Accredited by National Journal Accreditation (ARJUNA) Managed by
Ministry of Research, Technology, and Higher Education, Republic Indonesia with Second Grade (Peringkat 2, Sinta 2)
since year 2017 to 2021 according to the decree No. 10/E/KPT/2019

Published online on the journal's webpage: **http://jurnal.iaii.or.id**

JURNAL RESTI

(Rekayasa Sistem dan Teknologi Informasi)

Vol. 5 No. 4 (2021) 631 - 638 ISSN Media Elektronik: 2580-0760

Aspect Level Sentiment Analysis on Zoom Cloud Meetings App Review Using LDA

Janu Akrama Wardhana[1], Yuliant Sibaroni[2]
[1,2]Informatics, School of Computing, Telkom University
[1]januakramaw@student.telkomuniversity.ac.id, [2]yuliant@telkomuniversity.ac.id

Abstract

During the Covid-19 pandemic, almost all community activities are conducted from home. Therefore, video conference technology is needed for people to carry out their normal activities from home. One of the video conference applications is ZOOM Cloud Meetings. Applications certainly have been reviewed given by their users as a reference for new users and companies of the application to know the application's performance. However, in reviews, some constraints are the number of reviews as well as irregular. Therefore, a solution is needed with sentiment analysis that aims to classify the reviews of the application to be organized by categorizing positive or negative sentiment. In this study, aspect-based sentiment analysis was conducted on ZOOM Cloud Meetings app reviews from Google Play Store. The analysis's result of the review data obtained three aspects, namely aspects of usability, system, and appearance. The modeling topic used is the Latent Dirichlet Allocation (LDA) method and classification using the Support Vector Machine (SVM). This research resulted in the best performance with the best parameters resulting in the performance accuracy of usability aspect is 88.83%, system aspect with 91.2%, appearance aspect with 94.78%, and performance accuracy of all aspects 91.61%.

Keywords: LDA, SVM, review, aspect

1. Introduction

The development of technology today has been very rapidly developed in various fields coupled with the existence of the internet to provide convenience in supporting the use of technology. During the Covid-19 pandemic, technology played a big role in supporting all community activities. Almost all fields use technology because, with this condition, the government issued a policy to the public to do all activities from home to reduce the spread of Covid-19. This causes all circles such as students, office employees, businessmen, and others to do all normal activities that are usually done outside the home to be from home. And there is a solution from the development of technology that can be used by those people to do their normal activities from home using video conference technology. This condition proved video conference applications experienced a very significant increase in users.

One of the video conference applications that are widely used by all circles is the ZOOM Cloud Meetings application. ZOOM is a video communication application that can be used on a variety of mobile and desktop devices [1]. ZOOM also has supporting features for students and workers in performing activities that are usually done offline to be online. These features include scheduling features on the meeting to be held, screen record, share screen, team chat/chat. The ZOOM Cloud Meetings app can be downloaded through services such as the Google Play Store, which is a digital content provider service, or Android-specific applications owned by Google.

All applications must have advantages and disadvantages, including the ZOOM Cloud Meetings application. The user experience of this app is often written in the review column located on the Google Play Store. These reviews contain criticism, suggestions, and satisfaction from users. So, it takes sentiment analysis to identify the sentiment expressed in the text, then analyze it, and the goal is to find opinions, identify the sentiments they express, and then classify the polarity [2]. The study [3] has similarities because it conducted research using the data review application ZOOM Cloud Meetings, but this study has several drawbacks. Namely, this study only classifies with the data used not as much as the 1007 record, and the resulting review sentiment is

Diterima Redaksi: 19-05-2021 | Selesai Revisi: 11-08-2021 | Diterbitkan Online: 20-08-2021

still common, so it cannot know the elements or more specific aspects of the product being reviewed. The study compared the classifying performance of the Naïve Bayes and Support Vector Machine (SVM). And as a result, SVM has better accuracy than Naïve Bayes, with the accuracy of 81.22% and 74.37%.

