JCTA Journal of Computing Theories and Applications ISSN:3024-9104

Research Article

Aspect-Based Sentiment Analysis on E-commerce Reviews using BiGRU and Bi-Directional Attention Flow

De Rosal Ignatius Moses Setiadi [1],*, Warto [2], Ahmad Rofiqul Muslikh [3], Kristiawan Nugroho [4], and Achmad Nuruddin Safriandono [5]

[1] Faculty of Computer Science, Dian Nuswantoro University, Semarang, Central Java 50131, Indonesia; e-mail : moses@dsn.dinus.ac.id
[2] Informatics Department, Faculty of Dakwah, UIN Profesor Kiai Haji Saifuddin Zuhri, Purwokerto, Indonesia; e-mail: warto@uinsaizu.ac.id
[3] Faculty of Information Technology, University of Merdeka, Malang, East Java 65146, Indonesia; e-mail: rofickachmad@unmer.ac.id
[4] Department of Information Technology and Industry, StikubankUniversity, Semarang, Indonesia; e-mail: kristiawan@edu.unisbank.ac.id
[5] Faculty of Engineering, Sultan Fatah University, Demak, Central Java 59516, Indonesia; e-mail : udinozz@gmail.com
* Corresponding Author : De Rosal Ignatius Moses Setiadi

Abstract: Aspect-based sentiment Analysis (ABSA) is vital in capturing customer opinions on specific e-commerce products and service attributes. This study proposes a hybrid deep learning model integrating Bi-Directional Gated Recurrent Units (BiGRU) and Bi-Directional Attention Flow (BiDAF) to perform aspect-level sentiment classification. BiGRU captures sequential dependencies, while BiDAF enhances attention by focusing on sentiment-relevant segments. The model is trained on an Amazon review dataset with preprocessing steps, including emoji handling, slang normalization, and lemmatization. It achieves a peak training accuracy of 99.78% at epoch 138 with early stopping. The model delivers a strong performance on the Amazon test set across four key aspects: price, quality, service, and delivery, with F1 scores ranging from 0.90 to 0.92. The model was also evaluated on the SemEval 2014 ABSA dataset to assess generalizability. Results on the restaurant domain achieved an F1-score of 88.78% and 83.66% on the laptop domain, outperforming several state-of-the-art baselines. These findings confirm the effectiveness of the BiGRU-BiDAF architecture in modeling aspect-specific sentiment across diverse domains.

Keywords: Aspect-based sentiment analysis; Attention mechanism; BiDAF; E-commerce reviews analysis; Emoji handling; Lemmatization; Slang normalization.

Received: February, 12th 2025
Revised: March, 24th 2025
Accepted: March, 31st 2025
Published: April, 1st 2025

1. Introduction

The development of technology and the increasing use of e-commerce platforms have produced big data in the form of product reviews from users. Sentiment analysis is a popular approach in Natural Language Processing (NLP) which is used to classify user opinions or emotions towards a particular product, service, or topic into categories such as positive, negative, or neutral. This approach has been widely used to understand the general public perception of an entity[1]–[3]. However, conventional sentiment analysis approaches usually only provide an overall assessment of a single document or sentence without considering the context or specific aspects discussed in it. For example, a review such as "The price is good but the delivery was late" can give ambiguous results if analyzed globally, because it contains positive sentiment towards price but negative towards delivery[4]–[6].

To overcome these limitations, Aspect-Based Sentiment Analysis (ABSA) was developed as a more fine-grained approach to analyzing opinions. ABSA aims to identify specific entities or attributes (such as price, quality, delivery, or service) and evaluate sentiment towards these aspects independently[4], [7]–[10]. This approach has been shown to provide deeper and more relevant insights, especially in the context of customer opinion-based decision-making, and contributes to reducing customer churn[11] and increasing the effectiveness of recommendation and marketing strategies[12].

Several studies related to general sentiment analysis and ABSA use various traditional machine learning methods, Naive Bayes (NB)[2], [13], such as Support Vector Machine (SVM) and Random Forest(RF) are still widely used because of their ability to handle binary and multiclass classification with competitive performance on small datasets [14]. SVM is effective in high-dimensional feature spaces and is resistant to overfitting, but has limitations in capturing semantic context and word order in review text. Similarly, Random Forest excels in noise robustness and can handle non-linear features, but is less effective in representing complex linguistic structures in dynamic e-commerce text data.

As deep learning-based approaches develop, models such as Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and hybrid architectures with attention mechanisms [7], [15] are widely used due to their ability to learn long-term dependencies between words in a sentence. LSTM can capture sequential context well[16], [17], but is often slower and more complex to train. In contrast, GRU has a lighter structure, making it more efficient in training and still able to maintain contextual representations[18]. To improve understanding of bidirectional context, the Bidirectional GRU (BiGRU) model was developed, which processes data from both the front and back directions simultaneously, making it effective in capturing information before and after aspects in a sentence[7].

Although RNN-based models such as BiGRU are capable of handling word sequences, they still have limitations in identifying important words that are specifically related to an aspect[19]. Therefore, attention has shifted to the use of attention mechanisms, such as Bi-Directional Attention Flow (BiDAF), which allows the model to selectively focus attention on important words relevant to a particular aspect. BiDAF is effective in various text understanding tasks due to its ability to establish a bidirectional attention flow between the context and the target aspect, thereby improving the quality of aspect polarity classification[20]. The combination of BiGRU architecture with BiDAF enables synergy between word sequence processing and emphasis on important words, making it a promising approach in the ABSA domain[15].

