International Journal on Advanced Science Engineering Information Technology

Vol.13 (2023) No. 3
ISSN: 2088-5334

Observing the Performance of the TextRank Algorithm on Automatic Text Summarization for *Bahasa Indonesia*

Dani Gunawan [a,*], Deden Witarsyah [b], Dedy Syamsuar [c], Amalia Amalia [d], Abdurrohman [a], Romi Fadillah Rahmat [a]

[a] *Department of Information Technology, Universitas Sumatera Utara, Medan, 20155, Indonesia*
[b] *Department of Information System, Telkom University, Bandung, 40257, Indonesia*
[c] *Information Systems Department, School of Information Systems, Bina Nusantara University, Jakarta,11530, Indonesia*
[d] *Department of Computer Science, Universitas Sumatera Utara, Medan, 20155, Indonesia*
*Corresponding author: *danigunawan@usu.ac.id*

***Abstract*—The research about automatic text summarization is common in English text. According to the previous study, automatic text summarization in *Bahasa Indonesia* is still challenging due to research in this area, especially the research which discusses TextRank algorithm performance, which is still meagerly. Accordingly, this research observes the performance of the TextRank algorithm to summarize the text in *Bahasa Indonesia.* The TextRank algorithm summarizes a text by sorting out the essential words and relevant sentences regardless of the source language. This algorithm uses a vertex to represent a word. The similarity measurement process will calculate the overlapping words (the same word between two vertices). These overlapping words are represented by the edge, which connects the vertices. Thus, the text forms a graph. This research focuses on the similarity measurement process to determine relevant sentences in a text. As the similarity measurement is critical for the summarization result, this research switches the original process to the Levenshtein Distance algorithm and observes its performance. This research uses the human-produced summarized text by the expert in *Bahasa Indonesia* linguistics to evaluate the result. The evaluation method is conducted by using ROUGE-1 and ROUGE-2. The result shows that the average of ROUGE-1 and ROUGE-2 for the TextRank algorithm is 0.439 and 0.3186, respectively. Meanwhile, the modified TextRank obtains 0.3999 and 0.2805, respectively. Both of the algorithms have not shown satisfactory results as expected.**

Keywords*—Automatic text summarization; textrank; modified textrank; textrank performance; textrank *Bahasa Indonesia.

Manuscript received 19 Apr. 2021; revised 11 Jul. 2022; accepted 23 Feb. 2023. Date of publication 30 Jun. 2023.

I. INTRODUCTION

A shortened text version can be obtained automatically by distilling the key features of the sentences [1]. This method, automatic text summarization, is divided into two types [2]. This type summarizes the text by extracting its key features and combining them. Examples are the title word, content word, proper noun, sentence location, or sentence length. On the other hand, the second type extracts the text features and then recreates the new text by using those features. This summarization type is called abstractive. The extractive method is the most common method to summarize the text between both types.

Research about automatic text summarization is commonly conducted on English text. For example, context-based summarization [3] obtains the best summaries by applying different sentence scoring methods based on context. Meanwhile, personalized text summarization constructs personalized terms by the Singular Value Decomposition (SVD) and sentence matrix. This summarization obtained the terms by applying latent semantic analysis. Terms with the highest score compose the summarization text [4]. Another research utilizes two levels of Latent Dirichlet Allocation (LDA) and then automatically extracts a set of latent topics [5]. This method is called topic-based summarization.

TextRank, a text summarization algorithm, was claimed to work independently in any language because it measures the similarity among sentences using a graph-based model [6]. Thus, it has nothing to do with the language. Conversely, the research on automatic text summarization in *Bahasa Indonesia* is limited. Previous research utilized MEAD and modified IDF dictionaries for *Bahasa Indonesia* [7]. This research tested their method in eight articles and compared them with human-produced summarization. According to their result, the modified IDF dictionary for *Bahasa Indonesia* did not work well for the summarization in *Bahasa Indonesia*. Another research applies a different approach to extracting the similarity among text sentences. It utilized semantic analytics to extract sentences. Each sentence has vector values that are used as the calculation parameter [8]. The result of this research is confusing because the explanation and the presented data are different. Next, the research uses the term frequency to summarize text [9]. This research claims that 83.3 percent of the readers understand the meaning of the summarized text produced by the term frequency method [10]. Another method obtained text summarization by conducting the fuzzy model of the Mamdani inference system. This method was tested on 15 articles and compared to the human-produced summarization. The person who summarizes the text works as a teacher in *Bahasa Indonesia*. This method works slightly better than AutoSummarizer by Microsoft Office.

