Journal of Future Artificial Intelligence and Technologies

E-ISSN: 3048-3719

Research Article

Comprehensive Exploration of Machine and Deep Learning Classification Methods for Aspect-Based Sentiment Analysis with Latent Dirichlet Allocation Topic Modeling

De Rosal Ignatius Moses Setiadi [1],*, Dhendra Marutho [2], and Noor Ageng Setiyanto [1]

[1] Faculty of Computer Science, Dian Nuswantoro University, Semarang, Indonesia;
e-mail : moses@dsn.dinus.ac.id;

[2] Departement of Informatics, Universitas Muhammadiyah Semarang, Semarang, Indonesia;
e-mail : dhendra.vibiano@gmail.com

* Corresponding Author : De Rosal Ignatius Moses Setiadi

Abstract: This research explores the effectiveness of machine learning (ML) and deep learning (DL) classification methods in Aspect-Based Sentiment Analysis (ABSA) on product reviews, incorporating Latent Dirichlet Allocation (LDA) for topic modeling. Using the Amazon reviews dataset, this research tests models such as Naive Bayes (NB), Support Vector Machines (SVM), Random Forest (RF), Long Short-Term Memory (LSTM), and Gated Recurrent Units(GRU). Important aspects such as the product's quality, practicality, and reliability are discussed. The results show that the RF and DL models provide competitive performance, with the RF achieving an accuracy of up to 94.50% and an F1 score of 95.45% for the reliability aspect. The study's conclusions emphasize the importance of selecting an appropriate model based on specifications and data requirements for ABSA, as well as recognizing the need to strike a balance between accuracy and computational efficiency.

Keywords: Aspect-Based Sentiment Analysis; Aspect Extraction; LDA Topic Modelling; Natural Language Processing; Product Reviews Sentiment.

Received: April, 20th 2024
Revised: May, 16th 2024
Accepted: May, 21st 2024
Published: May, 22nd 2024
Curr. Ver.: July, 3rd 2024

1. Introduction

Aspect-based sentiment Analysis (ABSA) is an evolution of traditional sentiment analysis that offers a more detailed and specific approach to interpreting product reviews[1]. Traditional sentiment analysis generally only assesses overall polarity[2]. In the case of product reviews using ABSA allows sentiment analysis regarding specific aspects of the product, such as quality, price or customer service. The advantages of this technique include providing more specific information for product development and marketing strategies, increasing accuracy in understanding consumer sentiment, and the ability to target product improvement areas more effectively[2]–[5]. For example, if there is a review, "This camera has excellent resolution, but the battery drains quickly," the model will learn to identify "excellent resolution" as a positive sentiment related to the "camera" and "battery drains quickly" as a negative sentiment. which is related to the aspect of "battery." This approach allows companies to gain a more specific analysis of what consumers like and don't like about their products, providing more accurate and in-depth insights for product improvements and more effective marketing strategies.

One of the important stages in ABSA is topic modeling. This is an important stage to discover hidden topics or aspects in a large body of text. One popular method used is Latent Dirichlet Allocation (LDA)[6]–[9]. This method can reveal the main topics in the review, thus further increasing understanding of the various aspects discussed in the review. Next, experts are needed to determine further aspects to be researched. By integrating LDA topic modeling with ABSA, this research aims to provide a more comprehensive analysis that not only classifies sentiment but also identifies common topics in reviews.

Various models have been adopted to improve understanding of the nuances of sentiment regarding specific aspects of product reviews. Machine learning (ML) models such as

Naive Bayes (NB)[10]–[12], Support Vector Machines (SVM)[10]–[15], Random Forest (RF)[10]–[12], [16], as well as deep learning (DL) based on Recurrent Neural Networks (RNN)[14] such as Long Short-Term Memory (LSTM) and Gated Recurrent Units(GRU), and Transformer-based models have been widely used. NB is renowned for its speed and efficiency in handling large datasets, although it often suffers from a lack of accuracy when dependencies between features are significant. There are two types of NB, namely Gaussian NB (GNB) and Multinomial NB (MNB), two variants of the Naive Bayes classification algorithm that differ based on the assumed probability distribution of their features. GNB, the default NB option, works assuming that features follow a Gaussian distribution and is better suited to continuous data assuming a Gaussian distribution. In contrast, MNB can assume a multinomial distribution on features, particularly effective for data consisting of frequencies, such as words in text, making it a popular choice for document classification and sentiment analysis[17].

SVMs are highly regarded for their ability to clearly separate categories of data but can be inefficient on huge datasets. Random Forest improves this by providing more stable results and being able to handle data variability well, but it can be very complex and require large computing resources. LSTM and GRU provide advantages in recognizing context and long-term dependencies in text, which makes them ideal for tasks such as ABSA[18]–[23]. They allow models to understand more complex nuances of language but often require a lot of data and significant training time.

Finally, Transformer-based models, such as BERT, offer major advances in natural language processing by using attention mechanisms to understand the context of words in a sentence more effectively. These models have reached the state-of-the-art on many NLP tasks, including ABSA[2], [5], [24], [25]. Although very powerful, these models are often computationally resource-intensive and can overfit if not tuned carefully on smaller datasets. Each model has its place depending on the specific use case, data availability, and computing resources.

This research explores, compares, and analyzes various ML and DL methods on product review datasets, incorporating LDA for topic modeling to uncover hidden themes in the reviews.. Meanwhile, the transformer method is not used because it is related to the limited resources used. The remainder of this paper discusses the related works that inspired this research in section 2. The method and detailed explanation of the stages are in section 3, the results and discussion containing an explanation of the dataset, the results of the overall sentiment analysis based on aspects, and the conclusion presented in the last section.

2. Related Works

Research [3] developed an aspect-based sentiment analysis system for e-commerce product reviews, using unsupervised machine learning techniques such as Lexicon uni-gram and bi-gram, as well as supervised techniques such as SVM. The goal is to extract and classify reviews as positive or negative, allowing consumers and manufacturers to understand customer opinions of products better. SVM showed the best performance with an accuracy of 84%, standing out as the most effective method in assessing sentiment at the aspect level of reviews, which is beneficial in providing more detailed and relevant insights for purchasing decisions.