In other studies by Ekawati [4], aspect-based sentiment analysis was conducted on Indonesian restaurant reviews to display the sentiments of aspects contained in a review such as food, service, price, place. An example of a review is "The place remains comfortable, the cuisine remains delicious, and the service remains friendly". The sentence is included in 3 aspects, namely the place, food, and service of the restaurant that was reviewed. The study produced the highest F1-score was 0.823.

The study by Wahyudi [5] conducted an aspect-based sentiment analysis of E-Commerce user reviews using LDA and Lexicon sentiment. The study conducted three scenarios or experiments. Of all the experiments, the training data used is general training data and category /aspect training data. From the three scenarios obtained the best performance with an increase in accuracy worth 0.82% by doing a combination of general training data and data training category/aspect of the best results of the second and first experiments so that it proves that the combination of general training data and category training data in LDA modeling shows better accuracy than the use of training data respectively. Other LDA-related research conducted by Kaveh [6] discuss about topic modeling on customer complaints using LDA which explains that LDA is an algorithm to address unstructured data and is suitable for analyzing consumer behavior, product reviews, and other research.

In this study, sentiment analysis will be conducted on the ZOOM Cloud Meetings app review sourced from Google Play Store to find out the sentiment contained in t. Not only sentiment in general but more specific positive or negative sentiment by involving what aspects are contained in the ZOOM Cloud Meetings app review. It is useful to know the sentiment of users based on the reviews provided so that it can be used as a reference for new users and the service provider company to know the performance of the application. The classification method used is Support Vector Machine (SVM). SVM is used because it has better performance than other classifiers [7], in this study using Latent Dirichlet Allocation (LDA) topic modeling. This LDA can detect topics in the review data, and the LDA will group or classify by aspect. LDA is used because it can be widely implemented in various domains such as text, images, music, and others. LDA can also be used to perform classification tasks as well as topic grouping tasks. The data that will be used in this study comes from Play Store reviews collected by doing web scraping data.

The topic and limitations of the problem in this study are testing based on test scenarios to determine the performance of the model that has been designed. The limitation of the problem in this study was that aspect-level sentiment analysis was conducted using Support Vector Machine (SVM) with the Latent Dirichlet Allocation (LDA) topic modeling. The data used in the form of review data ZOOM Cloud Meetings application in Indonesian with the amount of 6000 data comes from the Google Play Store. The data is divided into three aspects, namely system aspects, appearance, and usability. These aspects are determined based on the analysis results of the review data used and these three aspects are aspects owned by the ZOOM Cloud Meetings application. Aspects of the review system discuss the performance of the system that the application has such as login features, registration, video quality, sound, and others. Aspects of appearance, reviews discuss the design of the appearance of the application such as attractive design, simple, easy to understand and others. and aspects of the last usability, reviews discuss the benefits and suitability of the needs of the application users, reviews such as this application helps in learning from home, this application is useful to conduct webinars and others. On the system aspect, appearance, and usability, each has three polarity classes, namely positive, negative, and neutral.

The purpose of this study was to create an aspect-level sentiment analysis system on the ZOOM Cloud Meetings app review data sourced from Google Play Store using Support Vector Machine with Latent Dirichlet Allocation topic modeling as well as measuring the performance of the system that has been created.

2. Research Method

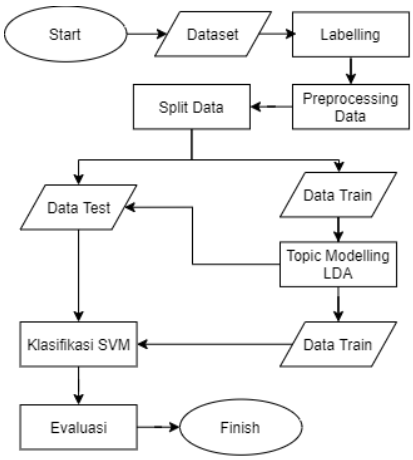


Figure 1. System Architecture

In this study, sentiment analysis was conducted using Latent Dirichlet Allocation (LDA) as topic modeling and then classifying sentiment using Support Vector

DOI: https://doi.org/10.29207/resti.v5i4.3143

Machine (SVM), which can be seen in figure 1, which is the flow of the research process.

