Deep sequential pattern mining for readability enhancement of Indonesian summarization

Dian Sa'adillah Maylawati[1,2], Yogan Jaya Kumar[1], Fauziah Kasmin[1], Muhammad Ali Ramdhani[2]
[1]Centre for Advanced Computing Technology, Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia
[2]Department of Informatics, Faculty of Science and Technology, UIN Sunan Gunung Djati Bandung, Bandung, Indonesia

Article Info

Article history:

Received May 11, 2023
Revised Sep 24, 2023
Accepted Oct 9, 2023

Keywords:

Automatic text summarization
Deep learning
Indonesian language
Readability evaluation
Sequential pattern mining

ABSTRACT

In text summarization research, readability is a great issue that must be addressed. Our hypothesis is readability can be accomplished by using text representations that keep the meaning of text documents intact. Therefore, this study aims to combine sequential pattern mining (SPM) in producing a sequence of a word as text representation with unsupervised deep learning to produce an Indonesian text summary called DeepSPM. This research uses PrefixSpan as an SPM algorithm and deep belief network (DBN) as an unsupervised deep learning method. This research uses 18,774 Indonesian news text from IndoSum. The readability aspect is evaluated by recall-oriented understudy for gisting evaluation (ROUGE) as a co-selection-based analysis; Dwiyanto Djoko Pranowo metrics, Gunning fog index (GFI), and Flesch-Kincaid grade level (FKGL) as content-based analysis; and human readability evaluation with two experts. The experiment result shows that DeepSPM yields better than DBN, with the F-measure value of ROUGE-1 enhanced to 0.462, ROUGE-2 is 0.37, and ROUGE-L is 0.41. The significance of ROUGE results also be tested using T-Test. The content-based analysis and human readability evaluation findings are conformable with the findings of co-selection-based analysis that generated summaries are only partially readable or have a medium level of readability aspect.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

Corresponding Author:

Dian Sa'adillah Maylawati
Centre for Advanced Computing Technology, Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka
Hang Tuah Jaya Street, Melaka, 76100, Malaysia
Email: diansm@uinsgd.ac.id

1. INTRODUCTION

Natural language processing (NLP) technology is developing rapidly in the era of big data today. NLP is an artificial intelligence technology that can process language in many types of language at the lexical, semantic, or syntactic level [1]. One popular NLP application is automatic text summarization, which produces a summary from document collections using an extraction or abstraction approach [2], [3]. A summary is a condensed version of a document's content that includes most of the original text(s) information. Text summarization involves content reduction and generalization based on the source text's relevant content to produce the summary. Automatic text summarizing research frequently discovers new techniques to construct summaries to meet the needs of various applications and users.

The great challenge regarding text summarization is how to produce a summary that is easy to read and maintains the meaning of the text. Of the many studies on Indonesian text summarization, research has yet to be conducted to measure the resulting summary's readability to the best of our knowledge. Most studies focus on the accuracy of the algorithm used to produce the summary [4]–[8]. An automatic text summarization should produce a readable summary. However, most automatic text summarization approaches focus less on producing a readable and understandable summary.

Text representation must be capable of preserving the meaning of the text data. There are many text representations used in NLP research, beginning from the basic text representation such as a bag of words [9], [10], multiple of words or N-gram [11], semantic text representation [12], [13], until word embedding representation [14], [15]. One multiple-word representation is the sequence of words (SoW), a text representation that pays attention to the relation between words, even sentences and paragraphs. Text mining applications such as text classification and clustering have been shown to benefit from the sequence of words [16], [17], information retrieval [18]–[20], and even text summarization for English [21] and Malay [22].

However, every language has a unique grammar structure that influences the text's meaning. For example, the structure is adjective-noun in English, Japanese, and Dutch. Nonetheless, the structure is noun-adjective in Indonesian and German, for example, such as "beautiful girl" in English but "*gadis* (girl) *cantik* (beautiful)" in Indonesian. The language itself heavily influences the meaning of the text. Because "Adhaya is taller than Shafa" and "Shafa is taller than Adhaya" are different. The sequence of words attempts to preserve the meaning of the text by focusing on the order in which the words appear. To find the sequential item, this study employs sequential pattern mining (SPM) (in this case, sequence of words) [18], [23]. SPM produces a sequence of words that can be implemented in any language, including Indonesian.

There are numerous methods for automatic text summarization, such as feature-based methods [24], [25], graph-based methods [26], and deep learning [27]–[31]. The latest research on Indonesian text summarization that produced Indonesian dataset benchmarks named IndoSum [32] and IndoLEM [33] did not consider the readability of the generated summary. IndoSum provides an Indonesian news dataset for text summarization research. Then, IndoLEM developed IndoSum research to provide more Indonesian news datasets and increase the performance in producing summary results compared with IndoSum. IndoSum and IndoLEM also use deep learning. IndoSum uses long short-term memory (LSTM), while IndoLEM uses bidirectional encoder representations from transformers (BERT). deep learning was discovered to outperform other methods in developing Indonesian automatic text summarization [32], [33]. Most of those previous research use supervised deep learning to generate automatic summaries, even though many studies use unsupervised deep learning for automatic text summarization with good results too [34]–[38]. Unfortunately, these studies focus on automatic text summarization in English.

Based on the findings discussed above, this study takes advantage of the opportunity to use unsupervised deep learning with deep belief network (DBN) for Indonesian automatic text summarization. This study will integrate SPM with DBN as unsupervised deep learning (named DeepSPM) to increase the quality of summary results. Furthermore, the focus and novelty of this study is the readability of the summary results. The summary results of previous research projects are generally measured by recall and precision utilizing ROUGE evaluation [39]–[41]. Unfortunately, only using co-selection-based analysis, such as ROUGE, to assess the readability level of the automatic summary results is insufficient, so this study includes content-based analysis and human readability evaluation. This study's evaluation of readability using content-based analysis and humans is challenging because the metrics and instruments used must be consistent with Indonesian language characteristics. Based on the problem statement, the objectives of this research are as follows: i) to integrate sequential pattern mining and deep learning (DeepSPM) as an enhanced model to produce Indonesian text summary, and ii) to evaluate the readability of Indonesian text summary using co-selection-based analysis, content-based analysis, and human readability evaluation.

2. METHOD

2.1. Research activities and materials

This study includes several activities, which are depicted in Figure 1. These activities include data collection and preparation, producing an Indonesian automatic text summary using DBN, producing an Indonesian automatic text summary using a combination of SPM and DBN (DeepSPM), and evaluating the Indonesian text summary result of DBN and DeepSPM using co-selection-based analysis, content-based analysis, and human readability evaluation. ROUGE is used for co-selection analysis, and Dwiyanto Djoko Pranowo metrics, gunning fog index (GFI), and Flesch-Kincaid grade level (FKGL) are used for content analysis and human readability evaluation with two Indonesian language experts. These evaluations aim to determine whether the proposed approach can generate a readable Indonesian text summary.

In gathering and preparing the dataset activity, this research uses the IndoSum dataset that contains news articles from CNN Indonesia, Kumparan, Merdeka.com [32]. This dataset is commonly used as a reference in Indonesian automatic text summarization research. The news article dataset is divided into six categories: Entertainment, Inspiration, Sport, Showbiz, Headline, and Technology. IndoSum news articles are comprised of 18,774 news documents that are published in JSON format.

The following process is text pre-processing, a significant step in most computational linguistics investigations. Pre-processing raw texts ensures an appropriate representation and can be used effectively in experiments. Numerous text pre-processing processes are used in this work, including sentence separation, case-folding, tokenizing, deleting regular expressions (non-letter characters), removing Indonesian stopwords, and stemming using the Nazief-Adriani algorithm for the Indonesian language. Furthermore, the Sastrawi Python library can prepare Indonesian text data during the pre-processing phase [42]. Sastrawi provides the Indonesian stop words list and Nazief-Adriani as a popular stemming process for Indonesian text [43].

