Infact: International Journal of Computers
Vol.9, No.01, Maret 2025
p-ISSN : 2527-8363, e-ISSN: 2829-0259
DOI:

Aspect-Based Sentiment Analysis Using Latent Dirichlet Allocation (LDA) and DistilBERT on Threads App Reviews

Andreas Noprianto Kambayo[1*], Sunneng Sandino Berutu[2], Jatmika[3] , Aristophane Nshimiyimana[4]
[1,2,3,] Informatika Department, Faculty of science and computer, Universitas Kristen Immanuel Yogyakarta
[4,] Department of Computer Science and Information Engineering, National Chin-Yi University of Technology, Taiwan

[1,2,3]Ukrim Rd Cupuwato, Purwomartani, Kalasan, Sleman, D.I Yogyakarta, 55571, Indonesia
[4]No.57, Sec. 2, Zhongshan Rd., Taiping Dist., Taichung 411030, Taiwan (R.O.C.)

E-mail: andreas.n19@student.ukrimuniversity.ac.id[1]

Submitted: 26/02/2025, Revision: 12/03/2025, Accepted: 17/03/2025

Abstract

Threads is a social media application that offers news services and user interaction, integrated with Instagram. Unlike other platforms, Threads does not have features like direct messaging (DM), trending topics, or advertisements. To understand users' opinions about this app, a sentiment analysis based on aspects was conducted on Threads reviews. The steps involved include applying web scraping techniques to collect reviews data from the Play Store. Aspect categories were identified using the Latent Dirichlet Allocation (LDA) algorithm. Sentiment labeling was then performed for positive and negative categories using the DistilBERT method. The results show that the LDA algorithm identified three aspects: Usage and Integration (with 3.147 positive and 8.173 negative reviews), Features and Comparisons (with 1.108 positive and 1.709 negative reviews), and User Experience and Satisfaction (with 3.529 positive and 2.208 negative reviews). The sentiment analysis results indicated 7,784 positive reviews and 12,090 negative reviews. Model performance evaluation using the Confusion Matrix showed an accuracy of 96.71%, precision of 97.24%, recall of 94.48%, and F1-score of 95.84%. Evaluation was also conducted for each aspect, with an accuracy of 96.99%, precision of 96.60%, recall of 92.85%, and F1-score of 94.69% for the Usage and Integration aspect; accuracy of 95.74%, precision of 94.11%, recall of 95.23%, and F1-score of 94.67% for the Features and Comparisons aspect; and accuracy of 96.74%, precision of 95.83%, recall of 99.06%, and F1-score of 97.42% for the User Experience and Satisfaction aspect.

Keywords:
Threads;
Aspect-Based Sentiment Analysis;
DistilBERT;
Latent Dirichlet Allocation;
Natural Language Processing;

*** Corresponding Author: Andreas Noprianto Kambayo**
E-mail: andreas.n19@student.ukrimuniversity.ac.id

1. Introduction

The advancement of the internet has created a virtual space for users to express opinions and reviews about various products and services, including digital applications [1]. Reviews provided by users not only impact potential new users but also serve as an important source of information for app developers in evaluating and improving the quality of their products [2]. One of the newest social media platforms is Threads, which allows users to share thoughts through text and is directly integrated with Instagram. However, this application has limitations compared to other platforms, such as the lack of direct messaging (DM), trending topics, and advertisements. On the other hand, Threads provides a feature called Hidden Words to block specific words or phrases as one of its advantages. The Google Play Store provides an app rating feature that consists of a score (rating) and text reviews. Scores are given on a scale of 1 to 5, while text reviews allow users to provide opinions as feedback on the app. However, there is often a mismatch between rating scores and text reviews, where users give positive reviews but low scores, or vice versa. This mismatch can make it difficult for developers to understand the aspects that need to be fixed or improved in their app.

Aspect-Based Sentiment Analysis (ABSA) aims to understand users' opinions regarding specific aspects of an application. Widiansyah et al. used LDA for topic modeling and IndoBERT for sentiment analysis on the M-Paspor application. The identified aspects were reliability, usability, and efficiency. With a sentiment classification accuracy of 94%, this study demonstrates the effectiveness of a machine learning-based approach in categorizing user opinions [3]. Furthermore, Roiqoh et al. analyzed the sentiment of the Jaminan Kesehatan Nasional (JKN) application using the LDA method for topic modeling, along with Naïve Bayes and lexicon-based approaches for sentiment analysis. The results revealed three main aspects: Service and Features, Registration and Login, and User Satisfaction. The Naïve Bayes method outperformed the lexicon-based approach, achieving an accuracy of 94.75% [4].

Although there have been many studies discussing aspect-based sentiment analysis on various applications, no research has specifically analyzed users' opinions on the Threads application. Furthermore, previous studies have used various methods, such as IndoBERT and Naïve Bayes, in sentiment analysis, but few have explored the use of DistilBERT, which is lighter yet still maintains high accuracy in sentiment classification. Therefore, this study aims to apply LDA-based topic modeling to detect the aspects discussed by users in their reviews of the Threads application and analyze user sentiment toward these aspects using the DistilBERT model (distilbert-base-uncased-sst-2-english).