Research regarding ABSA for product reviews was also carried out by [26]. This research proposes fastText word embedding to avoid the Out of Vocabulary problem in the dataset as well as GRU for aspect distribution detection. Sentiment classification on aspects using the Memory Network method. Experimental results show that aspect-based sentiment classification predictions have an accuracy of 83%, which is higher than the overall classification prediction of 78%, indicating that aspect-based sentiment analysis can improve model performance in product review classification predictions.

Another study [14] discusses the ABSA of hotel reviews using two machine learning approaches, namely Deep RNN and SVM. Both models are trained with lexical, word, syntactic, morphological, and semantic features. The research results show that the SVM approach performs better than RNN in three main tasks: aspect category identification, opinion target expression (OTE) extraction, and sentiment polarity identification. Numerically, the SVM achieved an F1 score of 93.4% for aspect category identification, 89.8% for OTE, and 95.4% accuracy for polarity identification. In comparison, the RNN recorded an F1 score of

48% for aspect and OTE category identification, as well as an accuracy of 87% for polarity identification. Although SVM is more accurate, RNN shows faster training and testing speed, especially for OTE extraction tasks. The research also discusses potential improvements to this approach by using other neural networks, such as LSTM, for future work.

Study [10] developed the ABSA model for film reviews using machine learning techniques. The dataset consists of 2800 film reviews collected from YouTube via the YouTube Data API. These reviews are then processed and labeled into positive or negative aspects by experts based on certain aspects of the reviewed film. Several machine learning algorithms are applied, including RF, Logistic Regression (LR), SVM, and MNB. RF and MNB show excellent performance with an accuracy of 88%, while LR and SVM achieve an accuracy of around 87%.

Based on the related studies, various ML methods, such as MNB, SVM, and RF, as well as DL, such as LSTM and GRU, have been applied in several ABSA studies. RF and MNB show good performance in [10], while SVM is also sufficient and used in research [3], [10], [14]. GRU has performed well in research [26], and LSTM is considered to have good potential in research [14]. So, this research proposes to explore and compare the five methods to compare their performance in ABSA product reviews.

3. Method

This research proposes to analyze three ML models and two DL models. In general, the proposed stages are illustrated in Figure 1.

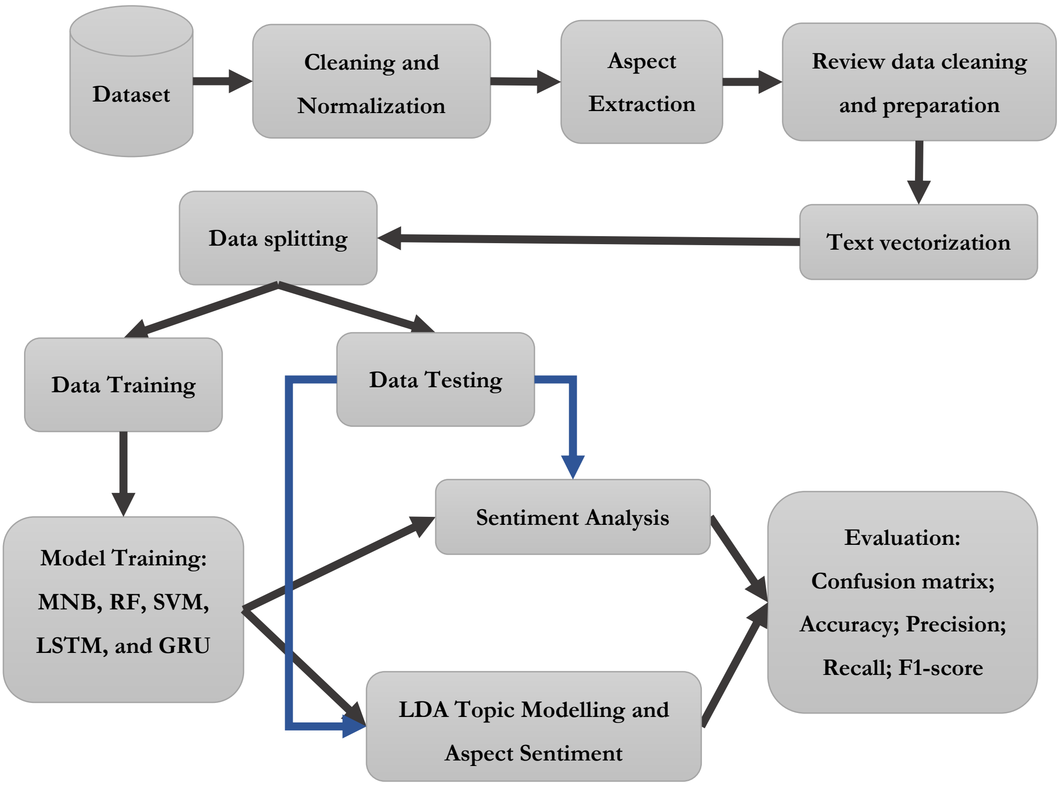

Figure 1. ABSA method stages in this research

Based on Figure 1, the stages above are explained in more detail as follows:

1. Text cleaning and normalization: The input dataset is read using a data frame, then a text cleaning process is carried out to clean the text from HTML tags, remove non-alphabetic characters, and carry out normalization by changing all text to lowercase and tokenization by breaking the text into words -say. These processes help to reduce noise and facilitate analysis. These commands are presented in Figure 2 as Python code snippets, where text is the output variable.

```
text = re.sub(r'<[^>]+>', '', text)
text = re.sub(r'[^a-zA-Z\s]', '', text, re.I|re.A)
text = text.lower().strip().split()
```

Figure 2. Command in cleaning text function.

