

A Hybrid of Syntactic Structure and Contextual Semantics for Enhanced Product Review Sentiment Analysis

Damrong Sattayawaksakul¹ and Pimpa Cheewaparakobkit^{2*}

¹Faculty of Business Administration, Asia-Pacific International University, Saraburi, Thailand

²Faculty of Information Technology, Asia-Pacific International University, Saraburi, Thailand

E-mail: damsat@apiu.edu, pimpa@apiu.edu*

Received: June 8, 2025 / Revised: January 7, 2026 / Accepted: January 30, 2026

Abstract—With the rapid growth of e-commerce platforms, analyzing customer reviews has become essential for understanding product perception and supporting data-driven decision-making. Traditional sentiment analysis methods often rely solely on sequential text representations, which may overlook the syntactic structures and contextual nuances present in customer feedback. This study aims to develop an interpretable hybrid framework that effectively captures both semantic meaning and syntactic dependencies to enhance sentiment prediction accuracy. The proposed model combines contextual embeddings from SentBERT with syntactic features extracted via Graph Feature Fusion. The process involves preprocessing customer reviews, extracting aspect-based semantic representations using SentBERT, and constructing syntactic graphs via dependency parsing. These features are merged in a fusion layer and refined using supervised contrastive learning to improve class separability in the sentiment space. SHAP Explainable AI is integrated to provide human-interpretable explanations for the sentiment predictions. The hybrid model achieves a sentiment classification accuracy of 97%, outperforming baseline methods including BERT (93%), GCN (88%), and classical machine learning algorithms. These findings highlight the effectiveness of integrating syntactic and contextual features in sentiment analysis. The framework can be applied to real-world e-commerce platforms to enhance automated review analysis, improve customer service insights, and support product development strategies.

Index Terms—Sentiment Analysis, Graph Feature Fusion, Product Review Analysis, Contextual Embedding

I. INTRODUCTION

In the modern digital landscape, e-commerce has revolutionized consumer behavior, with online product reviews becoming a significant factor in purchase decisions [1]. These reviews serve as a valuable source of information for both consumers and businesses, providing insights into customer satisfaction and product quality. However, the vast quantity of reviews on platforms like Amazon, Lazada, Twitter, and Shopee presents a challenge for extracting meaningful insights. The complexity of natural language, coupled with the sheer volume of user-generated content, necessitates the adoption of automated techniques such as sentiment analysis and review summarization powered by Artificial Intelligence (AI) [2], [3]. Sentiment analysis, a subfield of Natural Language Processing (NLP), has gained significant traction for its ability to automatically analyze, interpret, and categorize large volumes of text data [4]. Businesses increasingly rely on sentiment analysis to gauge customer opinions and facilitate data-driven decision-making. Classical machine learning algorithms, such as Naïve Bayes [5] and Support Vector Machine [6], excel at polarity detection but struggle with context-aware tasks. For instance, Karina and Sriani achieved 73% accuracy in classifying skincare reviews using Naïve Bayes, yet their model failed to capture mixed sentiments or attribute feedback to specific product aspects (e.g., packaging, texture) [7]. Similarly, SVM-based studies highlight trade-offs between precision and scalability, particularly when handling slang, abbreviations, or non-linear sentiment patterns [8]. These limitations are exacerbated in e-commerce, where reviews often contain conflicting opinions (e.g., “Great camera but poor battery life”) and domain-specific terminology. Recent breakthroughs in deep learning, particularly transformer-based models like BERT [9] and GPT

[3], have significantly advanced the state-of-the-art by leveraging self-attention mechanisms to capture long-range dependencies in text [10]. Nonetheless, these models often focus on holistic sentiment classification and may neglect syntactic structures and aspect-specific nuances, limiting interpretability and granular sentiment detection

Recent research has explored the integration of Graph Convolutional Networks (GCNs) into sentiment analysis frameworks. GCNs are adept at modeling complex relationships in data by representing text as graphs, where nodes correspond to words and edges capture syntactic dependencies. For instance, Li et al. proposed an edge-enhanced Graph Convolutional Network (GCN) model. This model enhances aspect-based sentiment classification by preserving feature integrity in syntactic graphs. While it improves syntactic structure preservation over traditional methods, its higher computational complexity and reliance on accurate syntactic parsing make it susceptible to noisy input. [11].

In another study by Hu and Li proposed a bi-channel GCN model that integrates phrase structure trees and hierarchical phrase matrices to enhance syntactic structural information. Although this approach improves sentiment classification accuracy, it introduces additional complexity to graph construction and may pose challenges for real-time applications [12].

To address these limitations, this study proposes a hybrid sentiment analysis framework that integrates semantic embeddings with syntactic structures to enhance both performance and interpretability. The key contributions of this research are as follows:

1) Context-Enriched Aspect Term Extraction: We proposed SentBERT for fine-grained identification of product-specific aspects (e.g., camera, battery life), allowing the model to distinguish sentiments directed at individual features.

2) Graph-Based Syntactic Modeling: A Graph Feature Fusion (GFF) module constructed via dependency parsing captures grammatical dependencies, facilitating a structural understanding of how sentiment-bearing words relate to specific product aspects.

