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) 672 - 679 ISSN Media Elektronik: 2580-0760

Sentiment Analysis of Public Opinion Related to Rapid Test Using LDA Method

Viny Gilang Ramadhan[1], Yuliant Sibaroni [2]
[1,2] Faculty of informatics, Telkom University
[1] gilangviny@students.telkomuniversity.ac.id, [2]yuliant@telkomuniversity.ac.id*

Abstract

In 2020 the world will be shocked by an outbreak of a disease that has developed tremendously. This disease is the Coronavirus. The Indonesian government, in overcoming conducted a Rapid early detection test in the spread of the Coronavirus. The steps of the Indonesian government have received rejection in several areas because people consume hoax news on social media. Indonesians widely use Twitter in conversations about the Coronavirus. Previous research was carried out using large-scale data, which affected the performance of the topic extraction method. The classification used resulted in poor accuracy using LDA to find the probability of topics in existing documents. LDA excels in large-scale data processing and is more consistent in generating the topic proportion value and word probability. Aspect-based sentiment analysis on public opinion regarding the rapid test on Twitter using LDA can determine aspects and public opinion on the rapid test. The test results of this study obtained 7000 tweets, four aspects of the results of topic using LDA, and getting the best accuracy using the RBF kernel by 95%. The sentiment of the Indonesian people towards the Rapid test is positive, with 4,305 sentiments.

Keywords: rapid test, Twitter, LDA, sentiment, corona.

1. Introduction

In early 2020 the world was shocked by an outbreak of a disease that developed extraordinarily, almost infecting the whole world. This disease is coronavirus or internationally called pandemic COVID-19 that causes conditions ranging from mild symptoms to cause death. In addressing the spike in deaths of almost 125 people per day in September 2020, the Government of Indonesia conducts rapid tests for early detection in the spread of the Coronavirus [1], [2]. However, the Indonesian government's move to reduce the death rate caused by the Coronavirus has been denied rapid tests in some areas because the public has been wrong in consuming hoax news circulating on social media [3]. With the corona pandemic globally, Twitter social media is recorded as the most widely used by Indonesians in discussions around the Coronavirus or 45.8%based on research from Zanroo Indonesia conducted on June 16 to 30, 2020.

According to Twitter's press release in 2019, there are already more than 500 million tweets by Twitter users per day, out of 500 million tweets used to upload user opinions and exchange information [4]. Opinions from tweets can be used to determine the sentiments that arise from exciting events such as the Coronavirus trending on Twitter. Computational studies of opinions, sentiments, and emotions expressed in the text are called sentiment analysis or opinion mining that focuses on classification issues. The polarity classification of text into sentences or documents to know opinions are positive, negative, or neutral [5].

In previous related research [6], there are weaknesses in classification accuracy using Naïve Bayes, which is not very good because it uses Tagging POS and little data resulting in an accuracy percentage of 50%. This also happens in conducted research [7] because it is not using weighting features to affect the performance of the SVM classification used, resulted in poor accuracy.

In previous related research conducted by [8], analyzing sentiment on social Twitter using the Method Latent Dirichlet Allocation (LDA) can produce words that often occur and organize them into different topics created by the LDA based on the highest coherence score. Using the same method in research conducted by [8] can extract sentiment and aspects of e-commerce applications to know reviews from users. The study produced four major topics made by LDA based on the highest

Accepted by editor: 18-05-2021 | Final Revision: 25-07-2021 | Online Publication: 20-08-2021

coherence score so that the topic is used as a reference for manually labelling aspects.

Research conducted by [9] conducted sentiment analysis by comparing KNN and SVM classification methods. In the study, KNN classification produced higher accuracy than SVM with 84% accuracy, while SVM produced 78% accuracy. But in the study SVM method excels in performance when using extensive data compared to the KNN method. Meanwhile, in the research conducted by [10], finding the best performance of KNN, SVM, and NB algorithms used in sentiment analysis related to Coronavirus on Twitter resulted from inaccuracy of 71%, 75%, and 76%. The SVM algorithm obtained the best results with 76%, with the amount of data amounting to 830 tweets that have done the preprocessing stage.