2. Aspect extraction: In this section, the extract_aspects function is created to search for keywords that determine aspects in text tokens and collect them as relevant aspects. These aspects are defined in the aspect_terms list.
3. Review data cleaning and preparation: The new cleaned_reviews column in the data frame is filled with reviews that have been cleaned. The aspects column contains the aspects extracted from each review.
4. Text vectorization: Cleaned review text is converted into a numeric vector using scikit-learn's CountVectorizer, which converts text into numeric features that a machine learning model can process. This vector describes the frequency of occurrence of words in a document, which is very useful for text classification models.
5. Data splitting and model training: Data is divided into training and testing sets with a composition of 80% and 20%. Five models are trained on training data, namely NB, RF, SVM, LSTM, and GRU. The configuration of each model is presented in Table 1.

Table 1. Models Configuration.

Model	Configuration
MultinomialNB	Default mode
Random Forest	n_estimators=100; random_state=42
SVM	kernel='linear'
LSTM	input_dim=1000; hidden_dim=50; num_layers=2; dropout_rate=0.5; optimizer=Adam; learning_rate= 0.001; batch_size=32
GRU	input_dim=1000; hidden_dim=50; num_layers=2; dropout_rate=0.5; optimizer=Adam; learning_rate= 0.001; batch_size=32

6. Topic modelling: Topic modeling with Latent Dirichlet Allocation (LDA) involves several main steps. First, the review data is loaded and cleaned of unwanted characters, converted to lowercase, and tokenized. The stop words were then removed. Next, a dictionary and corpus are created for the LDA model. A dictionary contains a unique list of words, while a corpus is a numerical representation of a document. The LDA model is then trained using the dictionary and corpus to identify key review topics. Once topics are identified, relevant aspects are extracted from them. These aspects were then used to identify key aspects in the review, and the frequency of occurrence of each aspect was calculated for further analysis. This process helps reveal hidden topic structures in large text datasets and can be used for aspect analysis in sentiment analysis. Next, words that are appropriate to the topic or aspect the expert wants to research are selected.
7. Evaluate the sentiment and aspect-based: testing is done by calculating the confusion matrix, accuracy, precision, recall, and F1 score. A confusion matrix measures the number of true positives, false positives, true negatives, and false negatives. Ac-curacy calculates the proportion of correct predictions to the total number of cases. Precision calculates the proportion of positive predictions that are truly positive. Recall/Sensitivity calculates the proportion of actual positives detected. F1-Score is the harmonic mean of precision and recall. For sentiment, it is calculated on test data, while aspect-based data uses a subset of data based on the presence of aspect keywords, then makes predictions and calculates evaluation metrics.

4. Results and Discussion

4.1. Dataset

This research was implemented using Python and Google Collabs as the IDE. Some important libraries used for data manipulation, statistical modeling, visualization, and text analysis are Pandas, sklearn, matplotlib, seaborn, and textblob. This research uses the Amazon reviews dataset, downloaded from the URL https://www.kaggle.com/datasets/lievgarcia/amazon-reviews. This dataset has been preprocessed to consist of two positive and negative sentiments, each with 10,500 reviews, with no missing values and duplicate data. Additionally, 30 product categories with 700 reviews indicate a balanced distribution across categories. Visualization of the dataset distribution is presented in Figures 3 and 4.

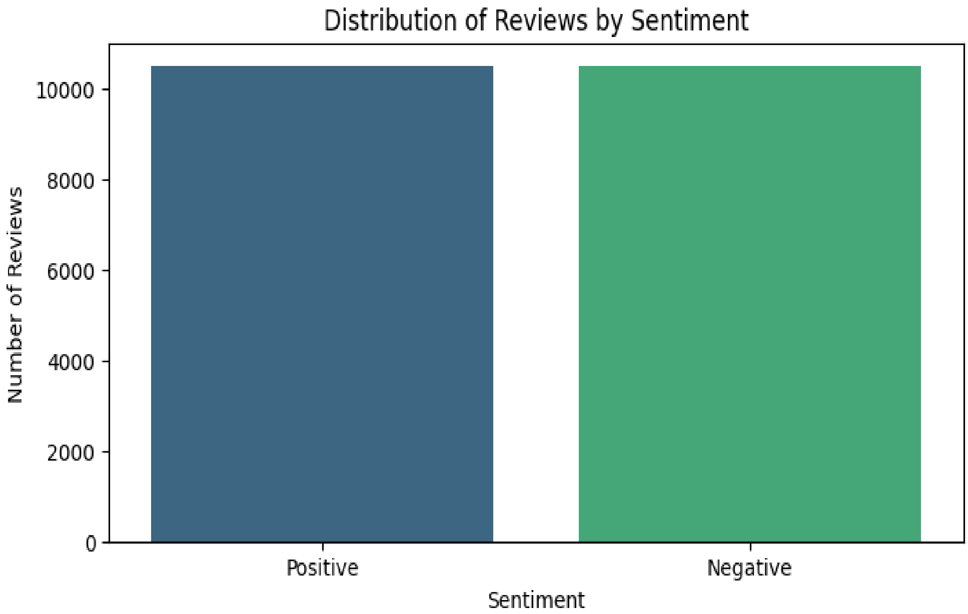

Figure 3. Distribution of Sentimen Reviews.