Recently, there have been several methods for summarizing text in *Bahasa Indonesia*. For example, previous research utilizes Latent Dirichlet Allocation (LDA) for determining the sentence features scoring and the Genetic Algorithm (GA) to determine the weight of sentence features [11]. This research claims that 69.7 percent of the readers stated that the summarized text represented the original text. Another research implements non-negative matrix factorization (NMF) to summarize 100 articles [12]. The result was compared to the human-produced summarization by the experts. However, the result was unsatisfactory because the summarized text from the experts is diverse.

Meanwhile, previous research by Hidayat compared Latent Dirichlet Allocation (LDA) [13] with K-Means and a Feature-based summarizer to summarize 398 articles. The research found that LDA is better at a compression rate of 40%. Another research implements text features and singular value decomposition (SVD) [14]. This research only calculates a single word as in [9] and does not consider multi-word expression (MWE) [15]. They found that the evaluation accuracy depends on the compression rate.

The most recently conducted research in 2017 offered different methods. It used sentence scoring and a decision tree algorithm to extract the sentences from the text [16]. This research extracts eight text features, such as upper case, similarity to the title, sentence position, TF-IDF, cue phrases, sentence length, proper noun, and numerical data. Fifty articles were tested by using this method and obtained an f-score of 0.58. Most research evaluates methods by comparing the results with the human-produced summarization. The human-produced summarization is considered the most appropriate reference for summarizing the text. However, most of that research does not compare the summarized text with one the expert in *Bahasa Indonesia* produced.

This research also relies on the human-produced summarized text. In order to achieve the high-quality data, we use the summarized text from an expert in *Bahasa Indonesia* linguistics. Meanwhile, another research uses TextRank as the reference of the summarized text [17] because the author of TextRank claims that the algorithm can be used in any language. However, there is no evidence yet that the TextRank algorithm performance to summarize the text in *Bahasa Indonesia*.

The latest study summarizes multi-document in *Bahasa Indonesia* using the TextRank algorithm with Maximal Marginal Relevance (MMR) [18-24]. They summarize similar articles from online newspapers. The MMR removes similar sentences that are selected after the text summarization process. The average f-score of ROUGE-1 and ROUGE-2 are 0.5103 and 0.4257, respectively. This result shows that summarizing multi-document using TextRank + MMR algorithm has not shown satisfactory results.

This research will utilize TextRank to provide the summarization. Previous research discusses the automatic text summarization in *Bahasa Indonesia* [19] and found much space for improvement in this area. Most algorithms to summarize the text in *Bahasa Indonesia* do not demonstrate exemplary performance. Therefore, we also modified the TextRank algorithm to achieve better text summarization performance. The evaluation method to measure the performance is Recall-Oriented Understudy of Gisting Evaluation (ROUGE) [25-30].

The organization of this paper is as follows. The opening section introduces and discusses the fundamental of the conducted research, previous research, and the objective of this research. The second section describes the material and methods that are utilized in this research. Meanwhile, the following section discusses the outcome of this research. This section describes the detailed results of our proposed method. Finally, the last section concludes the whole conducted research.

II. Material and Method

Automatic text summarization is usually used to distill the key sentences in a text to obtain the text digest. The summarization helps find the available facts or data in the text. Previous research introduces TextRank as an algorithm to summarize a text by automatically selecting the most important and related sentences without regarding the source language. In other words, this algorithm will work in any language. The source of this research is the article news from online newspaper websites such as liputan6.com, kompas.com, and detik.com. Those credible online newspaper publishers produce news in proper *Bahasa Indonesia*. This research collects a hundred articles from ten categories: technology, health, sport, economy, lifestyle, national, science, travel, automotive and general. Each article has an average length of 400-900 words. Those articles were obtained manually. Fig. 1 shows that the data source is online newspaper articles. However, these online newspaper publishers put unrelated items such as images, read more links and advertisements. The TextRank algorithm will not use those items.