By conducting this research can generate opinions or opinions on the news of COVID-19, an especially rapid test on Twitter, and know what aspects are in the tweets discussed by the public. This study will use the LDA Topic Modelling method, which has been done in previous research. This method was chosen because it can search and extract frequently discussed topics in documents that perform very well with large data to produce better accuracy, such as research conducted by [11]. Previous studies conducted by [6] did not use the LDA method in topic extraction, so it only produced an accuracy of 50%. This is also the case in research conducted by [7] because not using the extraction feature causes poor accuracy. Therefore, to increase the accuracy of this study, the TF-IDF feature extraction will be used because the method can produce accurate results in the research conducted by [9]. So in this study will also use the SVM method used to classify data because it has fast performance and has a good level of accuracy compared to other classification algorithms [10].

Based on the description above, in contrast to previous studies, this study used data taken on February 1 – 1 March 2021 on Twitter related explicitly to the Rapid test. The novelty in this study is in visualizing the topic results and using the LDA results to be interpreted into aspects for dataset labelling and using SVM methods in classification. Based on the research with the LDA method that has been described above, the method can solve problems related to modelling topics and grouping terms into specific topics. So it is expected that the results of this study can help understand what the public is talking about on Twitter related to the rapid test to be used for agencies in need.

2. Research Method

The research method in this study is collecting the dataset, preprocessing, topic modelling, feature extraction, classification, and evaluation. The following figure 1 above is a research method flow that is built:

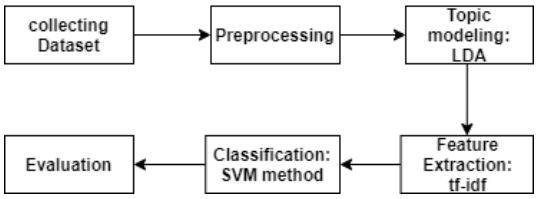


Figure 1. Research Method

2.1. Collecting dataset

In this study, the data used is in the form of primary data. The data source obtained in this study is a tweet that originated from Twitter related to Rapid test, a method to get the data by crawling data from Twitter. Crawling data by using search keywords rapid test, rapid antigen, rapid test free. Data is obtained in February 2021.

The number of tweets obtained from the crawling was 7000 tweets. The data is received by crawling the data using the python programming language. Data obtained from crawling is processed into the structure to be used as a model at the prepossessing stage.

2.2. Preprocessing

Preprocessing is the first step in sentiment text analysis. The use of suitable preprocessing techniques can also improve the performance of the model classifier [12]. Preprocessing data in this study was conducted Remove Punctuation, Case Folding, Tokenizing, Stopword Removal, and Stemming [13]

Remove Punctuation Marks This is to remove delimiters such as commas (,), period (.), All punctuation marks, numbers, and some typical components contained in tweets, namely username (@username), URL, and hashtag (#), because they do not affect anything in the sentiment analysis process. Then by removing that component in the tweet to reduce noise [12]. Examples of tweets and text processing results:

@Jokowi @kemenkes_ri Pelaksanaan Rapid tes petugas KPPS,GASTIB Ds. Kalirejo sebanyak 90 orang.#RapidTest#Covid19.

Pelaksanaan Rapid tes petugas KPPS GASTIB Ds Kalirejo sebanyak 90 orang.

Case folding is done to change the entire font size of the word to the same letter. Because not all tweet content is consistent in letters, this step converts the capital letter to lowercase. Examples of tweets and text processing results:

Personel Polsek Penawangan bersama Tim Gugus Tugas Covid-19 memantau Giat Rapid Tes di Ds Toko.

personel polsek penawangan bersama tim gugus tugas covid19 memantau giat rapid tes di ds toko.

Tokenizing is splitting text into chunks – pieces of words called tokens or single word pieces later processed for

DOI: https://doi.org/10.29207/resti.v5i4.3139

the next stage [14]. Examples of tweets and text processing results:

Pelaksanaan Rapid tes petugas KPPS GASTIB Ds Kalirejo sebanyak 90 orang

{'pelaksanaan', 'rapid', 'tes', 'petugas', 'kpps', 'gastib', ' Ds', 'kalirejo', 'sebanyak', '90', 'orang'.

Stopword removal is done to eliminate less important words or does not affect the analysis of sentiment. Examples of stopword in Indonesian are "*yang*", "*dan*", "*pun*", "*ke*", "*di*", etc. Examples of tweets and text processing results:

personel polsek penawangan bersama tim gugus tugas covid19 yang memantau giat rapid tes di ds toko

personel polsek penawangan tim gugus tugas covid19 memantau giat rapid tes ds toko.