2. Research Methods

LDA is adopted to identify aspects through topic modeling based on the distribution of word occurrences and its strong relationship with the dominant topics that frequently appear [5]. LDA is applied because it is considered superior in generating logical, easily interpretable topics with good predictive performance [6]. DistilBERT (distilbert-base-uncased-sst-2-english) is used for sentiment classification with positive or negative categories [7]. The research stages are presented in the following Figure 1.

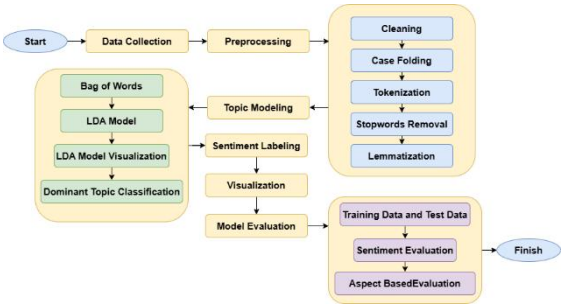

Figure 1. Research Flow Stages

2.1 Web Scraping

The data used in this study consists of user reviews of the Threads application available on the Play Store. The data was collected using scraping techniques [8] [9] with the Google Play Scraper library.

2.2 Preprocessing

The preprocessing steps in this study are as follows:

1) Cleaning is used to remove noise from the text data, such as punctuation, numbers, symbols, or distracting characters, leaving only alphabetic letters [10].
2) Case Folding is used to convert the text into lowercase to ensure a more uniform and consistent format [11].
3) Tokenization is used to split the text into several word units by removing non-alphabetic characters and separating words based on spaces [12].
4) Stopword removal removes words that are less significant in the text, such as "and," "is," and "also," to ensure that the analysis focuses more on important words [10].
5) Lemmatization is used to identify the base form of words by removing prefixes and suffixes [13].

2.3 Latent Dirichlet Allocation

LDA is a probabilistic model representing each topic as a distribution of words in the text, where

words strongly related to the dominant topics tend to appear more frequently. The LDA procedure produces two main distributions: the distribution of topics per word and the distribution of topics per document. This model has been proven to be superior in accuracy for topic modeling [5]. Documents can be understood as collections of hidden topics with different word distributions, which are grouped through a Dirichlet approach [14] [15].

2.4 Bag of Words

The Bag of Words (BoW) model is applied to extract word features in machine learning by representing text as a multiset of words without considering the order or grammar [16]. This model learns the vocabulary from all documents and represents each document based on the frequency of word occurrences [17].

2.5 DistilBERT

DistilBERT is an NLP model developed as a lighter and more efficient version of BERT, with 40% fewer parameters and 60% faster inference speed [7].

2.7 Confusion Matrix

The performance of the DistilBERT model is evaluated using the Confusion Matrix method [19]. This method produces several evaluation parameters, including accuracy, precision, recall, and F1-score.

3. Results and Discussion

3.1 Data Collection

The scraped review data for the Threads application consists of 20,000 reviews, where English reviews are the most relevant. The columns include username, content, score, and date. Table 1 shows the scraped review data of the Threads application on the Google Play Store.

Table 1. Data Scraping Results Review

No	userName	Content	Score	At
0	A Google user	Overall good, but has some issues it needs to ...	3	2024-09-02 09:51:17
1	A Google user	If I could give it 0 stars I would. It may be ...	1	2025-01-05 04:17:42
2	A Google user	Enjoy the easy of communicating with a timelin...	5	2024-12-07 15:05:50
…	…	…	…	…
19995	Bogdan Chiriac	Dead app , very few active real users.	1	2024-05-09 09:52:26
19996	Josh Avondoglio	Better than Twitter. Less trolls, less fake ac...	5	2023-10-17 23:33:48
19997	Dimas Christian	force close every try to upload photo	2	2023-07-07 08:20:14

3.2 Preprocessing

The results of the data preprocessing, which went through stages such as cleaning, case folding, tokenization, stopword removal, and lemmatization, are shown in Table 2, 3, 4.

Table 2. Result Cleaning and Case Folding

No	Content	Clean_content	Lower_content
0	Overall good, but it has some issues it need t...	Overall, it's good but has some issues it needs to w...	overall good but has some issues it needs to w...
1	If I could give it 0 stars I would . It may b...	If I could give it stars I would It may be the...	if i could give it stars i would it may be the...
2	Enjoy the easy of communicating with a timelin...	Enjoy the easy of communicating with a timelin...	enjoy the easy of communicating with a timelin...
…	…	…	…
19995	Dead app , very few active real users .	Dead app very few active real users	dead app very few active real users
19996	Better than Twitter. Less trolls, less fak...	Better than Twitter Less trolls less fake acco...	better than twitter less trolls less fake acco...
19997	force close every try to upload photo	force close every try to upload photo	force close every try to upload photo

Table 3. Result Tokenization and Stopwords Removal

	Tokens_content	Stopwords
0	[overall, good, but, has, some, issues, it, ne...	[overall, good, issues, needs, work, instance,...
1	[if, i, could, give, it, stars, i, would, it, ...	[could, give, stars, would, may, least, intuit...
2	[enjoy, the, easy, of, communicating, with, a,...	[enjoy, easy, communicating, timeline, copy, p...
…	…	…
19995	[dead, app, very, few, active, real, users]	[dead, app, active, real, users]
19996	[better, than, twitter, less, trolls, less, fa...	[better, twitter, less, trolls, less, fake, ac...
19997	[force, close, every, try, to, upload, photo]	[force, close, every, try, upload, photo]