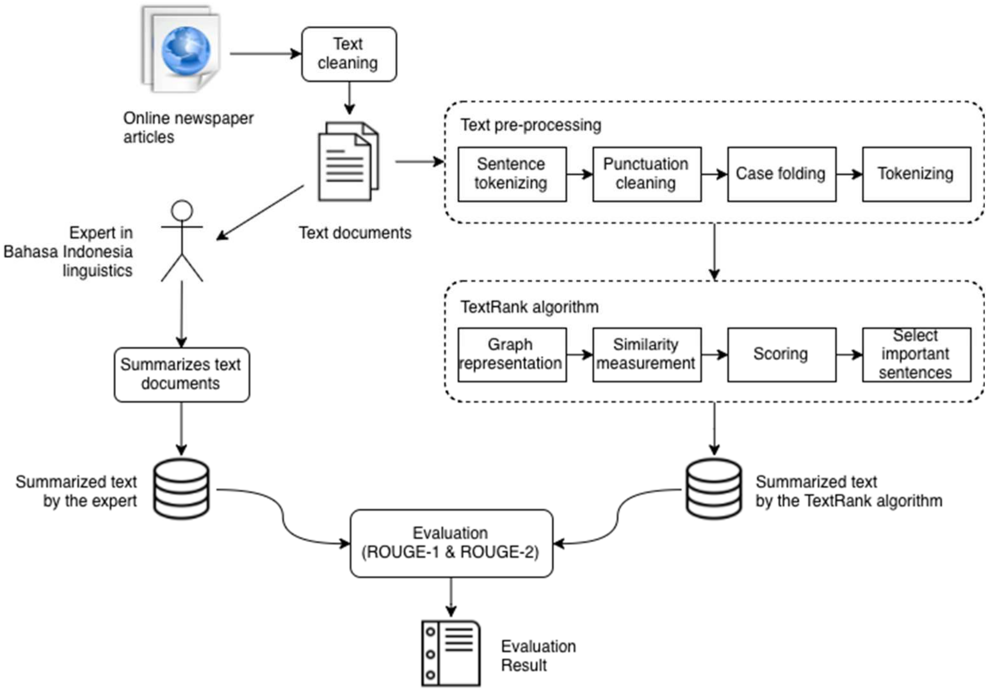
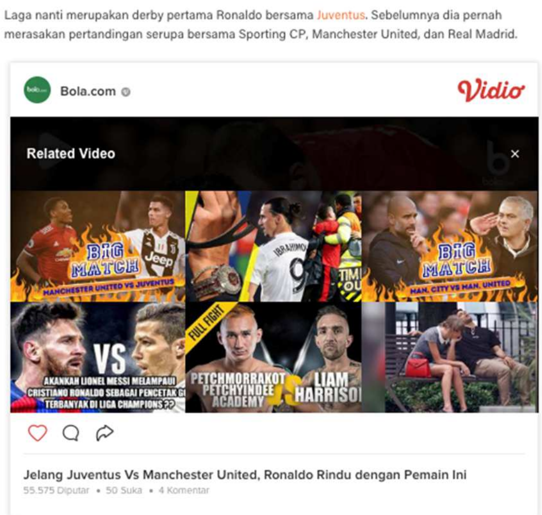


Fig. 1 General architecture of the research

Therefore, those items should be removed. The example of unrelated items is shown in Fig. 2. Fig. 2 (a) is the image in the middle of the article, and Fig. 2 (b) is the read-more links. Read more links have many variations and depend on the publishers. Fig. 2 (c) is the advertisement in the middle of the article.

(a) Image in the middle of the article

(b) Read more links in an article



(c) Advertisement in the middle of the text

Fig. 2 The example of unrelated items in an article

As shown in Fig. 1, the articles become text documents after the text-cleaning process. It means the texts are free from unrelated items. These documents are the source of human-produced (for evaluation purposes) and automatic text summarization. Furthermore, the automatic text summarization results will be compared to the human-produced summarization. The expert in *Bahasa Indonesia* linguistics will provide the human-produced summarization.

The expert imitates how the TextRank works by selecting the critical sentences. We create an online web application in order to record the human-produced summary. The expert only selects meaningful sentences by ticking the checkboxes, as shown in Fig. 3.