2.3. Topic Modelling

After the preprocessing phase, a topic modelling will be conducted with the LDA to find the topics being discussed in each document or sentence so that it can be used to classify, summarize, and evaluate the similarity and relevance of the topics contained in the document [15]. In LDA form, the document is represented in a random mix of each resulting topic, and the topic itself comes from the word distribution. Here is the form of the LDA shown in figure 2.

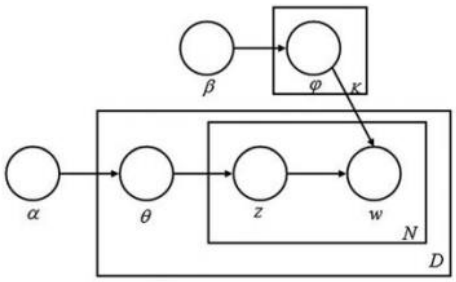


Figure 2. Visualization of LDA method

According to Figure 2 above, there are two levels in the LDA method. The α and β parameters are topic distribution parameters at the corpus level, i.e., the D-document. The αparameter is used to determine the distribution of topics in a document. The larger the alpha value in the document, the more topics discussed in the document [16]. Variable θ is the distributed topic of document D, ZN is the topic for the Nth word in document D, while WN defines the observed word. Variables β represent the probability of a word distribution in a topic. The β parameter is used to determine the distribution of words in the topic. The higher the beta value, the more words in the topic, and the lower the beta value, the fewer words in the topic, so the topic contains more specific words [16]. The variable represents the distribution of words in the topic -K.

According to [17] of some simplification, assumptions were made in distributing (latent) topics known to follow the distribution k Dirichlet. Second, the word probability is a matrix of β with a size of $k\ x\ V$ where $b_{ij} = p\ (w^j = 1 \mid z^i = 1)$. While the distribution of Dirichlet has a density function can be seen in equation (1) as follows:

$$p(\theta|a) = \frac{\tau(\sum_{i=1}^{k} \alpha_i)}{\prod_{i=1}^{k} \tau(\alpha_i)} \theta_1^{\alpha 1^{-1}} \dots \theta_k^{\alpha k^{-1}} \quad (1)$$

In equation (2) below, we can see that the combined distribution of mixture θ topics from N to topic Z and N to w with inputs α and β is as follows:

$$p(\theta, \boldsymbol{z}, \boldsymbol{w}|\alpha, \beta) = p(\theta|\alpha)\ p(w_n|z_n, \beta) \quad (2)$$

The marginal distribution form a document is obtain by integrating θ and summing z to produce the equation (3) below:

$$p(\boldsymbol{w}|\alpha, \beta) = \int(\theta|\alpha)\ \left(\prod_{n=1}^{N} \sum_{z_n} p(z_n|\theta)\ p(w_n|z_n, \beta)\right) d\theta \quad (3)$$

Finally, the equation for a document that will obtain a marginal probability of a corpus can be seen in equation (4) as follows:

$$p(\boldsymbol{D}|\alpha, \beta) = \prod_{d=1}^{M} \int p(\theta_d|a) \left(\prod_{n=1}^{N_d} \sum_{z_n} p\ (Z_{dn}|\theta_{dn}, \beta)\right) d\theta_d \quad (4)$$

So LDA works to include several parameters α and β, such as the number of documents, the number of words in the document, the number of topics, the number of iterations, and the coefficient of LDA. This is done to produce output in a list of topics, with weights for each document normalized according to the probability [18].

2.4. Feature Extraction

The weighting of features or words is done using the Term Frequency – Inverse Document Frequency (TF-IDF) method. TF-IDF is a feature extraction method for weighting that calculates the frequency of a word in a document using all the information or terms in the document [19]. This method is also famous for producing efficient, easy, and accurate results [20]. Term Frequency specifies the frequency term value that often appears in a document. If there is a high appearance value of the meal will have a significant impact on the value obtained. Even though the inverse document frequency (IDF) is A set of documents, the IDF can generate terms randomly [21]. TF-IDF can be formulated as follows [22] :

$$TF - IDF\ (t, d) = tf_{t,d}\ x\ log\ (\frac{N}{df_t}) \quad (5)$$

$Tf_{t,d}$ on formula (5) is the frequency of a *t*-word in the d -document, and N is the number of documents in the dataset. $d\ f_t$ is the number of documents in the dataset containing the word *t*. This study uses topic modeling

DOI: https://doi.org/10.29207/resti.v5i4.3139

with LDA and extraction features using TF-IDF. Using the method can produce a more accurate model in previous research. Topic modeling using LDA can overcome unlabeled data to create clusters of topics that can be used as labels. After that, the weighting feature is done using TF-IDF on the data that has been labeled from the results done using the LDA method [23].