Table 4. Result Lemmatization

	Lemmatization_content	reviews
0	[overall, good, issue, need, work, instance, c...	overall good issue need work instance click po...
1	[give, star, least, intuitive, app, ever, crea...	give star least intuitive app ever create comp...
2	[enjoy, easy, communicate, timeline, copy, pas...	enjoy easy communicate timeline copy paste tex...
…	…	…
19995	[dead, app, active, real, user]	dead app active real user
19996	[well, twitter, less, troll, less, fake, accou...	well twitter less troll less fake account well...
19997	[force, close, try, upload, photo]	force close try upload photo

3.3 Topic Modeling

3.3.1 Bag of Words

Tokens are represented in BoW to transform the text into a numerical vector. This vector contains a vocabulary of unique words along with their indices. The corpus in BoW form counts the frequency of each word's occurrence in the document, resulting in pairs of word indices and their frequencies. The results of the representation are presented in the following table.

Table 5. Bag of Words Results

Word	Word ID	Word Frequency
overall	0	1
good	1	2
issue	2	1
need	3	1
work	4	1
instance	5	1
click	6	1
post	7	1
read	8	1
sometimes	9	2
back	10	1
refresh	11	1
automatically	12	3
lose	13	1
forever	14	1
constantly	15	1
sssue	16	1
page	17	1
refreshe	18	1
midreade	19	2
whole	20	1
thing	21	1
go	22	1
irritating	23	1
still	24	1
consider	25	1
good	26	1
app	27	1
social	28	1
networking	29	1
well	30	1

3.3.2 LDA Model

The model was run with several topics ranging from three to fifteen to determine the optimal number of topics based on the coherence score. Each number of topics was tested by calculating the coherence score, and the results were compared to find the number of topics with the optimal value. Figure 2 shows the results of the number of topics and coherence value.

```
Num Topics = 3  has Coherence Value of 0.447
Num Topics = 4  has Coherence Value of 0.407
Num Topics = 5  has Coherence Value of 0.35
Num Topics = 6  has Coherence Value of 0.358
Num Topics = 7  has Coherence Value of 0.368
Num Topics = 8  has Coherence Value of 0.347
Num Topics = 9  has Coherence Value of 0.357
Num Topics = 10  has Coherence Value of 0.338
Num Topics = 11  has Coherence Value of 0.366
Num Topics = 12  has Coherence Value of 0.377
Num Topics = 13  has Coherence Value of 0.347
Num Topics = 14  has Coherence Value of 0.332
```

Figure 2. Results of Num Topics and Coherence Values

Figure 2 shows that with 3 topics (num topic), the coherence score obtained is 0.447. The relationship between the number of topics and the coherence score is illustrated in a graph in Figure 3. The optimal number of topics is determined by the highest coherence score value.

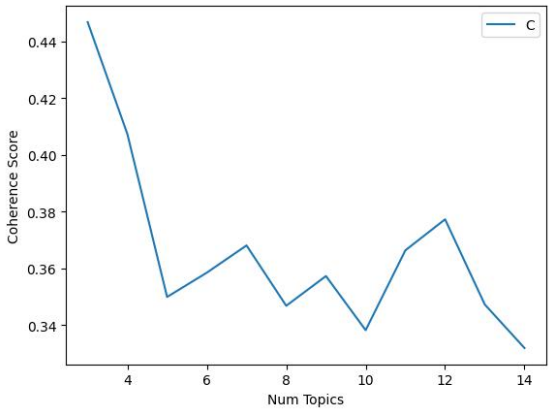

Figure 3. Coherence Score and Num Topics Graph

Figure 3 shows a line for 3 topics, which is then visualized using pyLDAvis to clearly observe the distribution and number of topics that emerge. This helps in analyzing the relationships between topics and understanding the representation of dominant words in each topic that is formed.

3.3.3 Visualization of LDA Model Topics

After the optimal number of topics is determined, the next step is to visualize the LDA model using the pyLDAvis library to assist in topic

interpretation through an interactive representation, where each topic consists of the most frequently occurring keywords that reflect the main aspects in the reviews. Figure 4 shows the graphical visualization of the LDA model topics.

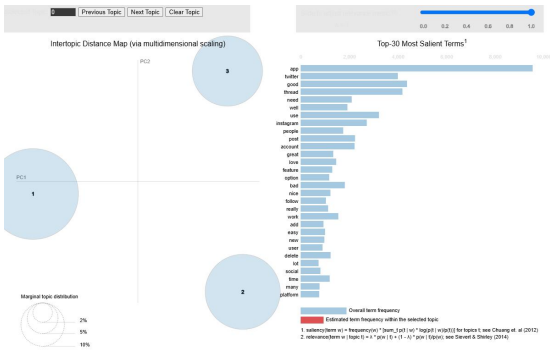

Figure 4. LDA Model Topic Visualization Graph

In Figure 4, there are 3 topics displayed. Each topic is represented by a bubble or blue circle, while the topic-term distribution is visualized in the form of blue bars, showing the contribution of words in the hidden topics. The most frequently occurring words across all topics provide an overview of the main focus of the user reviews.