2.1. Dataset

In this study, the data source used came from a review of the ZOOM Cloud Meetings application located on the Play Store. Data retrieval is done by scraping using the Google play scraper library in Python. WEB scraping is a method or way to get data in the form of documents from a WEB site. The way web scraper works starts with entering one or more URLs for scraping 16 Then, the scraper will extract the data on the selected web page specifically according to the given URL[8].

Researchers collected as many as 6000 data records. The data collected is Indonesian reviews and the location of app users located in Indonesia. Once the scraping process is complete, the data is saved in CSV format.

2.2 Labelling Data

After obtaining the data set to be used, the next process is to label the data set. This process is carried out by 3 college students who have previously discussed the content or discussion that is reviewed in every aspect. when 3 college students find reviews quite difficult to determine including what classes and what aspects will be discussed together. In this study, there are three aspects will be used, namely the system, appearance, and usability. In every aspect, there are three polarity classes to be classified, namely the positive class with a value of 1, the negative class with a value of -1, and the neutral class with a value of 0. In table 1, there are examples of data erosion results as well as labeling data.

Table 1. Example of Result Scraping and Label Data

Review	System	Appearance/ Design	Usability
Fitur dan tampilan sangat mudah digunakan serta dipahami	1	1	0
Aplikasi suaranya kadang kadang suka putus putus dan juga kadang kadang suaranya kecil😅😆😆	-1	0	0
Aplikasi ini sangat berguna untuk sy.jadi bisa mengajar anak murid sy	0	0	1
Terima kasih zoom sangat membantu untuk memenuhi syarat kartu prakerja	0	0	1

2.3 Preprocessing Data

After the previous stage, we already get the data. At this stage, researchers preprocess the data that has been obtained. Preprocessing is the stage of processing data that has been obtained into the ideal data to be processed. Preprocessing is done automatically using Python. Here are the preprocessing stages performed.

The first step is case folding. Case folding is to change all letters in a word and sentence to be equal by changing all letters to lowercase [9]. In table 2, there is an example of the case folding process.

Table 2. Case Folding

Before	After
Terimakasih @Zoom sdh mempermudah saya mengikuti WEBinar dengan baik	terimakasih @zoom sdh mempermudah saya mengikuti webinar dengan baik

The second step is tokenization. Tokenization is a preprocessing stage that aims to convert a sentence into words that will be processed for the next stage. in table 3 there is an example of tokenization process.

Table 3. Tokenization

Before	After
terimakasih @zoom sdh mempermudah saya mengikuti webinar dengan baik	‘terimakasih’, ‘@zoom’, ‘sdh, ‘mempermudah’, ‘saya’, ‘mengikuti’, ‘webinar’, ‘dengan’, ‘baik’

The third step is stopword removal. Stopword is a word that is not descriptive and appears in large numbers and has no connection with the information needed [10]. This process is done using the NLTK library to get a list of Indonesian stopwords. In table 4, there is an example of the stopword removal process.

Table 4. Stopword Removal

Before	After
‘terimakasih’, ‘@zoom’, ‘sdh, ‘mempermudah’, ‘saya’, ‘mengikuti’, ‘webinar’, ‘dengan’, ‘baik’	‘terimakasih’, ‘@zoom’, ‘sdh, ‘mempermudah’, ‘mengikuti’, ‘webinar’, ‘baik’

The fourth step is to remove mention, link/URL, hashtag, emoticon, punctuation. In the sentence data reviews, sometimes there are characters other than letters such as emoticons and words that are not needed by researchers, such as hashtags, links, punctuation, and others. So, it must be eliminated for the review data to be more efficient to process. in table 5, there is an example from the result in this process.

Table 5. Remove Mention, Link/URL, Hashtag, Emoticon, Punctuation

Before	After
‘terimakasih’, ‘@zoom’, ‘sdh, ‘mempermudah’, ‘mengikuti’, ‘webinar’, ‘baik’	‘terimakasih’, ‘zoom’, ‘sdh’, ‘mempermudah’, ‘mengikuti’, ‘webinar’, ‘baik’

The fifth step is normalization. This normalization is the process to resolve words that have errors such as typos, the presence of abbreviations. Then the words will be returned to the original word form. Words that have the error will be normalized based on the 7277 words of normalization that have been created by researchers. Table 6 can be seen the process of normalization results.