3) Unified Feature Representation via Fusion Layer: Semantic and syntactic features are merged into a unified vector space, ensuring a more comprehensive sentiment representation.

4) Supervised Contrastive Learning for Robust Classification: This strategy enhances class separation by minimizing intra-class variance and maximizing inter-class distance, leading to a more discriminative feature space.

5) Explainable AI with SHAP: The framework

incorporates SHAP to generate interpretable visual explanations, highlighting which words contribute most significantly to sentiment predictions, enhancing transparency and user trust.

The remainder of this paper is structured as follows. Section II reviews related work on sentiment analysis methodologies. Section III details the proposed hybrid framework. Section IV presents the experimental results and performance evaluation of the proposed method. Finally, Section V offers the conclusions of this study.

II. LITERATURE REVIEW

The analysis of customer reviews in e-commerce has been extensively studied, leading to various methodologies for sentiment analysis and review summarization. Early approaches focused on sentiment analysis, aiming to classify reviews as positive, negative, or neutral. For instance, Peal et al., proposed a method to summarize customer reviews by extracting pros and cons, providing a concise overview of product feedback [13]. However, this approach primarily relied on predefined sentiment lexicons, which may not capture the nuanced expressions in user reviews. Building upon sentiment analysis, researchers have explored topic modeling to enhance review summarization. Kumar et al. employed topic modeling alongside sentiment analysis to summarize online customer reviews, enabling a more structured understanding of user opinions [14]. While effective in identifying prevalent themes, this method may struggle with the dynamic and evolving nature of language used in reviews.

The integration of deep learning techniques has further advanced sentiment analysis in e-commerce. Yang *et al.* introduced a model combining a sentiment lexicon with a Convolutional Neural Network (CNN) and an attention-based Bidirectional Gated Recurrent Unit (BiGRU) to analyze Chinese e-commerce product reviews [15]. This approach improved the accuracy of sentiment classification by capturing complex linguistic patterns. However, deep learning models often require large labeled datasets and are computationally intensive.

Classical machine learning algorithms, such as Naive Bayes (NB), Support Vector Machines (SVM), Decision Tree (DT), and Logistic Regression (LR), have been widely applied to sentiment classification tasks. Suryani et al. demonstrated the use of Naive Bayes for sentiment analysis on Twitter data related to electric cars [15]. Similarly, Sulthana et al. applied SVM to classify movie reviews [16]. However, these methods often rely on bag-of-words representations, which ignore word order and contextual dependencies. While offering computational

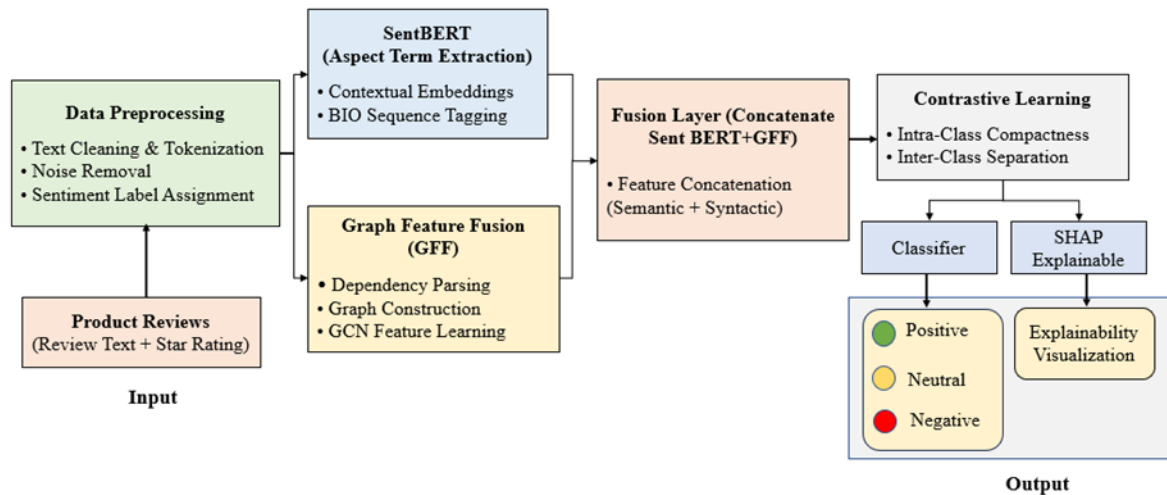


Fig. 1. The overview of the proposed method

While efficient and easy to implement, these models struggle with nuanced language, such as sarcasm, irony, and negation, which are prevalent in online reviews [17]. In the e-commerce domain, where reviews often contain complex product-specific jargon and multiple sentiments towards different aspects (e.g., “The battery life is great, but the camera is disappointing”), these methods often fall short. Furthermore, these models treat all words equally, failing to capture the relative importance of different terms in expressing sentiment. The challenge lies in adapting these techniques to understand the context and intricacies of online reviews.