There are two experiment scenarios: (1) produce an Indonesian text summary with DBN, and (2) produce an Indonesian text summary with a combination of DBN and SPM (DeepSPM). The experiment is in Python, utilizing multiple libraries, including natural language toolkit (NLTK), Sastrawi, and PrefixSpan for SPM. The DBN algorithm is then created based on earlier research that used DBN for extractive text summarization [44]. This model already incorporated the RBM algorithm. In accordance with the requirements of this study, the model has been updated to be suited for Indonesians. Aside from replacing the stopwords list and dictionary for the Indonesian language, the features have been minimized to meet the needs of summarizing the Indonesian text. This study employs the following features: sentence position, numerical token, named entity recognition, sentence length, TS-ISF, and sentence-to-centroid similarity. The final process is evaluation, including co-selection-based analysis, content-based analysis, and human readability evaluation.

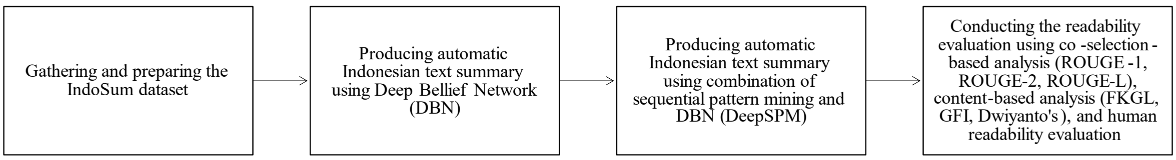


Figure 1. Research activities

2.2. Evaluation methods

2.2.1. Co-selection-based analysis

A co-selection-based analysis is a summary evaluation with a reference summary. The co-selection-based evaluation requires a document comparison summary based on the terms' co-occurrence in the system summary. The evaluation is performed by selecting frequent phrases from the system and reference summaries. ROUGE is a popular text summary evaluation tool. ROUGE is frequently used as a metric for summarizing results. ROUGE evaluates informativeness similarly to precision, recall, and F-measure. This research uses co-selection-based analysis with ROUGE-1, ROUGE-2, and ROUGE-L. ROUGE-1 and ROUGE-2 are a part of ROUGE-N, where N is 1, 2, 3, and 4. ROUGE-N score evaluation uses N-gram overlaps between the candidate and reference documents. ROUGE-N equation is available in (1) [40], [45].

$$ROUGE - N = \frac{\sum_{S\in\{Reference\ Summaries\}} \sum_{gram_n \in S} Count_{match}(gram_n)}{\sum_{S\in\{Reference\ Summaries\}} \sum_{gram_n \in S} Count(gram_n)} \quad (1)$$

In which *n* is the length of an N-gram, $gram_n$ is the maximum number of N-grams found in a candidate summary and a set of reference summaries, and $count_{match}(gram_n)$ is the maximum number of N-grams found in a candidate summary and a set of reference summaries. Conversely, ROUGE-L is based on the longest common subs. The longest common subsequence problem automatically determines the longest co-occurring in sequence N-grams by considering sentence-level structure similarities. The LCS has the advantage of not requiring consecutive matches but rather in-sequence matches that reflect sentence-level word order. It does not require a predefined N-gram length because it includes the sequence's longest common N-grams. The ROUGE-N equation is provided in (2).

$$ROUGE - L = \frac{LCS(gram_n)}{Count(gram_n)} \quad (2)$$

$LCS(gram_n)$ is the longest common subsequence between the reference and model outputs. Then, $gram_n$ is the maximum number of N-grams found in a set of candidate and reference summaries.

2.2.2. Content-based analysis

The content-based analysis is carried out without a reference summary and is used to assess the readability of the summary result. Text readability can be determined by the content and relatedness of sentence aspects [46]. The difficulty of the language's syntax and vocabulary and the relatedness between sentences (between previous and following sentences) determine the readability of the reference summary and the system's summary result. Readability can be evaluated using content-based analysis metrics evaluations, such as [46]–[49]: i) The FKGL evaluates text readability by using word complexity and sentence length, where difficult words and longer sentences influence the reader's focus to understand the meaning of the text; ii) GFI is a method for counting words with three or more syllables and determining grade levels based on the overall number of sentences; iii) The SMOG index (SMOGI) counts polysyllabic words in a defined number of sentences and provides an index of the text's relative complexity; iv) The Coleman-Liau index (CLI) uses length in characters in text to measure readability since, in this equation estimation, The length of a word in letters is superior to the length of a word in syllables; (5) The automated readability index (ARI) is a readability statistic used to assess text reading difficulty levels; and (6) relatedness with previous sentence (RPS) is a sentence readability measurement based on the cosine similarity calculation.

As previously stated, no one has concentrated on analyzing the readability of the summary outcomes in the Indonesian automatic text summarizing study. The study conducted that evaluated the readability of an Indonesian text using customer comments from the commercial website, not automated text summary [50]. Content-based analysis is required to assess the readability of Indonesian text. For English, SMOGI, GFI, and FKGL are utilized. Many studies employ readability metrics for English, while in Indonesian language study, FKGL, SMOGI, and GFI are appropriate [50]. Because evaluation using SMOGI may be done efficiently with a minimum of 30 sentences, the summary result should include more than 30 sentences. A GFI of 7 or 8 is ideal for readability. Most people find anything over the age of 12 too difficult to read. For example, the Bible, Shakespeare, and Mark Twain all have fog indexes of around 6. Time, Newsweek, and The Wall Street Journal average is around 11. While FKGL values between 0 and 6 are considered average, FKGL values between 6 and 12 are considered Skilled, and FKGL values between 12 and 18 are considered Skilled. Text written after the age of 18 is difficult to read. Equation (3) contains the FKGL equation, while (4) contains the GFI equation.

$$FKGL = 0.39\left(\frac{W}{|Sum|}\right) + 11.8\left(\frac{Syl}{W}\right) - 15.59 \tag{3}$$

$$GFI = 0.4\ \left[\left(\frac{W}{|Sum|}\right) + 100\left(\frac{cw}{W}\right)\right] \tag{4}$$

Several Indonesian language studies use English readability measurements (such as GFI, SMOGI, and FKGL) to assess the readability of Indonesian text, and most use surveys to do so [51]–[54]. Among the many Indonesian studies that employ surveys, questionnaires, and various measuring instruments for English, Dwiyanto Djoko Pranowo created a specific measuring instrument for Indonesian [50], [55]. Dwiyanto's measurements provide thirteen indications for evaluating the readability of Indonesian text. The thirteen indicators are classified as easy, medium, or tough to read. The readability of the Indonesian text can be determined by summing all indications depending on the range of criteria values. These indicators are as follows: i) the average number of paragraphs, ii) the average number of sentences per paragraph, iii) the sentence length, iv) the percentage of extension sentences, v) the percentage of compound sentences, vi) the percentage of sentences with polysemy, vii) the percentage of passive sentences, viii) the percentage of unfamiliar words, ix) the percentage of abstract words, x) the percentage of terms, xi) the percentage of conjunctions, and xii) the percentage of loan words. Dwiyanto's measurement conversion value is 13.0-21.7, which means easy; 21.8-30.5, which means medium; and 30.6-39, which indicates hard to read. Dwiyanto's stats are calculated using (5).

$$Dwiyanto's = \sum_{i=1}^{13} indicator_i \tag{5}$$

2.2.3. Human readability evaluation

Human evaluation is also necessary to ensure that the summary result is readable. Human readability is assessed using an expert/native/reader. Language experts should be involved in the evaluation, although this is rarely done. This is due to the limited time and professional resources available to test the readability of the text one by one, particularly when dealing with an enormous number of documents. Subjectivity can also have an impact on the outcomes of this human measurement. As a result, specialists who are odd in number perform human measurements best and have no personal stake in the outcome.

