study applied BERT and its variants like IndoBERT and IndoRoBERTa for detecting toxic comments in Indonesian social media, using pre-trained models tailored to the local language.

- Key Findings: The results showed that IndoBERT outperformed other models for toxic comment detection with an F1 score of 0.8897.

- Relevance to the project: This is closely related to the Thai-specific WangchanBERTa model we plan to use. The success of a localized model in handling sentiment-specific tasks, such as toxic comment detection, supports the application of WangchanBERTa for Thai comment analysis in our project.

- Key Point: Demonstrates how domain-specific models (IndoBERT) outperform general models in specific languages, justifying the use of WangchangBERTa for Thai sentiment analysis.

- Results: IndoBERT achieved the highest F1 score of 0.8897 for toxic comment classification, outperforming general models like SVM and demonstrating the effectiveness of localized BERT models.

2.9.3 Sentiment Analysis for Bengali Using Transformer-Based Models

- Objective: This paper compared multilingual BERT with XLM-R for sentiment analysis on Bengali datasets.

- Key Findings: XLM-R significantly outperformed multilingual BERT in some cases, especially when fine-tuned for specific tasks.

- Relevance to the project: It demonstrates the performance of XLM-R for sentiment analysis in low-resource languages, similar to our use of XLM-R for multilingual analysis of POP MART global comments.

- Key Point: Validates the use of XLM-R in multilingual environments, supporting its use alongside RoBERTa for the English and multilingual sentiment analysis in our study.

- Results: XLM-R achieved validation accuracy of 0.81 for binary classification and 0.68 for three-class classification on Bengali datasets, outperforming BERT in low-resource language tasks.

2.9.4 Applying Deep Learning Techniques for Sentiment Analysis to Assess Sustainable Transport

- Objective: Focused on the application of XLM-R and other transformer models to assess sentiment related to sustainable transportation.
- Key Findings: XLM-R showed excellent performance in binary and multiclass classification tasks, especially in handling noisy data.
- Relevance to the project: This paper demonstrates the robustness of XLM-R in handling noisy datasets, which is crucial when analyzing social media comments that may contain varying sentiment and linguistic structures.
- Key Point: XLM-R’s ability to handle noise and perform multiclass classification aligns with our project's need to classify sentiments from diverse social media data.
- Results: XLM-R achieved 97% accuracy for binary classification but slightly lower performance in multiclass classification tasks, making it ideal for noisy datasets like social media comments.

2.9.5 A Review on Sentiment Analysis from Social Media Platforms

- Objective: A comprehensive review of sentiment analysis techniques applied to social media platforms, including deep learning approaches.
- Key Findings: The paper discusses the limitations of current models in detecting sarcasm, humor, and complex inferences, and suggests improvements in model fine-tuning for better sentiment classification.
- Relevance to the project: Provides insights into the challenges of sentiment analysis on social media, which is directly relevant to our analysis of POP MART-related comments.
- Key Point: Offers a broader perspective on the state of sentiment analysis, highlighting areas for improvement that align with the objectives of our study.
- Results: The review identifies that models like BERT and RoBERTa are effective for sentiment analysis but struggle with detecting complex emotions such as sarcasm, necessitating further fine-tuning for more accurate sentiment detection.

2.9.6 A Fusion Architecture of BERT and RoBERTa for Enhanced Performance of Sentiment Analysis on Social Media Platforms

- Objective: This paper proposes a fusion of BERT and RoBERTa to improve the accuracy of sentiment classification on social media data.
- Key Findings: The fusion of these models improved overall performance compared to using them individually, particularly in handling imbalanced datasets.
- Relevance to the project: This study supports our use of multiple pretrained models (RoBERTa and XLM-R) to capture sentiment across different languages.
- Key Point: Demonstrates the potential benefits of combining models like BERT and RoBERTa for sentiment analysis, which could be an approach worth exploring in our project.
- Results: The fusion model outperformed individual models by improving precision and recall, particularly on imbalanced datasets, which is common in social media data.

2.9.7 Comparative Analysis of Transformer-Based Pre-Trained NLP Models

- Objective: A comparative study of BERT, RoBERTa, and ALBERT for sentiment analysis.
- Key Findings: BERT outperformed the other models with the highest F1 score, but RoBERTa was more efficient in terms of training time.
- Relevance to the project: This provides a comparative perspective that informs our decision to use RoBERTa and XLM-R for different linguistic contexts in the project.
- Key Point: Highlights the relative strengths and weaknesses of each model, helping to justify our choices of RoBERTa for English and XLM-R for multilingual comments.
- Results: BERT achieved the highest F1 score (0.85), while RoBERTa was more efficient in terms of training and inference times, suggesting trade-offs between accuracy and computational cost.

2.10 Model Comparison

Table 2.2 provides a comparative overview of three transformer-based models WangchanBERTa, RoBERTa, and XLM-R focusing on their core architectures, pretraining tasks, strengths, weaknesses, and their specific relevance to the sentiment analysis of POP MART and Labubu-related comments across different languages. The comparison aids in selecting the optimal model for each linguistic context in the project

Table 2.2 Comparative Evaluation of WangchanBERTa, RoBERTa, and XLM-R for Multilingual Sentiment Analysis in the Labubu Project

Model	Architecture	Pretraining Tasks	Strengths	Weaknesses	Relevance to the Project (Labubu Sentiment Analysis)
WangchanBERTa	Transformer (Encoder)	MLM, Thai-specific corpus	Fine-tuned for Thai, handles tokenization challenges, accurate Thai NLP tasks	Less effective in multilingual or mixed-language settings	Crucial for analyzing Thai comments on POP MART TH fan page. Its ability to handle Thai syntax is critical for sentiment classification in Thai.
RoBERTa	Transformer (Encoder)	MLM (No NSP)	Improved pretraining strategy, dynamic masking, outperforms BERT	Computationally expensive, large data requirements	Ideal for handling English-language comments on POP MART Global fan pages. Efficient and highly accurate for monolingual English sentiment tasks.
XLM-R	Transformer (Encoder)	MLM (Multilingual Corpus)	Best for multilingual tasks, strong cross-lingual transfer	High computational cost, may underperform on specific languages	Key for analyzing mixed-language comments on POP MART Global fan page. Excellent for cross-lingual tasks and multilingual sentiment analysis.

CHAPTER 3

RESEARCH METHODOLOGY

This research focuses on sentiment analysis of customer comments about Labubu products, aiming to evaluate and compare the performance of three transformer-based models: WangchangBERTa, RoBERTa and XLM-R. The study examines how effectively each model can classify sentiments in both Thai-specific and global multilingual contexts by testing them on two distinct datasets: The data was collected from the Popmart-th and Popmart-global Facebook pages by reading through comments. By using the evaluation criteria and results outlined in this paper, this research aims at identifying the advantages and disadvantages of each model in terms of sentiment classification in different languages. This summary workflow of research methodology is shown in figure 3.1

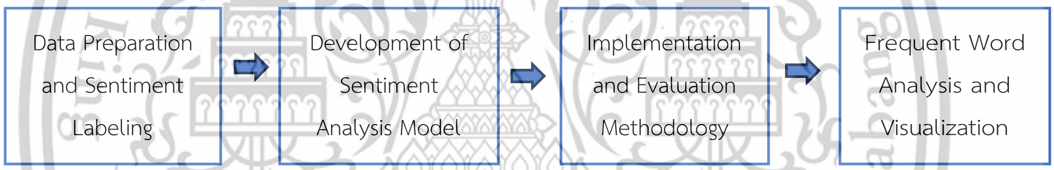

Figure 3.1 Workflow of Research Methodology

3.1 Data Preparation and Sentiment Labeling

3.1.1 Data Collection

The research data consists of comments collected from two primary Facebook sources: Popmart-th and Popmart-global pages. Thai dataset was collected from Popmart-th for testing the Thai language model, while Popmart-global data was collected for comparing different models in a multilingual environment. The collected comments will then be manually labeled into three sentiment categories:

1) Positive: Comments that express favorable opinions, excitement, or satisfaction toward Labubu.