Aspect-based sentiment analysis has also been a focal point in recent research. Wahyudi and Kusumaningrum investigated the performance of sentiment analysis on e-commerce user reviews using the Latent Dirichlet Allocation (LDA) algorithm, aiming to determine sentiment across different product aspects. [18]. While this approach allows for a more granular analysis, it may struggle to accurately identify aspects in reviews that contain implicit or ambiguous expressions.

Advancements in deep learning, particularly transformer-based models such as BERT [9] and its variants [3], have shown promising results in capturing contextual information to improve sentiment analysis. BERT’s bidirectional contextual understanding, achieved through self-attention mechanisms, allows it to model long-range dependencies in text. However, directly applying BERT to long documents, such as lengthy product reviews, can be computationally expensive due to the quadratic complexity of the attention mechanism.

Furthermore, while BERT captures contextual nuances, it does not inherently focus on specific product features, resulting in a less granular understanding of customer opinions. Recent studies have turned to Graph Convolutional Networks

(GCNs) to incorporate syntactic structure into sentiment analysis. GCNs treat sentences as graphs, where words are nodes and syntactic dependencies form the edges, allowing the model to propagate sentiment through syntactic paths.

III. PROPOSED METHOD

This research presents a hybrid framework for sentiment analysis of mobile phone product reviews that integrates semantic and structural information to improve classification accuracy and model interpretability. The proposed method begins with preprocessing raw review texts, followed by dual feature-extraction paths: semantic embeddings via SentBERT for aspect-term understanding and syntactic relationships via Graph Feature Fusion (GFF) using dependency parsing. These two complementary feature sets are then combined in a Fusion Layer to form a unified representation that captures both context and structure. To enhance the distinctiveness of sentiment categories—positive, neutral, and negative, a supervised contrastive learning strategy is applied, refining the feature space and improving classification robustness. A sentiment classifier then predicts the emotional polarity of the review, while a SHAP Explainable module generates interpretable explanations by highlighting which words most influenced the prediction. This architecture supports both high-performance sentiment analysis and human-centered explainability, making it well-suited for applications in e-commerce and customer feedback systems. The detailed workflow of the proposed method is illustrated in Fig. 1.

A. Data Preprocessing

The Data Preprocessing module transforms raw customer reviews into a clean, structured format for analysis. First, text in the “Comment” field undergoes standard NLP steps: Lowercasing, removal of

unnecessary punctuation, cleaning of HTML tags, special characters, and emojis, whitespace normalization, and optional spelling correction. Next, the “Stars” field is converted into sentiment labels—Positive (4-5 stars), Neutral (3 stars), and Negative (1-2 stars)—to serve as ground truth for training and evaluating the sentiment analysis model.

B. SentBERT (Aspect Term Extraction)

The SentBERT module, as shown in Fig. 2, plays a crucial role in the proposed framework by performing Aspect Term Extraction (ATE), which involves identifying specific entities or product features mentioned in customer reviews. In mobile phone reviews, common aspect terms include “battery life,”

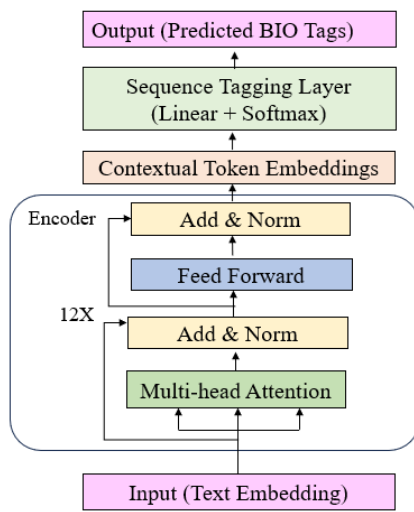


Fig. 2. The SentBERT architecture

“camera,” “price,” “display,” “performance,” and “design.” Accurate identification of these terms is essential for fine-grained sentiment analysis. SentBERT, a sentence-level variant of the BERT (Bidirectional Encoder Representations from Transformers) architecture [9], is adopted and fine-tuned for this sequence labeling task. Built on transformer encoder layers, SentBERT leverages multi-head self-attention, Add & Norm, and feed-forward sublayers to produce contextualized embeddings for each token in a sentence. These embeddings capture both semantic meaning and the relationships among words, making them ideal for identifying aspect terms in unstructured text. The self-attention mechanism, crucial for enabling the model to focus on important words within a sentence, computes attention weights using the following formula (1):

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

where Q , K , and V represent the query, key, and value matrices, respectively, and d_k is the dimensionality of the key vectors.

1) Problem Formulation as Sequence Tagging

To perform aspect term extraction using SentBERT, the input review text is prepared for token-level classification. This involves defining a structured tagging scheme and fine-tuning the model for sequence labeling. Given a customer review sentence $S=[w_1, w_2, \dots, w_n]$, where w_i represents the i -th token (word or sub-word unit), the objective of ATE is to assign a label l_i to each token w_i . This label indicates whether the token is part of an aspect term and, if so, its position within that term. We adopt the widely recognized BIO (Beginning, Inside, Outside) tagging scheme [19] for this purpose:

- **B-ASP**: Denotes the **B**eginning of an **ASP**ect term.