3.3.4 Classification of Dominant Topics

The topic modeling process is followed by the classification of the dominant topic for each review to determine the main topic. By implementing the Latent Dirichlet Allocation (LDA) model, each review is associated with the topic that has the highest probability value. This topic is classified based on the most frequently occurring main keywords. The information stored includes the dominant topic number (Dominant_Topic), the percentage contribution of the topic to the review (Perc_Contrib), the list of main keywords (Topic_Keywords), the content after lemmatization (Data_Lemmatized), the review content (Content), and the determined aspect (Aspect). Table 6, 7 shows the topic classification results.

Table 6. Topic Classification Results

No	Dominant Topic	Perc Contrib	Topic Keywords
0	0	0.628	app, thread, use, instagram, post, account, ba...
1	0	0.441	app, thread, use, instagram, post, account, ba...
2	0	0.593	app, thread, use, instagram, post, account, ba...
…	…	…	…
19869	2	0.517	good, people, love, nice, really, easy, new, u...
19870	2	0.399	good, people, love, nice, really, easy, new, u...
19871	0	0.668	app, thread, use, instagram, post, account, ba...

Table 7. Advanced Topic Classification Results

No	Data Lemmatized	Content	Aspect
0	['overall', 'good', 'issue', 'need', 'work', '...	overall good issue need work instance click po...	Usage and Integration
1	['give', 'star', 'least', 'intuitive', 'app', ...	give star least intuitive app ever create comp...	Usage and Integration
2	['enjoy', 'easy', 'communicate', 'timeline', '...	enjoy easy communicate timeline copy paste tex...	Usage and Integration
…	…	…	…
19869	['dead', 'app', 'active', 'real', 'user']	dead app active real user	User Experience and Satisfaction
19870	['well', 'twitter', 'less', 'troll', 'less', '...	well twitter less troll less fake account well...	User Experience and Satisfaction
19871	['force', 'close', 'try', 'upload', 'photo']	force close try upload photo	Usage and Integration

After the dominant topic for each review is classified, the next step is to group the topics into more specific aspects. This is done using a dictionary that links the dominant topic values with aspect categories. Through this approach, each review is categorized into an aspect, as shown in Table 8.

Table 8. Classification of Topics into Aspects

Topic	Keywords	Aspects
1	app (0.1071), thread (0.0471), use (0.0360), instagram (0.0306), post (0.0252), account (0.0250), bad (0.0205), work (0.0174), delete (0.0139), time (0.0134)	Usage and Integration
2	twitter (0.0752), need (0.0395), well (0.0362), great (0.0252), feature (0.0244), option (0.0221), follow (0.0195), add	Features and Comparisons

Topic	Keywords	Aspects
	(0.0176), meta (0.0142), lot (0.0139)	
3	good (0.0721), people (0.0288), love (0.0241), nice (0.0202), really (0.0186), easy (0.0165), new (0.0161), user (0.0148), social (0.0134), many (0.0128)	User Experience and Satisfaction

3.4 Sentiment Labeling

Sentiment labeling is performed on the topics that have been classified into aspects to determine user sentiment. The labeling is done using the DistilBERT model from the Transformers library, with the distilbert-base-uncased-sst-2-english (SST-2) model version from Hugging Face, which has been fine-tuned for sentiment classification into positive or negative for text in English Language. DistilBERT is chosen due to its efficiency in providing sentiment predictions with high accuracy, as well as generating probability scores that indicate the model's confidence level in its predictions. Table 9, 10, 11 shows the sentiment labeling results.

Table 9. Sentiment Labeling Results

No	Dominant Topic	Perc Contrib	Topic Keywords
0	0	0.628	app, thread, use, instagram, post, account, ba...
1	0	0.441	app, thread, use, instagram, post, account, ba...
2	0	0.593	app, thread, use, instagram, post, account, ba...
…	…	…	…
19869	2	0.517	good, people, love, nice, really, easy, new, u...
19870	2	0.399	good, people, love, nice, really, easy, new, u...
19871	0	0.668	app, thread, use, instagram, post, account, ba...

Table 10. Advanced Sentiment Labeling Results

No	Data Lemmatized	Content
0	['overall', 'good', 'issue', 'need', 'work', '...	overall good issue need work instance click po...
1	['give', 'star', 'least', 'intuitive', 'app', ...	give star least intuitive app ever create comp...
2	['enjoy', 'easy', 'communicate', 'timeline', '...	enjoy easy communicate timeline copy paste tex...
…	…	…
19869	['dead', 'app', 'active', 'real', 'user']	dead app active real user
19870	['well', 'twitter', 'less', 'troll', 'less', '...	well twitter less troll less fake account well...
19871	['force', 'close,' 'try,' 'upload', 'photo']	force close try upload photo

Table 11. Advanced Sentiment Labeling Results

No	Aspect	Sentiment	Score
0	Usage and Integration	Negative	0.983045
1	Usage and Integration	Negative	0.997752
2	Usage and Integration	Positive	0.973596
…	…	…	…
19869	User Experience and Satisfaction	Negative	0.999477
19870	User Experience and Satisfaction	Positive	0.990506
19871	Usage and Integration	Positive	0.949116

The distribution of sentiment labels for the reviews can then be seen in Table 12, and the distribution for each aspect can be seen in Table 13.