The factors for determining text readability are extremely diverse. It varies from one researcher to the next. However, automated text summarization research provides multiple grading criteria, including human evaluation. The experts will rate each summary result produced by the algorithm. The ratings are readable, partially readable, and non-readable, with numerous factors considered [47], [56]: i) the summary must be understandable, non-redundant, and focused on the main issue; ii) the summary's sentences must be complete and related to one another; and iii) the summary must not contain difficult language. If the summary meets all the above criteria, it is considered readable. The summary is classified as partially readable if it meets half of the requirements. Otherwise, it is considered unreadable.

Two Indonesian language experts are involved in the examination of human readability. The human assessment carried out by experts in measuring the readability of the automatic summary results does not actually have a standard format, but the evaluation can consider several aspects of readability such as: i) the question that aims to get the opinion about whether the summary result is easier to read or directly easy to understand the content; ii) the question that aims to know the opinion of the readability level of the summary result that has three options: unreadable, readable, or partially readable [46]; iii) the question that seeks to determine whether the assessors believe the summary result has content that is relevant to the main topic [46]; iv) the inquiry that seeks to examine if there is continuity between sentences in the summary results because continuity between sentences is one sign of a text's readability [46]; and v) the query that seeks to determine whether the summary result contains entire sentence qualities, resulting in a document that is both structurally sound and conveys information that is easily read and understood.

3. RESULTS AND DISCUSSION

3.1. Proposed deep sequential pattern mining (DeepSPM) for Indonesian text summarization

The DBN is a deep neural network comprising layers of restricted Boltzmann machines (RBMs) stacked on each other [57]. This technique learns all the top-down approaches and most significant generative weights layer by layer. These weights govern how variables in one layer affect variables in the layer above [58]. DBN can be used for unsupervised learning tasks like feature dimensionality reduction as well as supervised learning tasks like developing classification or regression models. A DBN is trained in two stages: layer-by-layer training and fine-tuning. After completing unsupervised training, fine-tuning refers to using error back-propagation methods to fine-tune the DBN's parameters. RBM networks are two-layer stochastic networks [59]. The RBM algorithm has two learning processes: gradient method and contrastive divergence, a quick learning method for RBM, and its described steps are in the passage [60].

This study's DBN model has six hidden layers that provide features for producing an Indonesian text summary:

a. The sentence position score determines a sentence's position within a text document. The opening and last phrases of a document are always the most important and contain the most information [36]. The feature of position is determined using (6).

$$Sentence_{position} = \begin{cases} 1, \; if\ first\ or\ last\ sentence \\ \frac{N-P}{N}, if\ others \end{cases} \tag{6}$$

b. The sentence length score determines how many words are in a sentence. Short sentences do not convey a lot of information. The feature score in (7) determines the important sentence based on its length.

$$Sentence_{length_i} = \frac{word\ in\ sentence_i}{words\ s\ in\ largest\ sentence} \tag{7}$$

c. Each word's appearance score is calculated using a numerical token score. The sum of the numerical values in a sentence is referred to as a numerical token. The equation is used to calculate it (8).

$$Numerical_{Token_i} = \frac{num\_numeric_{c_i}}{len} \tag{8}$$

where, $num\_numeric_{c_i}$ are the numerical tokens in i^{th} sentence, and len is the total words in i^{th} sentence.

d. The named entity recognition score reveals or detects important textual information. NER is the process of identifying and categorizing important information in a text. An entity is simply something that appears or is mentioned in the text several times. An organization, a person, a location, a date, a time, a model, a percentage, and other named entity types are examples. The entity's appearance in the sentence determines the NER score.

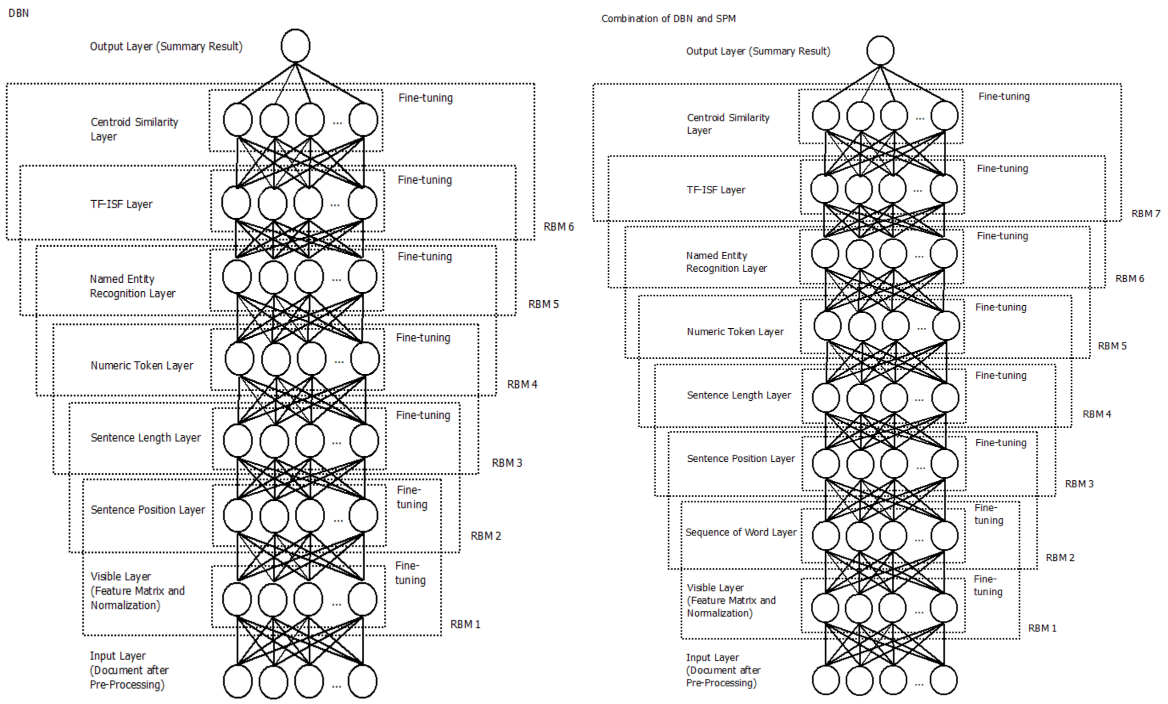
e. Term frequency and inverse sentence frequency score (TS-ISF) count how words appear in a sentence. For information retrieval systems, (TF-IDF) is necessary. In this study, text summarizing is done for a single document. TS-ISF calculation is available in (9). Where isf is the total occurrences of each term of i^{th} sentence in all other sentences, tf is the term frequency of each term in i^{th} sentence, and len is the total words in i^{th} sentence.

$$TFISF_{score} = \frac{\log(\sum_{all\_words} tf \times isf)}{len} \tag{9}$$

f. The centroid similarity score calculates the center of similarity between sentences. The cosine similarity between the centroid and the phrase is used to obtain this score. The centroid is the sentence with the highest TF-ISF. An equation calculates each sentence's cosine similarity to the centroid (10).

$$Cos_{sim_i} = \cos(sentence_i, centroid) = \frac{sentence_i \times centroid}{||sentence_i|| \times ||centroid||} \tag{10}$$

Figure 2(a) depicts the DBN model or architecture used to achieve this objective. Begin with the input layer, then move on to the visible layer, the seven hidden layers, and the output layer. There are six RBMs, each with its fine-tuning method. The PrefixSpan algorithm is used in this investigation [61] as one of the SPM methods. The sequential pattern layer process depiction will be inserted as a hidden layer in the DBN model. As a result, Figure 2(b) depicts the proposed architecture for combining DBN with SPM. As a result, there are eight hidden layers and seven RBM in the DBN and SPM procedure to produce an Indonesian test summary.



