JURNAL INFOTEL
Informatics - Telecommunication - Electronics
Website Jurnal : http://ejournal.st3telkom.ac.id/index.php/infotel
ISSN : 2085-3688; e-ISSN : 2460-0997

Multi-aspect sentiment analysis on netflix application using latent dirichlet allocation and support vector machine methods

Attala Rafid Abelard[1*], Yuliant Sibaroni[2]
[1,2] School of Computing, Telkom University
[1,2] Telekomunikasi no 1 Street, Bandung, 40257, Indonesia
*Corresponding email: attalarafid@student.telkomuniversity.ac.id

Received 24 May 2021, Revised 02 August 2021, Accepted 25 August 2021

Abstract — Among many film streaming platforms that have sprung up, Netflix is the platform that has the most subscribers compared to the other platforms. However, not all reviews provided by Netflix users are good reviews. These reviews will later be analyzed to determine what aspect user reviews based on reviews written on the Google Play Store, using the Latent Dirichlet Allocation (LDA) method. Then, the classification process using the Support Vector Machine (SVM) method will be carried out to determine whether each of these reviews is included in the positive or negative class (Sentiment Analysis). Two scenarios were carried out in this study. The first scenario resulted the best number of LDA topics to be used is 40. The second scenario resulted that the use of the filtering process in the preprocessing stage reduces the score of the f1-score. Thus, this study resulted in the best performance score on LDA and SVM testing with 40 topics, and without running the filtering process with score of 78.15%.

Keywords – LDA, Sentiment Analysis, SVM

Copyright © 2021 JURNAL INFOTEL
All rights reserved.

I. INTRODUCTION

In the new era of technology that is developing rapidly like today, all kinds of things can be accessed very easily, especially in the entertainment industry. Film has almost become a major need for everyone these days. The fact that many film streaming platforms have sprung up, makes it easier for people to watch various kinds of films in the palm of their hands. Among those many film streaming platforms, Netflix is currently ranked first as the platform with the highest number of subscribers [1]. Thus, Netflix has received many reviews given by its users. Those reviews can either be positive or negative. However, over the years, there have been many kinds of expression used by users in providing reviews. Those various expressions could cause a review to be uncertain or ambiguous, especially with words that contain both negative and positive meanings in one review.

The analysis process used is sentiment analysis. Sentiment analysis is a process carried out to determine whether a review contains a positive opinion or a negative opinion. Therefore, this study will discuss the process of analyzing the Netflix application reviews by proposing the use of LDA feature extraction modeling method and the SVM classification method to detect the value status of each of the reviews given. The reason behind the use of SVM as the classification method is because, as stated in research [2] and [3] that SVM can provide maximum results in sentiment analysis.

In a previous study [4], Faizal conducted a sentiment analysis process on the Ruangguru application review dataset using the Support Vector Machine (SVM) classifier method, and managed to get and accuracy value in the of 90%. Research [5] proves that the aspect-based classification process on movie reviews gives a higher average accuracy score than the one that does not classify the data based on the aspect with an accuracy of 79.372%.

In the following study about SVM use [6], Mathilda concluded that the more data used in the training process, the higher the F1-score will be. This study compared several kernels owned by SVM, namely Linear kernels, Polynomial kernels, and RBF kernels. The test resulted that Linear kernels to have the highest F1-score value. Research [7] also proves that linear kernel generates the best performance score for the classification process. Research [8] used SVM to classify sentiment on tourism destinations based on several aspects. The researcher explains that the clustering process is a way to classify data that class label is unknown into a certain number of groups according to their similarity.

In the study about LDA [9], Xianghua conducted a global topic search with the LDA model, and then searched for topics that were more specific, or also called as local topics. As a result, the study gave an accuracy value of 92.15%. In another research [10], Eko Wahyudi conducted an aspect-based sentiment analysis on e-commerce reviews using LDA and Sentiment Lexicon. Based on three experiments conducted, it was found that the accuracy value of each experiment had an increase in value of 0.82%. The increase in the accuracy value is caused by the use of different train data for each predetermined aspect. In addition, the comparison of the performance of sentiment analysis on e-commerce reviews using LDA shows that a model with a combination of general train data and aspect-based train data produces the best accuracy value.

According to research [11], the LDA method generates great results in aspect-based sentiment analysis. However, in research [12], sentiment analysis on movie review using SVM and Doc2Vec gives fairly low f1 score, which is 54.1872%. Therefore, this study combines the two methods, LDA and SVM, to determine the effectiveness of the combination and the resulting f1 score on the proposed model.