2.5. Classification

Classification in this study using the support vector machine (SVM) method. The support vector machine (SVM) is included in the Supervised learning class, a kernel-based machine learning method to classify and regression in linear and non-linear problems [10]. SVM can classify two classes in input space by determining the best hyperplane value [24].

The SVM illustration shown in figure 3 with a straight red line separates the two classes, class +1 and -1 called a hyperplane. Similarly, the blue dotted line indicates the margin, which is the closest distance to the hyperplane.

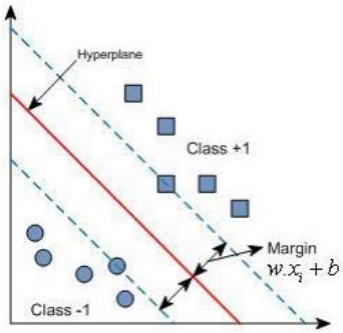


Figure 3. Support Vector Machine Illustration

The goal of SVM is to segment the datasets into groups to reach the Maximum Marginal Hyperplane (MMH).

2.6. Evaluation

The confusion matrix is used to test performance and evaluation with a matrix of predictions compared to the actual predictive result [9]. Calculations will be done based on the confusion matrix, namely precision, recall, f1-measure, and accuracy. Precision describes the accuracy of the requested data with the prediction results provided by the model (6). The recall represents the success of the model in rediscovering information (7). F1-measure harmonic combination of precision and recall directly proportional to the value of both (8). Accuracy is the percentage of text that is successfully classified appropriately by the system (9)

$$precision = \frac{TP}{TP+FP} \quad (6)$$

$$recall = \frac{TP}{TP+FN} \quad (7)$$

$$F1 - Measure = \frac{2(Precision\ x\ Recall)}{Precision+Recall} \quad (8)$$

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

True Positives (TP), namely the results of positive system predictions and under its positive targets. True Negatives (TN), namely the results of negative system predictions and according to its negative targets. False Positives (FP), namely the system's positive predictions, but the target results are negative. False Negatives (FN), namely the system's negative prediction, but the target results are positive.

3. Result and Discussion

The system testing results were carried out using a dataset resulting from crawling using an API key on Twitter social media and running using Python from 1 February - 1 March 2021. A total of 7,000 tweets were obtained. The dataset obtained will do the preprocessing process and continue the topic modeling using the LDA method to discuss the topic/aspect. The determination of the number of topics is determined by looking at the results of the coherence score graph that iterates 2 to 20 topics. Coherence score is one of the evaluation techniques in topic modeling to determine the optimal number of topics [25].

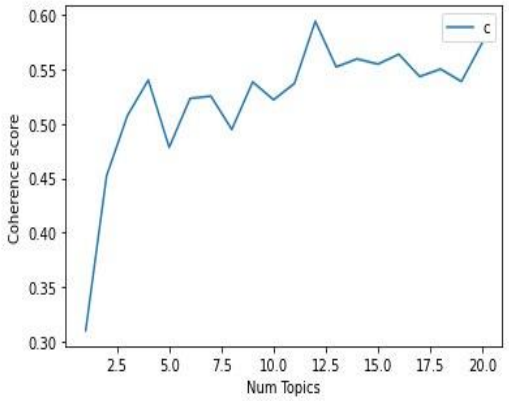


Figure 4. Graph of Coherence Values of 20 Topics

Figure 5. Visualization of 12 Topics

DOI: https://doi.org/10.29207/resti.v5i4.3139

In Figure 4, you can see a graph of coherence values for 20 topics, showing a graph that fluctuates throughout the iteration as the number of topics grows. Many topics can be selected by looking at the highest coherence score. Among the 20 iterations of topics performed, 12 topics had the highest coherence value. But from the figure 5 visualizations, LDA 12 topics using LDAvis precisely many clusters are intersecting. This indicates that intersecting or adjacent clusters can be merged into a single group so that subject clusters can be limited to fewer than 12 topic clusters.