(a) (b)

Figure 2. Proposed methods architecture: (a) DBN architecture and (b) DeepSPM architecture

3.2. Co-selection-based analysis of experimental results of DeepSPM for Indonesian text summarization

This section summarizes the DBN ROUGE-1, ROUGE-2, and ROUGE-L results and the combined DBN and SPM results. This evaluation aims to compare DBN's performance with and without SPM. Precision, recall, and F-measure score are all considered in ROUGE's evaluation. Table 1 and Figure 3 present the ROUGE evaluation results of DBN and DeepSPM for Indonesian text summarization. ROUGE-1, ROUGE-2 and ROUGE-L results show an intriguing trend. These findings indicate that the proposed technique, which combines DBN and SPM, outperforms DBN alone in precision, recall, and F-measure.

Although the difference in value is not statistically significant, the combination of DBN and SPM improves recall and F-measure scores over DBN alone. However, the combination of DBN and SPM has significant value for precision value. It is almost three times increased.

Table 1. ROUGE evaluation result of DBN and DeepSPM

Method	Evaluation	Precision	Recall	F-measure
DBN	ROUGE-1	0.6659	0.3431	0.4529
	ROUGE-2	0.5317	0.2720	0.3599
	ROUGE-L	0.5904	0.3046	0.4019
DeepSPM	ROUGE-1	0.6798	0.3501	0.4621
	ROUGE-2	0.5460	0.2791	0.3694
	ROUGE-L	0.6031	0.3107	0.4101

3.3. Content-based analysis for readability evaluation of DeepSPM summary result
3.3.1. Dwiyanto Djoko Pranowo metrics result

There has not been much research on the readability of Indonesian literature. No one has concentrated on measuring the readability of summary outcomes in Indonesian automatic text summarization technology. The lack of a readability assessment metric specific to the Indonesian language is due to the limited study on assessing the readability of Indonesian literature. Several kinds of research on Indonesian text readability have adopted FKGL, GFI, and SMOGI, which are still compatible with Indonesian characters [50]. Dwiyanto Djoko Pranowo discovered only one statistic for gauging the readability of Indonesian text throughout this research [55]. Dwiyanto's metrics include thirteen indications derived from three text components: paragraphs, phrases, and words. The thirteen indicators are classified as easy, medium, or tough to read. The readability of the Indonesian text can be determined by summing all indications depending on the range of criteria values.

Based on the evaluation results of Dwiyanto's metrics for the readability of DBN and Deep SPM summary results, which are presented in Table 2, the summary results of DBN and DeepSPM have the same level of readability. DeepSPM has a higher score than DBN according to Dwiyanto's metrics (DBN has a value of 21.35, while DeepSPM has a value of 20.94). The sentence length in the summary results has a category level that is difficult to read based on the value of each indicator. The average sentence contains a long word (more than 15 words). DeepSPM outperforms DBN in terms of indicator scores for terms and conjunctions. Similarly, loan words. This is influenced by data sources originating from news portals, where the average reader is older than a teenager, allowing for the use of many borrowed words and long sentences.

Table 2. Readability indicators and score result of DBN and DeepSPM using Dwiyanto metrics

Indicators	DBN		DeepSPM	
	Average Score	Average of Dwiyanto's Score	Average Score	Average of Dwiyanto's Score
1. Paragraph	1.00	1.00	1.00	1.00
2. Sentence Count	8.75	1.19	**8.78**	**1.00**
3. Sentence Length	15.16	2.95	**15.09**	2.95
4. Extension	5.15	1.00	**5.12**	1.00
5. Compound	5.29	1.00	**5.26**	1.00
6. Polysemy	**5.44**	3.00	5.41	3.00
7. Passive Sentence	4.42	1.00	**4.41**	1.00
8. Unfamiliar Word	1.17	1.00	**1.16**	1.00
9. Abstract Word	5.16	1.00	**5.14**	1.00
10. Terms	5.46	1.70	**5.43**	1.70
11. Conjunctions	5.11	1.88	**5.07**	**1.66**
12. Loan	4.99	2.38	**4.96**	2.38
13. Phrase	4.19	**2.24**	**4.17**	2.25
Total of Dwiyanto's Score	21.35		20.94	
Category	Easy		Easy	

Note: Bold means the lowest value, the lowest value is better except for sentence count and polysemy

3.3.2. Gunning fog index result

The Gunning fog index equation's underlying message is that short sentences in simple English outperform long words in sophisticated language. A fog index score of 7 or 8 is optimal for readability. Most people find anything above 12 too difficult to read. As a result, based on the GFI evaluation result in Figure 3, the summary result of DBN and DeepSPM is still understandable for readers, as the average value is still less than 12. DBN has a value of 10.15, while DeepSPM has a value of 10.10. DBN and SPM are not significantly different, including maximum, minimum, and summary results above 12. DeepSPM, on the other hand, produces more summary results that are easier to read than DBN. The number of summaries with GFI values less than 8 indicates this. According to GFI metric theory, the value around 11 represents publication text, such as news. It is related to a dataset obtained from a news portal and used in this study. DBN and DeepSPM, on the other hand, have a lower GFI score when compared to previous research [50].

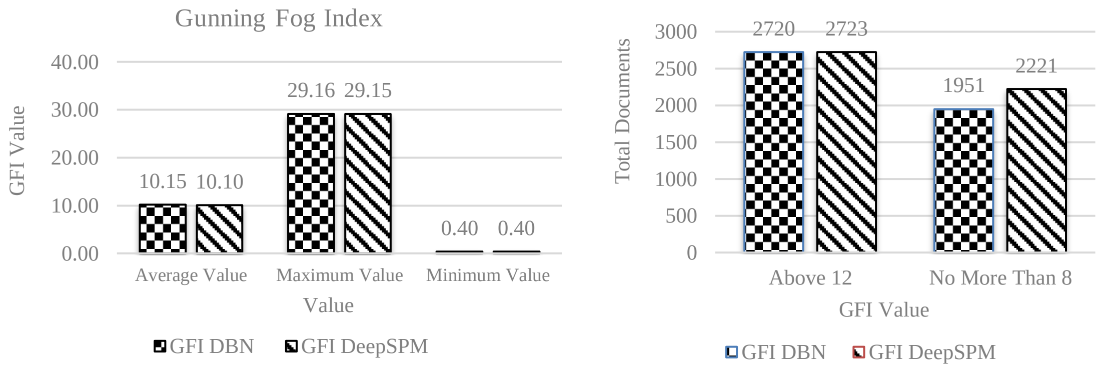


Figure 3. Result of GFI metrics for DBN and DeepSPM

3.3.3. Flesch-Kincaid grade level

The Flesch-Kincaid grade level, originally developed for the United States Navy, is best used in education. The US Department of Defense employs the Flesch-Kincaid grade level equation as a standard test. The procedure for analyzing the results is straightforward. The Flesch readability scores are the most widely used and evaluated readability scores. The reading level of the FKGL score is divided into three categories: easy, medium, and difficult to read. The easy category has FKGL values ranging from 0 to 6, medium has FKGL values ranging from 6 to 12, and hard has FKGL values ranging from 12 to 18 and above 18. Figure 4 shows that DBN and DeepSPM produce text summaries with average FKGL values of 19.19 for DBN and 19.12 for DeepSPM, based on references about the readability level using FKGL. It indicates that the Indonesian summary result is difficult to read because it is above 18. DeepSPM, on the other hand, has a lower FKGL maximum value than DBN. Then, both DBN and DeepSPM have a score of 7 for the FKGL minimum value. This means that DBN and DeepSPM can generate readable summaries.