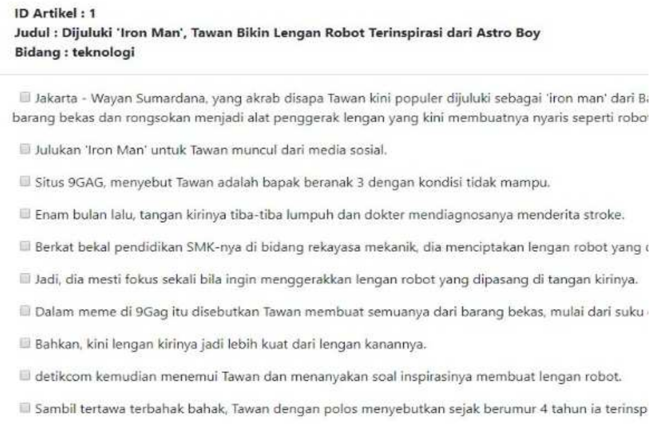


Fig. 3 A web application to record human-produced summary

This research utilizes the TextRank algorithm to implement automatic text summarization. In order to accomplish the summarization, there are a few steps that have to be done. After obtaining the text documents, the next step is text pre-processing. There are several processes in this step. The first process is sentence tokenization. Sentence tokenization means the text will be extracted into sentences. After sentence tokenizing, the text will be cleaned of punctuation. TextRank algorithm will not use these punctuations. Next, the text will be converted to lower or upper case (usually lower case). The conversion process is called case folding. The excerpt of the result is shown in Table 1.

TABLE I
THE EXCERPT OF THE EXTRACTED SENTENCES

No	Extracted Sentence
1	*Jakarta Kompas* (Translation: Jakarta Kompas)
2	*com di sela sela kunjungannya ke kantor menteri komunikasi dan…* (Translation: com on the sidelines of his visit to the minister's office communication and…)
3	*Setelah keluar dari ruangan keduanya segera menggelar jumpa pers rudiantara…* (Translation: after coming out from the room both immediately held a press conference rudiantara…)
4	*Saya senang berada di Indonesia negara digit ekonomi terbesar di…* (Translation: I am happy to be in Indonesia the country's largest economy digit in…)
5	*Punya juta pelanggan seluler lebih* (Translation: have more than million cellular subscribers)
6	*Indonesia sudah lompat ke generation technology* (Translation: Indonesia has jumped into generation technology)
7	*Tentu saja google akan senang hati mau membantu program digit ekonomi…* (Translation: of course, Google will be happy to help the digital economic program…)
8	*Apapun diantarkan misalnya kalau ingin makanan atau massage* (Translation: anything can be delivered for example if you want meal or massage)
9	*Membuat malas tapi juga memberi efisiensi* (Translation: make lazy but also gives efficiency)
10	*Ini masa ujarnya namun sebelum menteri melanjutkan kata…* (Translation: this is the time he said, but before the minister continued saying…)
11	*Bisa pesan massage juga* (Translation: can order massage too)
12	*Kalau begitu saya instal go jek sekarang menteri dan para hadirin pun…* (if so I will install go jek now the minister and the audience…)

The text pre-processing step yields sentence tokens ready to be included in the TextRank calculation. Then each sentence's weight to find the similarity among sentences in a text document. This process will determine the critical sentences. Initially, all the sentences will be assigned a score of -1. The TextRank calculation contains several processes, such as graph representation, similarity measurement, scoring each sentence, and selecting meaningful sentences. Those processes will be explained in detail as follow:

A. Graph Representation

Every token represents one vertex in a graph connected to obtain the similarity value. Fig. 4 shows the vertex 1-12, representing the sentences in Table I. The edge of the graph represents the similarity.

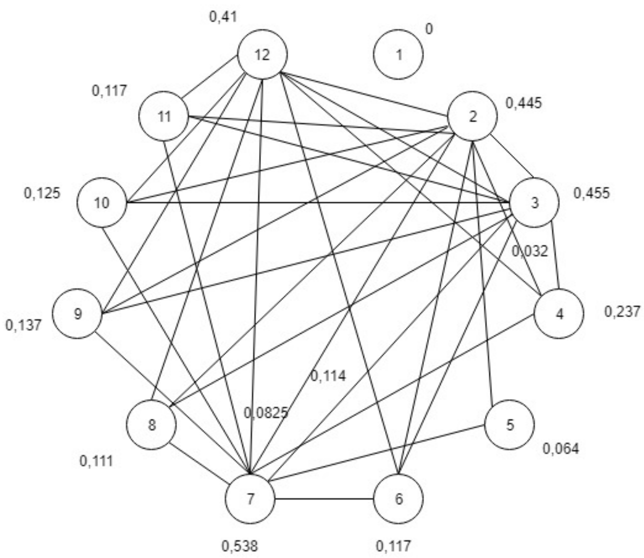


Fig. 4 Graph representation of extracted sentences

B. Similarity Measurement

Similarity measurement is done on the condition that every vertex has overlapping words (the same word between two vertexes) and normalized them. The normalization process divides the number of overlapped words by the length of every sentence. The vertexes with no overlapped words will be assigned 0 (zero) for the similarity score. Therefore, this vertex will not have an edge over other vertexes. As shown in Fig. 4, vertex one (1) does not have an edge because it has no overlapped words with other sentences.