DOI: https://doi.org/10.29207/resti.v5i4.3143

Table 6. Normalization

Before	After
'terimakasih', 'zoom', 'sdh', 'mempermudah', 'mengikuti', 'webinar', 'baik'	'terimakasih', 'zoom', 'sudah', 'mempermudah', 'mengikuti', 'webinar', 'baik'

The last step is stemming. Stemming aims to remove the words contained in a word. This stemming process is done using sastrawi stemmer. In table 7, there is an example of the stemming process.

Table 7. Stemming

Before	After
'terimakasih', 'zoom', 'sudah' 'mempermudah', 'mengikuti', 'webinar', 'baik'	'terimakasih', 'zoom', 'sudah' 'mudah', 'ikut', 'webinar', 'baik'

2.3. Split Data

Split data is a method used to evaluate the performance of machine learning models by dividing entire sets of data into train data used for fit models and test data used to evaluate the fit model of the data train.

This study will be divided from 6000 dataset records, will be used 70% or 4200 data records as training data and 30% or 1800 data records as data testing. The configuration is used because it increases the possibility of sharing all aspects, especially in the sharing of test data. therefore, use 30% configuration for test data. table 8 shows the results of class label distribution after data split.

Table 8. label distribution after data split

	Label	Aspect		
		Usability	System	Appearance
Data Train	Positif	3170	318	367
	Negatif	157	299	146
	Netral	873	3583	3687
Data Test	Positif	1357	137	161
	Negatif	70	121	72
	Netral	373	1542	1567

2.4. Latent Dirichlet Allocation

Next, perform topic modeling with Latent Dirichlet Allocation topic modeling. LDA is one of the Text Mining approaches used to find hidden text data and find relationships between text data [11]. The relationship between text data is determined by doing by reducing features or reducing dimensions by mapping them to topics, which in the topic of course there are words that belong to the entire text data set used [12]. The LDA groups text data based on the topics in the text data set, which is done in a clustering-like way that is grouped based on the similarity of each text data. This LDA is used to summarize, cluster, and connect or process large data because the LDA generates topics that exist in the processed text data. In the LDA form, the document is represented as a random mix for each resulting topic, while the topic is derived from the word distribution. Here is the form of the LDA [13].

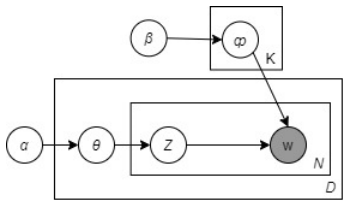


Figure 2. LDA

In figure 2, there are several parameters used in the LDA. β is Dirichlet parameters over the distribution of words to the topic so that the smaller the word will be discussed in the topic. α is a Dirichlet parameter so that if the greater the value of α then the more topics will be discussed. And also, on the contrary, if the less value α then the fewer topics will be discussed. φ is the distribution of words to topics in the corpus. K is a collection of topics, Wn is the observed word, N is a collection of words. M is a document, Zn is the topic for the Nth word in the document M. And θ is the distributed topic of the document M.

2.4. Support Vector Machine

After producing the model from the LDA process then the model will be trained using Support Vector Machine by using experiments on four kernels, namely Linear, Sigmoid, Polynomial, and Gaussian. It will then generate a model of the data train that will be used for the SVM classification process. Support Vector Machine (SVM) is one of the methods in supervised learning that is usually used for classification. SVM is used to find the best hyperplane by maximizing the distance between classes. The hyperplane is a function that can be used to separator between classes [14]. In figure 3 below, SVM has two classes, namely +1 (on white circles) and -1 (on black circles). The line between the dotted lines is called a hyperplane. This study will use class +1 for positive class and -1 for negative class.