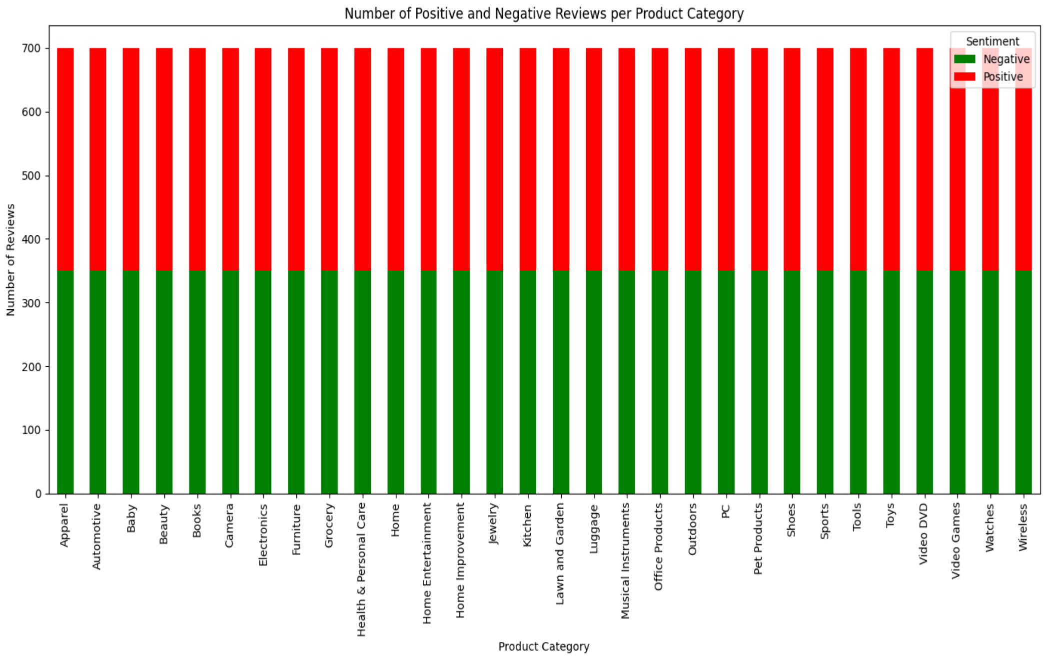

Figure 4. Distribution of product categories of Amazon reviews.

4.2. Training and Testing for Sentiment Analysis

As explained in section three, three machine learning models, namely MultinomialNB, RF, and SVM, were tested, and two deep learning models, LSTM and GRU, were also tested. The configuration of each model is described in Table 1. For training data, 80% of the total dataset was used. The training for the two DL models was conducted with 100 epochs. The LSTM achieved a training accuracy of 97.87% and a loss of 0.0582, while the GRU achieved an accuracy of 97.39% and a loss of 0.0675. The epoch training plots for LSTM and GRU can be viewed in Figures 5 and 6, respectively. For testing evaluation, 20% of the dataset was used, and the overall test results for accuracy, precision, recall, and F1-score of all five models are presented in Table 2, with the corresponding confusion matrices shown in Figure 7.

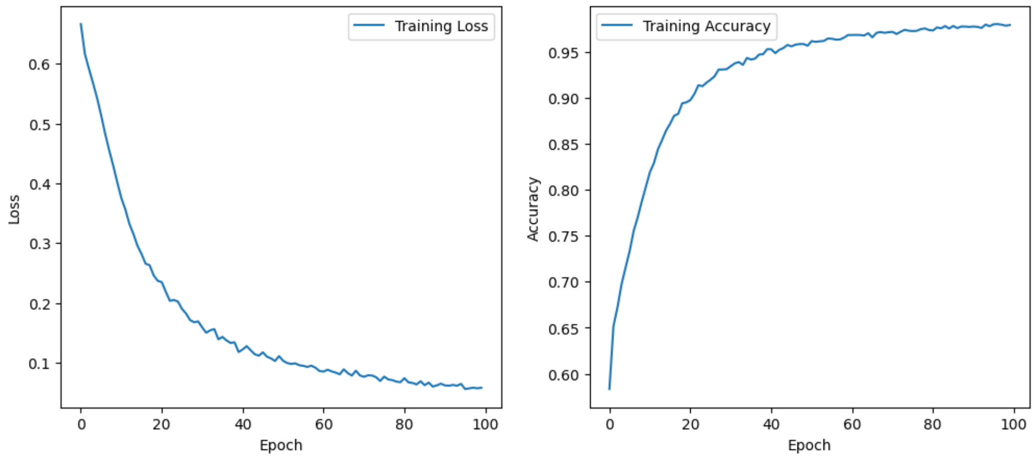

Figure 5. Training loss and accuracy plot of LSTM.

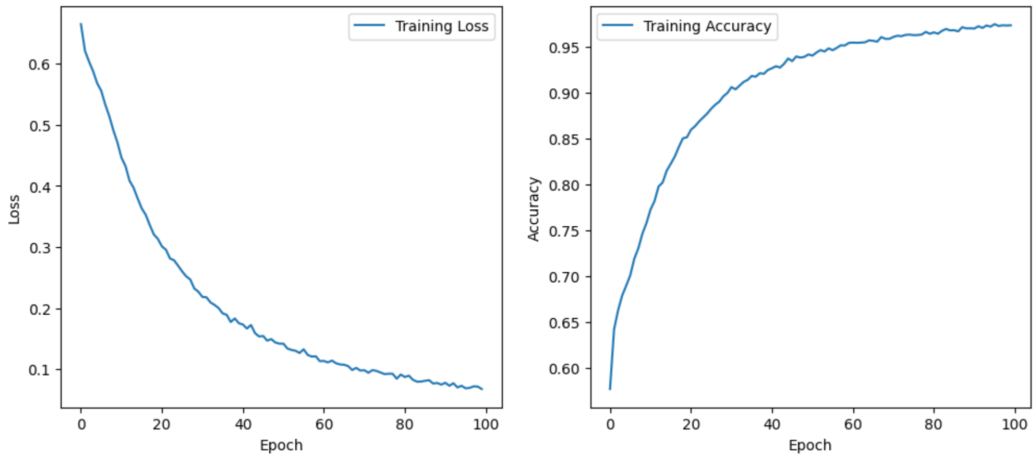

Figure 6. Training loss and accuracy plot of GRU.

Table 2. Results of Sentiment Analysis.