2) Neutral: Comments that are factual, descriptive, or do not express strong opinions in either direction

3) Negative: Comments that express dissatisfaction, criticism, or negative opinions about Labubu or related experiences.

The data collection process utilized a software tool from ExportComments, a platform for extracting comments from social media. The process involved the following steps:

1) Relevant Facebook post links from the Popmart Global and Popmart Thailand pages, specifically related to Labubu, were input into the ExportComments tool.

2) Posts were selected randomly from each fan page based on their relevance to Labubu and the inclusion of specific hashtags such as #LABUBU, #Themonsters, and #ลาบูบู้.

3) The tool exported the comments into an Excel file containing the comment context and other relevant data.

For the Popmart Thailand page:

- 15 posts were selected using hashtags such as #LABUBU, #Themonsters, and #ลาบูบู้.
- Data was collected during the period of July to December 2024.
- A total of 3,283 comments were extracted, categorized as follows:
    - Positive: 880 comments
    - Neutral: 1,829 comments
    - Negative: 574 comments

For the Popmart Global page:

- 15 posts were selected using the same hashtags: #LABUBU, #Themonsters, and #ลาบูบู้.
- Data was collected during the period of June to December 2024.
- A total of 3,517 comments were extracted, categorized as follows:
    - Positive: 1,237 comments
    - Neutral: 1,990 comments
    - Negative: 290 comments

All data collection procedures complied with Facebook’s terms of service and data privacy policies. Table 3.1 shows examples of collected comments with their sentiment labels.

Table 3.1 Representative Sample of Collected Comments with Sentiment Labels

Sentiment	Comment Text
3	Many bots already setup their timers. Once the item opens in milliseconds they are already gone before you proceed to payment page. No chance at all. then resellers skyrocket the price to 3x or more. That's reality.
2	Will this be released on Popmart Ph Flagship Store?
1	A monster become angel
3	Price goes up $110 from last zimomo not sure why but I pass.
2	How to order online
1	I WANT SO MUCHHHH

Note: Sentiment labels: 1 = Positive, 2 = Neutral, 3 = Negative

3.2 Development of Sentiment Analysis Models

3.2.1 Model Selection

This research employs three different transformer-based models for sentiment analysis of Labubu product comments:

3.2.1.1 Model Architecture Selection

For this research, three state-of-the-art transformer models were selected based on their unique characteristics and potential effectiveness for Thai and multilingual sentiment analysis:

3.2.1.2 WangchangBERTa

WangchangBERTa is the first transformer-based language model for Thai language. Trained on 78GB of raw Thai text data which includes social media posts and news articles and has 12 transformer layers. This model was chosen because of its focus on text processing that is important for Thai language due to the need to maintain space that is important for Thai chunk and sentence segmentation. The proposed model has shown better performance compared to other previous baselines and multilingual models for Thai language tasks and therefore is ideal for Thai text processing.

3.2.1.3 RoBERTa

RoBERTa (Robustly Optimized BERT Approach) is an extension and improvement of the original BERT model with several enhancements over it. The model incorporates dynamic masking techniques for a more effective training and increases the batch sizes during the training phase. RoBERTa also simplifies the architecture by eliminating the next sentence prediction task and increases the number of training data and iterations to outperform its predecessor.

3.2.1.4 XLM-R

XLM-R (XLM-RoBERTa) is a recently proposed multilingual masked language model fine-tuned on 2.5TB of CommonCrawl data across 100 languages. The model shows better performance than prior multilingual models such as mBERT and XLM in classification, sequence labeling and question answering. It is also fair in terms of performance between high resource and low resource languages and yields significant improvements in the later. This model was chosen for this study because it has great multilingualism and has been tested for cross-lingual sentiment analysis.

3.3 Implementation and Evaluation Methodology

3.3.1 Data Preparation and Model Training

The implementation uses contemporary deep learning libraries, primarily Python, HuggingFace Transformers for model architectures, pandas for data processing, and scikitlearn for the evaluation metrics. Data preparation begins with label mapping from 1, 2, 3 to zero-based indices 0, 1, 2. The authors used a 80:20 split for the training and testing sets. To make the experiments reproducible, all the random operations are set to use seed value of 42.

Text tokenization uses model-specific tokenizers from HuggingFace for Thai and multi-lingual text processing. Every comment is tokenized with the maximum sequence length of 128 tokens, and it is automatically padded and truncated if needed. These parameters were chosen with the average comment length in mind along with memory efficiency and the architecture of the model.

3.3.2 Hyperparameter Optimization and Training

The training process utilizes Optuna version 3.0 for hyperparameter optimization, implementing the Tree-structured Parzen Estimators (TPE) algorithm. The optimization process explores the following parameter spaces:

Table 3.2 Hyperparameter Search Configuration

Parameter	Search Range	Rationale
Learning Rate	log uniform [1e-5, 5e-5]	Optimal range for fine-tuning transformers
Batch Size	categorica l[8, 16, 32]	Memory optimization
Epochs	Int [3, 10]	Training convergence balance
Weight Decay	uniform [0.01, 0.1]	Regularization control
Warmup Ratio	Uniform [0.1, 0.3]	Training stability

The training implementation uses HuggingFace's Trainer API and is automatically compiled to run on GPU if one is available. The optimization process makes 30 trials, and for the primary optimization metric, F1-score is used because of the possible class imbalance. The training performance is evaluated by the accuracy, precision, recall, and the confusion matrix.

3.3.3 Evaluation Framework

The evaluation framework implements a comprehensive set of metrics to assess model performance:

Table 3.3 Core Evaluation Metrics

Metric	Formula	Purpose
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$	Overall classification accuracy
Precision	$\frac{TP}{TP+FP}$	Prediction precision measurement
Recall	$\frac{TP}{TP+FN}$	Coverage of actual positives
F1-score	$2\times\frac{Precision \times Recall}{Precision+Recall}$	Balanced performance metric

The confusion matrix structure for our three-class sentiment analysis follows:

Table 3.4 Confusion Matrix Organization Predicted

		Positive	Neutral	Negative
	Positive	n_{11}	n_{12}	n_{13}
Actual	Neutral	n_{21}	n_{22}	n_{23}
	Negative	n_{31}	n_{32}	n_{33}

These evaluation metrics and confusion matrix structure provide a comprehensive framework for assessing model performance across different aspects of sentiment classification. The combination of accuracy, precision, recall, and F1-score offer balanced insight into model effectiveness, while the three-class confusion matrix enables detailed analysis of classification patterns and potential areas for improvement. This evaluation approach ensures thorough understanding of each model's strengths and limitations in handling different sentiment categories.

3.4 Frequent Word Analysis and Visualization

To gain deeper insights into the linguistic patterns within customer feedback, we implement word frequency analysis with word cloud visualization for each sentiment category. This analysis complements our quantitative metrics by providing qualitative understanding of how customers express their opinions about Labubu products.