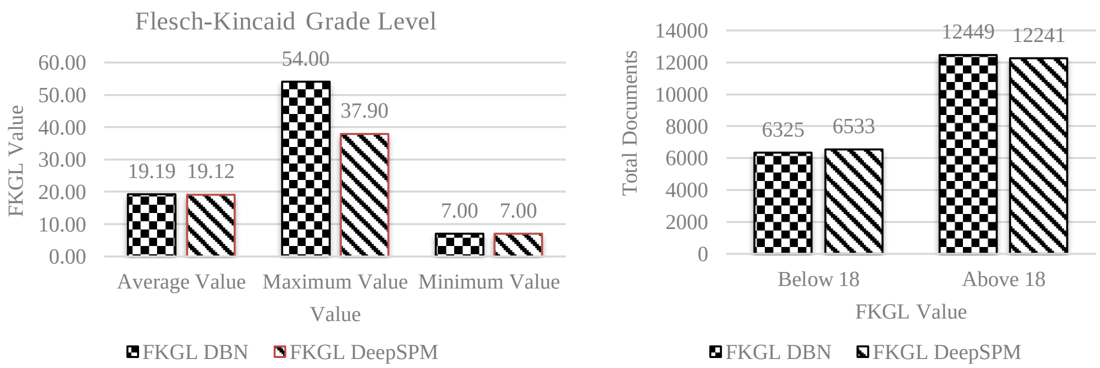


Figure 4. Result of FKGL metrics for DBN and DeepSPM

3.4. Human readability evaluation for DeepSPM summary result

This study asks two Indonesian language experts to evaluate the readability of the summary results. All evaluators willingly agreed to be one of the readability evaluators of the summary results for this study, with no conflict of interest. The evaluators will be required to read and complete the designed evaluation instrument. The first question becomes the main one in determining which DBN and DeepSPM summary is easier to read. Questions Q2 through Q5 will provide a more in-depth examination of the readability of the generated summary. The evaluation form includes evaluation objectives, instructions, summary results from

DBN and DeepSPM (a DBN and SPM combination), and questions about assessing the readability of an Indonesian text. The human readability evaluation results are classified into questions. Five questions about the readability of Indonesian text have been compiled from various sources. These questions are:
Q1: Which summary is easier to read?
Q2: What is the readability level of the summary you chose in question 1?
Q3: Does the summary result you chose in Question 1 focus on the main topic?
Q4: Is each sentence in your selected summary related to one another?
Q5: Does the summary result you choose in Question 1 contain incomplete sentences (not complete with Subject, Predicate, and Object)?

The first 1,878 sample summary results for each method were entered into the instrument by the evaluators (DBN and DeepSPM). Q1 through Q5 are answered in order. The first question becomes the main one in determining which DBN and DeepSPM summary is easier to read. Questions Q2 through Q5 will provide a more in-depth examination of the readability of the generated summary. When both evaluators (two evaluators) have a similar agreement when choosing DBN, DeepSPM, or both, the result of Q1 is taken. Based on the evaluation result of Q1 shown in Figure 5, two evaluators agree that five generated summaries from DBN are easier to read than 585 generated summaries from DeepSPM. The evaluators also agree that the 175 generated summaries from DBN and DeepSPM are both easy to read. This result demonstrates that DeepSPM outperforms DBN in producing more readable summaries.

An in-depth analysis was performed to evaluate the level and other readability indicators of DeepSPM's generated summary. The evaluation result of Q2 until Q5 are presented in Table 3. Based on the Q2 results, both evaluators agree that the readability level of the DeepSPM-generated summary is Partially Readable. Evaluators 1 and 2 rate DeepSPM-generated summaries as Partially Readable in 44.79% and 98.12% of cases, respectively. This means that while DeepSPM's generated summary is easier to read than DBN's, readers must exert more effort to comprehend the idea or event in the generated summary.

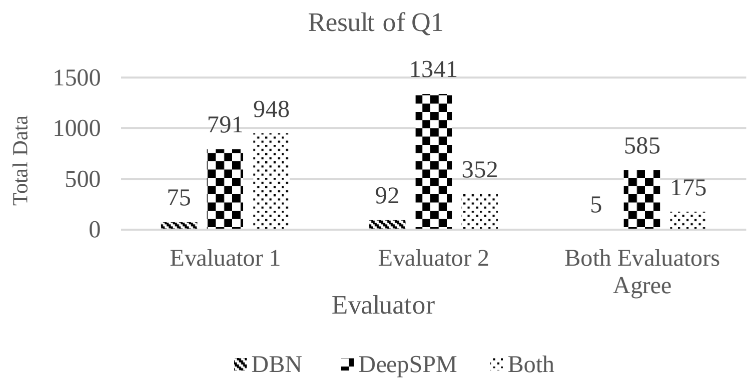


Figure 5. Result of Q1 for human evaluation

Table 3. Result of Q4-Q5 for human evaluation

ID	Choice	Evaluators	Total	Percentage	ID	Choice	Evaluators	Total	Percentage
Q1	Readable	Evaluator 1	302	51.62%	Q3	Yes	Evaluator 1	398	68.03%
		Evaluator 2	11	1.88%			Evaluator 2	582	99.49%
	Partially Readable	Evaluator 1	262	44.79%		No	Evaluator 1	187	31.97%
		Evaluator 2	574	98.12%			Evaluator 2	3	0.51%
	Unreadable	Evaluator 1	21	3.59%	Q4	Yes	Evaluator 1	398	68.03%
		Evaluator 2	0	0.00%			Evaluator 2	583	99.66%
Q2	Yes	Evaluator 1	398	68.03%		No	Evaluator 1	187	31.97%
		Evaluator 2	582	99.49%			Evaluator 2	2	0.34%
	No	Evaluator 1	187	31.97%	Q5	Yes	Evaluator 1	284	48.55%
		Evaluator 2	3	0.51%			Evaluator 2	11	1.88%
						No	Evaluator 1	301	51.45%
							Evaluator 2	574	98.12%

Then, the next evaluation (Q3) concerns whether the generated summaries from DeepSPM focus on the news topic category. There are six news topic categories that provide in IndoSum dataset: entertainment (*hiburan*), inspiration (*inspirasi*), sport (*olah raga*), showbiz, headline (*tajuk utama*) and technology (*teknologi*). The evaluators will choose "Yes" if the generated summary focuses on the news topic category and "No" if otherwise. The result shows that evaluators 1 and 2 agree that most of the generated summaries from DeepSPM are related to the news topic category. The next question, Q4, concerns relatedness between sentences in the generated summary from DeepSPM. Evaluators will choose "Yes" if the sentences in the generated summary are related to each other, while "No" for vice versa. The result shows that evaluators 1 and 2 agree that most of the generated summary from DeepSPM has sentences related to each other. The final question, Q5, determines whether the DeepSPM-generated summary result contains incomplete sentence attributes such as subject, predicate, and object. Evaluators will respond "Yes" if the generated summary contains an incomplete sentence and "No" if it does not. The evaluation result for Q5 shows that Evaluators 1 and 2 agree that most generated summaries have no incomplete attributes.

3.5. Discussion

This research proposes deep belief network (DBN) as an unsupervised learning approach in the Deep Learning method to generate Indonesian text summaries. As a result, this study used the IndoSum dataset to construct a summary of Indonesian texts using DBN as an unsupervised deep learning method. Then, SPM and DBN are combined as a proposed method called DeepSPM. SPM shows improvement in precision, recall, and F-measure values, according to the experiment results using ROUGE-1, ROUGE-2, and ROUGE-L. The recall, precision, and F-measure value have increased after DBN is combined with SPM. The precision of ROUGE-1, ROUGE-2, and ROUGE-L have increased from 0.6659, 0.5317, and 0.5904 to 0.6798, 54.6%, and 0.6031. Recall of ROUGE-1, ROUGE-2, and ROUGE-L have increased from 0.3431, 0.2720, and 0.3046 to 0.3501, 0.2791, and 0.3107. Then, the F-measure of ROUGE-1, ROUGE-2, and ROUGE-L also increased from 0.4529, 0.3599, and 0.4019 to 0.4621, 0.3694, and 0.4101. Seeing that SPM (using sequences of words as text representation) can improve the performance of summary results on unsupervised deep learning. Even though the result of DBN is lower than previous deep learning method for Indonesian text summarization, such as NeuralSum [32] and IndoBERT [33], but for further research combine SPM with those methods to enhance the performance.