In general, commonly used text preprocessing such as tokenization, stemming, and case folding have a fundamental role that directly affects the quality of feature representation and the performance of classification models[21]–[23]. However, raw text data on user reviews on e-commerce platforms tend to contain informal elements such as abbreviations, slang, spelling errors, and non-verbal expressions such as emoticons. If not handled properly, these elements can obscure semantic meaning, increase feature sparsity, and cause the model to fail to capture true sentiment polarity. Therefore, careful preprocessing not only improves the efficiency of vector representation but also enhances the model's ability to understand sentence context and aspect polarity more precisely.

Theoretically, emoji handling helps capture emotional expressions that are often explicit but not in the form of words, such as 😠 or 😍, which have strong sentiment value and are relevant in opinion classification[24], [25]. Slang normalization aims to align informal or non-standard terms such as "gr8", "luv", or "coz" into standard equivalents ("great", "love", and "because"), thereby reducing lexical heterogeneity and improving accuracy in the tokenization and feature extraction processes[26]. Meanwhile, lemmatization is the process of reducing words to their base form (e.g. "running", "ran" → "run"), which significantly reduces morphological variation in the data and strengthens semantic coherence between words in the vector space[5], [27].

In the context of this study, we apply three main preprocessing stages: emoji handling, slang normalization, and lemmatization. The three were chosen because they represent three dominant categories of challenges in e-commerce reviews: non-verbal emotional expressions, informal language use, and morphological variation. The implementation of this preprocessing is expected to improve the feature representation before being processed by the BiGRU-BiDAF model, as well as improve the model's ability to identify sentiment polarity towards aspects such as price, quality, service, and delivery more accurately and contextually. In addition to classifying sentiment polarity, this study also calculates the distribution of the number of sentiments towards predetermined aspects (e.g.: price, quality, service, delivery). This is in line with the approach used in several previous papers which also highlight the importance of knowing how many opinions are associated with each aspect, not just its polarity.

The rest of this paper is systematically organized to explain the approach and contributions of this research. Section 2 reviews related studies that form the basis of the model development, ranging from classical methods to modern deep-learning approaches. Section 3 presents the proposed method in detail, from preprocessing, and BiGRU-BiDAF model architecture, to the training scheme. Section 4 presents experimental results and model performance analysis based on evaluation metrics on Amazon and SemEval datasets. Finally, Section 5 summarizes the main findings and provides potential further research directions.

2. Related Work

Aspect-based sentiment analysis (ABSA) has been the focus of much research due to its ability to provide deeper sentiment analysis by identifying sentiment polarity towards certain aspects of a product or service. Early studies in ABSA generally used traditional machine learning approaches such as Naive Bayes, SVM, and Conditional Random Field (CRF), which relied on manual feature engineering and domain-specific lexicons. While effective in some cases, these approaches have limitations in capturing complex semantic contexts and dependencies between words in text.

With the advancement in deep learning, RNN-based models, such as LSTM, and GRU, have been widely used due to their ability to learn sequential patterns from text. Abdelgwad et al. [7], for example, proposed a combined architecture consisting of BiGRU-CNN-CRF for opinion target extraction, and IAN-BiGRU for sentiment polarity classification towards aspects. This model showed significant improvements in F1-score and accuracy on Arabic datasets, highlighting the importance of bidirectional context and attention mechanisms in ABSA tasks.

In the Indonesian context, Jayanto et al. [28] developed an optimized LSTM model for hotel reviews, with a customized hidden layer structure and activation function. The model successfully achieved a competitive F1-score and demonstrated that adaptation to domain characteristics and informal language is essential in improving analysis performance.

To overcome the limitations in explicit aspect extraction, Kabir et al. [29] proposed a hybrid method that combines frequency-based, dependency syntax, and CRF approaches to extract both explicit and implicit aspects, including aspects in the form of compound nouns. This method showed significant improvements in precision and recall on SemEval and Amazon datasets, highlighting the importance of a complex tagging strategy for implicit aspects.

Several studies have also explored topic modeling approaches in aspect extraction. Prakash and Sharma [5] used the Pachinko Allocation Model (PAM) to extract aspects from Amazon product reviews and classify sentiment polarity. Although efficient in discovering topical aspects, this approach does not capture word order dependencies as the RNN model does.

Rahin et al. [30] conducted a comparative study using SemEval and Amazon datasets and showed that the integration of additional information such as part-of-speech tagging, dependency parsing, and word embedding trained on a large corpus can improve the accuracy of aspect-sentiment classification, especially on small datasets. Meanwhile, Sivakumar and Uyyala [17] developed an LSTM-based model combined with fuzzy logic to classify sentiment intensity into four categories: very positive, positive, negative, and very negative. This multi-label approach aims to provide a more detailed representation of consumer opinions.

Overall, deep learning models such as BiGRU and LSTM offer strong capabilities in modeling word sequences. However, their effectiveness increases significantly when combined with attention mechanisms (e.g. Bi-Directional Attention Flow), syntactic features, or hybrid strategies. This study builds on these foundations by integrating BiGRU and BiDAF along with text preprocessing techniques tailored to the e-commerce domain to improve the accuracy of sentiment detection on aspects such as price, quality, service, and delivery.

3. Proposed Method

This study proposes ABSA on e-commerce reviews using a hybrid architecture of BiGRU and BiDAF. The model aims to identify sentiment polarity towards specific aspects such as price, quality, delivery, and service. The methodology consists of several main stages: text preprocessing with emoji handling, lemmatization, and slang normalization, feature representation, BiGRU-BiDAF model training, and sentiment analysis per aspect. Figure 1 illustrates the stages of the method, while detailed stages are presented in subsections 3.1 to 3.5.