Table 12. Number of Sentiment Distribution in Reviews

Sentiment	Number of Reviews
Positive	7.784
Negative	12.090

Table 13. Number of Sentiment Distribution in Reviews Based on Aspects

Aspects	Number of Reviews
Usage and Integration	11.320
User Experience and Satisfaction	5.737
Features and Comparisons	2.817

3.5 Visualisation

The reviews, classified by aspect and sentiment, is visualized to provide a clearer overview using bar charts.

3.5.1 Aspects Usage and Integration

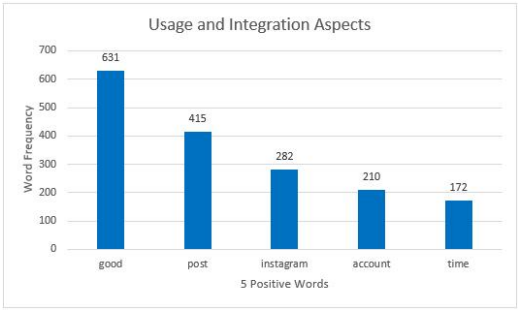

Figure 5. Frequency of Positive Words Usage and Integration Aspects

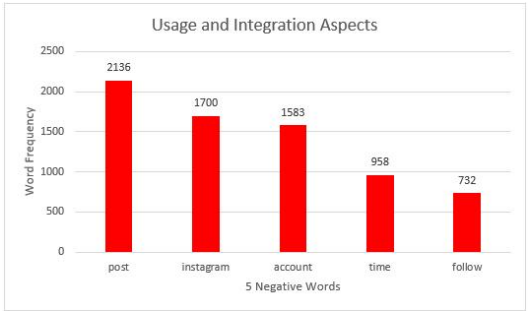

Figure 6. Frequency of Negative Words Usage and Integration Aspects

3.5.2 Aspects Feature and Comparisons

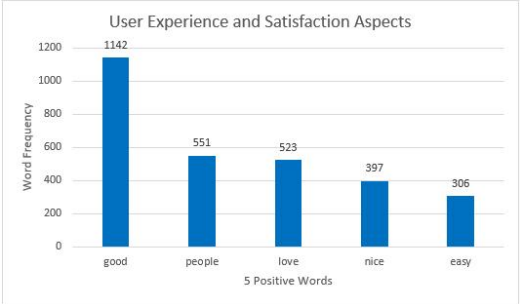

Figure 7. Frequency of Positive Words Features and Comparisons Aspects

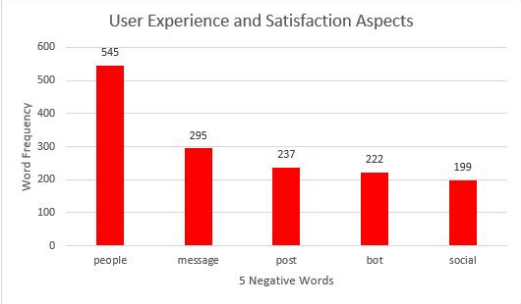

Figure 8. Frequency of Negative Words Features and Comparisons Aspects

3.5.3 Aspect User Experience and Satisfaction

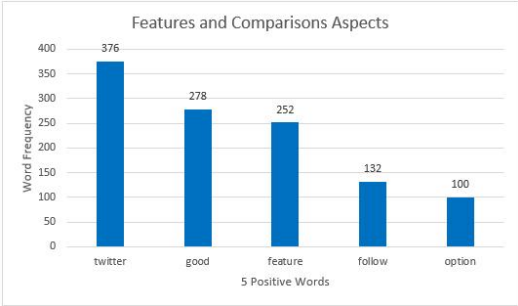

Figure 9. Frequency of Positive Words User Experience and Satisfaction Aspects

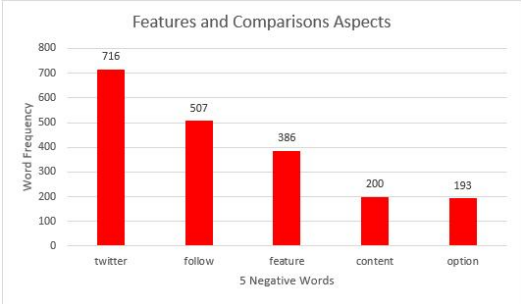

Figure 10. Frequency of Negative Words User Experience and Satisfaction Aspects

The results show that the majority of users have diverse opinions regarding the Usage and Integration aspect. While many appreciate the basic features of Threads, there are complaints about the integration with Instagram and the limitations of other features. In the User Experience and Satisfaction aspect, users generally feel satisfied with the app's interface, but some users complain about the discomfort in sending messages and using certain features. Meanwhile, in the Features and Comparisons aspect, many users compare Threads with Twitter, with most of the criticism focusing on the limited features of Threads compared to similar platforms.