Based on those studies, the motivation of this study is to conduct experiments to increase the performance score value of previous studies, by combining the LDA and SVM. In addition, the difference between the proposed research and previous studies is that this research runs several stages of scenarios in its testing, namely the use of several different topics in the LDA topic modelling process, and reducing certain preprocessing stages which are considered possible to reduce performance values. The dataset used for this study is the Netflix application reviews obtained from the Google Play Store. This research aims to build a sentiment analysis system on the Netflix application review using the SVM classification method and LDA feature extraction method. As well as to know the performance value of LDA and SVM combined in a multi-aspect calibration of sentiment analysis of the Netflix Application.

II. RESEARCH METHODS

The system to be built is a sentiment analysis process using LDA as a determinant of the number of aspects to be used, then the classification process of the sentiment using the SVM method. Figure 1 shows several stages of the whole system, starting with scraping the data from Google Play Store. The data collection process was carried out using a python library named google_play_scraper. The obtained data will later be labeled based on its sentiment polarity. After that, the labeled dataset will be preprocessed through 5 steps. Once the data is ready to be used, the data will go through the feature extraction stage using the LDA method, and later be classified into three different classes using the SVM classifier method. This study conducts two different scenarios. The first scenario runs the system using various numbers of topics on the LDA method with the number of topics of 20, 40, 60, and 80. After obtaining the number of topics that generate the best result, the second scenario uses that number of topics. It executes the system without using several preprocessing steps, i.e., filtering and stemming.

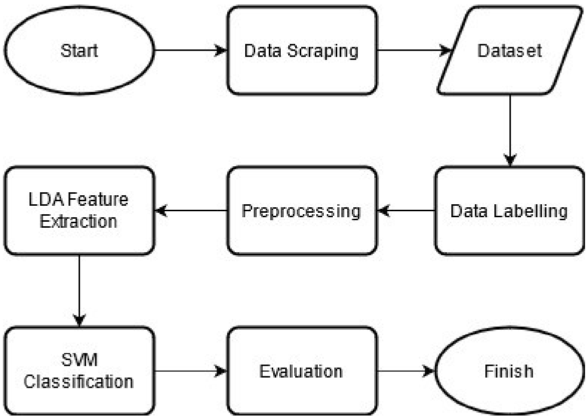

Fig. 1. System architecture

A. Data Scraping

Data scraping is the process of extracting data from a source that can be stored and processed. The data scraping process in this study is carried out using a library in python called *google_play_scraper* which has a special function to get data directly from the Google Play Store website. The data used in this study are 6000 reviews of the Netflix application that was obtained on February 12, 2021.

B. Data Labelling

The dataset is labeled manually. The first labelling stage is aspect identification. There are three aspects; *pembayaran* (payment), *sistem* (system), and film. From each of those aspects, several words represent each aspect. Those words work as a determiner for each review to which aspect it is categorized. The aspect identification can be seen in Table 1.

Table 1. Aspect identification

Aspect Category	Word(s)
Pembayaran	Mahal, Murah, Metode, Kartu

Aspect Category	Word(s)
Sistem	Mudah, Ribet, Lemot, Error, Susah
Film	Lengkap, Seru, Bosan, Banyak, Bagus

After that, the aspects of each review will be classified into three different classes, namely positive, neutral, and negative. A positive aspect is assigned a value of '1', neutral is assigned a value of '0', and negative is '-1'. Finally, if it is not discussed in a review, it will be given a value of '0'.

C. Preprocessing

In Natural Language Processing (NLP), information that is extracted has a random structure. Therefore, it is necessary to run a process that makes the data become structured for further steps. Data preprocessing is where the data is being cleaned and encoded to numerical values. After that, it is then given to the machine learning models [13]. Data preprocessing has several stages, which are:

a) Case Folding
The process aims to change all letters in the document to lowercase, and remove any other characters outside the alphabet, such as numbers and punctuation.

b) Tokenizing
This process aims to convert a sentence into words that will later be processed and analyzed.

c) Filtering
This process aims to remove words that usually appear in large numbers and are considered meaningless. This process uses an Indonesian stop words library derived from Natural Learning Toolkit (NLTK).

d) Normalization
The normalization process is used to homogenize words that have the same meaning but have different writing. This is usually caused by typos, and the use of non-standard language or 'slang' words.

e) Stemming
Stemming is the process of removing affixes to a word so that it changes it to its basic form.