- **I-ASP**: Denotes an **I**nside token of an **ASP**ect term.

- **O**: Denotes an **O**utside token, meaning it is not part of an aspect term.

For instance, consider the review sentence: “The **battery life** is excellent, but the **camera** is disappointing”. The corresponding BIO labels would be: O, B-ASP, I-ASP, O, O, O, O, B-ASP, O, O.

2) SentBERT as a Sequence Tagger

The BERT model is used as a sequence tagger because it generates rich, contextualized word embeddings by considering the entire input sequence. This capability is crucial for accurately identifying aspect terms, which often depend on surrounding words for their meaning. The operational flow of the SentBERT-based sequence tagger is as follows:

• Input Representation

Each review sentence is broken into words (or word parts) using BERT’s WordPiece tokenizer. Special markers, (CLS) at the start and (SEP) at the end, are added to help SentBERT understand the text. These words and markers are then converted into numerical embeddings that combine information about the words themselves, their segment (which indicates whether they belong to different sentences), and their order in the sentence.

• Contextual Embedding Generation

SentBERT processes the input words by feeding them into its multi-layer transformer encoder. This encoder uses a multi-head self-attention mechanism, enabling each word to “look” at all other words in the sentence to understand their deep contextual relationships. The output is a sequence of contextualized hidden states, where each state is a rich representation of its corresponding input word within the context of the entire sequence.

• Sequence Tagging Layer

After SentBERT processes the word contexts, a special “tagging” layer is added on top. This layer takes SentBERT’s output for each word and essentially “votes” on what kind of tag it should get (like “beginning of aspect (B-ASP),” “inside aspect

(I-ASP),” or “other O “). It calculates a probability for each possible tag, and the tag with the highest probability is chosen as the final label for that word. A linear layer followed by softmax is added to project each hidden state into label scores. The formula is shown in (2).

$$P(l_j|h_i) = \text{softmax}(Wh_i + b) \quad (2)$$

The variable W represents the weight matrix, a set of adjustable values that the model learns to combine information effectively. b is the bias term used to fine-tune the output. h_i refers to the hidden state (embedding) of token w_i , which represents the rich, contextual understanding of the word w_i that SentBERT has generated. $P(l_j|h_i)$ signifies the probability of assigning label l_j to token w_i , given its hidden state h_i . In simpler terms, it indicates how likely it is that this specific tag (l_j) should be given to this particular word (w_i), based on what the model knows about it (h_i). Finally, the label with the highest probability is selected for each token.

C. Graph Feature Fusion

The Graph Feature Fusion module plays a critical role in transforming each review sentence into a structured representation that captures the grammatical relationships between words. Once aspect terms are extracted, it becomes essential to understand how these terms relate to sentiment-bearing expressions within the sentence. To achieve this, the system utilizes dependency parsing, a syntactic analysis technique that identifies the grammatical structure of a sentence by determining which words are connected and the nature of their relationships—such as subject, object, or modifier.

In this process, each word in the sentence is treated as a node, and the syntactic dependencies between words are represented as directed edges connecting those nodes. For example, in the sentence “The battery life is amazing,” the parser may determine that “battery” modifies “life,” “life” is the subject of the verb “is,” and “amazing” is the complement of “is.” These relationships form a sentence-level dependency graph that preserves both the grammatical and semantic structure of the input.

The resulting sentence-level graph is fed into a Graph Convolutional Network (GCN), enabling the model to analyze sentiment propagation across the sentence structure. To enhance learning, each GCN layer applies the ReLU (Rectified Linear Unit) activation function, defined as $\text{ReLU}(x) = \max(0, x)$, which introduces non-linearity and helps the network focus on meaningful features [20].

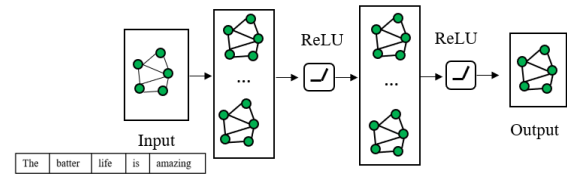


Fig. 3. Graph Feature Fusion

As shown in Fig. 3, this module visually maps syntactic connections between words, forming a graph that enables the model to relate opinions (e.g., “amazing”) to aspects (e.g., “battery life”) more effectively. This approach strengthens aspect-based sentiment classification by preserving grammatical context.

D. Fusion Layer

The Fusion Layer Module (Concatenate SentBERT+ GFF) serves as a critical integration point within the system, combining the rich semantic understanding from SentBERT with the structural grammatical insights from the Graph Feature Fusion (GFF) module. This module operates by concatenating the output representations from both SentBERT, which is responsible for aspect term extraction, and the Graph Feature Fusion module. The concatenation process creates a more comprehensive and robust feature vector for each data point or word. This unified representation is designed to capture both the deep contextual meaning of the text and the explicit relationships between words, as identified by the dependency parser. By bringing together these two distinct yet complementary types of information, the Fusion Layer provides richer input to subsequent stages, such as Contrastive Learning, enabling the model to achieve a more nuanced and accurate understanding of sentiment within the context of specific aspects of product reviews.