3.6 Model Development

In the model development stage, the training parameters were set as follows: learning rate of 1e-5, batch size of 32, maximum sequence length of 128, and a total of 3 epochs. The training process was carried out using a GPU to enhance computational efficiency, with the AdamW optimizer updating the model weights based on the loss calculation.

3.6.1 Splitting Dataset

The dataset is divided into training and testing data, with a proportion of 85% training data and 15% testing data to ensure accurate model predictions. The number of sentiment reviews in the training data (85%) includes 6,587 positive reviews and 10,305 negative reviews, totaling 16,892 reviews, while the sentiment in the testing data (15%) includes 1,197 positive reviews and 1,785 negative reviews, totaling 2,982 reviews. Table 14 shows the results of the training and testing data split.

Table 14. Number of Training Data and Testing Data

Sentiment	Training Data (85%)	Testing Data (15%)
Positive	6.587	1.197
Negative	10.305	1.785
Number of Reviews	16.892	2.982

3.6.2 Evaluation of Sentiment Model Performance

The trained DistilBERT model was tested to classify positive and negative sentiments. The evaluation was performed on the testing data, with the results shown in the confusion matrix in Figure 11.

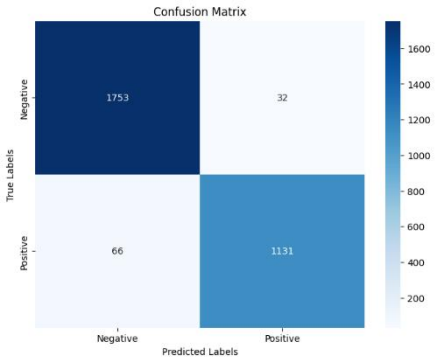

Figure 11. Confusion Matrix Sentimen

In Figure 11, the confusion matrix shows the model's performance in predicting sentiment. The model successfully classified 1,753 negative reviews (NR) and 1,131 positive reviews (PR). However, there were prediction errors, with 32 negative reviews predicted as positive (PR) and 66 positive reviews predicted as negative (PN). The results of these calculations are then presented in Table 15.

Table 15. Sentiment Evaluation Results

Metric	Value
Accuracy	96.71%
Precision	97.24%
Recall	94.48%
F1-Score	95.84%

A graph was then created for visualization, which can be seen in Figure 12.

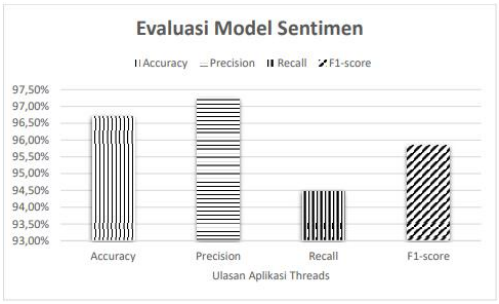

Figure 12. Sentiment Model Evaluation Results Graph

3.6.3 Aspect-Based Sentiment Model Evaluation

After evaluating the overall sentiment model, further evaluation was conducted based on the three main aspects that have been established: Usage and Integration, Features and Comparisons, and User Experience and Satisfaction. This evaluation aims to assess the model's ability to classify sentiment according to the defined aspects. The model was tested separately for each aspect, with the confusion matrices shown in Figures 13, 14, and 15.

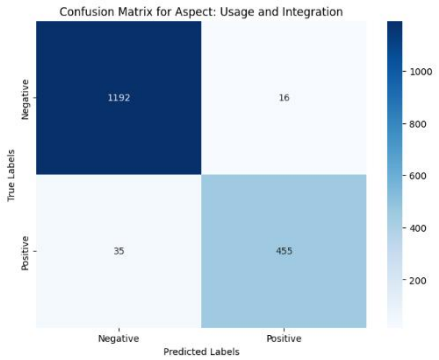

Figure 13. Confusion Matrix Aspek Usage and Integration

In Figure 13, the model successfully predicted 1,192 negative reviews (NR) and 455 positive reviews (PR) correctly. Prediction errors occurred, with 16 negative reviews classified as positive (PR) and 35 positive reviews classified as negative (NR).

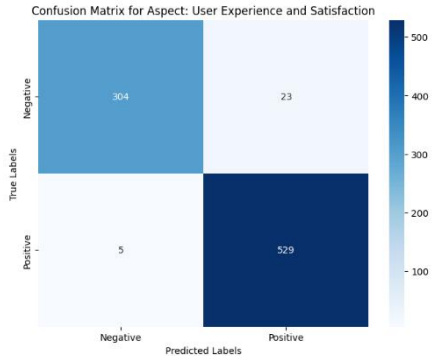

Figure 14. Confusion Matrix Aspek User Experience and Satisfaction

In Figure 14, the model correctly classified 304 negative reviews (NR) and 529 positive reviews (PR). Prediction errors occurred, with 23 negative reviews classified as positive (PR) and five positive reviews classified as negative (NR).