The evaluation of readability for the Indonesian text summary result is one of the main concerns of the present study. Only a little research on automatic text summarization focused on the readability of summaries. To the best of our knowledge, no automatic text summarization research for the Indonesian language considers the readability of the generated summary. Therefore, this study has proposed three evaluation methods to evaluate the readability of Indonesian texts summary: co-selection, content-based and human readability evaluation.

As previously stated, studies have yet to focus on evaluating the readability of the summary results in Indonesian automatic text summarization research. However, a study assessed the readability of an Indonesian language website [50]. Most automatic text summarization studies use co-selection-based analysis, such as ROUGE, to measure generated summaries. Co-selection based method alone is not enough to measure the readability of the summary since it is only based on the similarity score. Therefore, besides co-selection-based analysis, this research also evaluates the readability of generated summaries with content-based analysis and human readability evaluation. As a result, this research evaluated readability using co-selection-based analysis (ROUGE-1, ROUGE-2 and ROUGE-L); content-based analysis (Dwiyanto Djoko Pranowo metrics, gunning fog index (GFI) and Flesch-Kincaid grade level (FKGL)); and human readability evaluation by two Indonesian language experts.

Readability evaluation using GFI shows that the summary result of DBN and DeepSPM is applicable to read. In addition, based on the GFI value, DeepSPM has higher readability than DBN because its GFI value is lower than DBN. Besides that, more summary results have a GFI score below eight on DeepSPM, meaning more articles are easier to read than DBN. The readability evaluation using FKGL found that DeepSPM has a lower value than DBN, which means that DeepSPM is easier to read than DBN, as evidenced by the FKGL value. In addition, a human readability evaluation for Indonesian text summarization is performed. Overall, the human readability evaluation results show that DeepSPM is better in terms of ease of readability.

Further analysis into the readability of DeepSPM shows that most of the generated summaries are partially readable according to all evaluators. Two evaluators agree that the summary results focus on the news topic category, have connectivity between sentences, and have fairly complete sentence attributes. The interesting discovery comes from one of the evaluators who has a different opinion about the main idea, the text's cohesiveness, and the completeness of sentence attributes in the entire text. This indicates that further research can be done to measure other aspects of machine-generated text summaries' readability.

Overall, this research contributes to i) producing an extractive Indonesian text summarization that incorporates a sequence of words as a structured representation of text using sequential pattern mining to improve summary quality; ii) investigating the impact of combining sequential pattern mining with unsupervised deep learning, the DBN, in generating Indonesian automated text summarization; and

iii) measuring the readability of Indonesian text summary with readability measurement using co-occurrence analysis. This research has implications for the development of automatic text summarization research in the research community in Indonesia. It can be used as a baseline that the readability aspect is important in the summary results. It is not enough to evaluate using co-selection-based analysis but must also be equipped with content-based analysis and human readability evaluation. Likewise, the human readability evaluation instrument can be adapted for future research.

4. CONCLUSION

This study proposes DeepSPM, an unsupervised deep learning with deep belief network (DBN) combined with sequential pattern mining (SPM) method for generating Indonesian text summaries. Overall, the findings of this study show that SPM can improve summary quality. The ROUGE-1, ROUGE-2, and ROUGE-L evaluation results show that the combination of DBN and SPM (DeepSPM) has the highest precision, recall, and F-measure value, indicating that incorporating text representation into the deep learning model has a better effect on Indonesian text summary generation. The results of content-based analysis and human readability evaluation support the results of similarity measurement using ROUGE as a co-selection-based analysis. The readability evaluation of the summary results of Indonesian texts, specifically the results of Dwiyanto's, GFI, and FKGL, show that the average readability of the generated text is at the medium level. This is also consistent with the results of the human readability evaluation, which show that most of the generated summaries are partially readable according to all evaluators. The summary results are focused on the news topic category, have connectivity between sentences, and have complete sentence attributes, according to two evaluators. Because most generated summary results are still partially readable, this lays the groundwork for better understanding reader expectations when it comes to judging the readability of machine-generated summary. This result shows that the proposed DeepSPM method produced better results in the aspect of readability.

This research has also revealed a few issues that need further investigation. As a result, several potential future studies to improve the current work are suggested. It is hoped that this research has paved the way for future research. Therefore, for further work, SPM can be used for multiple document summarization. Furthermore, it is intriguing to learn how SPM influences abstractive text summarization because abstract summarization produces a summary that differs from the original text. According to the human readability evaluation, the aspect of text cohesion and coherence should be given more attention to producing better Indonesian text summaries. Because it focuses on the relationship between words, clauses, sentences, and paragraphs, cohesion is a micro-level feature in a text. Word references, repetition of words or phrases, ellipsis, pronouns, conjunctions, and lexical aspects such as synonyms, antonyms, hyponyms, or collocations must all be used as cohesion tools. While text coherence is a macro-level feature in a text due to the relationship between paragraphs in the text, where the paragraph framework is concerned. Then, combining SPM with other deep learning methods, such as SumBasic, latent semantic analysis, LexRank, TextRank, Bayesian, hidden Markov model, NeuralSum, and IndoBERT, for Indonesian text summarization research can be considered. Although using a sequence of words as a text representation can improve the quality of generated Indonesian text summaries, evaluating readability remains difficult and requires further research. Other text readability metrics include SMOGI, CLI, ARI, and RPS. However, more research is needed to determine whether it is appropriate for Indonesian texts.

ACKNOWLEDGMENTS

The author would like to thank you for supporting the Center for Advanced Computing Technology, Universiti Teknikal Malaysia Melaka, so this article can be published. Thank you for the support from the UIN Sunan Gunung Djati Bandung academic community. Special thanks to Dr. Hj. Yeti Heryati, S.Ag., M. Pd; Irfan Rosyidin, S. Pd; and Hendi Sudarman, S. Pd as discussion colleagues of this study.

REFERENCES

[1] V. K. Pandey and P. Rajput, "Review on natural language processing," *Journal of Critical Reviews*, vol. 7, no. 10, pp. 1170–1174, 2020, doi: 10.31838/jcr.07.10.230.
[2] W. S. El-Kassas, C. R. Salama, A. A. Rafea, and H. K. Mohamed, "Automatic text summarization: a comprehensive survey," *Expert Systems with Applications*, vol. 165, Art. no. 113679, Mar. 2021, doi: 10.1016/j.eswa.2020.113679.
[3] Y. Kumar, K. Kaur, and S. Kaur, "Study of automatic text summarization approaches in different languages," *Artificial Intelligence Review*, vol. 54, no. 8, pp. 5897–5929, Feb. 2021, doi: 10.1007/s10462-021-09964-4.
[4] D. Gunawan, A. Pasaribu, R. F. Rahmat, and R. Budiarto, "Automatic text summarization for Indonesian language using TextTeaser," *International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, vol. 3, Art. no. 12048, Apr. 2016, doi: 10.1088/1757-899X/190/1/012048.

[5] P. M. Sabuna and D. B. Setyohadi, “Summarizing Indonesian text automatically by using sentence scoring and decision tree,” in *Proceedings - 2017 2nd International Conferences on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2017*, Nov. 2018, vol. 2018-Janua, pp. 1–6, doi: 10.1109/ICITISEE.2017.8285473.

[6] W. Yulita, S. Priyanta, and A. SN, “Automatic text summarization based on semantic networks and corpus statistics,” *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 13, no. 2, Art. no. 137, Apr. 2019, doi: 10.22146/ijccs.38261.

[7] R. Adelia, S. Suyanto, and U. N. Wisesty, “Indonesian abstractive text summarization using bidirectional gated recurrent unit,” *Procedia Computer Science*, vol. 157, pp. 581–588, 2019, doi: 10.1016/j.procs.2019.09.017.

[8] P. M. Hanunggul and S. Suyanto, “The impact of local attention in LSTM for abstractive text summarization,” in *2019 2nd International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2019*, Dec. 2019, pp. 54–57, doi: 10.1109/ISRITI48646.2019.9034616.