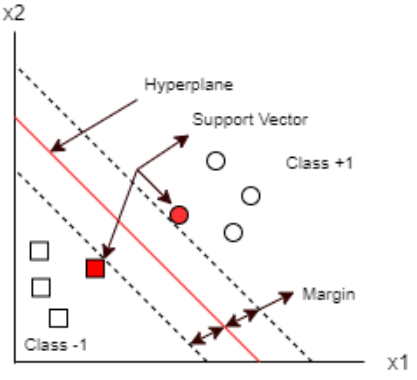


Figure 3. SVM

SVM tries to find the best hyperplane as a separator between two kinds of objects. The best hyperplane is a hyperplane located in the middle between two sets of objects from two classes, as shown in figure 3, obtained

DOI: https://doi.org/10.29207/resti.v5i4.3143

by maximizing the margin or distance between two sets of objects from different classes. There is a hyperplane that supports classes -1 and +1 can be seen in equations 1 and 2.

$$w.x_i + b \leq -1 \quad (1)$$

$$w.x_i + b \geq +1 \quad (2)$$

The margin between the two classes can be calculated by finding the distance between the two hyperplanes supporting both classes, using formula 3 [14]:

$$\frac{1}{2}\|w\|^2 \quad (3$$

Data problems cannot be linearly separated in the input space, SVM soft margins cannot find hyperplanes, so they cannot have good accuracy and do not generalize properly. The kernel is needed to transform data into higher dimensional spaces. Formulas 4, 5, and 6 are formulas from several kernels.

$$\text{Polynomial: } k(x_i x_j) = (x_i x_j + 1)^d \quad (4$$

$$\text{Gaussian: } k(x,y) = \exp(-\frac{\|x-y\|^2}{2\sigma^2}) \quad (5$$

$$\text{Sigmoid: } k(x,y) = \tanh(\alpha x^T y + c) \quad (6$$

2.5. Evaluation

Evaluation is done to measure the performance of the system that has been built. The evaluation stage using the confusion matrix after that measuring the performance of the system is done by calculating accuracy, precision, recall, and f1-score [15].

Table 9. Evaluation

Actual Class	Predicted	
Positive	TP	FP
Negative	FN	TN

Information for table 9:

True Positive (TP): data predicted positive and originally positive.

True Negative (TN): data predicted negative and original negative.

False Negative (FN): negative and originally predicted positive data.

False Positive (FP): data predicted positive and original negative.

Accuracy is the ratio of the number of accurate predictions per document with the total number of all predictions classified. Formula 7 is an accuracy formula.

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

Precision is the ratio of the number of relevant documents to the total number of documents found by the classifier. Formula 8 is a precision formula.

$$Precision = \frac{TP}{TP + FP} \quad (8$$

The recall is the ratio of the number of documents recovered by a classifier to the total number of relevant documents. Formula 9 is a recall formula.

$$Recall = \frac{TP}{TP + FN} \quad (9$$

F1-score is an average harmonic combination of precision and recall that is directly proportional to the value of both. Formula 10 is an F1-score formula.

$$F1 - Score = \frac{2(Precision \times Recall)}{(Precision + Recall)} \quad (10$$

3. Result and Discussion

The testing in this study used ZOOM Cloud Meetings app review data as much as 6000 data sourced from Google Play Store. Data sets that have gone through the preprocessing process will be divided into 70% train data and 30% test data. And then, the topic modeling process will be done using LDA to get the model to be used for SVM classification. There are 3 test scenarios.

3.1. Scenario 1

In this first test scenario, the experiment will be conducted by assuming in the parameters the number of topics in the LDA as many as 5 different topics in the form of 5 topics, 10 topics, 20 topics, 40 topics, and 60 topics. For parameters α worth 0.001 for topic distribution in documents/reviews, and β worth 0.001 for word distribution in topics. After obtaining the model produced from LDA with the values of parameters that have been determined, then the model will be classified using SVM with testing four kernels, namely Linear, Sigmoid, Polynomial, and Gaussian. The results in this first scenario are in table 10.