Method	Accuracy	Precision	Recall	F1 score	Training Time (s)
MNB	**63.93%**	62.83%	**69.46%**	**65.98%**	**0.013**
RF	63.52%	**64.90%**	60.05%	62.38%	59.223
SVM	60.17%	59.95%	62.98%	61.42%	1099.573
LSTM	60.91%	60.50%	60.58%	60.54%	415.802 (100 epoch)
GRU	61.04%	61.68%	61.55%	61.62%	255.309(100 epoch)

Before analyzing the results in Table 2, it is important to consider that the dataset used is balanced. This allows for a greater bias towards accuracy metrics as they effectively reflect the model's capabilities. Apart from that, the F1 Score is also relevant because it combines precision and recall, which is important in evaluating models in balanced dataset conditions, where each class has the same weight in the overall evaluation. The results presented in Table 2 show that no one method stands out significantly, but several important observations can be concluded. MNB and RF both show more effective performance based on accuracy and F1 Score. However, MNB has a very fast time processing 21,000 data records, followed by

RF, which has a training time of less than 1 minute. SVM takes a much longer time compared to other methods, and its performance in sentiment analysis is relatively weak. The performance of DL methods such as LSTM and GRU, in this case, is also not better than that of RF and MNB, but all the metrics are relatively balanced. The training process per epoch is also only around 4 seconds for LSTM and 2 seconds for GRU. Nevertheless, none of the methods achieves a very high level of accuracy, indicating room for further improvement and optimization.

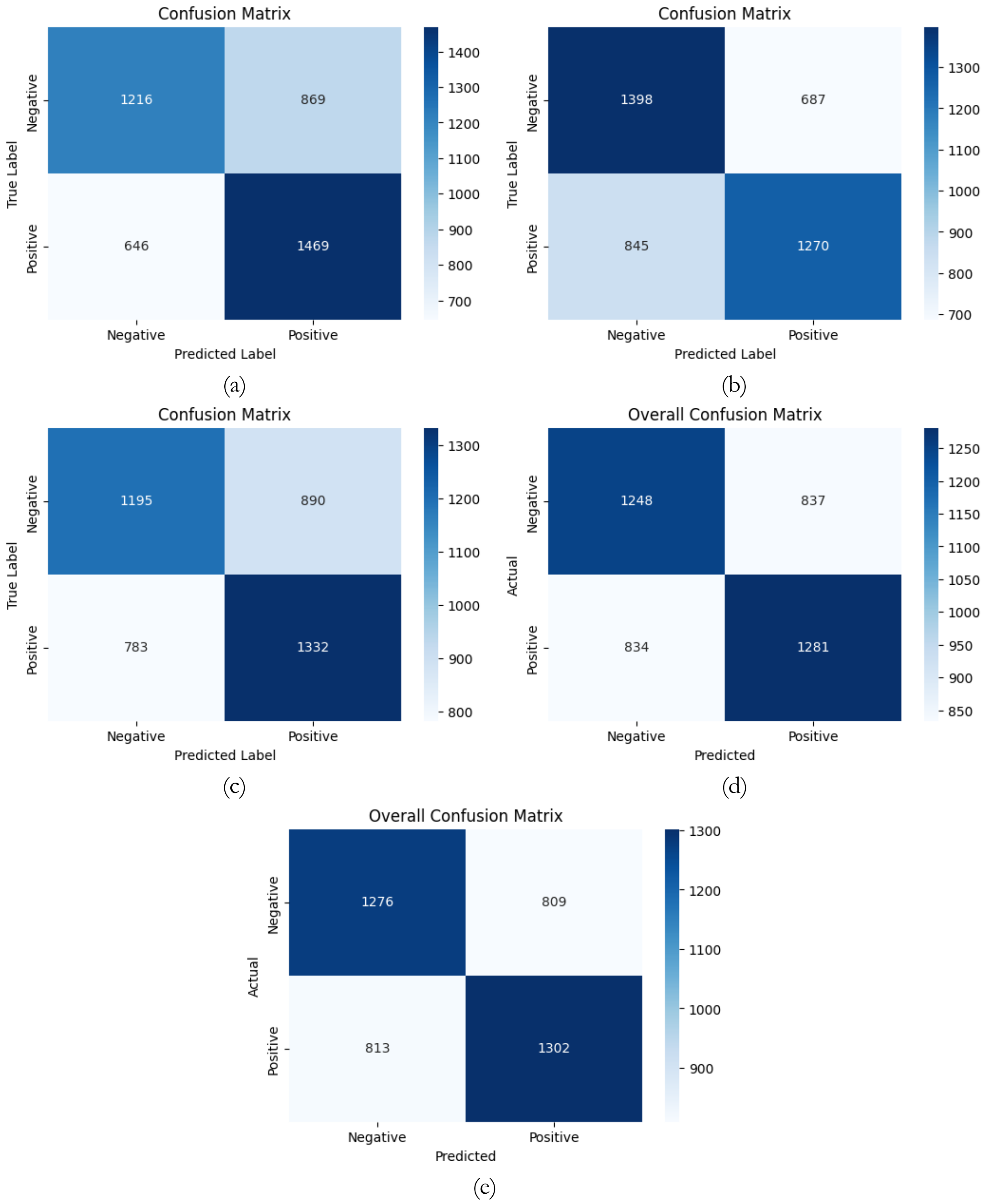

Figure 7. Confusion matrix of sentiment analysis (**a**) Multinomial Naïve Bayes; (**b**) Random Forests; (**c**) SVM; (**d**) LSTM; (**e**) GRU.

4.2. Aspect Analysis Comparison Results

As explained in section three, topic modeling was conducted using LDA, resulting in the identification of the following aspects: 'first', 'black', 'flavor', 'water', 'air', 'loves', 'power', 'one', 'speaker', 'would', 'tube', 'light', 'ink', 'read', 'story', 'case', 'sink', 'taste', 'gun', 'hot', 'like', 'use', 'really', 'im', 'works', 'ice', 'skin', 'recommend', 'chain', 'kitchen', 'bar', 'mouse', 'printer', 'game', 'tea', 'size', 'small', 'food', 'bought', 'little', 'shower', 'oil', 'dont', 'ring', 'cup', 'games', 'tv', 'nice', 'luggage', 'phone', 'quality', 'shoe', 'well', 'good', 'camera', 'get', 'bag', 'time', 'bulb', 'temperature', 'movie', 'drink', 'watch', 'weather', 'battery', 'gold', 'bottle', 'print', 'play', 'sound', 'book', 'love', 'bluray', 'weight', 'price', 'great', 'screen', 'old', 'coffee', 'clip', 'clean', 'easy', 'product', 'fit'. LDA visualization presented in Figure 8.