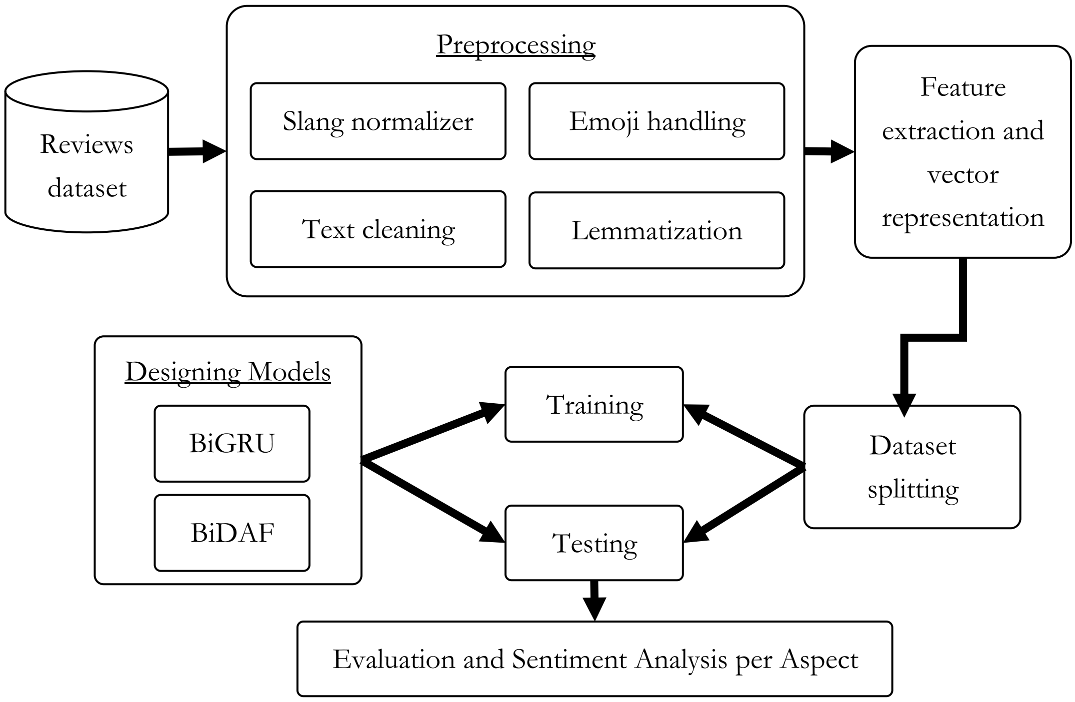

Figure 1. Illustration of the proposed method.

3.1. Preprocessing Texts

The preprocessing stages are carried out in stages to improve the quality of feature representation:

1. Slang normalization is carried out to convert non-standard words with formal equivalents using a defined slang dictionary presented in Table 1.

Table 1. Slang Dictionaries.

Acronym	Formal words	Acronym	Formal words	Acronym	Formal words
u	you	idk	i do not know	omw	on my way
lol	laugh	brb	be right back	smh	shaking my head
gr8	great	omg	oh my god	afaik	as far as i know
luv	love	dunno	do not know	imo	in my opinion
coz	because	cuz	because	ttyl	talk to you later
b4	before	gonna	going to	bff	best friend forever
ok	okay	bday	birthday	lmao	laughing my ass off
ur	your	tbh	to be honest	fyi	for your information
bc	because	ikr	i know right	imho	in my humble opinion
ya'll	you all	wanna	want to	wtf	what the fuck
thx	thanks	gotta	got to	wth	what the hell
plz	please	ain't	is not	nvm	never mind
r	are	lemme	let me	btw	by the way
tho	though	gimme	give me		

2. Each emoji expression is converted to text form using emoji.demojize, for example 😍 becomes :smiling_face_with_heart_eyes: to maintain non-verbal emotional expressions. At this stage, the emoji library from Python is used.
3. Text Cleaning consists of removing HTML tags, and non-alphabetic characters done with regular expressions, converting to lowercase, and removing excess spaces.

4. After cleaning, the text is broken into words (tokens) and then stopwords are removed from common words that are not informative based on the ENGLISH_STOP_WORDS list.
5. Lemmatization is the last stage, where words are returned to lemma form using the WordNetLemmatizer function taken from the nltk.stem library. This is done to reduce morphological variation.

3.2. Feature Extraction and Vector Representation

The preprocessed corpus is converted into a feature vector using CountVectorizer with a limit of 1,000 most frequently occurring features with the command or formula below.

$$X = CountVectorizer(max\_features = 1000).fit\_transform(D) \tag{1}$$

Where D is a list of strings resulting from concatenating the review tokens per line. The binary sentiment labels are obtained from the LABEL column with the transformation below.

$$y = \begin{cases} 1 \text{ if } LABEL = \_label1\_ \\ 0 \quad\quad\quad\quad\quad \text{others} \end{cases} \tag{2}$$

The data was then divided into training and testing sets in a ratio of 80:20.