The results of the first test scenario with a difference of five parameters of the number of LDA topics in this study obtained performance results in the form of ever-increasing accuracy. In a test scenario using an experiment's five parameters the number of different topics is 5 topics, 10 topics, 20 topics, 40 topics, and 60 topics. This can be seen in figure 4. In this test scenario, performance accuracy would be better if the number of topics on the LDA was higher on the data used in this study. This happens because of the reduction in LDA dimensions by making the topic a feature, in contrast to other methods such as TF-IDF or bag of words that map

DOI: https://doi.org/10.29207/resti.v5i4.3143

directly to the token or word to be displayed. So, the accuracy performance will certainly be reduced due to the decrease in dimensions [16]. But that does not mean that with a higher number of topics, the more optimal the performance for any other data. optimization of the number of topics depending on the data used. The first trial scenario with five different topics with 5, 10, 20, 40, and 60, produced the best accuracy performance with 60 topics.

Table 10. Result Scenario 1

Number of Topics	Kernel	*Accuracy*
5	Linear	86.33%
	Sigmoid	82.37%
	Polynomial	86.41%
	Gaussian	86.34%
10	Linear	87.56%
	Sigmoid	84.01%
	Polynomial	87.62%
	Gaussian	88.09%
20	Linear	88.88%
	Sigmoid	86.47%
	Polynomial	89.01%
	Gaussian	89.88%
40	Linear	90.29%
	Sigmoid	87.38%
	Polynomial	89.7%
	Gaussian	91.17%
60	Linear	90.74%
	Sigmoid	88.57%
	Polynomial	89.74%
	Gaussian	91.61%

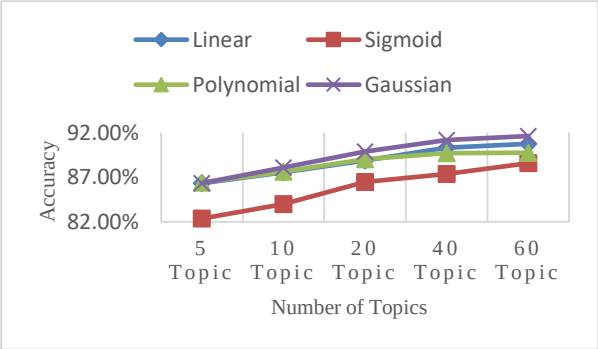


Figure 4. Comparison of Scenario 1 Accuracy Results

3.2. Scenario 2

The second test scenario is to implement SVM parameter tuning to get the best parameters and to test the tuning of SVM parameters in linear, sigmoid, polynomial, and Gaussian kernels with the parameters of the number of topics i.e. 5 topics, 10 topics, 20 topics, 40 topics, and 60 topics, Resulting in the best performance results. In table 11 is the parameter value that will be used for tuning the parameter.

The most optimal results are obtained from the second test scenario by tuning SVM parameters using the Gaussian kernel with a parameter value of C = 1 as well as the number of topics of 60 topics. The choice of the Gaussian kernel because it has the best accuracy performance compared to other kernels, proves that the data used in this study is non-linear. because the accuracy performance of the Gaussian kernel is better than that of the Linear kernel. Can be seen in table 12 can be seen the best results of each parameter number of topics and kernel. With Gaussian kernel with a value of C = 1 results in the best accuracy of 91.61% with a total of 60 features.

Table 11. Range Parameter

Parameter	Value
C	0.1, 1, 10, 100, 1000
Kernel	Linear, Sigmoid, Polynomial, Gaussian

Table 12. Result Scenario 2

Number of Topics	Kernel	*C*	*Accuracy*
5	Linear	10 dan 100	86.36%
	Sigmoid	0.1	84.82%
	Polynomial	1 dan 10	86.41%
	Gaussian	1	86.34%
10	Linear	1	87.56%
	Sigmoid	0.1	86.19%
	Polynomial	10	87.78%
	Gaussian	1	88.09%
20	Linear	100 dan 1000	89.05%
	Sigmoid	0.1	88.47%
	Polynomial	10	89.31%
	Gaussian	1	89.88%
40	Linear	10	90.52%
	Sigmoid	0.1	89.03%
	Polynomial	10	90.23%
	Gaussian	1	91.17%
60	Linear	100	91.08%
	Sigmoid	0.1	89.28%
	Polynomial	10	90.71%
	Gaussian	1	91.61%

In figure 5, the results of the experiment of five parameters number of topics with the best use of parameters of each kernel results from tuning parameters showed results like the previous scenario of improved performance results on each parameter the number of topics used in this test.