Then experts selected these aspects from this list to ensure relevance and significance. The chosen aspects for detailed analysis are ‘quality’, ‘great’, ‘good’, ‘use’, ‘easy’, and ‘works’.” Here experts divide them into several main aspects, namely quality, practicality or usability, and reliability. The quality aspect is generally reflected by words such as “quality”, “great”, and “good”. The practicality or usability aspect is reflected by words like “use” and “easy”, and the aspect of reliability is reflected by words like “works”. Evaluation of the frequency of appearance of these aspects is presented in Figure 9.

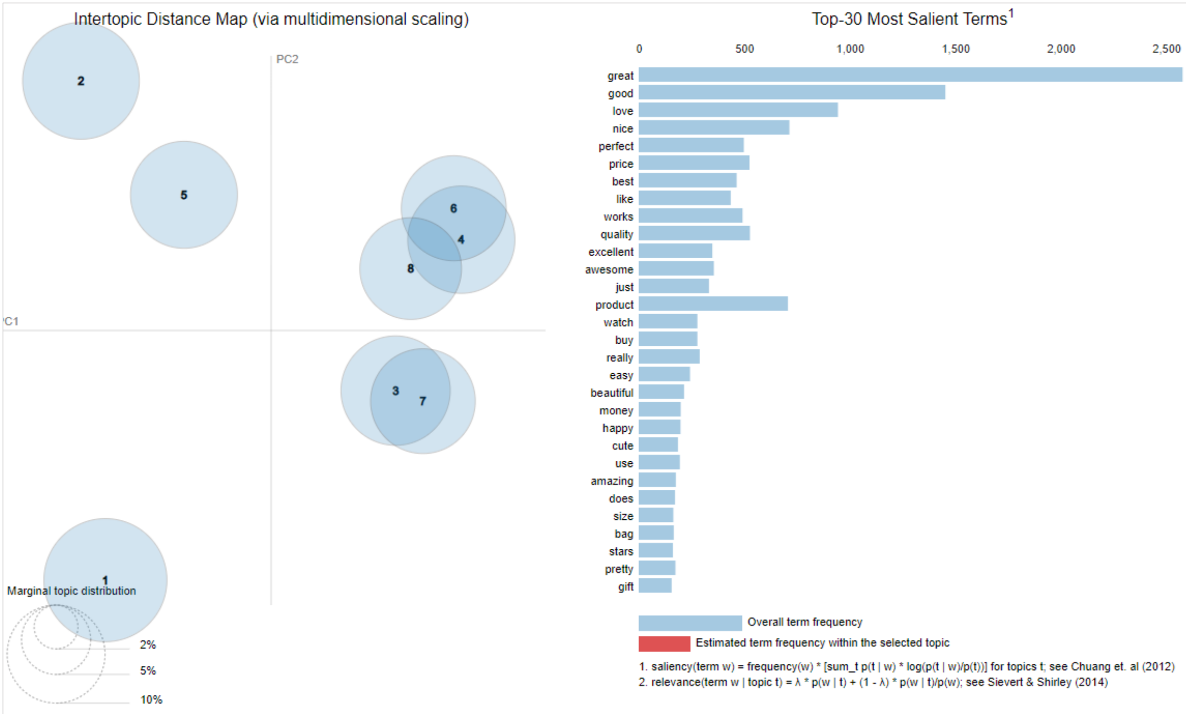

Figure 8. LDA topic modelling visualization.

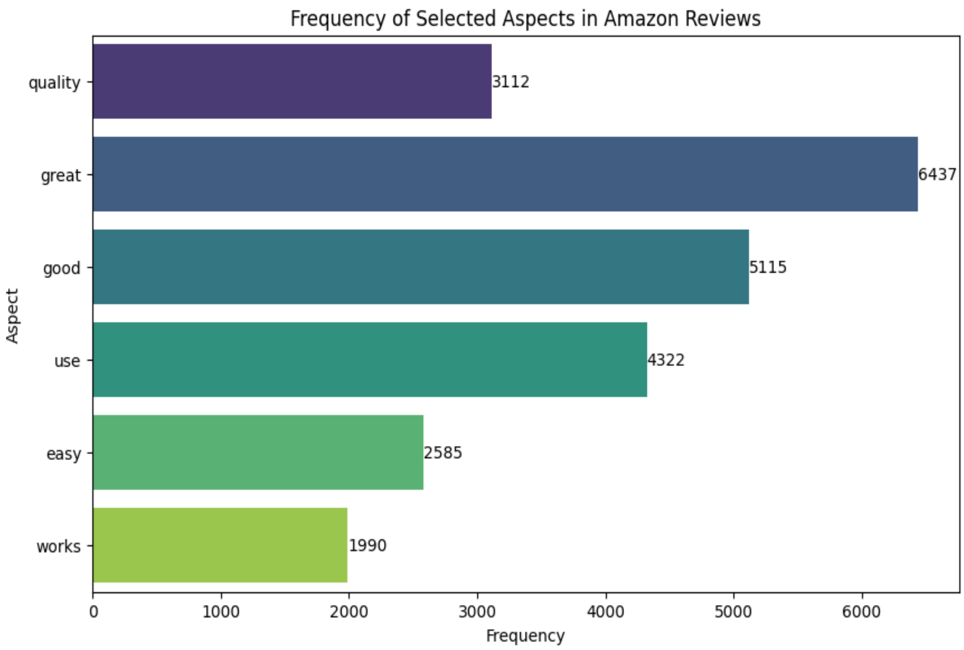

Figure 9. Aspect frequency in Amazon reviews.

Table 3. ABSA results for "quality" for the quality aspect.

Method	Accuracy	Precision	Recall	F1 score
MNB	81.54%	79.38%	92.46%	85.42%
RF	**93.51%**	93.54%	95.50%	94.51%
SVM	89.36%	89.48%	92.71%	91.07%
LSTM	<u>92.73%</u>	<u>93.98%</u>	**93.56%**	**93.77%**
GRU	92.66%	**94.09%**	<u>93.30%</u>	<u>93.70%</u>

Table 4. ABSA results for "great" for the quality aspect.