D. LDA Feature Extraction

LDA is a probabilistic model for collecting discrete data (one data is unrelated to other data), such as a corpus. There are two different types of process in the LDA method, which are generative LDA and interference LDA. Generative LDA is a process of implementing a data corpus, which is derived from its latent variables. On the other hand, interference LDA is a process used to find the probability value in a topic, which then reports its value [14]. LDA's basic idea is that documents are represented as random mixes of latent topics, where each topic is determined by the word distribution[15]. It can be stated that the purpose of LDA is to find out how many topics are in a corpus, and the word distribution on each of the topics. Figure 2 shows the graph model of LDA.

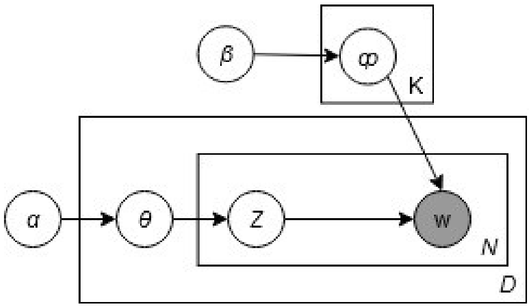

Fig. 2. LDA model illustration

Following the formal rules of the graphical model, nodes represent random variables and arrows indicate possible dependencies between nodes. The dark-colored node is the observed random variables, and the light colored nodes represent the latent random variable (likely to occur). Meanwhile, the box that is surrounding the nodes is a plate that shows replication. The outer plate represents a review and the inner plate represents the word. D and N indicate the number of product reviews and the number of words that are in a review. The main parameters used to build the model are the topic number alpha (*α)* and beta (*β).* These parameters work as determiners for the number of topics to find hidden topics. If the number of topics generated is higher, more topics will be obtained, and vice versa [16].

E. SVM Classification

After the process of feature extraction, SVM is used to predict the sentiment contained in each aspect of the review. SVM is one of the classification techniques that aims to find a hyperplane with the largest margin. A good hyperplane is obtained by maximizing the margin value. Margin is the distance between the hyperplane and the support vector (the point closest to the hyperplane) [12].

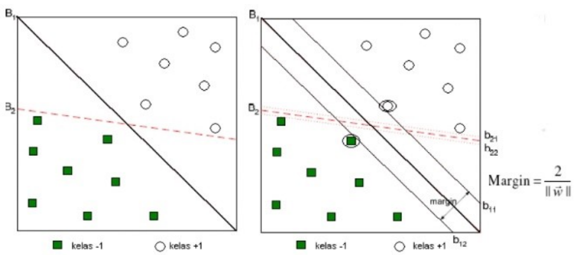

Fig. 3. SVM hyperplane formation

Figure 3 shows a hyperplane formation on SVM. The SVM structure consists of two classes, namely class -1 and class +1. The two data classes are separated by a hyperplane. The hyperplane value must be determined first in order to maximize the margin value [17] with the formula as seen in (1), With the condition of (2),

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

$$w.x_i + b = 0 \quad (2)$$

The formulas (3)(4) for predicting the maximum margin are as follows [18],

$$w.x_i + b \ \leq -1 \quad (3)$$

$$w.x_i + b \ \geq \ +1 \quad (4)$$

Based on the research [7] and [8] that have been mentioned earlier, SVM with linear kernel can generate great performance score. Thus, this study uses linear kernel on the sklearn library in python to classify the data.

F. Evaluation

Evaluation is the last step of the whole process. In this step, the performance score on the classification system is being evaluated. Sklearn library in python is used to show the performance scores for every functional aspect.

III. RESULTS

In this study, in order to get a model with the best accuracy, two scenarios were carried out by using 6000 reviews divided into 1800 (30%) data test and 4200 (70%) data train. The first scenario was performed by executing the whole process for each aspect using various numbers of topics on the LDA method.

After getting the number of topics that generate the best f1 score, the second scenario was performed by re-executing the process without using several preprocessing steps, i.e., filtering and stemming. This scenario aims to prove the effectiveness of the preprocessing steps mentioned.

A. Scenario 1

In this first scenario, system testing was performed by using various numbers of topics on the LDA method with the number of topics of 20, 40, 60, and 80 as the topic parameters. In the following model, the alpha (*α)* parameter is set to 0.001 for the distribution of internal topics documents. The beta (*β)* parameter is also set to 0.001 for the word distribution on the topic. Tables 2, 3, and 4 show the detailed f1 scores for each aspect.