TABLE II
SIMILARITY BETWEEN VERTEX 7 AND 12

Vertex	Similar Words After Sentence Extraction
7	xxxxx xxxx xxxxx xxxx xxxxxx xxxx xxx xxxxxxx xxxxxxx **digital** **ekonomi** **indonesia** xxxxxxx xx xxxx xxxx **gojek** xxxxxxxx xxxxxx xxxxxxxxx **seperti** **gojek** xxxxxx xxxx xxxxx xxxxx xx xxxxx **dia** xxxxxxxxxx xxxx xxxxx **aplikasi** xxx **merupakan** xxxxxx xxxx **gojek** xxx xxxx xxxxx xxxxxxx xxxxx xxxx xxxx xxxxx
12	yyyyy yyyyy yyyy yyyyyy **gojek** yyyyyyy yyyyyyy yyy yyyy yyyyyy yyy **dia** yyyyyy yyyyy yyyy **aplikasi** **seperti** **gojek** **merupakan** yyyy yyyyyy **ekonomi** **digital** **indonesia** yyy yyyyyyy yyyy yyyyy yyyy yyyy yyyyyyy

Table II shows the similarity between vertex 7 and 12. It is represented by the overlapped words between vertex 7 and 12. Table II shows that the overlapped words are printed in bold + underline. The letter "x" in the extracted sentence column (vertex 7) and the letter "y" (vertex 12) represent the words that have no relation between both vertexes. According to Table II, the number of the overlapped word (Wk) is 9. The similarity is measured based on equation (1). *log* ($|S_7|$) + *log* ($|S_{12}|$) are the length of the words in that vertex.

$$Similarity(S_7, S_{12}) = \frac{|\{Wk\ |Wk \in S_7\ \&\ Wk \in S_{12}\}|}{log(|S_7|)+log(|S_{12}|)} \quad (1)$$

Therefore, the similarity score between vertex 7 and 12 is:

$$Similarity(S_7, S_{12}) = \frac{9}{76+33} = \frac{9}{109} = 0{,}0825$$

This research also modifies similarity measurement by implementing the Levenshtein distance algorithm. We will observe the summarization performance after the TextRank similarity measurement is modified.

C. *Scoring*

After the similarity score between two vertexes in all sentences has been obtained, the following process assigns the final score to each vertex. The final score is assigned by summing up the edges toward a particular vertex. Table III shows all the vertex with the final score.

TABLE III
FINAL SCORING

Vertex	Final Score
1	0
2	0.445
3	0.455
4	0.237
5	0.064
6	0.117
7	0.538
8	0.111
9	0.137
10	0.125
11	0.117
12	0.410

D. *Select Important Sentences*

The final similarity score represents the critical sentences. The most important sentence is the one with the highest final similarity score. According to Table III, the order based on its rank from the highest to the lowest final similarity score is vertex 7, 3, 2, 12, 4, 9, 10, 6, 11, 8, 5, and the last is 1.

III. RESULTS AND DISCUSSION

This research yields the summarized text by utilizing the TextRank algorithm. We also observed the TextRank performance by modifying similarity measurements. The original similarity measurement by TextRank is replaced with the Levenshtein distance algorithm. The result of both TextRank and modified TextRank will be compared to the user-generated summary. They all summarize 100 articles from several categories originating from online newspapers. The summarization result is shown in Table IV.

TABLE IV
FINAL SCORING

No	NW	NA	Text Rank		The Expert	
			AWS	PWS	AWS	PWS
1	>800	22	373	46%	256	30%
2	700-800	12	341	45%	203	27%
3	600-699	11	304	46%	178	27%
4	500-599	22	269	48%	123	22%
5	<500	33	143	39%	78	17%

Table IV shows that column NW means the number of words in the original articles after the text cleaning result. Next, column NA means the number of articles we collected according to the number of words. Column AWS means the average word after the summarization process. Meanwhile, PWS means a percentage of the compressed articles after summarization.