3.3. BiGRU-BiDAF Model Architecture

3.3.1. BiGRU Model

BiGRU processes sequential input x_t by storing previous and subsequent information through hidden states h_t, as in Equation (3). The detailed configuration of BiGRU uses input_dim = 1000; hidden_dim = 50; num_layers = 2, and dropout_rate = 0.5.

$$h_t = \text{GRU}\left(x_t, \vec{h}_{t-1}\right); \text{GRU}\left(x_t, \overleftarrow{h}_{t+1}\right) \tag{3}$$

3.3.2. BiDAF Model

BiDAF is an attention mechanism designed to build a two-way interaction between the representation of the input sequence and the query or target. The attention module used in the proposed BiDAF is Context2Query attention to find the part of the context that is most relevant to a particular aspect, and Query2Context to highlight the part of the context that is very similar or often appears repeatedly. Furthermore, the attention score is calculated using the similarity function shown in Equation (4).

$$\alpha_{ij} = \text{sim}\left(h_i, h_j\right) = w^{\intercal}\left[h_i; h_j; h_i \odot h_j\right] \tag{4}$$

where h_i and h_j as hidden state vectors, $\odot$ as element-wise multiplication, and w is the trainable parameter of the attention layer. Then, the attention result vector is obtained by Equation (5).

$$a_i = \sum_j \text{softmax}\left(\alpha_{ij}\right) \cdot h_j \tag{5}$$

3.3.3. Classification

The result of the attention is fed to the Linear layer and activated with the sigmoid function, such as Equation (6).

$$\hat{y} = \sigma(W^{\intercal}a + b) \tag{6}$$

The sigmoid activation function is chosen because in this case it is a binary classification. The Linear layer produces a single scalar output that reflects the probability of label 1 (positive). Furthermore, this classification layer is optimized using Binary Cross Entropy Loss which is calculated by Equation (7).

$$\mathcal{L} = -[y.\log(\hat{y}) + (1 - y) \cdot \log(1 - \hat{y})] \tag{6}$$

3.4. Training Model

The model is trained with a maximum of 250 epochs using the Adam optimizer (lr=0.001) with a mini-batch size of 32. Early stopping is applied with a patience of 10 epochs if there is no improvement in the loss value. The implementation of early stopping is done by monitoring the validation loss value and stopping training if there is no improvement for 10 epochs.

3.5. Evaluation and Sentiment Analysis per Aspect

Aspect analysis was conducted by matching keywords in the reviews against four main categories defined manually as in Table 2.

Table 2. Aspect Details.

Aspek	Kata Kunci
Price	price, cost, expensive, cheap, value, affordable
Quality	quality, durable, sturdy, good, bad, excellent
Service	service, support, helpful, rude, friendly, customer
Delivery	delivery, shipping, fast, slow, on time, late

For each aspect, reviews containing extracted keywords are performed. These are classified by positive-negative polarity and measured by performance with evaluation metrics: accuracy, precision, recall, and f1-score. Label distribution is calculated with the threshold presented in Equation (6).

$$\text{label} = \begin{cases} \text{positive}, \hat{y} > 0.5 \\ \text{negative}, \hat{y} \leq 0.5 \end{cases} \quad (6)$$

Where the threshold is set to 0.5, following the general convention of binary classification. Evaluation is performed on the subset of data containing the aspect keywords. Evaluation metrics: accuracy, precision, recall, and F1-score are calculated using sklearn.metrics.

4. Results and Discussion

This research was implemented using the Google Colaboratory (Colab) platform with several main libraries used in this study including PyTorch, scikit-learn, NLTK, emoji, and Matplotlib. All codes were executed in the Python 3 runtime provided by Google Colab. The dataset used in this study was obtained from Kaggle with the link https://www.kaggle.com/datasets/lievgar-cia/amazon-reviews. The data set contains 21,000 e-commerce product reviews from the Amazon platform, with a two-label format, namely positive and negative, whose sentiment distribution is presented in Figure 2.

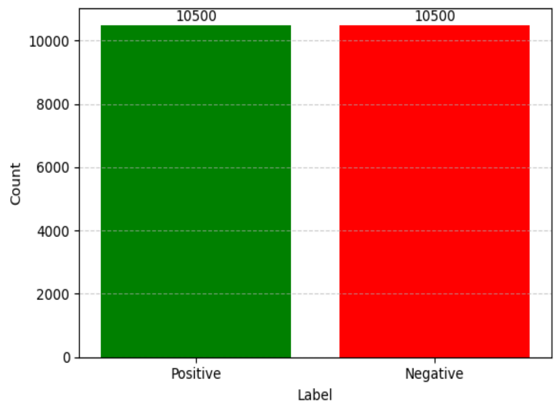

Figure 2. Global sentiment distribution of Amazon reviews dataset.

Furthermore, the distribution based on the main aspects such as price, quality, service, and delivery is presented in Figure 3. These four aspects were chosen because they are crucial elements in the consumer decision-making process and the perception of value towards a product or service. In the marketing literature, price reflects the perception of value and sensitivity to cost; quality reflects satisfaction with product performance; service relates to customer experience; and delivery becomes an important factor in the context of e-commerce which depends on the speed and reliability of delivery.[31]–[34].