3.4.1 Word Cloud Generation and Analysis

The analysis begins with careful text preprocessing using PyThaiNLP's word tokenization engine ('newmm'), specifically optimized for Thai language processing. Each comment undergoes preprocessing to ensure quality visualization:

- Text cleaning and standardization:
  - Removal of single-character tokens to reduce noise
  - Thai character filtering using regular expressions (-)
  - Empty content handling through null value management

- Visualization configuration:
  - Thai font support (THSarabun.ttf)
  - High-resolution output (800x400 pixels)
  - White background for clarity
  - Word sizing based on frequency distribution

The visualization process generates separate word clouds for positive, neutral, and negative sentiments, allowing direct comparison of linguistic patterns across sentiment categories. Each word cloud is saved as a high-resolution image (300 DPI) for detailed analysis.

3.4.2 Frequency Pattern Analysis

For each sentiment category, the analysis identifies the top 10 most frequent words, providing quantitative metrics for word usage patterns. This approach enables:

- Direct comparison of word frequencies across sentiment categories
- Identification of sentiment-specific vocabulary
- Understanding of common themes in customer feedback
- Recognition of key product-related concerns or praise

By combining visual representation through word clouds with specific frequency counts, this analysis reveals distinctive patterns in how customers express different sentiments. The insights gained help understand customer communication styles and identify recurring themes in product feedback, supp

3.5 Research Tools

The research tools used in this study include:

3.5.1 Development Hardware

MacBook with Apple M2 chip was used as the primary development system with specifications:

- Apple M2 Processor
- Neural Engine for ML acceleration
- Unified Memory Architecture
- macOS operating system

3.5.2 Cloud Computing Platform

Kaggle platform was utilized for:

- Model training and experimentation
- GPU-accelerated computing
- Notebook development environment
- Dataset management and storage

3.5.3 Scraping platform

The data scraping process was facilitated using ExportComments, a platform specialized in extracting comments from social media, ensuring efficient and structured data collection.

CHAPTER 4

MAIN RESULTS AND DISCUSSION

This chapter provides the experimental outcomes and discussion on the three transformer-based models in sentiment classification of Labubu product comments. The evaluation includes both the technical performance and the linguistic pattern analysis, which offer an overall picture of the model efficiency in various language environments. For technical assessment, we compare model performance based on the standard measures such as accuracy, precision, recall, and F1-score and, in addition, provide more comprehensive error analysis through the confusion matrices. The linguistic analysis includes word frequency distributions and patterns, presented in the form of word clouds, to provide further insight into the characteristics of customer communication in sentiment categories.

4.1 Thai Language Dataset Performance Analysis

The performance evaluation of WangchangBERTa, RoBERTa, and XLM-R models on Thai language comments demonstrates varying levels of effectiveness. Each model was tested using the same Thai dataset from Popmart-th to ensure fair comparison.

4.1.1 Model Performance Comparison

Table 4.1 Performance Metrics on Thai Dataset

Model	Accuracy	F1 Score	Precision	Recall
WangchanBERTa	0.8737	0.8751	0.8777	0.8737
RoBERTa	0.7686	0.7701	0.7731	0.7686
XLM-R	0.8402	0.8416	0.8439	0.8402

4.1.2 Analysis of Results

The experimental results reveal several key findings:

- WangchanBERTa Performance: Achieved the highest performance across all metrics (Accuracy: 87.37%, F1 Score: 87.51%) among the three models, demonstrating the advantage of a Thai-specific pre-trained model for Thai sentiment analysis.

- XLM-R Performance: Showed strong performance (Accuracy: 84.02%, F1 Score: 84.16%) despite being a multilingual model, indicating effective cross-lingual capabilities for Thai text analysis.

- RoBERTa Performance: While achieving reasonable results (Accuracy: 76.86%, F1 Score: 77.01%), it showed lower performance compared to the other models, suggesting limitations in handling Thai language nuances.

The enhanced efficiency of WangchangBERTa could be caused by the fact that this model was pre-trained on Thai dataset, so WangchangBERTa possibly better captures Thai language and its context. It is most apparent in the F1-score, which was 87.51% for the proposed method and 77.01% for RoBERTa and 84.16% for XLM-R.

4.1.3 Detailed Performance Analysis of WangchangBERTa

To provide deeper insights into WangchangBERTa's performance on Thai comments, we analyze its confusion matrix results. Figure 4.1 shows the confusion matrix where the rows represent true labels and columns represent predicted labels (0: Negative, 1: Neutral, 2: Positive).

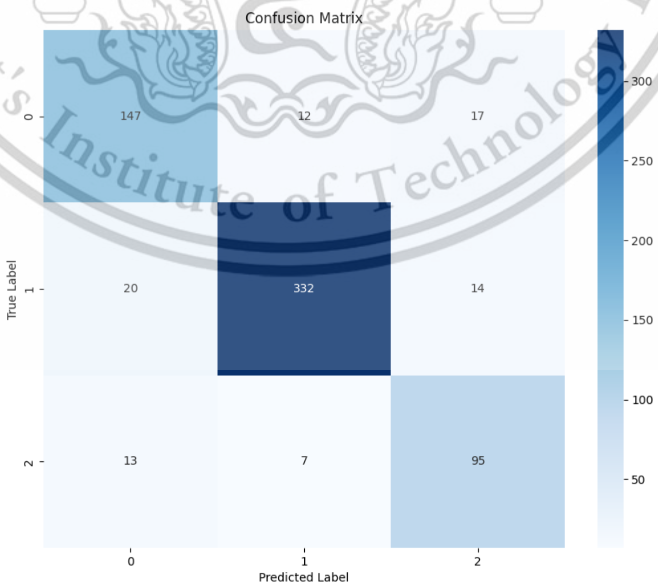

Figure 4.1 Confusion Matrix of WangchanBERTa on Thai Dataset

The confusion matrix reveals detailed classification patterns:

- Neutral Sentiment (1): Shows the strongest performance with 332 correct predictions, indicating high accuracy in identifying neutral comments. Only 20 neutral comments were misclassified as negative and 14 as positive.

- Negative Sentiment (0): Achieved good accuracy with 147 correct predictions, though with some confusion - 12 negative comments were misclassified as neutral and 17 as positive.

- Positive Sentiment (2): Demonstrated reasonable performance with 95 correct predictions, with minimal misclassification - only 13 positive comments were incorrectly labeled as negative and 7 as neutral.

These results indicate that WangchanBERTa is especially efficient in distinguishing neutral sentiments in Thai language comments while keeping high accuracy for all kinds of sentiments. A low percentage of misclassification between the two extreme classes of sentiments (negative to positive and vice versa) shows that the proposed model has a good grasp of sentiment polarity for Thai text.

4.2 Global Dataset Performance Analysis

Following the Thai language evaluation, the models were tested on the global dataset from Popmart-global to assess their performance on multilingual comments.

Table 4.2 Performance Metrics on Global Dataset

Model	Accuracy	F1 Score	Precision	Recall
WangchangBERTa	0.8043	0.8045	0.8088	0.8043
RoBERTa	0.8295	0.8283	0.8296	0.8295
XLM-R	0.8450	0.8465	0.8495	0.8450

4.2.1 Comparative Analysis

The global dataset testing revealed interesting patterns:

- XLM-R Superiority: XLM-R demonstrated the best performance (Accuracy: 84.50%, F1 Score: 84.65%) on the global dataset, leveraging its multilingual pretraining advantage.