Method	Accuracy	Precision	Recall	F1 score
MNB	77.11%	73.21%	87.49%	79.71%
RF	**92.45%**	**92.86%**	92.42%	**92.64%**
SVM	88.04%	87.03%	90.17%	88.58%
LSTM	91.86%	91.66%	**92.63%**	92.15%
GRU	<u>92.25%</u>	<u>92.52%</u>	<u>92.44%</u>	<u>92.48%</u>

Table 5. ABSA results for "good" for the quality aspect.

Method	Accuracy	Precision	Recall	F1 score
MNB	79.88%	76.43%	87.40%	81.55%
RF	**93.22%**	**93.94%**	**92.64%**	**93.28%**
SVM	88.76%	88.30%	89.81%	89.05%
LSTM	92.41%	93.26%	<u>91.65%</u>	92.45%
GRU	<u>92.60%</u>	<u>93.84%</u>	91.41%	<u>92.61%</u>

Table 6. ABSA results for "use" for the practicality aspect.

Method	Accuracy	Precision	Recall	F1 score
MNB	81.25%	78.95%	83.93%	81.37%
RF	**93.10%**	**93.92%**	91.80%	**92.84%**
SVM	90.12%	88.96%	91.04%	89.98%
LSTM	92.43%	92.30%	<u>92.01%</u>	92.15%
GRU	<u>92.95%</u>	<u>92.80%</u>	**92.57%**	<u>92.69%</u>

Table 7. ABSA results for "easy" for the practicality aspect.

Method	Accuracy	Precision	Recall	F1 score
MNB	80.04%	77.30%	86.47%	81.62%
RF	**93.15%**	**94.74%**	<u>91.73%</u>	**93.21%**
SVM	88.95%	88.45%	90.23%	89.33%
LSTM	92.26%	92.68%	**92.19%**	<u>92.43%</u>
GRU	<u>92.35%</u>	<u>93.83%</u>	91.06%	<u>92.43%</u>

Table 8. ABSA results for "works" for the reliability aspect.

Method	Accuracy	Precision	Recall	F1 score
MNB	78.89%	75.08%	82.85%	78.77%
RF	<u>93.69%</u>	<u>94.14%</u>	92.41%	<u>93.26%</u>
SVM	89.66%	88.89%	89.30%	89.09%
LSTM	93.57%	93.85%	<u>92.51%</u>	93.17%
GRU	**94.19%**	**94.78%**	**92.86%**	**93.81%**

The results shown in Tables 3 to 8 indicate several important findings. RF consistently shows excellent performance, with high accuracy and F1 scores across all analyzed aspects, confirming its effectiveness in dealing with variations in review data, which is also reflected in the visualization in Figure 8. Specifically, RF achieved the highest accuracy and F1 scores in almost all aspects, demonstrating its robustness and reliability in ABSA. DL models such as LSTM and GRU also show competitive results, underscoring their ability to understand the deeper context and nuances of the text, which is critical in sentiment analysis. The LSTM and GRU models exhibited high accuracy and F1 scores, particularly in the "quality" and "practicality" aspects, indicating their effectiveness in capturing complex patterns in the data. GRU, in particular, showed the highest F1 score in the "works" aspect.

Meanwhile, although MNB performs better in sentiment analysis with speedy training times, its performance in aspect analysis is lower compared to the RF and DL models in several aspects. MNB performed adequately in the "use" and "easy" aspects but lagged behind RF and DL models in terms of accuracy and F1 scores for the "quality" and "works" aspects. SVM does not always achieve the best results compared to other methods, but it shows reasonable performance in specific aspects. SVM demonstrated reasonable accuracy and F1 scores but required significantly longer training times than other models.

In general, although no one method dominates in all conditions, these findings indicate that the choice of model must be tailored to specific needs and data conditions. RF and DL models stand out for their complexity in capturing text nuances, while MNB offers an efficient solution for situations that require speed. This underscores the importance of balancing accuracy and computational efficiency in method selection for ABSA.

5. Conclusions

This research succeeded in identifying and comparing the performance of various ML and DL methods in the context of ABSA, with the addition of LDA for topic modeling. The integration of LDA provided deeper insights by uncovering hidden themes in the reviews, which enhanced the aspect extraction process. The results show that Random Forest and DL models such as LSTM and GRU consistently perform well, reflecting their effectiveness in analyzing aspect-based sentiment by understanding the context and nuances of text more deeply. However, no method is significantly superior in all situations, indicating the need to adapt methods based on specific needs and data conditions. Additionally, these results offer insight into the importance of balancing accuracy and computational efficiency in selecting analytical methods. This research also recognizes some limitations, such as excluding the Transformer model due to limited resources, and recommends further investigation with more diverse models to optimize aspect-based sentiment analysis.

Author Contributions: Conceptualization: D.R.I.M.S and D.M.; Methodology, D.R.I.M.S and D.M.; Software: D.R.I.M.S.; Validation: D.R.I.M.S, D.M., and N.A.S; Formal analysis: N.A.S.; Investigation: D.M.; Resources: D.R.I.M.S.; Data curation: D.R.I.M.S.; Writing—original draft preparation: D.R.I.M.S.; Writing—review and editing: D.M., and N.A.S; Visualization: D.R.I.M.S and D.M.; Supervision: D.R.I.M.S.; Project administration: N.A.S; Funding acquisition: All.

Funding: This research received no external funding.

Data Availability Statement: The Amazon reviews dataset is a public dataset that can be downloaded from the URL https://www.kaggle.com/datasets/lievgarcia/amazon-reviews.

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

References

[1] M. Hoang, O. Alija Bihorac, and J. Rouces, "Aspect-Based Sentiment Analysis Using BERT," *Proc. 22nd Nord. Conf. Comput. Linguist.*, pp. 187–196, 2019, [Online]. Available: https://aclanthology.org/W19-6120

[2] 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.