To illustrate the effect of the Fusion Layer, consider the review sentence “The battery life is excellent, but the camera is disappointing.” SentBERT generates contextual semantic embeddings that capture the positive polarity of “excellent” and the negative polarity of “disappointing,” while the GFF module encodes syntactic dependencies that link these sentiment-bearing words to their corresponding aspects (“battery life” and “camera”). After concatenation, the resulting fused representation jointly preserves sentiment polarity, aspect context, and grammatical structure within a single feature vector, enabling downstream modules to distinguish aspect-level sentiments more effectively than using semantic or syntactic features alone.

E. Contrastive Learning

The Contrastive Learning Module is designed to improve the model's ability to learn more discriminative and robust feature representations for sentiment classification. After extracting semantic features using SentBERT and structural features via Graph Feature Fusion (GFF), the concatenated embeddings are passed into this module, which applies a supervised contrastive loss. The objective is to pull embeddings of reviews with the same sentiment label closer together in the feature space, while pushing those with different sentiment labels apart. This technique enhances the separability of sentiment classes—positive, neutral, and negative—especially when reviews contain ambiguous or conflicting expressions. By reinforcing class-specific similarity and inter-class distinction, the Contrastive Learning Module helps the classifier make more accurate predictions and generalize better to unseen data.

The Supervised Contrastive Loss [21] is used to train the model by comparing the similarity of samples in a batch of size N . The overall loss for the batch is calculated by summing the individual losses for each sample as defined in (3).

$$L_{batch} = \sum_{i=1}^N L_i \quad (3)$$

Here, L_i represents the loss for the i -th sample, also called the anchor sample. The individual loss L_i is calculated as defined in (4).

$$L_i = \sum_{j \in P(i)} -\log \left[\frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{k \in A(i)} \exp(\text{sim}(z_i, z_k)/\tau)} \right] \quad (4)$$

Where z_i represent the embedding (vector representation) of the i -th sample. y_i is its ground-truth sentiment label of the i -th sample (e.g., positive, neutral, negative). The function $\text{sim}(u, v)$ measures the similarity between two embedding vectors u and v . Typically, cosine similarity is used, defined as in (5).

$$\text{sim}(u, v) = \frac{u \cdot v}{\|u\| \cdot \|v\|} \quad (5)$$

The parameter $\tau > 0$ is a temperature scalar that scales the similarity scores. Smaller τ values make the model more sensitive to differences between samples. $P(i)$ denotes the set of positive samples. $A(i)$ represents the set of all other samples in the batch.

F. SHAP Explainable

The SHAP Explainable Module enhances the interpretability of the model's sentiment predictions by attributing importance scores to individual input

tokens. After the sentiment classifier generates an output—whether positive, neutral, or negative—this module applies SHAP (Shapley Additive Explanations) [22] to analyze how much each word in the product review contributed to the final prediction. Grounded in game theory, SHAP calculates the marginal contribution of each word by comparing the model's predictions with and without that word across various subsets. This results in a ranked list of tokens, each associated with a SHAP value indicating its positive or negative influence. The output is then visualized using color-coded summary plots, heatmaps, or bar charts, providing users with transparent and intuitive explanations. By incorporating this module, the framework supports explainable AI, allowing stakeholders to better understand the rationale behind each sentiment decision, which is especially valuable in customer feedback analysis and decision support systems.

For example, in the review “The battery life is terrible, but the camera is great,” SHAP may assign strong negative importance to “terrible” and strong positive importance to “great,” revealing their opposing effects on sentiment prediction. These attributions are visualized using heatmaps or bar plots to help users understand which terms drive the model's decisions. This module enhances model transparency, enabling better trust, error analysis, and actionable insights from product review data. The predictive formula constructed by the SHAP algorithm [22] is illustrated in (6).

$$y_i = y_{base} + y(x_i^1) + y(x_i^2) + \dots + y(x_i^j) \quad (6)$$

Where y_i is the i^{th} predicted value, y_{base} is the base-line value, while y_j is the SHAP value of the j^{th} feature in the i^{th} sample.

G. Implementation Details

The proposed framework was implemented in Python 3.7 and executed on an NVIDIA GeForce RTX 4060 GPU. Semantic features were extracted using a fine-tuned SentBERT (BERT-base, 12 layers, 768 hidden units), trained for aspect term extraction with a BIO tagging scheme using a learning rate of $2e-5$ and a batch size of 32. Syntactic features were obtained via dependency parsing, where each sentence was represented as a graph and processed using a layer Graph Convolutional Network with ReLU activation and 128 hidden units. The semantic and syntactic embeddings were concatenated in a fusion layer and optimized using supervised contrastive learning with cosine similarity and a temperature parameter of 0.07. Final sentiment classification was performed using a softmax layer, and SHAP was applied to provide word-level interpretability of the model's predictions.

IV. EXPERIMENTAL RESULTS

This section presents the experimental setup and results for the proposed hybrid sentiment analysis framework, designed to evaluate product reviews on Amazon. The experiments aimed to assess the model's ability to classify customer sentiment across multiple mobile phone brands using key performance metrics, including accuracy, precision, recall, and F1 Score. Results were compared against various baseline machine learning and deep learning models to demonstrate the effectiveness of the proposed method.