- RoBERTa Improvement: RoBERTa showed strong performance (Accuracy: 82.95%, F1 Score: 82.83%), performing notably better on the global dataset compared to its Thai-only results.

- WangchanBERTa Adaptation: While showing good performance (Accuracy: 80.43%, F1 Score: 80.45%), WangchanBERTa's effectiveness decreased compared to its Thai dataset results, reflecting its specialization in Thai language processing.

The findings show how the structural differences of the models and their pretraining strategies affect their ability to work in various language environments. The improvements of XLM-R on the global data set support the design of the model for multilingual sentiment analysis; conversely, the relatively poorer performance of WangchanBERTa is consistent with the fact that it was optimized for Thai.

4.2.2 Cross-Dataset Performance Comparison

Comparing the performance across both datasets reveals important insights about model versatility:

1) Language Specialization Trade-off: WangchanBERTa shows superior performance on Thai content but demonstrates reduced effectiveness on multilingual data.

2) Multilingual Capability: XLM-R maintains consistent high performance across both datasets, demonstrating robust multilingual understanding.

3) Adaptability: RoBERTa shows improved performance on the global dataset, suggesting better handling of diverse language patterns.

4.2.3 Detailed Performance Analysis of XLM-R on Global Dataset

A detailed examination of XLM-R's performance on the global dataset is provided through its confusion matrix analysis. Figure 4.2 presents the confusion matrix where the rows represent true labels and columns represent predicted labels (0: Negative, 1: Neutral, 2: Positive).

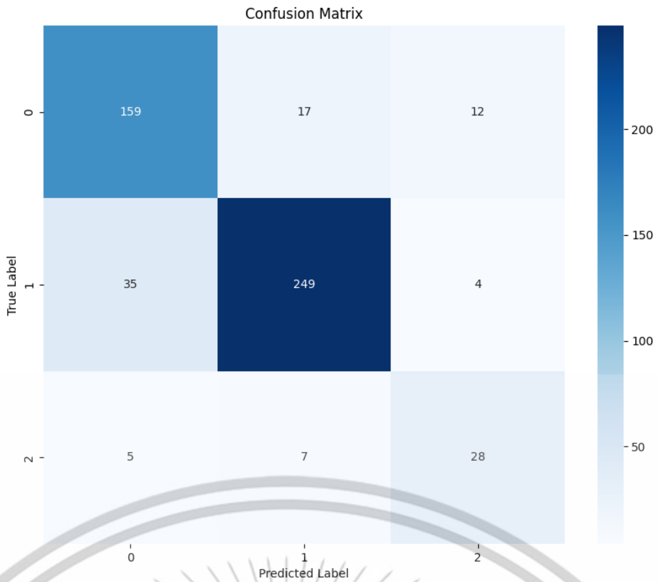

Figure 4.2 Confusion Matrix of XLM-R on Global Dataset

The confusion matrix reveals several important patterns in XLM-R's multilingual sentiment classification:

- Neutral Sentiment (1): Demonstrates strongest classification accuracy with 249 correct predictions out of 288 neutral comments. Only 35 neutral comments were misclassified as negative and 4 as positive, showing excellent discrimination of neutral content across languages.
- Negative Sentiment (0): Shows robust performance with 159 correct predictions out of 188 negative comments. Misclassifications were relatively balanced with 17 comments incorrectly labeled as neutral and 12 as positive.
- Positive Sentiment (2): Exhibits reasonable accuracy with 28 correct predictions, though showing some challenges with only 40 total positive samples. Misclassifications were minimal with 5 positive comments incorrectly labeled as negative and 7 as neutral.

The results show that XLM-R has good performance in multilingual sentiment classification especially in the identification of neutral and negative sentiments. The model reaches high accuracy in detecting the content which is not associated with either positive or negative sentiment, which is 86.5%, and good discrimination between positive and negative sentiments across multiple languages. This performance is particularly important because it demonstrates that across languages, there is little confusion between positive and negative sentiment polarity. The results support the use of XLM-R for multilingual sentiment analysis to show that it can sustain performance across languages and capture the sentiment expression.

4.3 Word Distribution Analysis of Customer Feedback

This section presents a comprehensive analysis of word frequency patterns across Thai and global datasets, revealing distinctive customer communication characteristics and sentiment expressions in different markets. The analysis combines visual representations through word clouds with quantitative frequency analysis to provide actionable insights for customer experience improvement.

4.3.1 Thai Market Analysis

4.3.1.1 Positive Sentiment Comments

Figure 4.3 Word Cloud of Thai Positive Sentiment Comments

Thai positive comments reveal high purchase intent and product enthusiasm:

- Interactive engagement terms dominate, with "kod" (press/click, 76 occurrences) indicating active participation in purchases
- Brand recognition is strong with "boo" (Labubu, 69 occurrences) frequently mentioned
- Patient engagement shown through "ror" (wait, 52 occurrences)

- Emotional connection evident in "narak" (cute, 33 occurrences)

4.3.1.2 Neutral Sentiment Comments

Figure 4.4 Word Cloud of Thai Neutral Sentiment Comments

Purchase process inquiries dominate neutral comments:

- Queue management ("queue", 77 occurrences) emerges as primary concern
- Time-related queries ("rob"/round, 51 occurrences) indicate scheduling interest
- Community support evident through "hiu" (help buy, 48 occurrences)
- System interaction terms ("longtalien"/register, 44 occurrences) show process focus

4.3.1.3 Negative Sentiment Comments

Figure 4.5 Word Cloud of Thai Negative Sentiment Comments

Purchase friction points emerge in negative feedback:

- Timing issues predominate ("rob"/round, 176 occurrences)
- Technical difficulties ("kod"/click, 96 occurrences) cause frustration
- Human factor concerns ("kon"/person, 49 occurrences)
- Distribution issues ("khai"/sell, 43 occurrences)

4.3.2 Global Market Analysis

4.3.2.1 Positive Sentiment Comments

Figure 4.6 Word Cloud of Global Positive Sentiment Comments

International positive sentiment centers on product appeal:

- Strong brand presence ("labubu", 241 occurrences)
- Seasonal collection enthusiasm ("halloween", 183 occurrences)
- Multilingual appreciation ("cute": 100, "": 59 occurrences)
- Theme engagement ("trick": 35, "treat": 36, "costume": 33)

4.3.2.2 Neutral Sentiment Comments

Figure 4.7 Word Cloud of Global Neutral Sentiment Comments

Global neutral comments reflect market diversity:

- Brand search patterns ("labubu": 53, "popmart": 28)
- Digital engagement ("https": 51)
- Regional market activity (Philippines: "ph": 28, Vietnam: "nguyen": 28)
- Platform exploration ("mart": 28, "pop": 31)