[3] A. P. Rodrigues and N. N. Chiplunkar, "Aspect Based Sentiment Analysis on Product Reviews," *14th Int. Conf. Inf. Process. Internet Things, ICInPro 2018 - Proc.*, 2018, doi: 10.1109/ICINPRO43533.2018.9096796.

[4] W. Zhang, H. Xu, and W. Wan, "Weakness Finder: Find product weakness from Chinese reviews by using aspects based sentiment analysis," *Expert Syst. Appl.*, vol. 39, no. 11, pp. 10283–10291, 2012, doi: 10.1016/j.eswa.2012.02.166.

[5] D. Marutho, Muljono, S. Rustad, and Purwanto, "Optimizing aspect-based sentiment analysis using sentence embedding transformer, bayesian search clustering, and sparse attention mechanism," *J. Open Innov. Technol. Mark. Complex.*, vol. 10, no. 1, p. 100211, Mar. 2024, doi: 10.1016/j.joitmc.2024.100211.

[6] H. Jelodar *et al.*, "Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey," *Multimed. Tools Appl.*, vol. 78, no. 11, pp. 15169–15211, Jun. 2019, doi: 10.1007/s11042-018-6894-4.

[7] E. S. Negara, D. Triadi, and R. Andryani, "Topic Modelling Twitter Data with Latent Dirichlet Allocation Method," in *2019 International Conference on Electrical Engineering and Computer Science (ICECOS)*, Oct. 2019, pp. 386–390. doi: 10.1109/ICECOS47637.2019.8984523.

[8] C. B. Asmussen and C. Møller, "Smart literature review: a practical topic modelling approach to exploratory literature review," *J. Big Data*, vol. 6, no. 1, p. 93, Dec. 2019, doi: 10.1186/s40537-019-0255-7.

[9] Z. Tong and H. Zhang, "A Text Mining Research Based on LDA Topic Modelling," in *Computer Science & Information Technology (CS & IT)*, May 2016, pp. 201–210. doi: 10.5121/csit.2016.60616.

[10] O. G. Horsa and K. K. Tune, "Aspect-Based Sentiment Analysis for Afaan Oromoo Movie Reviews Using Machine Learning Techniques," *Appl. Comput. Intell. Soft Comput.*, vol. 2023, 2023, doi: 10.1155/2023/3462691.

[11] D. Chehal, P. Gupta, and P. Gulati, "Evaluating Annotated Dataset of Customer Reviews for Aspect Based Sentiment Analysis," *J. Web Eng.*, Dec. 2021, doi: 10.13052/jwe1540-9589.2122.

[12] S. Haque *et al.*, "Aspect Based Sentiment Analysis in Bangla Dataset Based on Aspect Term Extraction," in *Cyber Security and Computer Science*, 2020, pp. 403–413. doi: 10.1007/978-3-030-52856-0_32.

[13] H. Benarafa, M. Benkhalifa, and M. Akhloufi, "An Enhanced SVM Model for Implicit Aspect Identification in Sentiment Analysis," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 5, pp. 42–53, 2023, doi: 10.14569/IJACSA.2023.0140505.

[14] M. Al-Smadi, O. Qawasmeh, M. Al-Ayyoub, Y. Jararweh, and B. Gupta, "Deep Recurrent neural network vs. support vector machine for aspect-based sentiment analysis of Arabic hotels' reviews," *J. Comput. Sci.*, vol. 27, pp. 386–393, Jul. 2018, doi: 10.1016/j.jocs.2017.11.006.

[15] K. Aurangzeb, N. Ayub, and M. Alhussein, "Aspect Based Multi-Labeling Using SVM Based Ensembler," *IEEE Access*, vol. 9, pp. 26026–26040, 2021, doi: 10.1109/ACCESS.2021.3055768.

[16] M. Imani and S. Noferesti, "Aspect extraction and classification for sentiment analysis in drug reviews," *J. Intell. Inf. Syst.*, vol. 59, no. 3, pp. 613–633, Dec. 2022, doi: 10.1007/s10844-022-00712-w.

[17] 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.

[18] S. Kalbhor and D. Goyal, "Survey on ABSA based on machine learning, deep learning and transfer learning approach," in *AIP Conference Proceedings*, 2023, p. 020041. doi: 10.1063/5.0154549.

[19] H. Gandhi and V. Attar, "Extracting Aspect Terms using CRF and Bi-LSTM Models," *Procedia Comput. Sci.*, vol. 167, pp. 2486–2495, 2020, doi: 10.1016/j.procs.2020.03.301.

[20] Y. Ma, H. Peng, and E. Cambria, "Targeted Aspect-Based Sentiment Analysis via Embedding Commonsense Knowledge into an Attentive LSTM," *Proc. AAAI Conf. Artif. Intell.*, vol. 32, no. 1, Apr. 2018, doi: 10.1609/aaai.v32i1.12048.

[21] 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.

[22] T. U. Tran, H. T. Thi Hoang, and H. X. Huynh, "Aspect Extraction with Bidirectional GRU and CRF," in *2019 IEEE-RIVF International Conference on Computing and Communication Technologies (RIVF)*, Mar. 2019, pp. 1–5. doi: 10.1109/RIVF.2019.8713663.

[23] S. Rani and A. Jain, "Aspect-based sentiment analysis of drug reviews using multi-task learning based dual BiLSTM model," *Multimed. Tools Appl.*, vol. 83, no. 8, pp. 22473–22501, Aug. 2023, doi: 10.1007/s11042-023-16360-3.

[24] P. N. Andono, S. Sunardi, R. A. Nugroho, and B. Harjo, "Aspect-Based Sentiment Analysis for Hotel Review Using LDA, Semantic Similarity, and BERT," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 5, pp. 232–243, Oct. 2022, doi: 10.22266/ijies2022.1031.21.

[25] 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.

[26] H. T. Ismet, T. Mustaqim, and D. Purwitasari, "Aspect Based Sentiment Analysis of Product Review Using Memory Network," *Sci. J. Informatics*, vol. 9, no. 1, pp. 73–83, May 2022, doi: 10.15294/sji.v9i1.34094.