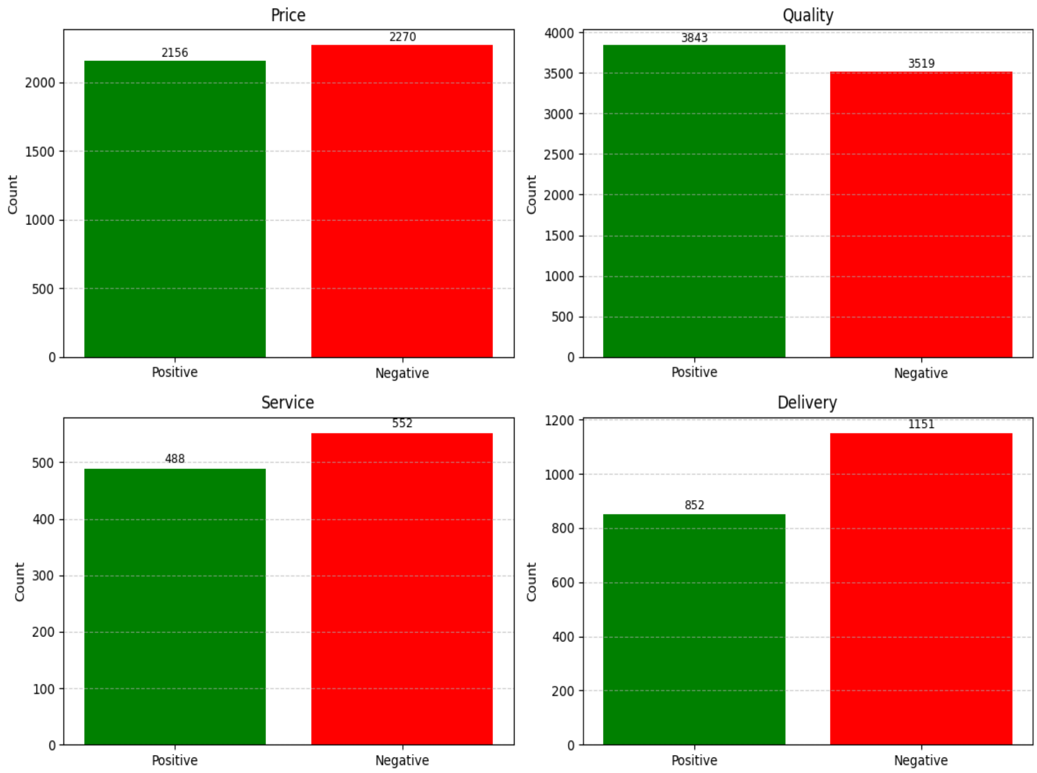

Figure 3. Sentiment distribution per aspects of Amazon reviews dataset.

Based on Figure 2 shows that the dataset is balanced with 10,500 data each for positive and negative labels. But in Figure 3 the distribution of the number of positive and negative reviews for each aspect varies. The aspects of price, delivery, and service have more negative opinions, while for the quality aspect, positive sentiment is more dominant. These plots support the initial understanding that consumer opinions are not only globally different but also vary depending on the aspects discussed.

The BiGRU-BiDAF model training process was carried out for 250 epochs, but with the application of the early stopping mechanism and optimization using the Adam algorithm, the training process stopped at 138 epochs, see Figure 4. The training results show a consistently decreasing loss curve and a significantly increasing accuracy curve, especially in the first 20 epochs. In the early training phase, the loss dropped from 0.65 to 0.13 and the accuracy increased from 59% to 94%, indicating that the model quickly recognizes basic patterns. In the middle epochs (21–60), the loss decreases slowly but steadily, while the accuracy continues to increase to almost 99%, reflecting effective convergence. In the final phase, the loss approaches zero and the accuracy stabilizes above 99%, indicating that the model has reached its optimal condition without overfitting. Overall, the training process shows that the BiGRU-BiDAF architecture can generalize very well to the data, with the loss and accuracy curves showing stability and no significant fluctuations.

Table 3 presents the results of testing the BiGRU-BiDAF model on four main aspects in the Amazon e-commerce dataset, namely price, quality, service, and delivery. In general, the model shows very good performance with an F1-score above 0.90 for all aspects, indicating that the model can classify sentiment polarity accurately for each aspect. Interestingly, the delivery aspect recorded the highest accuracy of 92.96%, while the service aspect had the lowest accuracy value of 90.97%, although it is still in the very good category. Judging from the recall value, the service aspect showed the lowest value (89.07%), indicating that the model tends to miss some negative or positive opinions about the service.

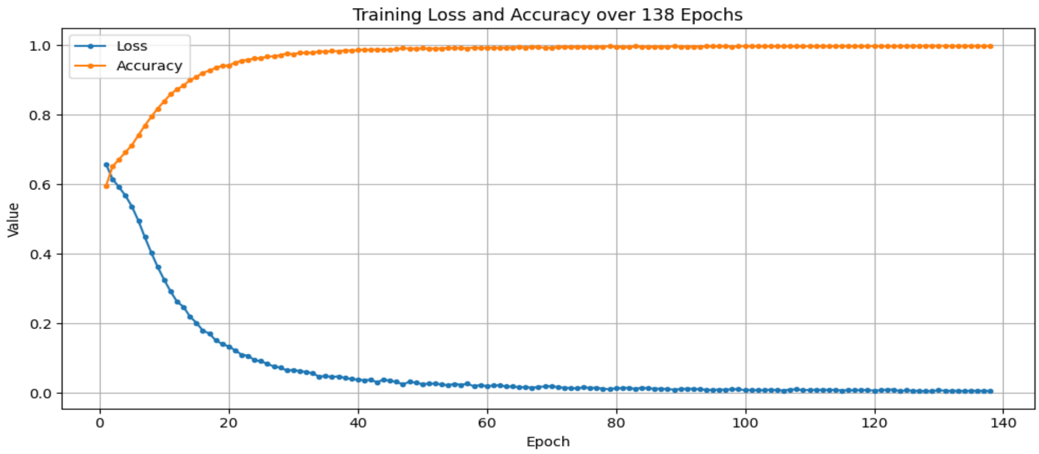

Figure 4. The plot of loss and accuracy of training proposed model for Amazon dataset.

Table 3. Testing result of the proposed method for Amazon dataset.

Aspect	Accuracy	Precision	Recall	F1- Score
Price	0.9167	0.9199	0.9127	0.9163
Quality	0.9204	0.9272	0.9198	0.9235
Service	0.9097	0.9157	0.8907	0.9030
Delivery	0.9296	0.9182	0.9241	0.9211

Note that in the context of ABSA, precision and recall play a more important role than accuracy, especially when the class distribution is imbalanced. For example, if the data is dominated by positive sentiment towards a certain aspect, the model may produce high accuracy by only classifying the majority but fail to detect the minority class. Therefore, the main focus of evaluation in ABSA should be directed at the F1-score, as a harmonic metric between precision and recall.[35], [36].

