Aspect-Based Sentiment Classification of iPhone 15 YouTube Reviews Using VADER-Augmented LSTM

Hasna Rafida Alya, Yuliant Sibaroni*

School of Computing, Informatics, Telkom University, Bandung, Indonesia
Email: [1]hasnarafidaalya@student.telkomuniversity.ac.id, [2,*]yuliant@telkomuniversity.ac.id
Email Penulis Korespondensi: hasnarafidaalya@student.telkomuniversity.ac.id
Submitted: **99/99/9999**; Accepted: **99/99/9999**; Published**: 99/99/9999**

Abstract−This research investigates the effectiveness of the Long Short-Term Memory (LSTM) model in performing aspect-based sentiment classification on English-language reviews of the iPhone 15 sourced from the YouTube platform. The study focuses on five key product aspects frequently mentioned by users: charger port, camera, screen, design, and battery. To evaluate the model's performance, two distinct labeling strategies were employed. The first involved manual annotation, where human annotators identified both the relevant aspects and the associated sentiment in each review. The second strategy integrated additional sentiment cues derived from a lexicon-based method, Valence Aware Dictionary and sEntiment Reasoner (VADER). In this approach, the polarity output from VADER was prepended to each review to enrich the input with emotional context. The experimental results demonstrate that supplementing review texts with sentiment polarity information from VADER contributes to a modest but measurable improvement in sentiment classification accuracy. Specifically, using the micro-average accuracy metric, defined as the ratio of correct predictions to the total number of test instances, the model's performance improved from 67% under the manual only annotation to 68% with VADER enhanced input. Additionally, aspect classification remained consistently strong, showing a slight improvement from 90% to 91% after incorporating VADER. Furthermore, based on macro-average accuracy an evaluation metric that calculates the mean performance across all classes regardless of class distribution, accuracy improvements were observed in several aspects, particularly the camera, screen, and design. However, a minor decline in performance was noted for the battery and charger port aspects. These results suggest that enriching review data with sentiment polarity information derived from lexicon-based tools like VADER can enhance the model's ability to comprehend emotional nuance, leading to more accurate identification of user sentiments within aspect-specific reviews.

Keywords: Sentiment Analysis; Aspect-Based; iPhone 15; LSTM; VADER

1. INTRODUCTION

One of the leading companies in the technology industry is Apple, which first introduced the iPhone on January 9, 2007 [1]. Since then, the iPhone has grown to become one of the leading smartphone brands, known for its unique iOS ecosystem. Every year, Apple consistently releases new models, including the iPhone 15, which was launched in September 2023. According to data from Counterpoint [2], the iPhone 15 became the best-selling smartphone in the third quarter of 2024, with a market share of 3.5%, surpassing the iPhone 14, which dominated the previous year. This indicates that improvements in the camera, performance, and other features have successfully attracted consumer interest. However, consumer reviews of various aspects of the iPhone, such as the charging port, camera, display, design, and battery, vary significantly, ranging from positive, negative, to neutral reviews.

Sentiment analysis aims to identify various opinions expressed in a text or comment in order to determine the underlying sentiment trend, whether it is positive, negative, or neutral. Through sentiment analysis, companies can gain deeper insights into consumer perceptions of their products, which can serve as a basis for evaluating and improving product quality to enhance customer satisfaction [3].

Various studies have been conducted on sentiment analysis using different methods. For example, a study conducted by Diny Wahyuni et al. explored the application of deep learning methods, specifically Long Short-Term Memory (LSTM), to classify sentiment in Indonesian-language hotel reviews. The study showed that the LSTM model was effective in understanding and analyzing the aspects and sentiment of hotel review sentences, achieving a relatively high level of accuracy [4].

Furthermore, research by Wahyuni et al. demonstrated that the combination of the Long Short-Term Memory (LSTM) and lexicon-based methods in sentiment analysis of TikTok app user reviews achieved an accuracy of 90.05%, with a precision of 92.14%, recall of 97.35%, and an F1-score of 98.66%. The results of this analysis revealed that the TikTok app tends to receive negative sentiment, with 59.5% of the reviews being negative, compared to 30.0% positive and 10.5% neutral. These findings confirm that the LSTM and lexicon-based methods are effective in accurately classifying and interpreting user opinions, as well as providing an in-depth picture of public sentiment towards the TikTok platform [5].

Meanwhile, Johan Setiawan et al. [6] analyzed sentiments related to tourist destinations in Labuan Bajo using data collected from Instagram via the hashtag "labuanbajo." This study aimed to explore feedback from both local and foreign tourists and to identify the most popular tourist destinations in the area. The analysis results showed that the VADER method achieved an accuracy of 72%.

Marvin Gultom et al. [7] conducted sentiment analysis on English-language text using the VADER sentiment algorithm implemented on a website platform. The study categorized text into four classes: positive, negative, neutral, and mixed, using data from 50 Twitter comments on the YouTube channel "Simplilearn." The final results showed that the system achieved an accuracy of approximately 80%.

Arum Prabowo G. et al. [8] examined the use of the XGBoost algorithm for aspect-based sentiment analysis of iPhone 14 Pro video review comments on YouTube. The aim was to classify reviews into positive and negative sentiments and identify the strengths and weaknesses of the iPhone 14 Pro. The results showed an accuracy ranging from 89% to 97%, with identified strengths including brand image and performance, and weaknesses in design and specifications.

Most previous studies have focused on general sentiment analysis without considering specific aspects. Furthermore, there have been few studies that specifically compare manual labeling approaches with lexicon-based methods in the context of aspect-based sentiment analysis, particularly for iPhone 15 reviews. Based on this gap, this study aims to evaluate the performance of the Long Short-Term Memory (LSTM) model in classifying sentiment toward five main aspects: charger port, camera, screen, design, and battery.

This study compares two data processing scenarios. In the first scenario, aspect and sentiment labels are manually annotated by human annotators through direct reading and understanding of English-language iPhone 15 reviews collected from the YouTube platform. In the second scenario, sentiment labels are generated using the Valence Aware Dictionary and sEntiment Reasoner (VADER) method by applying a threshold value for automatic sentiment classification. The VADER-generated sentiment information is inserted at the beginning of the review text as additional input for the model. However, the target sentiment labels remain based on manual annotation and serve as the ground truth for evaluating model performance. This study aims to compare the two scenarios and determine the extent to which the inclusion of VADER-generated sentiment information at the beginning of the text affects the LSTM model's performance in aspect-based sentiment classification.