4.3.2.3 Negative Sentiment Comments

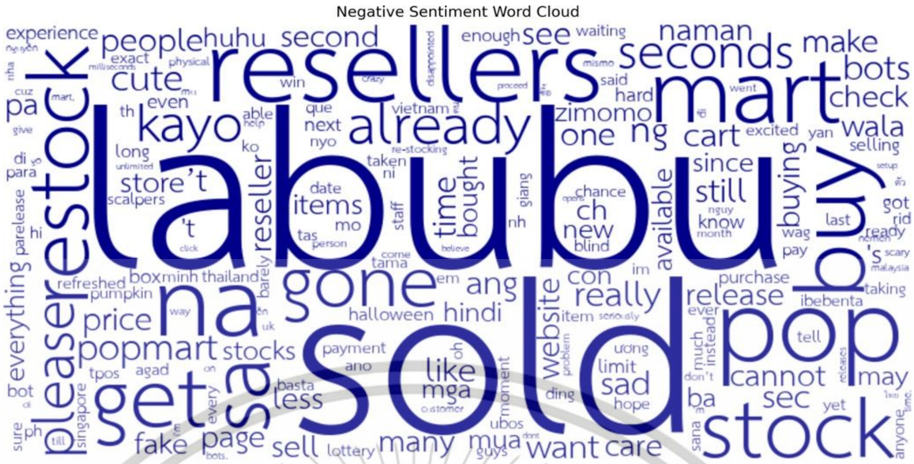

Figure 4.8 Word Cloud of Global Negative Sentiment Comments

International market friction points:

- Availability concerns ("sold": 41, "stock/restock": 40)
- Secondary market issues ("resellers": 33)
- Purchase difficulties ("buy": 26, "get": 21)
- Brand loyalty despite issues ("labubu": 45)

4.3.3 Strategic Implications

This comprehensive analysis reveals several actionable insights:

4.3.3.1 Purchase Process Optimization

- Need for improved queue management systems
- Enhanced technical infrastructure for high-traffic periods
- Better registration process flow

4.3.3.2 Market-Specific Adaptations

- Tailored regional release strategies
- Local language support enhancement
- Market-specific inventory management

4.3.3.3 Customer Experience Enhancement

- Proactive communication about releases
- Reseller policy enforcement
- Cross-market coordination for global releases

These findings indicate that although product appreciation has been shown to be more or less consistent across markets, optimization of the purchase processes and the adaptation of these processes to the specific markets could add much more value to the customer satisfaction. The analysis also emphasises the continuing importance of technical support and sustaining and developing the requisite technical base while simultaneously serving the needs of regional markets.

CHAPTER 5

CONCLUSION AND SUGGESTIONS

5.1 Conclusion

This research focused on sentiment analysis of customer comments about Labubu, a popular product by POP MART, utilizing three advanced transformer-based models: WangchanBERTa, RoBERTa, and XLM-R. Through the analysis of Thai-specific and global datasets, several key insights were revealed, which highlight the performance of the models, the sentiment distribution among consumers, and practical implications for POP MART's marketing strategies.

5.1.1 Performance of Models

5.1.1.1 WangchanBERTa

WangchanBERTa achieved the highest accuracy (87.37%) and F1-score (87.51%) on the Thai-specific dataset. These results underscore its effectiveness in understanding the complexities of the Thai language, such as its lack of word boundaries and unique syntax.

The model's Thai-specific pretraining allowed it to outperform multilingual models like XLM-R and RoBERTa in Thai sentiment classification. However, its performance diminished significantly on the multilingual global dataset, with an F1-score of 80.45%, indicating its limitations in handling diverse languages.

5.1.1.2 RoBERTa

RoBERTa demonstrated moderate performance on the Thai dataset (F1-score: 77.01%) but excelled on the global dataset (F1-score: 82.83%).

Its dynamic masking and large-scale pretraining made it particularly effective for English-language sentiment analysis, but the model struggled with Thai's unique linguistic features.

5.1.1.3 XLM-R

XLM-R showcased consistent performance across both datasets, achieving an F1-score of 84.16% for Thai comments and 84.65% for the global dataset. Its multilingual training on over 100 languages equipped it with robust cross-lingual capabilities.

The model's ability to generalize across languages made it the best choice for multilingual sentiment classification, particularly for datasets with mixed-language content.

5.1.2 Sentiment Patterns and Insights

5.1.2.1 Positive Sentiments

- Positive comments in Thai emphasized emotional attachment to Labubu, with frequent use of terms like "cute" and "adorable." This highlights the product's appeal among younger demographics who value aesthetics and emotional connection.
- Global positive sentiments similarly focused on themes of excitement and brand loyalty, with references to seasonal collections such as Halloween driving engagement.

5.1.2.2 Neutral Sentiments

- Neutral comments largely consisted of procedural inquiries, such as how to purchase, queueing systems, and stock availability. These highlight the importance of clear communication and efficient operational processes in driving consumer satisfaction.
- For global audiences, neutral comments included regional-specific inquiries, such as shipping policies and localized market availability, reflecting diverse consumer needs.

5.1.2.3 Negative Sentiments

- Negative feedback in Thai primarily revolved around technical issues, such as difficulties during online purchases and dissatisfaction with stock shortages.
- Global negative comments highlighted frustrations with reseller markups and limited product availability, reflecting concerns about fairness and accessibility.

5.1.3 Practical Implications

1) The superior performance of WangchanBERTa on Thai data emphasizes the importance of language-specific models for markets like Thailand, where cultural and linguistic nuances significantly influence consumer behavior.

2) XLM-R’s robust multilingual capabilities make it an ideal choice for analyzing comments from diverse language groups, offering actionable insights for POP MART’s global marketing strategies.

3) Sentiment-specific language patterns provide a foundation for tailoring marketing campaigns, enhancing customer engagement, and addressing operational inefficiencies.

5.1.4 Analysis of Model Efficiency Current Study vs. Literature Review

5.1.4.1 WangchanBERTa

The Present Research: Achieved the highest performance on Thai datasets with an F1 score of 87.51%.

Literature Review: Similar to IndoBERT’s performance (F1 score of 88.97%) in detecting toxic comments on Indonesian social media.

Comparison: WangchanBERTa's performance aligns closely with IndoBERT, both being fine-tuned for specific languages (Thai and Indonesian, respectively). WangchanBERTa's 87.51% F1 score is slightly lower, which might be due to dataset or linguistic differences.

Variation (%): IndoBERT performed approximately 1.64% higher than WangchanBERTa.

5.1.4.2 XLM-R

The Present Research: Achieved an F1 score of 84.16% on Thai data and 84.65% on multilingual global data

Literature Review: Performed well in Bengali analysis with validation accuracy of 81% for binary classification and 97% for noisy binary tasks

Comparison: XLM-R’s consistent multilingual capabilities are validated in both cases, though the results from this study are slightly lower in noisy binary settings compared to literature (97%). On Thai and multilingual tasks, the results align closely.

Variation (%): XLM-R in noisy binary tasks (97%) outperformed by approximately 12.35%, but for multilingual tasks, the results are similar.

5.1.4.3 RoBERTa

The Present Research: Achieved an F1 score of 82.83% on global data

Literature Review: Outperformed BERT in several studies, achieving an F1 score of 85% in general sentiment classification

Comparison: The findings from this research are close to the reported literature, demonstrating RoBERTa's efficiency in English sentiment tasks. The difference might arise from dataset complexity or the training setup.

Variation (%): Literature results (85%) are about 2.61% higher than this study.

5.2 Suggestions

Based on the findings, the following recommendations are proposed to enhance sentiment analysis methodologies and address the operational and strategic challenges identified:

5.2.1 Addressing Consumer Concerns

1) Operational Improvements

Upgrade queue management systems and enhance technical infrastructure to handle high-traffic periods, addressing frustrations expressed in negative comments.

Implement clearer communication channels to inform consumers about stock availability and release schedules, reducing neutral sentiment queries.

2) Reseller Regulation

Develop policies to limit reseller activity, such as implementing purchase limits or introducing direct-to-consumer sales channels.