A. Data Collection

This study used a publicly available dataset from Amazon to perform sentiment analysis (SA). This dataset contains 10,500 customer reviews across eight mobile phone brands: Apple, Samsung, Motorola, OnePlus, Xiaomi, Google, LG, and Huawei. It comprises 40 models, including the Apple iPhone 15, Samsung Galaxy S24 Ultra, and Google Pixel 8. The dataset includes several attributes, such as brand, model, star rating, and review text, forming a robust basis for comprehensive analysis in this research. The dataset can be accessed via the following link: <https://rb.gy/rnrv3e>

B. Experimental Evaluation

The dataset was split into an 80% training set and a 20% test set. Baseline models—including Naïve Bayes, Decision Tree, Logistic Regression, Random Forest, Support Vector Machine, BERT-only, and GCN-only—were implemented using standard configurations, and deep learning baselines were fine-tuned under the same training and evaluation conditions to ensure fair comparison. Model performance was evaluated using accuracy, F1-score, recall, and precision. Accuracy measures the overall correctness of a model's predictions. High accuracy indicates that the model makes correct predictions most of the time. F1-score provides a balanced measure of precision and recall. A high F1-score indicates that the model is achieving a good balance between precision and recall. Precision measures the model's ability to correctly identify only positive instances. High precision means that when the model predicts a positive instance, it is likely to be correct. Recall measures the model's ability to correctly identify all positive instances. A high recall means the model can identify most positive instances. These formulas are shown in (7)-(10).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$\text{F1-Score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (9)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (10)$$

The performance comparison shown in Table I demonstrates that the proposed method significantly outperforms traditional and deep learning-based baselines. Models such as Decision Tree, Naive Bayes, Random Forest, Support Vector Machine, Logistic Regression, GCN-only, and BERT-only yielded lower performance scores. In contrast, the proposed method achieved 97% in accuracy, precision, recall, and F1-score, reflecting superior sentiment classification capability. These results confirm that integrating SentBERT, Graph Feature Fusion, contrastive learning, and SHAP-based explainability yields a more robust and interpretable sentiment analysis model.

TABLE I
PERFORMANCE COMPARISON OF THE PROPOSED
METHOD WITH OTHER APPROACHES

Models	Accuracy	Precision	Recall	F1-Score
Decision Tree	0.84	0.84	0.84	0.84
Naive Bayes	0.84	0.85	0.84	0.81
Random Forest	0.88	0.89	0.88	0.87
GCN-only	0.88	0.88	0.88	0.88
Support Vector Machine	0.89	0.89	0.89	0.89
Logistic Regression	0.91	0.91	0.91	0.91
BERT-only	0.93	0.93	0.93	0.93
Proposed Method	0.97	0.97	0.97	0.97

Fig. 4 and Fig. 5 illustrate the output of the SHAP Explainable Module, which visualizes the importance of different product attributes across sentiment classes.

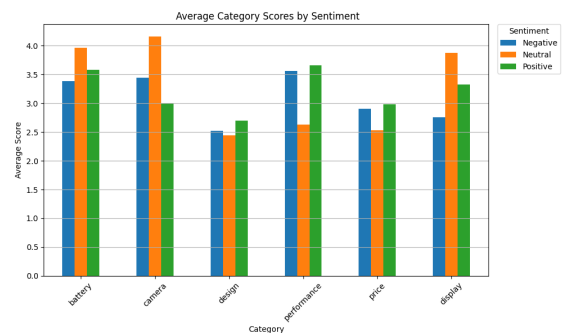


Fig. 4. Average Aspect Scores by Sentiment Category

In Fig. 4, a grouped bar chart shows the average sentiment scores—positive, neutral, and negative—for six key aspects of mobile phones: battery, camera, design, performance, price, and display. This visualization highlights how specific features are associated with different emotional polarities. For instance, camera and display tend to receive higher scores in neutral reviews, while design shows relatively low scores across sentiments, indicating common user dissatisfaction or indifference.

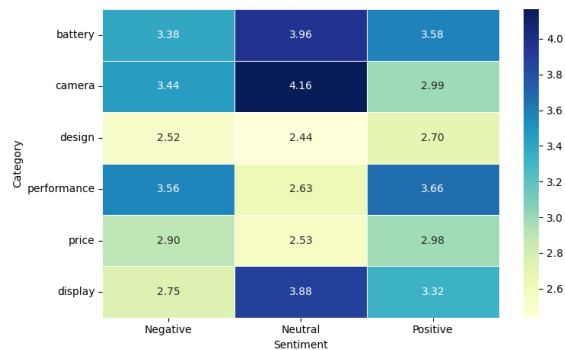


Fig. 5. Heatmap of average aspect scores across sentiment categories

Complementing this, Fig. 5 presents a heatmap showing the same average category scores but in a more intuitive, color-coded format. The deeper the color intensity, the higher the average sentiment score. This allows users and analysts to quickly identify which aspects significantly influence sentiment predictions and how those influences vary across different categories. Together, these visualizations provide interpretable insights into how feature-level contributions affect the model's decisions, thereby supporting transparency and trust in sentiment analysis outcomes.