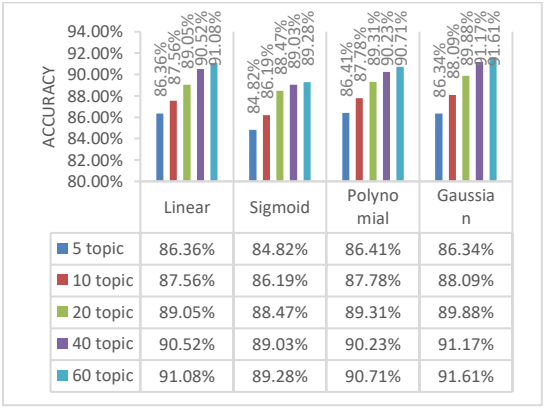
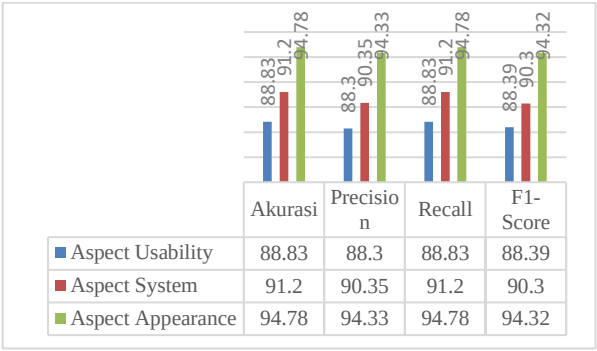


Figure 5. Comparison of Scenario 2 Accuracy Results

DOI: https://doi.org/10.29207/resti.v5i4.3143

3.3. Scenario 3

After getting the parameters C, kernel, and the best number of topics from the test scenario that has been done, the performance results of each aspect that has been determined in this study. In figure 6 shows the best performance results in the form of accuracy, precision, recall, and f1-score with the best Gaussian kernel parameters used.



	Akurasi	Precision	Recall	F1-Score
Aspect Usability	88.83	88.3	88.83	88.39
Aspect System	91.2	90.35	91.2	90.3
Aspect Appearance	94.78	94.33	94.78	94.32

Figure 6. Result Scenario 3

It can be seen in figure 6 the results of performance accuracy from the aspect of usability are 88.83%, f1-score with 88.39. For aspect System produces an accuracy of 91.2% as well as 90.3 for f1-score. The appearance gets 94.78% accuracy and an f1-score of 94.32.

4. Conclusion

Based on the results of the tests and analysis that has been done, it can be concluded that the system designed is aspect level sentiment analysis using Latent Dirichlet Allocation and SVM classification can be built using 6000 ZOOM Cloud Meetings application review data sourced from Google Play Store. In this study, five experiments were conducted for the parameters of the number of topics, namely as many as 5 topics, 10 topics, 20 topics, 40 topics, and 60 topics, and conducted test scenarios using linear kernels, Sigmoid, Polynomial, and Gaussian. This study resulted in the performance of the most optimal system accuracy of 91.61%, consisting of the performance of 3 aspects, namely aspects of usability with 88.83%, aspects of the system with 91.2%, and aspects of appearance with 94.78%. The most optimal result is obtained by using parameter C worth 1, 60 topics and use the Gaussian kernel. This, at the same time, proves that from the five experiments to the parameters, the number of different topics will result in increased performance from the experiment number of 5 topics as the smallest to the number of topics 60. In addition to being influenced by the number of topics or features, it can occur because it is influenced by the distribution of words on topics where the interrelationship between words in a topic also affects the distribution of topics in the document.

In further research, it was suggested to combine Dirichlet's Latent Allocation method with other methods such as TF-IDF or Word2vec and conduct experiments using classifiers other than Support Vector Machine such as Naïve Bayes, K-Nearest Neighbors and others.