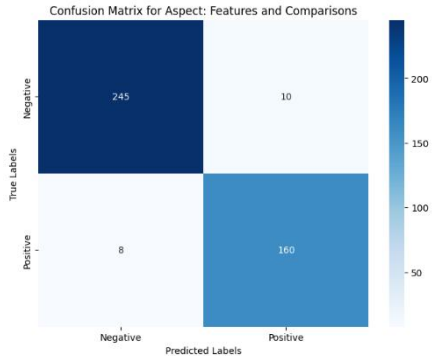

Figure 15. Confusion Matrix Aspek Features and Comparisons

In Figure 15, the model successfully predicted 245 negative reviews (NR) and 160 positive reviews (PR) correctly. Prediction errors occurred, with 10 negative reviews classified as positive (PR)

and eight positive reviews classified as negative (NR).

The performance evaluation results of the model for each aspect are then presented in a table, which can be seen in Table 16.

Tabel 16. Results of Aspect-Based Sentiment Model Evaluation

Aspect	Metrics	Score
Usage and Integration	Accuracy	96.99%
	Precision	96.60%
	Recall	92.85%
	F1-Score	94.69%
User Experience and Satisfaction	Accuracy	96.74%
	Precision	95.83%
	Recall	99.06%
	F1-Score	97.42%
Features and Comparisons	Accuracy	95.74%
	Precision	94.11%
	Recall	95.23%
	F1-Score	94.67%

It is then visualized in the form of a graph, which can be seen in Figure 16.

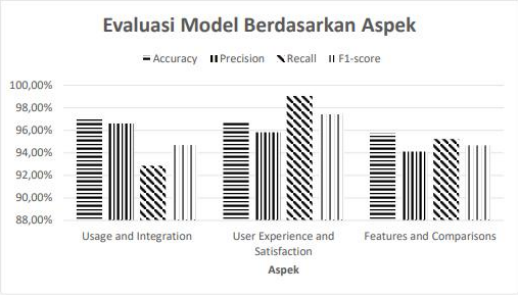

Figure 16. Graph of Aspect-Based Sentiment Model Evaluation Results

4 Conclusions

The aspects of user reviews for the Threads application have been identified using the LDA method, consisting of Usage and Integration, Features and Comparisons, and User Experience and Satisfaction. The distribution of review counts for each aspect is as follows: Usage and Integration with 11,320 reviews (3.147 positive and 8.173 negative), User Experience and Satisfaction with 5,737 reviews (3.529 positive and 2.208 negative), and Features and Comparisons with 2,817 reviews (1.108 positive and 1.709 negative). Meanwhile, sentiment labeling of the reviews resulted in 12,090 negative reviews and 7,784 positive reviews. The evaluation of the sentiment classification model performance using the confusion matrix showed the following metrics for overall sentiment: accuracy of 96.71%, precision of 97.24%, recall of 94.48%, and F1-score of 95.84%. For the aspect-based model performance evaluation, the results were: accuracy of 96.99%, precision of 96.60%, recall of 92.85%, and F1-score of 94.69% for the Usage and Integration aspect; accuracy of 95.74%, the precision of 94.11%, recall of 95.23%, and F1-score of 94.67% for the Features and Comparisons aspect; and accuracy of 96.74%, precision of 95.83%, recall of 99.06%, and F1-score of 97.42% for the User Experience and Satisfaction aspect.

In future research, the author will add an algorithm for detecting non-standard words during the data preprocessing stage to optimize the sentiment analysis results.

References:

[1] A. G. Gani, “Pengenalan Teknologi Internet Serta Dampaknya,” JSI (Jurnal Sistem Informasi), vol. 2, no. 2, pp. 71-86, 2015, doi: https://doi.org/10.35968/jsi.v2i2.49.

[2] S. R. Yustihan, P. P. Adikara, and I. Indriati, “Analisis Sentimen Berbasis Aspek Terhadap Data Ulasan Rumah Makan Menggunakan Metode Support Vector Machine (SVM),” Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, vol. 5, no. 3, pp. 1017-1023, Maret 2021, doi: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/8710.

[3] M. Widiansyah, F. F. Az-Zahra, and A. Pambudi, “Fine-Tunning Model IndoBERT (Indonesian Bidirectional Encoder Representations from Transformers) Untuk Analisis Sentimen Berbasis Aspek Pada Aplikasi M-Paspor," Journal of Informatic Engineering (JOUTICA), vol. 9, no. 2, pp. 184-195, September 2024, doi: https://doi.org/10.30736/informatika.v9i2.1310.

[4] S. Roiqoh, and B. Zaman, “Analisis Sentimen Berbasis Aspek Ulasan Aplikasi Mobile JKN dengan Lexicon Based dan Naïve Bayes,” Jurnal Media Informatika Budidarma, vol. 7, no. 3, pp. 1582-1592, Juli 2023, doi: https://doi.org/10.30865/mib.v7i3.6194.

[5] M. Y. Febrianta, S. Widiyanesti, and S. R. Ramadhan,”" Analisis Ulasan Indie Video Game Lokal pada Steam Menggunakan Analisis Sentimen dan Pemodelan Topik Berbasis Latent Dirichlet Allocation,” Journal of Animation & Games Studies, vol. 7, no. 2, pp. 117-144, Oktober 2021, doi: https://doi.org/10.24821/jags.v7i2.5162.