V. DISCUSSION AND CONCLUSION

This study successfully developed and evaluated a novel hybrid framework for sentiment analysis of mobile phone reviews, effectively integrating semantic embeddings from SentBERT with syntactic structures extracted through Graph Feature Fusion (GFF). The proposed method demonstrated a significant performance advantage over traditional machine learning and deep learning baselines, consistently achieving 97% across key evaluation metrics—accuracy, precision, recall, and F1-score. This superior outcome highlights the effectiveness of combining diverse linguistic representations for robust sentiment classification.

The achieved performance can be attributed to several synergistic architectural components. Firstly, the SentBERT module's capability for Aspect Term Extraction (ATE) was instrumental in enabling a fine-grained understanding of product-specific sentiments.

This aspect-level granularity is crucial in analyzing e-commerce reviews, which often contain complex expressions or conflicting opinions about different features. Secondly, the Graph Feature Fusion (GFF) module leveraged dependency parsing to capture the intricate grammatical relationships and syntactic dependencies between words. This structural understanding enabled the model to effectively map sentiment-bearing terms to specific product aspects, a common challenge for methods that rely solely on sequential text representations. The subsequent Fusion Layer then seamlessly integrated these rich semantic and structural features, forming a comprehensive input for the classifier. Furthermore, the application of supervised contrastive learning significantly enhanced the model's discriminative power by refining the feature space and promoting greater separability among positive, neutral, and negative sentiment classes. Finally, the integration of SHAP Explainable AI provided critical interpretability, offering transparent insights into which words and aspects most significantly influenced the sentiment predictions, as visually corroborated by Fig. 4 and Fig. 5. This transparency not only fosters user trust but also provides actionable insights for businesses. This framework holds significant potential for real-world applications in e-commerce, offering enhanced automated review analysis capabilities, deeper insights for customer service, and valuable support for product development strategies. Despite its demonstrated high performance and interpretability, the proposed framework is not without limitations. Its dependence on accurate dependency parsing renders it susceptible to grammatical errors or informal language prevalent in user-generated content, potentially leading to inaccuracies in graph construction. Additionally, the computational demands of combining SentBERT, GCN-based processing and supervised contrastive learning are substantial, potentially limiting real-time deployment in large-scale applications. Lastly, while highly effective for mobile phone reviews, the model's domain specificity suggests that its generalizability to other diverse domains (e.g., hospitality, software reviews) may require further adaptation and retraining. Future research will therefore focus on optimizing model efficiency to support real-time sentiment analysis by reducing inference latency and computational cost. Cross-domain and cross-platform evaluations will be conducted to assess generalizability and robustness across diverse review types.

Further extensions may include multilingual sentiment analysis, dynamic and adaptive aspect extraction for emerging product features, and lightweight graph representations to improve scalability. Incorporating online or continuous learning mechanisms to adapt to evolving language

use in customer reviews is also a promising direction for future work.

Funding: This research was supported by Asia-Pacific International University, Thailand, under grant number: RRDC 2024-222.

B Econ in Business Economics. His research interests focus on business administration, management, marketing, and finance.