Strengthen monitoring systems to ensure fair pricing practices and improve consumer trust.

5.2.2 Market-Specific Strategies

1) Localized Marketing Campaigns

Leverage insights from positive sentiments to design campaigns that emphasize Labubu's emotional appeal and exclusivity in Thai markets.

For global audiences, focus on seasonal themes and collaborations to drive excitement and engagement.

2) Cultural Sensitivity

Tailor product releases to align with regional preferences, incorporating language-specific promotions and culturally relevant themes.

Enhance local language support for global campaigns to foster inclusivity and accessibility.

REFERENCES

Bhowmick, A., & Jana, A. (2021). Sentiment Analysis for Bengali Using Transformer-Based Models. *Proceedings of the 18th International Conference on Natural Language Processing,* 481-486.

Jiang, R. (2022). Analysis of Trend Culture from the Perspective of Social Currency-Taking the POP MART as an Example. *BCP Business & Management BEMS, 18*, 183-190.

Kruekaew, B., & Kimpan, W. (2022). Multi-Objective Task Scheduling Optimization for Load Balancing in Cloud Computing Environment Using Hybrid Artificial Bee Colony Algorithm with Reinforcement Learning. *IEEE Access, 10,* 17803-17815.

Kumar, B. V. P., & Sadanandam, M. (2024). A Fusion Architecture of BERT and RoBERTa for Enhanced Performance of Sentiment Analysis of Social Media Platforms. *International Journal of Computing and Digital Systems, 15*(1), 51-66.

Melton, C. A., White, B. M., Davis, R. L., Bednarczyk, R. A., & Shaban-Nejad, A. (2022). Fine-Tuned Sentiment Analysis of COVID-19 Vaccine–Related Social Media Data: Comparative Study. *Journal of Medical Internet Research, 24*(10), e40408.

Nabiilah, G. Z., Prasetyo, S. Y., Izdihara, Z. N., & Girsanga, A. S. (2023). BERT Base Model for Toxic Comment Analysis on Indonesian Social Media. *Procedia Computer Science, 216,* 714-721.

Rodríguez-Ibáñez, M., Casánez-Ventura, A., Castejón-Mateos, F., & Cuenca-Jiménez, P.-M. (2023). A Review on Sentiment Analysis from Social Media Platforms. *Expert Systems with Applications, 223,* 119862.

Serna, A., Soroa, A., & Agerri, R. (2021). Applying Deep Learning Techniques for Sentiment Analysis to Assess Sustainable Transport. *Sustainability, 13*(4), 2397.

Singla, S., & Ramachandra, N. (2020). Comparative Analysis of Transformer Based Pre-Trained NLP Models. *International Journal of Computer Sciences and Engineering, 8*(11), 40-44.

Wu, P. (2020). *Marketing Analysis of POP MART*. Chulalongkorn University Theses and Dissertations.

APPENDICES

APPENDIX A

DATA COLLECTION EXAMPLES

A.1 Example of Data Collection Platform

The data collection process used ExportComments, a web-based tool designed for extracting comments from social media platforms. By inputting specific post URLs related to Labubu from Facebook pages, the tool efficiently exported relevant comments into an Excel file for project analysis. Figure A.1 shows the interface of this website, and Figure A.2 illustrates an example of the exported data saved to an Excel file.

Figure A.1 Export Comments Website

content	author	post_url	Post Date	collected_at
Cute	Plush Krush	https://www.facebook.com/popmartglobal/posts/pfbid0w66WtB27Ex6gR	19/11/2024	05/12/2024 14:35
I want you. I'll get you. I'll have you. WAHSHSAHS	Kyla Angeles	https://www.facebook.com/popmartglobal/posts/pfbid0w66WtB27Ex6gR	19/11/2024	05/12/2024 14:35
งืออออ อยากได้	Sukanya Yui	https://www.facebook.com/popmartglobal/posts/pfbid0w66WtB27Ex6gR	19/11/2024	05/12/2024 14:35
A monster becomes an angel lol	리하	https://www.facebook.com/popmartglobal/posts/pfbid0w66WtB27Ex6gR	19/11/2024	05/12/2024 14:35
A monster become angel	Neko Miko	https://www.facebook.com/popmartglobal/posts/pfbid0w66WtB27Ex6gR	19/11/2024	05/12/2024 14:35
I WANT SO MUCHHHH	Lander Palma	https://www.facebook.com/popmartglobal/posts/pfbid0w66WtB27Ex6gR	19/11/2024	05/12/2024 14:35
Yes im ready	Chnisakan Chnsk	https://www.facebook.com/popmartglobal/posts/pfbid0w66WtB27Ex6gR	19/11/2024	05/12/2024 14:35
我要～	Yi Ping Zhang	https://www.facebook.com/popmartglobal/posts/pfbid0w66WtB27Ex6gR	19/11/2024	05/12/2024 14:35
It is a beauty	Suzen Chan	https://www.facebook.com/popmartglobal/posts/pfbid0w66WtB27Ex6gR	19/11/2024	05/12/2024 14:35
Angel Zimomo plays cupid	Divina Gracia Galang	https://www.facebook.com/popmartglobal/posts/pfbid0w66WtB27Ex6gR	19/11/2024	05/12/2024 14:35
I need Zimomo in my life	Ava Lee Weblog	https://www.facebook.com/popmartglobal/posts/pfbid0w66WtB27Ex6gR	19/11/2024	05/12/2024 14:35
Nguyễn Phương Thảo mua cho đủ cặp	Vi Nguy	https://www.facebook.com/popmartglobal/posts/pfbid0w66WtB27Ex6gR	19/11/2024	05/12/2024 14:35
May น่ารักอยู่แต้	Dreamsqweez Mch	https://www.facebook.com/popmartglobal/posts/pfbid0w66WtB27Ex6gR	19/11/2024	05/12/2024 14:35
Why soooo cute	Chufts	https://www.facebook.com/popmartglobal/posts/pfbid0w66WtB27Ex6gR	19/11/2024	05/12/2024 14:35
มาจ้ อิแม่ก็ตายไปเลยค์ Zem Zem Kampanat	Piyawan Rueangchirawat	https://www.facebook.com/popmartglobal/posts/pfbid0w66WtB27Ex6gR	19/11/2024	05/12/2024 14:35
Here maganda! Carouzel S. Tienzo	Caroline Baca	https://www.facebook.com/popmartglobal/posts/pfbid0w66WtB27Ex6gR	19/11/2024	05/12/2024 14:35
KimGayle may bago ule!!!!!!!	Wendy Chu	https://www.facebook.com/popmartglobal/posts/pfbid0w66WtB27Ex6gR	19/11/2024	05/12/2024 14:35
มาแล้ว Narumon Pankamnerd	Switta CH	https://www.facebook.com/popmartglobal/posts/pfbid0w66WtB27Ex6gR	19/11/2024	05/12/2024 14:35
Tram Anh coá mào khum thì mê	Khánh Phạm	https://www.facebook.com/popmartglobal/posts/pfbid0w66WtB27Ex6gR	19/11/2024	05/12/2024 14:35
Hoài An xinh k	Luong Giang	https://www.facebook.com/popmartglobal/posts/pfbid0w66WtB27Ex6gR	19/11/2024	05/12/2024 14:35
Tanawat Rueangchatri สนมั้ย วันที่ 25 นี้	Jannyy Chanyanuch	https://www.facebook.com/popmartglobal/posts/pfbid0w66WtB27Ex6gR	19/11/2024	05/12/2024 14:35
Super cuteee	Sakura Chan	https://www.facebook.com/popmartglobal/posts/pfbid0w66WtB27Ex6gR	19/11/2024	05/12/2024 14:35
Jezzel Ignacio OMG	Angel Jose-Opiniano	https://www.facebook.com/popmartglobal/posts/pfbid0w66WtB27Ex6gR	19/11/2024	05/12/2024 14:35
จะเอาๆๆๆๆ	Panisara Panyacharoensri	https://www.facebook.com/popmartglobal/posts/pfbid0w66WtB27Ex6gR	19/11/2024	05/12/2024 14:35
CUTE	Noeyz Nattharin	https://www.facebook.com/popmartglobal/posts/pfbid0w66WtB27Ex6gR	19/11/2024	05/12/2024 14:35