Furthermore, this study also conducted testing on another dataset, namely the SemVal dataset[37], which specifically discusses ABSA with the topics of Laptops and Restaurants. This is done to test the robustness of the proposed method. Figure 5 presents the training epoch plot for the Laptop and Restaurant topics. The laptop dataset produced a final accuracy of 0.9771 at epoch 91, while the restaurant was 0.9600 at epoch 80. Both datasets also stopped early and did not need to reach 250 epochs because early stopping was activated. Table 4 presents the results of testing and comparison with several state of the art.

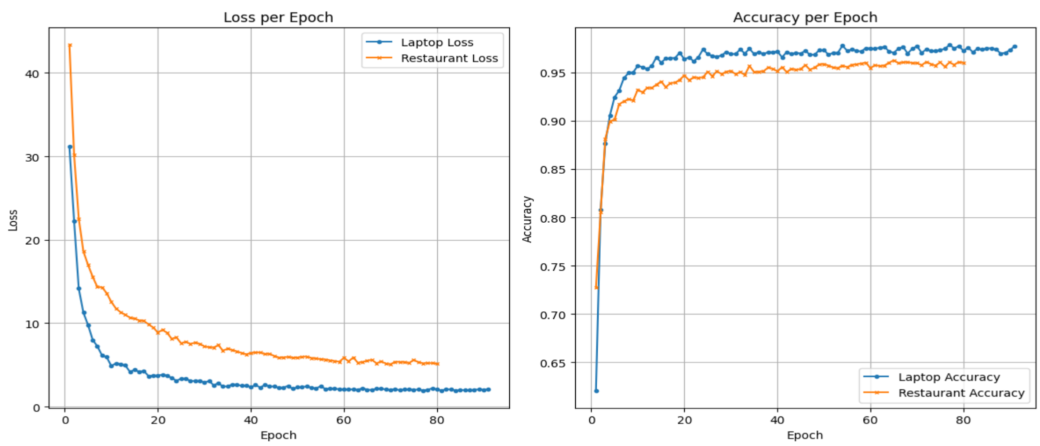

Figure 5. The plot of loss and accuracy of training proposed model for SemVal dataset.

Table 4 compares the performance of the proposed method with several approaches from previous studies on two different domains, namely Restaurant and Laptop from the SemEval dataset. In the Restaurant domain, the proposed method records the highest F1-

score of 88.78, although its accuracy (84.01%) is slightly lower than the method [6] (86.88%). However, the higher F1-score indicates that the model is more balanced in handling precision and recall, and is better able to detect minority opinions, a crucial factor in ABSA on imbalanced data. Meanwhile, in the Laptop domain, the proposed method also excels with an F1-score of 83.66, surpassing other reference methods. The consistently high performance on both domains confirms the effectiveness of the BiGRU-BiDAF hybrid architecture in understanding the semantic relations between aspects and opinions, while also demonstrating strong generalization capabilities across different domains.

Table 4. Testing result of proposed method and comparison for SemVal dataset.

Method	Restaurants				Laptops			
	Accuracy	**Precision**	**Recall**	**F1- Score**	**Accuracy**	**Precision**	**Recall**	**F1- Score**
Ref [9]	<u>84.95</u>	-	-	76.96	78.07	-	-	75.08
Ref [10]	84.29	-	-	76.79	78.53	-	-	75.15
Ref [6]	**86.88**	-	-	<u>81.16</u>	<u>80.56</u>	-	-	<u>77.00</u>
Ours	84.01	90.60	87.03	**88.78**	**82.21**	80.48	87.11	**83.66**

5. Conclusions

This study proposes a hybrid BiGRU-BiDAF model to perform aspect-based sentiment classification on e-commerce reviews. The main objective of this study is to identify sentiment polarity towards specific aspects such as price, quality, service, and delivery more accurately. Experimental results show that the model successfully achieves a training accuracy of up to 99.78% with an F1 score above 90% for each aspect of the Amazon dataset. In addition, the model is also tested on the SemEval 2014 benchmark dataset and shows competitive performance with an F1-score of 88.78% for the restaurant domain and 83.66% for the laptop domain.

Although the results show excellent performance, this model still has some limitations. One of them is the reliance on keyword matching in the aspect extraction stage, which can lead to the loss of semantic context or implicit aspects. In addition, this model does not consider time dynamics or changes in user opinions over time. As a direction for further research, development can be focused on the integration of transformer-based models to enrich contextual representation. Handling implicit aspects and using unsupervised techniques for automatic aspect extraction are also relevant potential areas of development, especially to improve model generalization in broader domains and more unstructured data.

Author Contributions: Conceptualization: DRIMS; methodology: DRIMS.; software: X.X.; validation: W, ARM, and KN; formal analysis: W, ARM; investigation: W, ARM, KN, and ANS.; resources: ANS.; data curation: W, ARM, and KN; writing—original draft preparation: DRIMS; writing—review and editing: W, ARM, KN, and ANS; visualization: KN; supervision: DRIM; project administration: ANS.; funding acquisition: All. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The dataset used in the study is publicly accessible at the URL https://www.kaggle.com/datasets/lievgar-cia/amazon-reviews and https://www.kaggle.com/datasets/charitarth/semeval-2014-task-4-aspectbasedsentimentanalysis

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] K. Nugroho, E. Winarno, D. R. I. M. Setiadi, and O. Farooq, "Enhanced multi-lingual Twitter sentiment analysis using hyperparameter tuning k-nearest neighbors," *Bull. Electr. Eng. Informatics*, vol. 13, no. 6, pp. 4327–4334, Dec. 2024, doi: 10.11591/eei.v13i6.7265.