According to Table IV, the TextRank algorithm can summarize an average of 44.8 percent of the original text. Meanwhile, the expert can summarize an average of 24.6 percent of the original text. However, TextRank only successfully yields 67 shorter than the original text. The rest of the summarization result is similar to the original text because the TextRank algorithm obtains the same score in every token when measuring similarity. Thus, all the sentences are considered essential. However, modified TextRank (with the Levenshtein algorithm) yields 98 shorter texts than the original.

This research is evaluated by comparing the TextRank-generated and user-generated summaries. One hundred text documents will be compared among them. The evaluation measurement uses the Recall-Oriented Understudy of Gisting Evaluation (ROUGE) method. ROUGE has several measurements. However, this research only utilizes ROUGE-N (N-gram), which consists of ROUGE-1 and ROUGE-2. The difference between both measurements lies in the accuracy calculation. ROUGE-1 compares word by word, generated by the TextRank or by the user.

On the other hand, ROUGE-2 compares all the summarized text by the TextRank and by the expert. The evaluation result using ROUGE-1 of both original TextRank and modified TextRank is shown in Table V. The F-score 1 means the summarization result is the same between the TextRank and the expert. Conversely, the f-score of 0 means both summarized texts differ. Table V shows that 36 percent of the summarized text has an f-score between 0.5-1. On the other hand, only 23 percent of summarized text has an f-score between 0.5-1. According to the results, the original TextRank algorithm yields better-summarized text than the modified TextRank.

TABLE V
ROUGE-1 EVALUATION RESULT

F-Score	Original TextRank		Modified TextRank	
	Frequency	NP	Frequency	NP
1	-	0%	-	0%
0.9	-	0%	-	0%
0.8	2	2%	-	0%
0.7	3	3%	1	1%
0.6	8	8%	7	7%
0.5	23	23%	16	16%
0.4	23	23%	29	29%
0.3	20	20%	19	19%
<0.3	21	21%	29	29%

Table VI shows that the original TextRank algorithm only has 11 percent for an f-score of 0.5-1 for the ROUGE-2 evaluation method. This result is slightly better than the modified TextRank, with only 9% for the same f-score. Therefore, both algorithms yield summarized texts that are not very similar to expert summaries.

TABLE VI
ROUGE-2 EVALUATION RESULT

F-Score	Original TextRank		Modified TextRank	
	Frequency	NP	Frequency	NP
1	-	0%	-	0%
0.9	-	0%	-	0%
0.8	-	0%	-	0%
0.7	3	3%	1	1%
0.6	2	2%	2	2%
0.5	6	6%	6	6%
0.4	22	22%	12	12%
0.3	20	20%	22	22%
<0.3	47	47%	57	57%

There is still space for improvement in the modification of the TextRank algorithm in order to escalate the summarization performance. The summary of the f-score of each algorithm is shown in Table VII.

TABLE VII
THE AVERAGE OF EVALUATION RESULT

	Original TextRank	Modified TextRank
ROUGE-1	0.439	0.3999
ROUGE-2	0.3186	0.2805

IV. CONCLUSION

Generally, summarizing text automatically extracts the key features of the sentences in a text document and combines the result to create a short version of the original text. Unlike the research in the summarization field in English, the research in this area in *Bahasa Indonesia* is still limited. However, some of the previously conducted research does not use the summarized text from the expert in *Bahasa Indonesia*. Therefore, to achieve high-quality data, this research uses human-produced summarized texts generated by expert in *Bahasa Indonesia* linguistics. Thus, the result of this research can be considered as a valid result.

Furthermore, previously conducted research compared the summarized text with the text generated by the TextRank algorithm. However, no evidence supports the TextRank algorithm's performance for summarizing text in *Bahasa Indonesia*. This research found that the original TextRank algorithm and modified TextRank algorithm with Levenshtein Distance does not perform well in summarizing text in *Bahasa Indonesia*. Therefore, it supports the claim that there is much space for improvement in automatic text summarization for *Bahasa Indonesia*.