Figure 6. Visualization of 4 Topics

When viewed on the number of topics = 4 in figure 6 visualization LDA 4 topics, resulting in a separate cluster, the words of each cluster can be interpreted topics. Although the coherence value for topic = 4 is 0.54, words can already be used in each cluster to get the topic spoken about. The top terms result of each topic cluster result is shown as shown in the table below:

Table 1. Word Distribution of Each Topic

Topic	example of each topic's word	interpretation of topic
1	di puskesmas juga bisa layani swab kok coba Tanya dipuskesmas terdekat	service
2	takut lihat orang rapid test kemarin ternyata sakit colok hidungnya pas rapid.	convenience
3	keluar masuk kota semarang harus membawa surat terang hasil negatif guna mendukung pemerintah melakukan ppkm.	regulation
4	info daftar rapid test dibandara buka jam 7 waktu setempat harap patuhi peraturan pemerintah karena menjadi syarat terbang.	informative

The vocabulary taken from each topic in table 1 can be interpreted, discussing what it is about. The words in the example column of each topic's word define topic interpretation more than any other word in each cluster. From table 1, four aspects result from the interpretation of the topic, namely service, convenience, regulation, and informative, which will be used as aspects for labelling the dataset.

3.1. Data Condition

Based on the dataset consisting of 4305 positive sentiments, 1935 negative sentiments, and 760 neutral sentiments generated after manually labeling, data visualizations will be performed, as shown below:

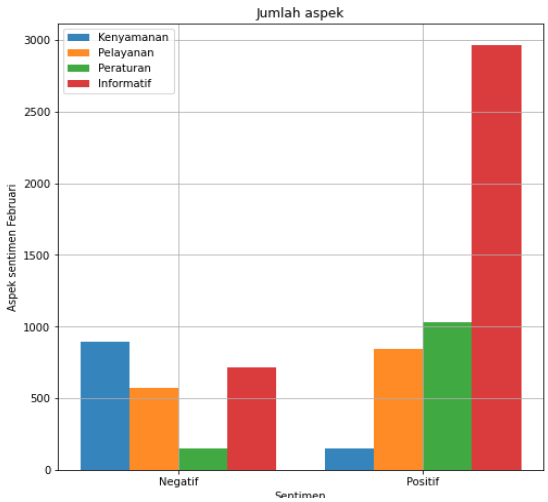
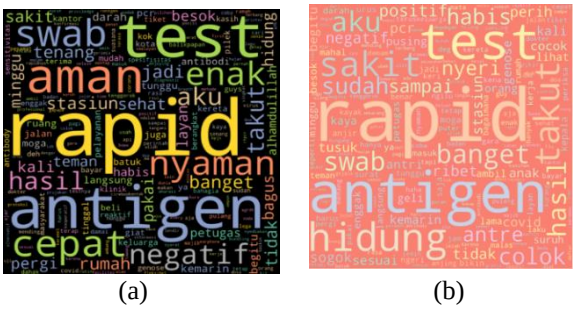


Figure 7. Number of Aspect Labels by Sentiment

In figure 7, the number of aspect labels based on sentiment shows informative aspects with positive sentiment. This is also because the data dominated by positive sentiment has the most number, while in negative sentiment, most have the convenience aspect.

3.2. Data visualization and interpretation

Data visualization according to the data that has been obtained using the wordcloud on sentiment in every aspect, as shown in the following figure 8:

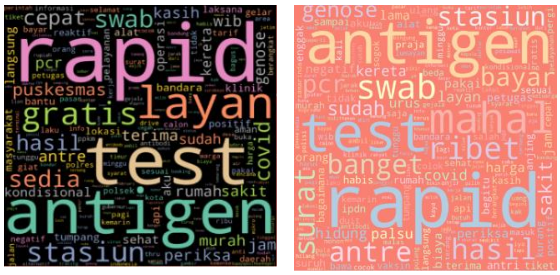


(a) (b)

Figure 8. Wordcloud Convenience Sentiment Aspects
(a) Positive and (b) Negative

DOI: https://doi.org/10.29207/resti.v5i4.3139

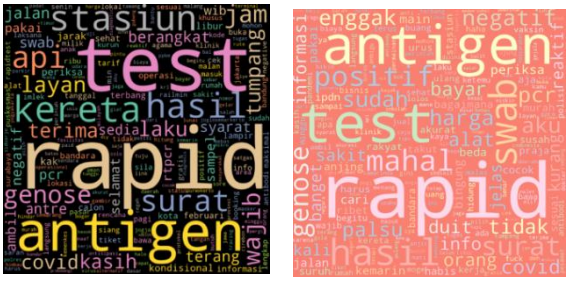
Figure 8 shows positive sentiment on the comfort aspect, showing that rapid testing feels comfortable, safe, fast, and not afraid. While in negative sentiment assessed by the public is when doing rapid tests, swab test feels discomfort such as pain, soreness, pain in the nose, making it afraid to do rapid tests again.