[6] E. Puspita, D. F. Shiddieq, and F. F. Roji, Pemodelan Topik pada Media Berita Online Menggunakan Latent Dirichlet Allocation (Studi Kasus Merek Somethinc),” Institut Riset dan Publikasi Indonesia (IRPI), MALCOM: Indonesian Journal of Machine Learning and Computer Science, vol. 4, no. 2, pp: 481-489, April 2024, doi: https://doi.org/10.57152/malcom.v4i2.1204.

[7] N. R. A. Putri, T. Trimono, and A. T. Damaliana, “Sentiment Analysis on Digital Korlantas POLRI Application Reviews Using the Distilbert Model,” Journal of Renewable Energy, Electrical, and Computer Engineering, vol. 4, no. 2, pp. 83-89, September 2024, doi: https://doi.org/10.29103/jreece.v4i2.17197.

[8] H. C. Morama, D. E. Ratnawati, and I. Arwani, "Analisis Sentimen berbasis Aspek terhadap Ulasan Hotel Tentrem Yogyakarta menggunakan Algoritma Random Forest Classifier," Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, vol. 6, no. 4, pp. 1702-1708, April 2022, available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/10908. [9] W. Parasati, F. A. Bachtiar, and N. Y. Setiawan, "Analisis Sentimen Berbasis Aspek pada Ulasan Pelanggan Restoran Bakso President Malang dengan Metode Naïve Bayes Classifier," Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, vol. 4, no. 4, pp. 1090-1099, April 2020, available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/7134.

[10] F. A. J. Ayomi, and K. Evita, "Analisis Emosi Pada Media Sosial Twitter Menggunakan Multinomial Naïve Bayes dan Synthetic Minority Oversampling Technique," KOMPUTA: Jurnal Ilmiah Komputer dan Informatika, vol. 12, no. 2, pp. 9-19, Oktober 2023, doi: https://doi.org/10.34010/komputa.v12i2.9454.

[11] M. U. Albab, Y. Karuniawati. P, and M. N. Fawaiq, "Optimization of the Stemming Technique on Text preprocessing President 3 Periods Topic," Jurnal TRANSFORMATIKA, vol. 20, no. 2, pp. 1-10, Januari 2023, available: https://journals.usm.ac.id/index.php/transformatika/.

[12] R. Lamhot, "Implementasi Metode Generalized Vector Space Model Pada Aplikasi Information Retrieval untuk Pencarian Informasi Pada Kumpulan Dokumen Teknik Elektro Di UPT BPI LIPI," J. Ilm. Komput. dan Inform. (KOMPUTA), 2014.

[13] Y. Miftahuddin, J. Pardede, and R. Dewi, "Penerapan Algoritma Lemmatization pada Dokumen Bahasa Indonesia," MIND Journal, vol. 3, no.2, pp. 47-56, Desember 2018, doi: https://doi.org/10.26760/mindjournal.v3i2.47-56.

[14] K. R. Hilal, N. Y. Setiawan, and D. E. Ratnawati, "Analisis Sentimen Berbasis Aspek Untuk Pengguna PLN Mobile Pada Google Playstore Menggunakan Metode Support Vector Machine (SVM)," Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, vol.8, no.7, pp. 1-13, Juli 2024, available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/13927.

[15] Z. N. Muna, B. D. Setiawan, and R. S. Perdana, Penerapan Pemodelan Topik Komentar Melalui Media Sosial Twitter Menggunakan Latent Dirichlet Allocation (Studi Kasus: Pemerintah Kota Malang)," Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, vol. 8, no. 7, pp. 1-10, 2024, available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/13979.

[16] R. F. R. Pohan, D. E. Ratnawati, and I. Arwani, "Implementasi Algoritma Support Vector Machine dan Model Bag-of-Words dalam Analisis Sentimen mengenai PILKADA 2020 pada Pengguna Twitter," Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, vol. 6, no. 10, pp. 4924-4931, Oktober 2022, available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/11715.

[17] W. T. H. Putri, and R. Hendrowati, "Penggalian Teks Dengan Model Bag of Words Terhadap Data Twitter," Jurnal Muara Sains, Teknologi, Kedokteran, dan Ilmu Kesehatan, vol. 2, no. 1, pp. 129-138, April 2018, available: https://www.journal.untar.ac.id/index.php/jmistki/article/view/1560/1389.

[18] D. P. Santoso, and W. Wibowo, "Analisis Sentimen Ulasan Aplikasi Buzzbreak Menggunakan Metode Naïve Bayes Classifier pada Situs Google Play Store," Jurnal Sains dan Seni ITS, vol. 11, no. 2, pp.190-196, 2022, doi: https://doi.org/10.12962/j23373520.v11i2.72534.

[19] H. Apriyani, and K. Kurniati, "Perbandingan Metode Naïve Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus," Journal of Information Technology Ampera, vol. 1, no. 3, pp. 133-143, Desember 2020, doi: https://doi.org/10.51519/journalita.

[20] S. Clara, D. L. Prianto, R. A. Habsi, E. F. Lumbantobing, and N. Chamidah, "Implementasi Seleksi Fitur Pada Algoritma Klasifikasi Machine Learning Untuk Prediksi Penghasilan Pada Adult Income Dataset," Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA), vol.2, no.1 pp. 741-747, April 2021, available: https://conference.upnvj.ac.id/index.php/senamika/article/view/1417.