[9] K. Juluru, H. H. Shih, K. N. K. Murthy, and P. Elnajjar, “Bag-of-words technique in natural language processing: A primer for radiologists,” *Radiographics*, vol. 41, no. 5, pp. 1420–1426, Sep. 2021, doi: 10.1148/rg.2021210025.

[10] Q. Liu, L. Chen, Y. Yuan, and H. Wu, “History reuse and bag-of-words loss for long summary generation,” *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 29, pp. 2551–2560, 2021, doi: 10.1109/TASLP.2021.3100281.

[11] S. J. Putra, M. N. Gunawan, and A. Suryatno, “Tokenization and N-gram for indexing Indonesian translation of the Quran,” in *2018 6th International Conference on Information and Communication Technology, ICoICT 2018*, May 2018, pp. 158–161, doi: 10.1109/ICoICT.2018.8528762.

[12] Y. Zhou *et al.*, “Enhancing word-level semantic representation via dependency structure for expressive text-to-speech synthesis,” *arxiv.org/abs/2104.06835*, Apr. 2021.

[13] J. Zhang and L. Zhu, “Citation recommendation using semantic representation of cited papers’ relations and content,” *Expert Systems with Applications*, vol. 187, p. 115826, Jan. 2022, doi: 10.1016/j.eswa.2021.115826.

[14] S. Selva Birunda and R. Kanniga Devi, “A review on word embedding techniques for text classification,” in *Lecture Notes on Data Engineering and Communications Technologies*, vol. 59, Springer Singapore, 2021, pp. 267–281, doi: 10.1007/978-981-15-9651-3_23.

[15] S. Hasanzadeh, S. M. Fakhrahmad, and M. Taheri, “Review-based recommender systems: A proposed rating prediction scheme using word embedding representation of reviews,” *The Computer Journal*, vol. 65, no. 2, pp. 345–354, Feb. 2022, doi: 10.1093/comjnl/bxaa044.

[16] A. Hoonlor, M. Goldberg, M. J. Zaki, and W. A. Wallace, “Sequential patterns and temporal patterns for text mining,” Theses, Rensselaer Polytechnic Institute, Troy, NY, 2011.

[17] D. Rahmawati, G. A. Putri Saptawati, and Y. Widyani, “Document clustering using sequential pattern (SP): Maximal frequent sequences (MFS) as SP representation,” in *Proceedings of 2015 International Conference on Data and Software Engineering, ICODSE 2015*, Nov. 2016, pp. 98–102, doi: 10.1109/ICODSE.2015.7436979.

[18] Q. Wang, V. S. Sheng, and X. Wu, “Keyphrase extraction with sequential pattern mining,” *31st AAAI Conference on Artificial Intelligence, AAAI 2017*, vol. 31, no. 1, pp. 5003–5004, Feb. 2017, doi: 10.1609/aaai.v31i1.11075.

[19] F. Xie, X. Wu, and X. Zhu, “Efficient sequential pattern mining with wildcards for keyphrase extraction,” *Knowledge-Based Systems*, vol. 115, pp. 27–39, Jan. 2017, doi: 10.1016/j.knosys.2016.10.011.

[20] Q. Wang, V. S. Sheng, and C. Hu, “Keyphrase extraction using sequential pattern mining and entropy,” in *Proceedings - 2017 IEEE International Conference on Big Knowledge, ICBK 2017*, Aug. 2017, pp. 88–95, doi: 10.1109/ICBK.2017.20.

[21] J. P. Qiang, P. Chen, W. Ding, F. Xie, and X. Wu, “Multi-document summarization using closed patterns,” *Knowledge-Based Systems*, vol. 99, pp. 28–38, May 2016, doi: 10.1016/j.knosys.2016.01.030.

[22] S. Alias, S. K. Mohammad, G. K. Hoon, and T. T. Ping, “A Malay text summarizer using pattern-growth method with sentence compression rules,” in *2016 3rd International Conference on Information Retrieval and Knowledge Management, CAMP 2016 - Conference Proceedings*, Aug. 2017, pp. 7–12, doi: 10.1109/INFRKM.2016.7806326.

[23] S. Alias, S. K. Mohammad, G. K. Hoon, and T. T. Ping, “A text representation model using sequential pattern-growth method,” *Pattern Analysis and Applications*, vol. 21, no. 1, pp. 233–247, Jun. 2018, doi: 10.1007/s10044-017-0624-9.

[24] A. Sharaff, M. Jain, and G. Modugula, “Feature based cluster ranking approach for single document summarization,” *International Journal of Information Technology (Singapore)*, vol. 14, no. 4, pp. 2057–2065, Jan. 2022, doi: 10.1007/s41870-021-00853-1.

[25] G. Sharma, S. Gupta, and D. Sharma, “Extractive text summarization using feature-based unsupervised RBM method,” in *Lecture Notes in Networks and Systems*, vol. 370, Springer Nature Singapore, 2022, pp. 105–115, doi: 10.1007/978-981-16-8664-1_10.

[26] G. Garmastewira, M. L. Khodra, and M. Leylia Khodra, “Summarizing Indonesian news articles using graph convolutional network,” *Journal of Information and Communication Technology*, vol. 18, no. 3, pp. 345–365, 2019, doi: 10.32890/jict2019.18.3.6.

[27] B. Muthu *et al.*, “A framework for extractive text summarization based on deep learning modified neural network classifier,” *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 20, no. 3, pp. 1–20, May 2021, doi: 10.1145/3392048.

[28] P. G. Magdum and S. Rathi, “A survey on deep learning-based automatic text summarization models,” in *Advances in Intelligent Systems and Computing*, vol. 1133, Springer Nature Singapore, 2021, pp. 377–392, doi: 10.1007/978-981-15-3514-7_30.

[29] R. S. Shini and V. D. A. Kumar, “Recurrent neural network based text summarization techniques by word sequence generation,” in *Proceedings of the 6th International Conference on Inventive Computation Technologies, ICICT 2021*, Jan. 2021, pp. 1224–1229, doi: 10.1109/ICICT50816.2021.9358764.

[30] T. Ma, Q. Pan, H. Rong, Y. Qian, Y. Tian, and N. Al-Nabhan, “T-BERTSum: topic-aware text summarization based on BERT,” *IEEE Transactions on Computational Social Systems*, vol. 9, no. 3, pp. 879–890, Jun. 2022, doi: 10.1109/TCSS.2021.3088506.

[31] T. Shi, Y. Keneshloo, N. Ramakrishnan, and C. K. Reddy, “Neural abstractive text summarization with sequence-to-sequence models,” *ACM/IMS Transactions on Data Science*, vol. 2, no. 1, pp. 1–37, Jan. 2021, doi: 10.1145/3419106.

[32] K. Kurniawan and S. Louvan, “IndoSum: A new benchmark dataset for Indonesian text summarization,” in *Proceedings of the 2018 International Conference on Asian Language Processing, IALP 2018*, Nov. 2019, pp. 215–220, doi: 10.1109/IALP.2018.8629109.

[33] F. Koto, J. H. Lau, and T. Baldwin, “Liputan6: A large-scale Indonesian dataset for text summarization,” *arxiv.org/abs/2011.00679*, Nov. 2020.

[34] M. Yousefi-Azar and L. Hamey, “Text summarization using unsupervised deep learning,” *Expert Systems with Applications*, vol. 68, pp. 93–105, Feb. 2017, doi: 10.1016/j.eswa.2016.10.017.

[35] N. T. Mohamed N. Sweilam, “Hybridization of DBN with SVM and its impact on performance in multi-document summarization,” *Machine Learning and Applications: An International Journal*, vol. 8, no. 3, pp. 37–51, Sep. 2021, doi: 10.5121/mlaij.2021.8304.