(a) (b)

Figure 9. Wordcloud Service Sentiment Aspects
(a) Positive and (b) Negative

Based on figure 9, people are happy with providing services when conducting rapid tests that can be done in health centers, hospitals, and even airports and provide free of charge. While on negative sentiment, people complain of expensive, complicated, long queues and offer fake mail services.



(a) (b)

Figure 10. Wordcloud Informative Sentiment Aspects
(a) Positive and (b) Negative

Wordcloud informative aspect in figure 10 on positive sentiment shows that the public uses Twitter to share rapid test, rapid test service, and rapid test information on the station. While on negative sentiment, people share information about rapid test antigens, swabs cost much money. There is still much lack of information about rapid tests in some areas.

(a) (b)

Figure 11. Wordcloud Regulation Sentiment Aspects
(a) Positive and (b) Negative

Wordcloud regulation aspect in figure 11 on positive sentiment shows that the community has complied with established regulations such as conducting rapid tests, implementing PPKM, and carrying out rapid test practice letters to travel. In negative sentiment, some people still do not want to do rapid tests. There are groups of vehicles that pass the inspection and do not follow the regulations set by the government.

3.3. Analysis of Test Results

In this study, testing based on SVM kernel functions was conducted using different kernel methods, namely linear, Gaussian radial base function, and polynomial. An analysis is done with the parameters that have been set and then implemented against the kernel functions used, then compared the accuracy value of each parameter.

Table 2. Linear Kernel Accuracy On Every Aspect

Value C	Aspect			
	Convenience	Service	Regulation	Information
1	94%	87%	90%	75%
10	93 %	85%	88%	72%
100	92%	83%	87%	68%
1000	92%	82%	87%	67%

Based on the accuracy results shown in table 2 above, you can see a comparison of accuracy in the SVM linear kernel using the value C as a comparison. It can be said that using the constant value C=1.0 gets the best accuracy results on every aspect. For example, on the Convenience aspect, get an accuracy of 94%. In this test, parameter C affects the accuracy from classification. The greater the value of C, the lower the accuracy value produced. This can be because the trade-off between margin and error is getting bigger.

Table 3. RBF Kernel Accuracy In Every Aspect

Parameter Value	Convenience	Service	Regulation	Inform -ative
c:1,g:1	94%	88%	90%	75%
c:1,g:10	90%	84%	86%	61%
c:1,g100	90%	84%	86%	59%
c:10,g:1	95%	90%	90%	75%
c:10,g:10	90%	84%	87%	62%
c:10,g:100	90%	84%	87%	59%
c:100,g:1	95%	89%	91%	75%
c:100,g:10	90%	84%	87%	62%
c:100,g:100	90%	84%	87%	59%

Based on the accuracy results as shown in table 3, it can be seen that the accuracy value of the SVM Gaussian RBF kernel uses the c (constant) and g (gamma) values as comparisons. The gamma value determines the hyperplane shape, where the smaller the gamma value, the more linear the hyperplane shape will be. Based on table 4, the best classification accuracy for each aspect is 95% using a constant value of C = 10 and gamma = 1 in the Convenience aspect. It can be said that the greater the value of C, the results of the accuracy do not have a

DOI: https://doi.org/10.29207/resti.v5i4.3139

significant change and even look constant. So it takes a larger C parameter.

Table 4. Polynomial Kernel Accuracy In Every Aspect

Parameter Value	Convenience	Service	Regu-lation	Informa-tive
g:1,d:1	94%	88%	90%	75%
g:1,d:2	94%	88%	89%	75%
g:1,d:3	93%	85%	88%	74%
g:10,d:1	93%	85%	88%	72%
g:10,d:2	94%	88%	90%	74%
g:10,d:3	93%	86%	89%	74%
g:100,d:1	92%	83%	87%	68%
g:100,d:2	94%	88%	90%	74%
g:100,d:3	93%	86%	90%	74%

Based on the analysis results as shown in table 4, it can be seen the accuracy value of the polynomial kernel classification using the (g) gamma and (d) degree values. Following table 4, it can be seen that the results of the polynomial kernel classification experience a change in the accuracy value is very constant in almost every aspect, and the highest accuracy value is at the gamma value = 1 for the degree value = 1. Based on the above results, the smaller degree value, the small accuracy value. Therefore, an enormous gamma value is needed to obtain more excellent accuracy results.