[2] H. A. Santoso, E. H. Rachmawanto, A. Nugraha, A. A. Nugroho, D. R. I. M. Setiadi, and R. S. Basuki, "Hoax classification and sentiment analysis of Indonesian news using Naive Bayes optimization," *TELKOMNIKA (Telecommunication Comput. Electron. Control.*, vol. 18, no. 2, p. 799, Apr. 2020, doi: 10.12928/telkomnika.v18i2.14744.
[3] P. Sundarreson and S. Kumarapathirage, "SentiGEN: Synthetic Data Generator for Sentiment Analysis," *J. Comput. Theor. Appl.*, vol. 1, no. 4, pp. 461–477, Apr. 2024, doi: 10.62411/jcta.10480.
[4] W. Zhang, X. Li, Y. Deng, L. Bing, and W. Lam, "A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 11, pp. 11019–11038, Nov. 2023, doi: 10.1109/TKDE.2022.3230975.
[5] Y. Prakash and D. K. Sharma, "Aspect Based Sentiment Analysis for Amazon Data Products using PAM," in *2023 6th International Conference on Information Systems and Computer Networks (ISCON)*, Mar. 2023, pp. 1–5. doi: 10.1109/ISCON57294.2023.10112193.
[6] G. Zhao, Y. Luo, Q. Chen, and X. Qian, "Aspect-based sentiment analysis via multitask learning for online reviews," *Knowledge-Based Syst.*, vol. 264, p. 110326, 2023, doi: 10.1016/j.knosys.2023.110326.
[7] M. M.Abdelgwad, T. H. A Soliman, A. I.Taloba, and M. F. Farghaly, "Arabic aspect based sentiment analysis using bidirectional GRU based models," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 9, pp. 6652–6662, Oct. 2022, doi: 10.1016/j.jksuci.2021.08.030.
[8] K. K. Yusuf, E. Ogbuju, T. Abiodun, and F. Oladipo, "A Technical Review of the State-of-the-Art Methods in Aspect-Based Sentiment Analysis," *J. Comput. Theor. Appl.*, vol. 1, no. 3, pp. 287–298, Feb. 2024, doi: 10.62411/jcta.9999.
[9] H. Xu, B. Liu, L. Shu, and P. Yu, "BERT Post-Training for Review Reading Comprehension and Aspect-based Sentiment Analysis," in *Proceedings of the 2019 Conference of the North*, 2019, pp. 2324–2335. doi: 10.18653/v1/N19-1242.
[10] B. Zeng, H. Yang, R. Xu, W. Zhou, and X. Han, "LCF: A Local Context Focus Mechanism for Aspect-Based Sentiment Classification," *Appl. Sci.*, vol. 9, no. 16, p. 3389, Aug. 2019, doi: 10.3390/app9163389.
[11] R. E. Ako *et al.*, "Effects of Data Resampling on Predicting Customer Churn via a Comparative Tree-based Random Forest and XGBoost," *J. Comput. Theor. Appl.*, vol. 2, no. 1, pp. 86–101, Jun. 2024, doi: 10.62411/jcta.10562.
[12] A. Imtiaz, N. Pathirana, S. Saheel, K. Karunanayaka, and C. Trenado, "A Review on the Influence of Deep Learning and Generative AI in the Fashion Industry," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 3, pp. 201–216, Oct. 2024, doi: 10.62411/faith.3048-3719-29.
[13] A. Angdresey, L. Sitanayah, and I. L. H. Tangka, "Sentiment Analysis for Political Debates on YouTube Comments using BERT Labeling, Random Oversampling, and Multinomial Naïve Bayes," *J. Comput. Theor. Appl.*, vol. 2, no. 3, pp. 342–354, Jan. 2025, doi: 10.62411/jcta.11668.
[14] D. R. I. M. Setiadi, D. Marutho, and N. A. Setiyanto, "Comprehensive Exploration of Machine and Deep Learning Classification Methods for Aspect-Based Sentiment Analysis with Latent Dirichlet Allocation Topic Modeling," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 1, pp. 12–22, May 2024, doi: 10.62411/faith.2024-3.
[15] L. Zhu, M. Xu, Y. Bao, Y. Xu, and X. Kong, "Deep learning for aspect-based sentiment analysis: a review," *PeerJ Comput. Sci.*, vol. 8, p. e1044, Jul. 2022, doi: 10.7717/peerj-cs.1044.
[16] "Sentiment Analysis for User Reviews Based on Improved Binarization Aquila Optimization with Self-Attention Bi-LSTM Model," *Int. J. Intell. Eng. Syst.*, vol. 17, no. 5, pp. 813–824, Oct. 2024, doi: 10.22266/ijies2024.1031.61.
[17] M. Sivakumar and S. R. Uyyala, "Aspect-based sentiment analysis of mobile phone reviews using LSTM and fuzzy logic," *Int. J. Data Sci. Anal.*, vol. 12, no. 4, pp. 355–367, Oct. 2021, doi: 10.1007/s41060-021-00277-x.
[18] D. R. I. M. Setiadi, S. Widiono, A. N. Safriandono, and S. Budi, "Phishing Website Detection Using Bidirectional Gated Recurrent Unit Model and Feature Selection," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 2, pp. 75–83, Jul. 2024, doi: 10.62411/faith.2024-15.
[19] M. R. R. Rana, A. Nawaz, T. Ali, A. M. El-Sherbeeny, and W. Ali, "A BiLSTM-CF and BiGRU-based Deep Sentiment Analysis Model to Explore Customer Reviews for Effective Recommendations," *Eng. Technol. Appl. Sci. Res.*, vol. 13, no. 5, pp. 11739–11746, Oct. 2023, doi: 10.48084/etasr.6278.
[20] W. Gu, W. Aishanl, W. Sun, and Q. Yang, "Relational Graph Attention Network Based on Bi-Directional Attention Flow for Aspect-based Sentiment Analysis," in *2022 European Conference on Natural Language Processing and Information Retrieval (ECNLPIR)*, Jul. 2022, pp. 69–77. doi: 10.1109/ECNLPIR57021.2022.00025.
[21] M. Arief and M. B. M. Deris, "Text Preprocessing Impact for Sentiment Classification in Product Review," in *2021 Sixth International Conference on Informatics and Computing (ICIC)*, Nov. 2021, pp. 1–7. doi: 10.1109/ICIC54025.2021.9632884.
[22] M. IŞIK and H. DAĞ, "The impact of text preprocessing on the prediction of review ratings," *TURKISH J. Electr. Eng. Comput. Sci.*, vol. 28, no. 3, pp. 1405–1421, May 2020, doi: 10.3906/elk-1907-46.
[23] P. H. Hussan and S. M. Mangj, "BERTPHIURL : A Teacher-Student Learning Approach Using DistilRoBERTa and RoBERTa for Detecting Phishing Cyber URLs," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 4, 2025, doi: 10.62411/faith.3048-3719-71.
[24] P. Kralj Novak, J. Smailović, B. Sluban, and I. Mozetič, "Sentiment of Emojis," *PLoS One*, vol. 10, no. 12, p. e0144296, Dec. 2015, doi: 10.1371/journal.pone.0144296.
[25] G. Zhao, Z. Liu, Y. Chao, and X. Qian, "CAPER: Context-Aware Personalized Emoji Recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 9, pp. 3160–3172, Sep. 2021, doi: 10.1109/TKDE.2020.2966971.
[26] C. Biancalana, F. Gasparetti, A. Micarelli, and G. Sansonetti, "An approach to social recommendation for context-aware mobile services," *ACM Trans. Intell. Syst. Technol.*, vol. 4, no. 1, pp. 1–31, Jan. 2013, doi: 10.1145/2414425.2414435.
[27] A. Tabassum and R. R. Patil, "A Survey on Text Pre-Processing & Feature Extraction Techniques in Natural Language Processing," *Int. Res. J. Eng. Technol.*, vol. 7, no. 6, pp. 4864–4867, 2020, [Online]. Available: www.irjet.net
[28] R. Jayanto, R. Kusumaningrum, and A. Wibowo, "Aspect-based sentiment analysis for hotel reviews using an improved model of long short-term memory," *Int. J. Adv. Intell. Informatics*, vol. 8, no. 3, p. 391, Nov. 2022, doi: 10.26555/ijain.v8i3.691.
[29] M. M. Kabir, Z. A. Othman, and M. R. Yaakub, "A Hybrid Frequency Based, Syntax, and Conditional Random Field Method for Implicit and Explicit Aspect Extraction," *IEEE Access*, vol. 12, no. May 2024, pp. 72361–72373, 2024, doi: 10.1109/ACCESS.2024.3403479.
[30] S. A. Rahin, T. Hasib, and M. Hassan, "Aspect-Based Sentiment Analysis Using SemEval and Amazon Datasets," in *2022 Fifth International Conference of Women in Data Science at Prince Sultan University (WiDS PSU)*, Mar. 2022, pp. 85–90. doi: 10.1109/WiDS-PSU54548.2022.00029.