Table 2. F1 score on pembayaran aspect

Number of Topics	F1-Score
20	76.92%
40	**81.67%**
60	79.83%
80	79.88%

Table 3. F1 score on sistem aspect

Number of Topics	F1-Score
20	68.81%
40	68.42%
60	**70%**
80	69.68%

Table 4. F1 score on film aspect

Number of Topics	F1-Score
20	82.31%
40	82.65%
60	**82.79%**
80	82.77%

After obtaining the performance scores of each aspect, the scores are summed up and the average score is calculated. The result shows that the number of topics of 40 generates the best f1 score. The result of the first scenario is shown in Table 5.

Table 5. Scenario 1 result

Number of Topics	F1-Score
20	76.02%
40	**77.61%**
60	77.56%
80	77.44%

B. Scenario 2

The first scenario shows that 40 is the best number of topics to be used in this experiment. Thus, scenario two will also use the number of topics of 40.

Table 6. Scenario 2 result

Scenario	F1-Score
Full Preprocessing	77.61%
No Stemming	74.65%
No Filtering	**78.15%**
No Stemming and Filtering	71.26%

Based on the second scenario shown in Table 6, the scenario that does not use the filtering process generates the best f1 score. It shows that the filtering process might erase some words that could be useful to determine the real meaning of the sentence. As in Table 7, the raw review is ‘film nya kurang lengkap’, which means Netflix’s film stock is not complete, and it should be classified as negative. However, the scenario that uses full preprocessing steps, including filtering, removes the word ‘kurang’ and changes the sentence to ‘film lengkap’. Thus, instead of classifying the review as negative, it classifies the review as positive. That makes the labelling process of that review to encounters an error.

Table 7. Preprocessing error

Data	Review	Sistem	Film
Raw	film nya kurang lengkap tapi aplikasi menonton ternyaman sejauh ini	1	-1
Full preprocessing	film lengkap aplikasi tonton nyaman	1	1
No Filtering	Film kurang lengkap aplikasi menonton ternyaman	**1**	**-1**

IV. DISCUSSION

Based on scenario 1, it can be seen in Tables 2, 3, and 4 that the film aspect generates the best f1 score among other aspects in every number of topics used. It occurs because of the data distribution between positive and negative on the film aspect is the most balanced one. However, the score itself can still be categorized as low. Thus, it can be concluded that unbalanced data can affect performance score. One of the reasons for the data to be imbalance is the selection of the review region, where users from Indonesia tend to experience still some difficulties on accessing the Netflix application due to some several reasons (i.e. the lack of information on how to gain access to the application using given payment methods), which leads them to leave some negative reviews.

After calculating the average number of all the f1 scores from each aspect, the result shows that the number of topics of 40 produces the highest f1 score, which is 77.61%. This shows that the higher the number of topics is used, it does not necessarily mean that the greater value will be generated. Then, by executing the system without using the filtering process results in the highest score which is 78.15%. Higher than the one that does not use stemming process and the one that does not use both filtering and stemming processes, which are 74.65% and 71.26%.

V. CONCLUSSION

The use of LDA and SVM methods to classify the sentiment polarity on Netflix application reviews can generate an f1 score with the best result obtained by the value of 78.15%, with the number of topics of 40, and by eliminating the filtering process at the preprocessing stage. This shows that the higher the number of topics used, it does not necessarily mean that the results given will be better, and not all preprocessing stages will always give better results, especially for in this case, filtering stage. In further research, another feature extraction method can be added to the system, so that the classification process can be more accurate, and the dataset can be changed with the English version review of the Netflix application, or other streaming platform, and fixing the word dictionary on the filtering process to reduce the possibility of data imbalance.

REFERENCES

[1] T. Bean, “Yes, Netflix Is The #1 Streaming Platform—But That Is Changing Quickly Thanks To This New Player,” *Forbes Magazine*, 2020. https://www.forbes.com/sites/travisbean/2020/10/16/yes-netflix-is-the-1-streaming-platform-but-that-is-changing-quickly-thanks-to-this-new-player/?sh=33c2e1cf6a7b (accessed Jan. 15, 2021).

[2] B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up? Sentiment Classification using Machine Learning Techniques,” vol. 31, no. 9, pp. 481–482, 2002, doi: 10.3115/1118693.1118704.