The performance of the TextRank algorithm is determined in the process of obtaining related sentences. Thus, future research might consider some approaches to improve the accuracy of the related sentences. For example, the following research might reckon the multi-word expression (MWE) in the text pre-processing. Therefore, the TextRank will not only rely on a single word to calculate the related sentences. For instance, "*kotak hitam*" has two meanings. The first is a black box, and the second is an electronic device in an aircraft that records the flight data. In the original TextRank, "*kotak hitam*" will be considered as two different words, a "*kotak*" (box) and "*hitam*" (black), not a compound word. Another approach that might improve the accuracy of the related sentences is using the synonym dictionary. Therefore, words with similar meanings will be treated as the exact words. In the original TextRank, every word is treated as different. Moreover, machine learning or deep learning might be considered to measure sentence similarity.

REFERENCES

[1] D. Jurafsky and J. H. Martin, *Speech and Language Processing (2nd Edition)*, 2nd ed. Prentice-Hall, Inc., 2009.

[2] V. Gupta and G. Singh Lehal, "A Survey of Text Summarization Extractive Techniques," 2010.

[3] R. Ferreira *et al.*, "A Context Based Text Summarization System," in *2014 11th IAPR International Workshop on Document Analysis Systems*, 2014, pp. 66–70.

[4] R. Moro and M. Bielikov', "Personalized Text Summarization Based on Important Terms Identification," in *2012 23rd International Workshop on Database and Expert Systems Applications*, 2012, pp. 131–135.

[5] S. Liu, M. X. Zhou, S. Pan, W. Qian, W. Cai, and X. Lian, "Interactive, topic-based visual text summarization and analysis," in *Proceeding of the 18th ACM conference on Information and knowledge management - CIKM '09*, 2009, p. 543.

[6] R. Mihalcea and P. Tarau, "TextRank: Bringing Order into Text," *Proc. 2004 Conf. Empir. Methods Nat. Lang. Process.* , 2004.

[7] B. Prasetyo, T. Uliniansyah, and O. Riandi, "SIDoBI: Indonesian Language Document Summarization System," in *International Conference on Rural Information and Communication Technology 2009*, 2009, pp. 378–382.

[8] P. P. Tardan, A. Erwin, K. I. Eng, and W. Muliady, "Automatic text summarization based on semantic analysis approach for documents in Indonesian language," in *2013 International Conference on Information Technology and Electrical Engineering (ICITEE)*, 2013, pp. 47–52.

[9] M. Fachrurrozi, N. Yusliani, and R. U. Yoanita, "Frequent Term based Text Summarization for Bahasa Indonesia," in *International Conference on Innovations in Engineering and Technology (ICIET'2013)*, 2013, pp. 30–32.

[10] A. Ridok and T. C. Romadhona, "Peringkas dokumen otomatis menggunakan metode fuzzy model sistem inferensi Mamdani," in *Seminar Nasional Teknologi Informasi dan Multimedia*, 2013, pp. 19–24.

[11] Silvia, P. Rukmana, V. R. Aprilia, D. Suhartono, R. Wongso, and Meiliana, "Summarizing Text for Indonesian Language by Using Latent Dirichlet Allocation and Genetic Algorithm," *Proceeding*

Electr. Eng. Comput. Sci. Informatics, vol. 1, no. 1, pp. 148–153, Aug. 2014.

[12] A. Ridok, “Peringkasan Dokumen Bahasa Indonesia Berbasis Non-Negative Matrix Factorization (NMF),” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 1, p. 39, Jul. 2014.

[13] E. Y. Hidayat, F. Firdausillah, K. Hastuti, I. N. Dewi, A. Azhari, and A. Azhari, “Automatic Text Summarization Using Latent Drichlet Allocation (LDA) for Document Clustering,” *Int. J. Adv. Intell. Informatics*, vol. 1, no. 3, p. 132, Dec. 2015.

[14] F. E. Gunawan, A. V. Juandi, and B. Soewito, “An automatic text summarization using text features and singular value decomposition for popular articles in Indonesia language,” in *2015 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, 2015, pp. 27–32.

[15] D. Gunawan, A. Amalia, and I. Charisma, “Automatic extraction of multiword expression candidates for Indonesian language,” in *2016 6th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, 2016, pp. 304–309.

[16] P. M. Sabuna and D. B. Setyohadi, “Summarizing Indonesian text automatically by using sentence scoring and decision tree,” in *2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, 2017, pp. 1–6.

[17] D. Gunawan, A. Pasaribu, R. F. Rahmat, and R. Budiarto, “Automatic Text Summarization for Indonesian Language Using TextTeaser,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 190, no. 1, p. 012048, Apr. 2017.