4. Conclusion

Based on the test results in this study, several aspects refer to the opinion of the Indonesian people towards the Rapid test based on tweets written by users on Twitter, where each has positive and negative sentiments. These results are based on tweet data February 1st to March 1st, 2021, which obtained 7000 tweets. By using topic clustering using LDA, the number of aspects to be used is obtained. There are four aspects, namely convenience, service, regulations, and informative. The topic results from the LDA results are used for reference in the aspect annotation process.

On the aspect of convenience, the public still expressed negative sentiments with a total of 894 compared to 147 positive sentiments, and it can be concluded that the comfort in carrying out a rapid test is still lacking because there are still many rapid test users who feel pain and discomfort after carrying out a rapid test. In the service aspect, it is pretty good. It can be seen from the resulting public sentiment with 846 positive sentiments and 569 negative sentiments, where services in the implementation of rapid tests are good enough with services that are already available at health centers, hospitals, train stations, and sometimes provided free of charge, by the government. In the regulatory aspect, Indonesian society is very good with 1032 positive sentiments. The community is quite obedient in obeying the regulations that apply during a pandemic, such as carrying out health protocols, PPKM, and carrying negative results letters for traveling. For the informative aspect, the Indonesian people often share positive information regarding the rapid test and prevention of Covid-19, such as reminding each other continuously to keep your distance, share rapid test information, and share information on Covid-19 prevention for the community. This can be seen with a positive sentiment of 2968. So it can be concluded that the sentiment of the Indonesian people is positive towards the Rapid test, which was carried out as early detection of the spread of the COVID-19 virus.

The kernel functions contained in SVM can be used in this analysis sentiment research, namely linear kernel, RBF kernel, and polynomial kernel. Of the three kernels, it can be known that the RBF kernel has the highest accuracy average for each aspect. As in the convenience aspect, the average accuracy is as significant.

Classification accuracy uses the SVM kernel in every aspect. The highest accuracy is 95% in the comfort aspect using the RBF kernel at the gamma value of 1 and constant C above 10. The best accuracy is 90% for the service aspect using the RBF kernel at the gamma value of 1 and constant C = 10. Regulatory aspects get the best accuracy of 91% on RBF kernels at gamma = 1 and C values above 100. While in the informative aspect, the best accuracy is 75% for RBF kernels using gamma one and for each value C ≥1. Based on these results, the classification using the RBF kernel can produce maximum accuracy for this study.

The suggestion needed from this research to develop the following system is to use more datasets to get a larger set of words to help find the topics spoken. It addresses data imbalances first before classifying, using data resampling methods, and using other classification methods that have not been used in this study.

References

[1] J. Akbar, "1.254 Orang di Indonesia Meninggal Akibat Corona dalam 10 Hari, Ini Saran Epidemiolog"," *Kompas.com*, 2020. https://www.kompas.com/tren/read/2020/09/23/173300365/1.254-orang-di-indonesia-meninggal-akibat-corona-dalam-10-hari-ini-saran?page=all (accessed Oct. 14, 2020).

[2] satuan tugas penanganan covid, "rapid test massal," 2020. https://covid19.go.id/p/berita/gugus-tugas-berlakukan-rapid-test-massal-identifikasi-penyebaran-covid-19 (accessed Oct. 15, 2020).

[3] Replubika, "Menolak Rapid Test," 2020. https://www.republika.id/posts/7700/menolak-rapid-test (accessed Nov. 27, 2020).

[4] Y. Lin, "10 Twitter Statistics Every Marketer Should Know in 2019 [Infographic]," *Oberlo Blog*, 2019.

[5] A. Fahmi, I. Ramadhan, P. Studi, S. Informasi, and F. I. Komputer, "Analisis Sentiment Masyarakat Selama Bulan Ramadhan Dalam Menghadapi Pandemi Covid-19," vol. 1, no. 1, pp. 608–617, 2020.

[6] N. S. Hari, "Analisis Sentimen Berbasis Aspek terhadap Ulasan Masyarakat pada Google Maps," 2020.