Figure A.2 Exported Excel data file collected from Exportcomments.com

A.2 Importing Data into Google Colab

After collecting the Excel files from ExportComments and labeling the data into sentiment categories (Positive, Neutral, Negative), the datasets were imported into Google Colab for further analysis and processing. The 2 primary uses were as follows:

A2.1 Training Sentiment Analysis Models

The dataset was converted to CSV format and imported into Google Colab using Python's pandas library see in figure A.3 Sentiment labels were numerically encoded for training. The data was then used to train models such as WangchanBERTa and XLM-R, utilizing libraries like transformers and sklearn. To ensure reproducibility, a random seed was set, GPU availability was verified for faster processing, and the dataset was split into an 80:20 ratio for training and testing.

```python
import pandas as pd
import numpy as np
from datasets import Dataset, DatasetDict
from transformers import (
    AutoTokenizer,
    AutoModelForSequenceClassification,
    TrainingArguments,
    Trainer
)
from sklearn.metrics import accuracy_score, precision_recall_fscore_support, confusion_matrix
import torch
from sklearn.model_selection import train_test_split
import optuna
import os
import json
import matplotlib.pyplot as plt
import seaborn as sns

# ปิด wandb
os.environ["WANDB_DISABLED"] = "true"

# ตั้งค่า seed และ device
SEED = 42
np.random.seed(SEED)
torch.manual_seed(SEED)

device = torch.device('cuda' if torch.cuda.is_available() else 'cpu')
print(f"Using device: {device}")

# โหลดข้อมูล
df = pd.read_csv('/kaggle/input/thai-labubu/thai_page_labubu_comments.csv')
df['label'] = df['label'].map({1: 0, 2: 1, 3: 2})
```

Figure A.3 Python setup for sentiment analysis

A2.2 Creating Word Cloud Visualizations Figure

The dataset was loaded into Google Colab and processed using the pandas and pythainlp libraries, as shown in Figure A.4 Sentiment labels were mapped to descriptive categories (e.g., Positive, Neutral, Negative) to facilitate visualization. Stop words were removed using the pythainlp corpus, and word frequencies were analyzed. The WordCloud library was used to generate visual representations of the most frequently used words for each sentiment category, providing insights into consumer sentiment trends visually.

```
import pandas as pd
from wordcloud import WordCloud
import matplotlib.pyplot as plt
from pythainlp import word_tokenize
from pythainlp.corpus import thai_stopwords
from collections import Counter

# Load data
df = pd.read_csv('thai_page_labubu_comments.csv')

# Map sentiment labels
sentiment_map = {1: 'Positive', 2: 'Neutral', 3: 'Negative'}
df['sentiment'] = df['label'].map(sentiment_map)
```

Figure A.4 Mapping sentiment labels and preparing data for word cloud visualization

APPENDIX B

MODEL PERFORMANCE TESTING EXAMPLES

B.1 Example of Target Class Assignment

```python
# แบ่ง train/test
train_df, test_df = train_test_split(
    df,
    test_size=0.2,
    random_state=SEED,
    stratify=df['label']
)

# แปลงเป็น Hugging Face datasets
dataset_dict = DatasetDict({
    'train': Dataset.from_pandas(train_df),
    'test': Dataset.from_pandas(test_df)
})

# โหลด tokenizer
tokenizer = AutoTokenizer.from_pretrained("airesearch/wangchanberta-base-att-spm-uncased")

def tokenize_function(examples):
    return tokenizer(
        examples['content'],
        padding='max_length',
        truncation=True,
        max_length=128,
        return_tensors=None
    )
```

Figure B.1 Initial process of assigning sentiment labels to the dataset

Figure B.1 Figures illustrate how target classes are assigned for sentiment analysis. The dataset is divided into three categories: Positive, Neutral, and Negative. Each comment is carefully reviewed and labeled based on specific sentiment criteria.

```python
# tokenize datasets
tokenized_datasets = dataset_dict.map(
    tokenize_function,
    batched=True,
    remove_columns=dataset_dict["train"].column_names
)

# เพิ่ม labels กลับเข้าไป
tokenized_datasets["train"] = tokenized_datasets["train"].add_column("labels", dataset_dict["train"]["label"])
tokenized_datasets["test"] = tokenized_datasets["test"].add_column("labels", dataset_dict["test"]["label"])
```

Figure B.2 Examples of labeled data categorized as Positive, Neutral, or Negative

Figure B.2 provides examples of labeled data and their respective classes. These labels serve as the correct answers (ground truth) for training and testing the model.

B.2 Example of Model Training, Testing, and Hyperparameter Optimization

```python
# tokenize datasets
tokenized_datasets = dataset_dict.map(
    tokenize_function,
    batched=True,
    remove_columns=dataset_dict["train"].column_names
)

# เพิ่ม labels กลับเข้าไป
tokenized_datasets["train"] = tokenized_datasets["train"].add_column("labels", dataset_dict["train"]["label"])
tokenized_datasets["test"] = tokenized_datasets["test"].add_column("labels", dataset_dict["test"]["label"])
```

Figure B.3 Data preparation process, including training and testing splits and tokenization

Figure B.3 demonstrates the process of preparing the data for training and testing. First, the dataset is split into 2 parts: one for training the model and the other for testing its performance. The data is then converted into a format suitable for Hugging Face, a popular NLP framework. Next, the WangchanBERTa tokenizer is used to break Thai text into tokens for the model to process.

```python
def objective(trial):
    # กำหนด hyperparameters ที่จะ tune
    hyperparameters = {
        'learning_rate': trial.suggest_float('learning_rate', 1e-5, 5e-5, log=True),
        'per_device_train_batch_size': trial.suggest_categorical('per_device_train_batch_size', [8, 16, 32]),
        'num_train_epochs': trial.suggest_int('num_train_epochs', 3, 10),
        'weight_decay': trial.suggest_float('weight_decay', 0.01, 0.1),
        'warmup_ratio': trial.suggest_float('warmup_ratio', 0.1, 0.3),
    }

    training_args = TrainingArguments(
        output_dir=f"./results/trial-{trial.number}",
        evaluation_strategy="epoch",
        save_strategy="no",  # ไม่เซฟระหว่าง trial
        metric_for_best_model="f1",
        report_to="none",
        **hyperparameters
    )
```

Figure B.4 Hyperparameter optimization using Optuna to find the best settings

Figure B.4 highlights the process of finding the best settings for the model using a tool called Optuna. The training pipeline experiments with different settings such as learning rate (how fast the model learns), batch size (number of samples processed at once), number of training rounds (epochs), and more. The best settings are saved to ensure the model is optimized for the task.