[18] D. Gunawan, S. H. Harahap, and R. Fadillah Rahmat, “Multi-document Summarization by using TextRank and Maximal Marginal Relevance for Text in Bahasa Indonesia,” in *Proceeding - 2019 International Conference on ICT for Smart Society: Innovation and Transformation Toward Smart Region, ICISS 2019*, 2019.

[19] D. Gunawan and A. Amalia, “Review of the recent research on automatic text summarization in Bahasa Indonesia,” in *Proceedings of the 3rd International Conference on Informatics and Computing, ICIC 2018*, 2018.

[20] C.-Y. Lin, “Rouge: A package for automatic evaluation of summaries,” in *Proc. ACL workshop on Text Summarization Branches Out*, 2004.

[21] I Ketut Gede Darma Putra,Rahmat Fauzi,Deden Witarsyah and I Putu Deva Jayantha Putra,"Classification of Tomato Plants Diseases Using Convolutional Neural Network," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 10, no. 5, pp. 1821-1827, 2020, doi: 10.18517/ijaseit.10.5.11665.

[22] I Ketut Gede Darma Putra,Deden Witarsyah,Muhardi Saputra and Putu Jhonarendra,"Palmprint Recognition Based on Edge Detection Features and Convolutional Neural Network," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 11, no. 1, pp. 380-387, 2021, doi: 10.18517/ijaseit.11.1.11664.

[23] L. C. Hao *et al.*, "Mobile Malaysian Sign Language Application," *2022 International Conference Advancement in Data Science, E-learning and Information Systems (ICADEIS)*, Bandung, Indonesia, 2022, pp. 1-5, doi: 10.1109/ICADEIS56544.2022.10037539.

[24] Oka Sudana,Deden Witarsyah,Adhitya Putra and Sunia Raharja,"Mobile Application for Identification of Coffee Fruit Maturity using Digital Image Processing," International Journal on Advanced Science, Engineering and Information Technology, vol. 10, no. 3, pp. 980-986, 2020, doi: 10.18517/ijaseit.10.3.11135.

[25] A. Rahmatulloh, R. I. Gunawan, I. Darmawan, R. Rizal and B. Z. Rahmat, "Optimization of Hijaiyah Letter Handwriting Recognition Model Based on Deep Learning," 2022 International Conference Advancement in Data Science, E-learning and Information Systems (ICADEIS), Bandung, Indonesia, 2022, pp. 1-7, doi: 10.1109/ICADEIS56544.2022.10037496.

[26] H. S. Zhou et al., "2D Mobile Vocab Library Learning Application," 2022 International Conference Advancement in Data Science, E-learning and Information Systems (ICADEIS), Bandung, Indonesia, 2022, pp. 1-5, doi: 10.1109/ICADEIS56544.2022.10037521.

[27] A. Priandhika Izzulhaq, R. Fauzi, S. Suakanto, A. Kadir Hassan Disina, H. Mahdin and I. Anka Salihu, "Development of User Management in Ihya Digital Ecosystem Using Iterative Incremental Method," *2022 International Conference Advancement in Data Science, E-learning and Information Systems (ICADEIS)*, Bandung, Indonesia, 2022, pp. 01-06, doi: 10.1109/ICADEIS56544.2022.10037391.

[28] N. K. Trivedi, R. G. Tiwari, A. Anand, V. Gautam, D. Witarsyah and A. Misra, "Application of Machine Learning for Diagnosis of Liver Cancer," *2022 International Conference Advancement in Data Science, E-learning and Information Systems (ICADEIS)*, Bandung, Indonesia, 2022, pp. 1-5, doi: 10.1109/ICADEIS56544.2022.10037379.

[29] Ananthapadmanabha M V,Dhanesh Kumar A C,Sabariraju S,Eswar M and Mathi Senthilkumar,"Cluster Ensemble Method and Convolution Neural Network Model for Predicting Mental Illness," International Journal on Advanced Science, Engineering and Information Technology, vol. 13, no. 1, pp. 392-398, 2023, doi: 10.18517/ijaseit.13.1.17498.

[30] Udoinyang G. Inyang,Funebi F. Ijebu,Francis B. Osang,Aderenle A. Afoluronsho,Samuel S. Udoh and Imo J. Eyoh,"A Dataset-Driven Parameter Tuning Approach for Enhanced K-Nearest Neighbour Algorithm Performance," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 13, no. 1, pp. 380-391, 2023, doi: 10.18517/ijaseit.13.1.1676.