Reference

[1] Zoom, "Zoom Help Center," 2021. [Online]. Available: https://support.zoom.us/hc/en-us (accessed Aug. 16, 2021).

[2] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, 2014, doi: 10.1016/j.asej.2014.04.011.

[3] N. Herlinawati, Y. Yuliani, S. Faizah, W. Gata, and S. Samudi, "Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes dan Support Vector Machine," *CESS (Journal Comput. Eng. Syst. Sci.*, vol. 5, no. 2, pp. 293–298, 2020, doi: 10.24114/cess.v5i2.18186.

[4] D. Ekawati and M. L. Khodra, "Aspect-based sentiment analysis for Indonesian restaurant reviews," in *ICAICTA 2017,International Conference on Advanced Informatics: Concepts, Theory and Applications International Conference on Advanced Informatics: Concepts, Theory and Applications*, 2017, pp. 1–6, doi: 10.1109/ICAICTA.2017.8090963.

[5] E. Wahyudi and R. Kusumaningrum, "Aspect Based Sentiment Analysis in E-Commerce User Reviews Using Latent Dirichlet Allocation (LDA) and Sentiment Lexicon," in *ICICOS 2019 - 3rd International Conference on Informatics and Computational Sciences: Accelerating Informatics and Computational Research for Smarter Society in The Era of Industry 4.0, Proceedings*, 2019, pp. 1–6, doi: 10.1109/ICICoS48119.2019.8982522.

[6] K. Bastani, H. Namavari, and J. Shaffer, "Latent Dirichlet allocation (LDA) for topic modeling of the CFPB consumer complaints," *Expert Syst. Appl.*, vol. 127, pp. 256–271, 2019, doi: 10.1016/j.eswa.2019.03.001.

[7] Suhardjono, W. Ganda, and H. Abdul, "Prediksi Kelulusan Menggunakan Svm Berbasis Pso," *Bianglala Inform.*, vol. 7, no. 2, pp. 97–101, 2019.

[8] Trias Ismi, "Web Scraping: Pengertian dan Apa Saja Manfaatnya Bagi Bisnis," *glints*, 2021. [Online]. Available: https://glints.com/id/lowongan/web-scraping-adalah/#.X7YvE2gzaMp (accessed Mar. 11, 2021).

[9] H. Manning, Christopher D. and Raghavan, Prabhakar and Schutze, *Introduction to Modern Information Retrieval*. USA: Cambridge University Press, 2008.

[10] K. SetyoNugroho, "Dasar Text Preprocessing dengan Python," *Medium.com*, 2019. [Online]. Available: https://medium.com/@ksnugroho/dasar-text-preprocessing-dengan-python-a4fa52608ffe (accessed Nov. 12, 2020).

[11] 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, 2019, doi: 10.1007/s11042-018-6894-4.

[12] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, no. 4–5, pp. 993–1022, 2003, doi: 10.1016/b978-0-12-411519-4.00006-9.

[13] C. Doig, "Introduction to Topic Modeling in Python," *CONTINUUM analytics*, 2015. [Online]. Available: http://chdoig.github.io/pygotham-topic-modeling/#/ (accessed Nov. 16, 2020).

[14] J. P. Jiawei Han, Micheline Kamber, *Data mining: Data mining concepts and techniques, third edition*. San Francisco: Morgan Kaufmann Publishers, 2012.

[15] I. Mathilda Yulietha and S. Al Faraby, "Klasifikasi Sentimen Review Film Menggunakan Algoritma Support Vector Machine," *e-Proceeding Eng.*, vol. 4, no. 3, pp. 4740–4750, 2017.

DOI: https://doi.org/10.29207/resti.v5i4.3143

[16] X. Ma, "Dimensionality-Reduction with Latent Dirichlet Allocation," *towards data science*, 2019. [Online]. Available: https://towardsdatascience.com/dimensionality-reduction-with-latent-dirichlet-allocation-8d73c586738c (accessed May 21, 2021).

DOI: https://doi.org/10.29207/resti.v5i4.3143