B.3 Example of Model Saving and Best Hyperparameter Selection

```
# ค้นหา hyperparameters ที่ดีที่สุด
print("กำลังค้นหา hyperparameters ที่ดีที่สุด...")
study = optuna.create_study(direction='maximize')
study.optimize(objective, n_trials=30)

best_params = study.best_params
print("\nBest hyperparameters:", best_params)
print("Best F1 score:", study.best_value)

# บันทึก best parameters
with open('best_params.json', 'w') as f:
    json.dump(best_params, f)

print("\nกำลังเทรนโมเดลด้วย hyperparameters ที่ดีที่สุด...")
# เทรนโมเดลใหม่ด้วย hyperparameters ที่ดีที่สุด
model = AutoModelForSequenceClassification.from_pretrained(
    "airesearch/wangchanberta-base-att-spm-uncased",
    num_labels=3
).to(device)

training_args = TrainingArguments(
    output_dir="./final_model",
    evaluation_strategy="epoch",
    save_strategy="no",  # เซฟแค่โมเดลสุดท้าย
    metric_for_best_model="f1",
    report_to="none",
    **best_params
)
```

Figure B.5 Saving the trained model and retraining it with optimal parameters

Figure B.5 illustrates the process of saving the trained model along with its optimal settings for future applications. Once the best settings are identified, the model is retrained using these parameters to achieve peak performance. This ensures the results are reproducible and simplifies the deployment of the model in practical scenarios.

APPENDIX C

EXAMPLES OF MODEL EVALUATION RESULTS

C.1 Code Example for Model Evaluation Using F1-Score

Figures C.1 to C.3 illustrate the code and results used for evaluating the model's performance. The evaluation process includes computing accuracy, precision, recall, and F1-score metrics using the scikit-learn library. The confusion matrix provides a detailed view of model predictions compared to actual labels.

```
def compute_metrics(eval_pred):
    predictions, labels = eval_pred
    predictions = np.argmax(predictions, axis=1)

    # คำนวณ metrics ด้วย scikit-learn
    acc = accuracy_score(labels, predictions)
    precision, recall, f1, _ = precision_recall_fscore_support(labels, predictions, average='weighted')

    # สร้าง confusion matrix
    cm = confusion_matrix(labels, predictions)

    # Plot confusion matrix
    plt.figure(figsize=(10,8))
    sns.heatmap(cm, annot=True, fmt='d', cmap='Blues')
    plt.title('Confusion Matrix')
    plt.ylabel('True Label')
    plt.xlabel('Predicted Label')
    plt.savefig('confusion_matrix.png')
    plt.close()

    return {
        'accuracy': acc,
        'f1': f1,
        'precision': precision,
        'recall': recall
    }
```

Figure C.12 Code for computing evaluation metrics and generating a confusion matrix.

Epoch	Training Loss	Validation Loss	Accuracy	F1	Precision	Recall
1	No log	0.430917	0.828006	0.830989	0.835997	0.828006
2	No log	0.426503	0.841705	0.848592	0.868165	0.841705
3	No log	0.456734	0.855403	0.854932	0.859804	0.855403
4	0.534300	0.476067	0.866058	0.867987	0.872483	0.866058
5	0.534300	0.511858	0.873668	0.875126	0.877668	0.873668

Figure C.2 Performance metrics table for training and validation across epochs.

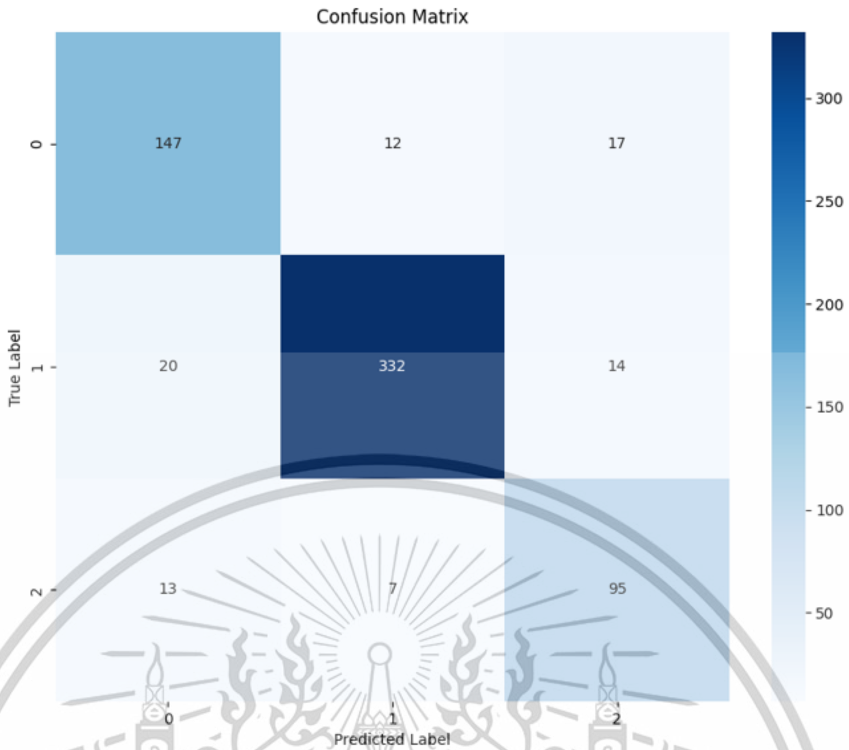

Figure C.3 Confusion matrix showing model prediction results for each class.

C.2 Code Example for Summarizing Model Performance on Test Dataset

Figures C.4 and C.5 illustrate how to summarize model evaluation results on the test dataset. This section highlights the importance of presenting key metrics such as accuracy, F1-score, precision, and recall in a clear and concise format for interpretation.

```
print("\nผลการทดสอบบน Test Set:")
print(f"Accuracy: {results['eval_accuracy']:.4f}")
print(f"F1 Score: {results['eval_f1']:.4f}")
print(f"Precision: {results['eval_precision']:.4f}")
print(f"Recall: {results['eval_recall']:.4f}")
```

Figure C.4 Code for summarizing evaluation metrics on the test dataset.

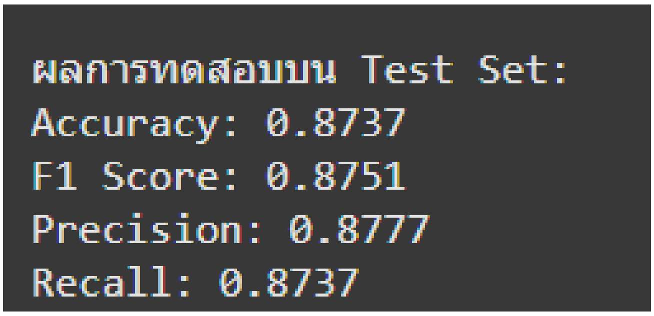

Figure C.5 Summarized evaluation results displayed as accuracy, F1-score, precision, and recall

AUTHOR BIOGRAPHY

Name: Kittinun Poonsawat

Date of Birth: August 7, 1998

Current Address: 9/129 Mantana Ramintra Wongweang, Khan Na Yao Sub-district, Khan Na Yao District, Bangkok Province, Thailand

Education:

2023 Bachelor of Engineering in Chemical Engineering (International Program), King Mongkut's Institute of Technology Ladkrabang, GPA: 3.02