[7] N. Monarizqa, L. E. Nugroho, and B. S. Hantono, "Penerapan Analisis Sentimen Pada Twitter Berbahasa Indonesia Sebagai Pemberi Rating," *J. Penelit. Tek. Elektro dan Teknol. Inf.*, vol. 1, pp. 151–155, 2014.

[8] S. P. Astuti, "Analisis sentimen berbasis aspek pada aplikasi

DOI: https://doi.org/10.29207/resti.v5i4.3139

tokopedia menggunakan lda dan naïve bayes," 2020.
[9] M. Rezwanul, A. Ali, and A. Rahman, "Sentiment Analysis on Twitter Data using KNN and SVM," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 6, pp. 19–25, 2017, doi: 10.14569/ijacsa.2017.080603.
[10] R. ; N. A. ; M. K. Risnantoyo, "JITE (Journal of Informatics and Telecommunication Engineering) Initial Centroid Optimization of K-Means Algorithm Using," *J. Informatics Telecommun. Eng.*, vol. 3, no. 2, pp. 224–231, 2020.
[11] S. W. Kim and J. M. Gil, "Research paper classification systems based on TF-IDF and LDA schemes," *Human-centric Comput. Inf. Sci.*, vol. 9, no. 1, 2019, doi: 10.1186/s13673-019-0192-7.
[12] S. Symeonidis, D. Effrosynidis, and A. Arampatzis, "A comparative evaluation of pre-processing techniques and their interactions for twitter sentiment analysis," *Expert Syst. Appl.*, vol. 110, pp. 298–310, 2018, doi: 10.1016/j.eswa.2018.06.022.
[13] I. Mentaruk, A. Herdiani, and D. Puspandari, "Analisis Sentimen Twitter Transportasi Online Berbasis Ontologi (Studi Kasus : Go-Jek)," *e-Proceeding Eng.*, vol. 6, no. 1, pp. 2029–2047, 2019.
[14] R. KURNIAWAN and A. APRILIANI, "Analisis Sentimen Masyarakat Terhadap Virus Corona Berdasarkan Opini Dari Twitter Berbasis Web Scraper," *Jurnal INSTEK (Informatika Sains dan Teknologi)*, vol. 5, no. 1. p. 67, 2020, doi: 10.24252/instek.v5i1.13686.
[15] I. M. K. B. Putra and R. P. Kusumawardani, "Analisis Topik Informasi Publik Media Sosial Di Surabaya Menggunakan Pemodelan Latent Dirichlet Allocation (Lda) Topic Analysis of Public Information in Social Media in Surabaya Based on Latent Dirichlet Allocation (Lda) Topic Modelling," *J. Tek. Its*, vol. 6, no. 2, pp. 2–7, 2017.
[16] M. Maryamah, A. Z. Arifin, R. Sarno, and R. W. Sholikah, "Enhanced topic modelling using dictionary for questions and answers problem," *Proc. 2019 Int. Conf. Inf. Commun. Technol. Syst. ICTS 2019*, pp. 219–223, 2019, doi: 10.1109/ICTS.2019.8850986.
[17] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, no. Jan, pp. 993–1022, 2003.
[18] 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.
[19] S. Qaiser and R. Ali, "Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents," *Int. J. Comput. Appl.*, vol. 181, no. 1, pp. 25–29, 2018, doi: 10.5120/ijca2018917395.
[20] A. Rafiqi, "Penerapan Algoritma Fuzzy," *ADLN Univ. Airlangga*, [Online]. Available: repository.unair.ac.id/29371/3/15 BAB II.pdf.
[21] 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.
[22] V. Amrizal, "Penerapan Metode Term Frequency Inverse Document Frequency (Tf-Idf) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Hadits Shahih Bukhari-Muslim)," *J. Tek. Inform.*, vol. 11, no. 2, pp. 149–164, 2018, doi: 10.15408/jti.v11i2.8623.
[23] D. Kim, D. Seo, S. Cho, and P. Kang, "Multi-co-training for document classification using various document representations: TF–IDF, LDA, and Doc2Vec," *Inf. Sci. (Ny).*, vol. 477, pp. 15–29, 2019, doi: 10.1016/j.ins.2018.10.006.
[24] 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.
[25] S. J. Blair, Y. Bi, and M. D. Mulvenna, "Aggregated topic models for increasing social media topic coherence," *Appl. Intell.*, vol. 50, no. 1, pp. 138–156, 2020, doi: 10.1007/s10489-019-01438-z.

DOI: https://doi.org/10.29207/resti.v5i4.3139