[31] J.-Y. Ho, G. Ju, S. Hong, J. An, and C. C. Lee, "Factors influencing customer satisfaction with AR shopping assistant applications in e-commerce: an empirical analysis utilizing text-mining techniques," *Aslib J. Inf. Manag.*, vol. 77, no. 2, pp. 239–259, Mar. 2025, doi: 10.1108/AJIM-03-2023-0089.

[32] E. Sung, W. Y. Chung, and D. Lee, "Factors that affect consumer trust in product quality: a focus on online reviews and shopping platforms," *Humanit. Soc. Sci. Commun.*, vol. 10, no. 1, p. 766, Nov. 2023, doi: 10.1057/s41599-023-02277-7.

[33] W. Deng, T. Su, Y. Zhang, and C. Tan, "Factors Affecting Consumers' Online Choice Intention: A Study Based on Bayesian Network," *Front. Psychol.*, vol. 12, Oct. 2021, doi: 10.3389/fpsyg.2021.731850.

[34] B. Daroch, G. Nagrath, and A. Gupta, "A study on factors limiting online shopping behaviour of consumers," *Rajagiri Manag. J.*, vol. 15, no. 1, pp. 39–52, Apr. 2021, doi: 10.1108/RAMJ-07-2020-0038.

[35] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," *WIREs Data Min. Knowl. Discov.*, vol. 8, no. 4, Jul. 2018, doi: 10.1002/widm.1253.

[36] M. Pontiki *et al.*, "SemEval-2016 Task 5: Aspect Based Sentiment Analysis," in *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, 2016, pp. 19–30. doi: 10.18653/v1/S16-1002.

[37] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "SemEval-2014 Task 4: Aspect Based Sentiment Analysis," in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, 2014, no. SemEval, pp. 27–35. doi: 10.3115/v1/S14-2004.