REFERENCE

- [1] K. Qiu and L. Zhang, "How online reviews affect purchase intention: A meta-analysis across contextual and cultural factors," *Data Inf. Manag.*, vol. 8, no. 2, p. 100058, Jun. 2024.
- [2] V. Shrirame, J. Sabade, H. Soneta, and M. Vijayalakshmi, "Consumer behavior analytics using machine learning algorithms," in *Proc. 2020 IEEE Int. Conf. Electron., Comput. Commun. Technol. (CONECCT)*, 2020, pp. 1-6, <https://doi.org/10.1109/CONECCT50063.2020.9198562>.
- [3] T. Brown et al., "Language models are few-shot learners," in *Proc. 34th Int. Conf. Neural Inf. Process. Syst. (NeurIPS)*, 2020, pp. 1877-1901.
- [4] J. J. Rahman, T. M. A. Riaz, M. Partha, K. M. Mohsin, N. Kamruddin, and M. F. Mridha, "Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review," *Nat. Lang. Process. J.*, p. 100059, Mar. 2024.
- [5] C. Dewi, R.-C. Chen, H. J. Christanto, and F. Cauteruccio, "Multinomial naïve Bayes classifier for sentiment analysis of internet movie database," *Vietnam J. Comput. Sci.*, vol. 10, no. 4, pp. 485-498, Jan. 2023, <https://doi.org/10.1142/S2196888823500100>.
- [6] I. A. K. Shaikh, R. P. Pujar, S. P. Kishore, S. Ragamayi, P. V. Krishna, and A. B. Nadaf, "A novel approach for e-commerce system for sale prediction with denoised autoencoder and SVM-based approach," in *Proc. 2023 Int. Conf. Sustain. Comput. Smart Syst. (ICSCSS)*, 2023, pp. 1684-1689.
- [7] N. Karina and Sriani, "Sentiment analysis of skincare products using the naïve Bayes method," *J. Inf. Syst. Informatics*, vol. 6, no. 3, pp. 1663-1676, Sep. 2024, <https://doi.org/10.51519/journalisi.v6i3.817>.
- [8] T. A. Khan, R. Sadiq, Z. Shahid, M. M. Alam, and M. B. Mohd Su'ud, "Sentiment analysis using support vector machine and random forest," *J. Informatics Web Eng.*, vol. 3, no. 1, pp. 67-75, Feb. 2024, <https://doi.org/10.33093/jiwe.2024.3.1.5>.
- [9] J. Devlin, M. -W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, 2019, pp. 4171-4186.
- [10] A. Vaswani et al., "Attention is all you need," in *Proc. 30th Int. Conf. Neural Inf. Process. Syst. (NeurIPS)*, 2017, pp. 5998-6008.
- [11] C. Li, D. Cheng, and Y. Morimoto, "Leveraging deep neural networks for aspect-based sentiment classification," *arXiv*, pp. 1-19, Mar. 2025, <https://doi.org/10.48550/arXiv.2503.12803>.
- [12] J. Hu and Y. Li, "Aspect-level sentiment analysis of bi-graph convolutional networks based on enhanced syntactic structural information," *J. Comput. Commun.*, vol. 13, no. 1, pp. 72-89, Jul. 2025, <https://doi.org/10.4236/jcc.2025.131006>.
- [13] M. Peal, M. S. Hossain, and J. Chen, "Summarizing consumer reviews," *J. Intell. Inf. Syst.*, vol. 59, no. 1, pp. 193-212, Feb. 2022, <https://doi.org/10.1007/s10844-022-00694-9>.
- [14] K. S. Kumar, M. Priyanka, M. Rishitha, D. D. Teja, and N. Madhuri, "Text summarization with sentiment analysis," *Int. J. Innov. Res. Comput. Sci. Technol.*, vol. 9, no. 4, Jul. 2021.
- [15] L. Yang, Y. Li, J. Wang, and R. S. Sherratt, "Sentiment analysis for e-commerce product reviews in Chinese based on sentiment lexicon and deep learning," *IEEE Access*, vol. 8, pp. 23522-23530, Jan. 2020.
- [16] A. R. Sulthana, K. Jaithunbi, H. Harikrishnan, and V. Vijayakumar, "Sentiment analysis on movie reviews dataset using support vector machines and ensemble learning," *Int. J. Inf. Technol. Web Eng.*, vol. 17, pp. 1-23, Jan. 2022.
- [17] D. Šandor and M. Bagić Babac, "Sarcasm detection in online comments using machine learning," *Inf. Discov. Deliv.*, vol. 52, no. 2, pp. 213-226, Mar. 2024, <https://doi.org/10.1108/IDD-01-2023-0002>.
- [18] E. Wahyudi and R. Kusumaningrum, "Aspect-based sentiment analysis in e-commerce user reviews using latent Dirichlet allocation (LDA) and sentiment lexicon," in *Proc. 3rd Int. Conf. Informatics Comput. Sci. (ICICoS)*, 2019.
- [19] L. Luo, C. -H. Wei, P. -T. Lai, R. Leaman, Q. Chen, and Z. Lu, "AIONER: All-in-one scheme-based biomedical named entity recognition using deep learning," *Bioinformatics*, vol. 39, no. 5, pp. 1-9, May 2023, <https://doi.org/10.1093/bioinformatics/btad310>.
- [20] M. A. Sejan, M. H. Rahman, M. A. Aziz, J. -I. Baik, Y. -H. You, and H. -K. Song, "Graph convolutional network design for node classification accuracy improvement," *Mathematics*, vol. 11, no. 17, p. 3680, Aug. 2023.
- [21] P. Khosla et al., "Supervised contrastive learning," *arXiv*, 2020. [Online]. Available: <https://arxiv.org/abs/2004.11362> [Accessed: Jun. 8, 2025].
- [22] X. Zhao, P. -F. Zhang, Q. Zhao, D. Zhang, Y. Tuerxunmaiti, and H. Cao, "ASHAP algorithm-based prediction of the interlaminar shear strength degradation of G/BFRP bars embedded in concrete exposed to marine environment," *Case Stud. Constr. Mater.*, vol. 22, pp. 1-19, Jul. 2025.



Damrong Sattayawaksakul is an Assistant Professor at the Faculty of Business, Asia-Pacific International University, Thailand. He holds a PhD in Business Administration. He obtained an MBA in management, a BBA in Finance and Banking, and another BEcon in Business Economics. His research interests focus on business administration, management, marketing, and finance.



Pimpa Cheewaprakobkit is an Assistant Professor at the Faculty of Information Technology, Asia-Pacific International University, Thailand. She holds a Ph.D. in Computer Science and Information Engineering from National Central University, Taiwan. She also earned an M.S. in Computer Science and Information Systems and a B.Sc. in Computer Science. Her research interests include computer vision, object tracking, big data analysis, deep learning, object detection, and object recognition.