[3] A. Agarwal, B. Xie, Ll. Vovsha, O. Rambow, and R. Passonneau, “Sentiment analysis of twitter data,” *Proc. Work. Lang. Soc. Media (LSM 2011),"*, 2011.

[4] F. F. Irfani, “Analisis Sentimen Review Aplikasi Ruangguru Menggunakan Algoritma Support Vector Machine,” *JBMI (Jurnal Bisnis, Manajemen, dan Inform.*, vol. 16, no. 3, pp. 258–266, 2020, doi: 10.26487/jbmi.v16i3.8607.

[5] V. Parkhe and B. Biswas, “Aspect based sentiment analysis of movie reviews: Finding the polarity directing aspects,” *Proc. - 2014 Int. Conf. Soft Comput. Mach. Intell. ISCMI 2014*, pp. 28–32, 2014, doi: 10.1109/ISCMI.2014.16.

[6] 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.

[7] S. Al Faraby, E. R. R. Jasin, A. Kusumaningrum, and Adiwijaya, “Classification of hadith into positive suggestion, negative suggestion, and information,” *J. Phys. Conf. Ser.*, vol. 971, no. 1, 2018, doi: 10.1088/1742-6596/971/1/012046.

[8] F. F. Rahmawati and Y. Sibaroni, “Multi-Aspect Sentiment Analysis pada Destinasi Pariwisata Yogyakarta Menggunakan Support Vector Machine dan Particle Swarm Optimization sebagai Seleksi Fitur,” 2019.

[9] F. Xianghua, L. Guo, Y. Guo, and W. Zhiqiang, “Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon,” *Knowledge-Based Syst.*, vol. 37, no. ISSN 0950-7051, pp. 186–195, 2013, [Online]. Available: https://doi.org/10.1016/j.knosys.2012.08.003.

[10] E. Wahyudi and R. Kusumaningrum, “Aspect Based Sentiment Analysis in E-Commerce User Reviews Using Latent Dirichlet Allocation (LDA) and Sentiment Lexicon,” *ICICOS 2019 - 3rd Int. Conf. Informatics Comput. Sci. Accel. Informatics Comput. Res. Smarter Soc. Era Ind. 4.0, Proc.*, pp. 1–6, 2019, doi: 10.1109/ICICoS48119.2019.8982522.

[11] Y. Yiran and S. Srivastava, “Aspect-based Sentiment Analysis on mobile phone reviews with LDA,” *ACM Int. Conf. Proceeding Ser.*, pp. 101–105, 2019, doi: 10.1145/3340997.3341012.

[12] W. C. Widyaningtyas, A. Adiwijaya, and S. Al Faraby, “Klasifikasi Sentiment Analysis Pada Review Film Berbahasa Inggris Dengan Menggunakan Metode Doc2vec Dan Support Vector Machine (svm),” *eProceedings Eng.*, vol. 5, no. 1, pp. 1570–1578, 2018.

[13] K. S. Nugroho, “Dasar Text Preprocessing dengan Python,” 2019. https://medium.com/@ksnugroho/dasar-text-preprocessing-dengan-python-a4fa52608ffe (accessed Nov. 17, 2020).

[14] V. Octriany, “Aspect Based Sentiment Analysis With

Combination Feature Extraction LDA and Word2vec," 2021. Unpublished.
[15] D. M. Blei, A. Y. Ng, and M. T. Jordan, "Latent dirichlet allocation," *Adv. Neural Inf. Process. Syst.*, no. January 2001, 2002.
[16] F. E. Cahyanti, Adiwijaya, and S. Al Faraby, "On the Feature Extraction for Sentiment Analysis of Movie Reviews Based on SVM," *2020 8th Int. Conf. Inf. Commun. Technol. ICoICT 2020*, 2020, doi: 10.1109/ICoICT49345.2020.9166397.
[17] W. P. Ramadhan, N. Astri, and S. Casi, "Analisis Sentimen Menggunakan Support Vector Machine dan Maximum Entropy," *e-Proceeding Eng.*, vol. 4, no. 2, pp. 13–14, 2017.
[18] I. Subagyo, L. D. Yulianto, W. Permadi, and A. W. Dewantara, "Sentiment Analisis Review Film Di IMDB Menggunakan Algoritma SVM Sentiment Analysis of Film Review at IMDB using SVM algorithm Abstrak Pendahuluan Metode Penelitian," vol. 8, no. 1, pp. 47–56, 2019.