[36] N. S. Shirwandkar and S. Kulkarni, “Extractive text summarization using deep learning,” in *2018 Fourth International*

Conference on Computing Communication Control and Automation (ICCUBEA), IEEE, 2018, pp. 1–5, doi: 10.1109/ICCUBEA.2018.8697465.
[37] A. Rezaei, S. Dami, and P. Daneshjoo, "Multi-document extractive text summarization via deep learning approach," in *2019 IEEE 5th Conference on Knowledge Based Engineering and Innovation, KBEI 2019*, Feb. 2019, pp. 680–685, doi: 10.1109/KBEI.2019.8735084.
[38] B. Sharma, M. Tomer, and K. Kriti, "Extractive text summarization using F-RBM," *Journal of Statistics and Management Systems*, vol. 23, no. 6, pp. 1093–1104, Aug. 2020, doi: 10.1080/09720510.2020.1808353.
[39] N. Nazari and M. A. Mahdavi, "A survey on automatic text summarization," *Journal of AI and Data Mining*, vol. 7, no. 1, pp. 121–135, 2019, doi: 10.22044/JADM.2018.6139.1726.
[40] M. Gambhir and V. Gupta, "Recent automatic text summarization techniques: a survey," *Artificial Intelligence Review*, vol. 47, no. 1, pp. 1–66, Mar. 2017, doi: 10.1007/s10462-016-9475-9.
[41] F. Kiyani and O. Tas, "A survey automatic text summarization," *Pressacademia*, vol. 5, no. 1, pp. 205–213, Jun. 2017, doi: 10.17261/pressacademia.2017.591.
[42] H. A. Robbani, "Sastrawi," *MIT*, 2016, https://pypi.org/project/Sastrawi/
[43] J. Asian, H. E. Williams, and S. M. M. Tahaghoghi, "Stemming Indonesian," *Conferences in Research and Practice in Information Technology Series*, vol. 38, no. 4, pp. 307–314, Dec. 2005, doi: 10.1145/1316457.1316459.
[44] S. Verma and V. Nidhi, "Extractive Summarization using deep learning," *arxiv.org/abs/1708.04439*, Aug. 2017.
[45] C. Y. Lin, "Rouge: A package for automatic evaluation of summaries," in *Proceedings of the workshop on text summarization branches out (WAS 2004)*, Jul. 2004, no. 1, pp. 25–26.
[46] P. Verma, S. Pal, and H. Om, "A comparative analysis on Hindi and English extractive text summarization," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 18, no. 3, pp. 1–39, May 2019, doi: 10.1145/3308754.
[47] P. Verma and H. Om, "A novel approach for text summarization using optimal combination of sentence scoring methods," *Sadhana - Academy Proceedings in Engineering Sciences*, vol. 44, no. 5, pp. 1–15, May 2019, doi: 10.1007/S12046-019-1082-4/METRICS.
[48] R. Rautray and R. C. Balabantaray, "An evolutionary framework for multi document summarization using Cuckoo search approach: MDSCSA," *Applied Computing and Informatics*, vol. 14, no. 2, pp. 134–144, Jul. 2018, doi: 10.1016/j.aci.2017.05.003.
[49] S. Ayoub, E. Tsui, T. Mohammed, and J. Tseng, "Readability assessment of online uveitis patient education materials," *Ocular Immunology and Inflammation*, vol. 27, no. 3, pp. 399–403, Dec. 2019, doi: 10.1080/09273948.2017.1413396.
[50] M. K. Biddinika, R. P. Lestari, B. Indrawan, K. Yoshikawa, K. Tokimatsu, and F. Takahashi, "Measuring the readability of Indonesian biomass websites: The ease of understanding biomass energy information on websites in the Indonesian language," *Renewable and Sustainable Energy Reviews*, vol. 59, pp. 1349–1357, Jun. 2016, doi: 10.1016/j.rser.2016.01.078.
[51] Y. N. Fadziah and E. F. Rahman, "Application of the enhanced confix stripping algorithm in measuring text readability using gunning fog index," *(in Bahasa) JATIKOM: Jurnal Aplikasi dan Teori Ilmu Komputer*, vol. 1, no. 1, pp. 15–24, 2018.
[52] U. Mursyadah, "Readability level of electronic school books (BSE) for class x SMA/MA biology lessons," *(in Bahasa) TEACHING: Jurnal Inovasi Keguruan dan Ilmu Pendidikan*, vol. 1, no. 4, pp. 298–304, Dec. 2021, doi: 10.51878/teaching.v1i4.774.
[53] S. D. Utami, I. N. Dewi, and I. Efendi, "Readability level of scientific collaborative based flexible learning teaching materials," *(in Bahasa) Bioscientist: Jurnal Ilmiah Biologi*, vol. 9, no. 2, Dec. 2021, doi: 10.33394/bioscientist.v9i2.4246.
[54] M. P. Sari and H. Herri, "Content analysis and readable level of the mission statement and its influence on Indonesian banking performance," *(in Bahasa) Menara Ilmu: Jurnal Penelitian dan Kajian Ilmiah*, vol. 14, no. 1, Jul. 2020, doi: 10.31869/MI.V14I1.2003.
[55] D. D. Pranowo, "Instrument of indonesian texts readability," Universitas Negeri Yogyakarta, 2011.
[56] P. Verma and H. Om, "MCRMR: Maximum coverage and relevancy with minimal redundancy based multi-document summarization," *Expert Systems with Applications*, vol. 120, pp. 43–56, Apr. 2019, doi: 10.1016/j.eswa.2018.11.022.
[57] H. Liu, "Single-point wind forecasting methods based on reinforcement learning," in *Wind Forecasting in Railway Engineering*, Elsevier, 2021, pp. 177–214, doi: 10.1016/b978-0-12-823706-9.00005-3.
[58] S. S. Nisha, M. M. Sathik, and M. N. Meeral, "Application, algorithm, tools directly related to deep learning," in *Handbook of Deep Learning in Biomedical Engineering: Techniques and Applications*, Elsevier, 2020, pp. 61–84, doi: 10.1016/B978-0-12-823014-5.00007-7.
[59] Y. Hua, J. Guo, and H. Zhao, "Deep belief networks and deep learning," in *Proceedings of 2015 International Conference on Intelligent Computing and Internet of Things, ICIT 2015*, Jan. 2015, pp. 1–4, doi: 10.1109/ICAIOT.2015.7111524.
[60] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Computation*, vol. 14, no. 8, pp. 1771–1800, Aug. 2002, doi: 10.1162/089976602760128018.
[61] J. Pei *et al.*, "Mining sequential patterns by pattern-growth: The PrefixSpan approach," *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 11, pp. 1424–1440, Nov. 2004, doi: 10.1109/TKDE.2004.77.

BIOGRAPHIES OF AUTHORS

Dian Sa'adillah Maylawati is a lecturer in Department of Informatics at UIN Sunan Gunung Djati Bandung, Indonesia. She got a Ph.D. from Centre for Advanced Computing Technology, Faculty of Information and Communication Technology in Universiti Teknikal Malaysia Melaka (UTeM), Malaysia. She got a master's degree from Bandung Technology Institute, Indonesia. Her current research interests focus on software engineering, machine learning, text mining, and natural language processing. She can be contacted at email: diansm@uinsgd.ac.id.

Yogan Jaya Kumar is a lecturer at Centre for Advanced Computing Technology, Faculty of Information and Communication Technology in Universiti Teknikal Malaysia Melaka (UTeM), Malaysia. He got a doctoral degree from Universiti Teknologi Malaysia Johor. His current position is a Centre for Advanced Computing Technology, Faculty of Information and Communication Technology in Universiti Teknikal Malaysia Melaka (UTeM), Malaysia. His current research interests focus on soft computing, text mining, and information retrieval. He can be contacted at email: yogan@utem.edu.my.

Fauziah Kasmin is a senior lecturer at Centre for Advanced Computing Technology, Faculty of Information and Communication Technology in Universiti Teknikal Malaysia Melaka (UTeM), Malaysia. She got doctoral degree from Universiti Kebangsaan Malaysia. Her current research interest focuses on statistics and image processing. She can be contacted at email: fauziah@utem.edu.my.

Muhammad Ali Ramdhani is a research of information technology professor in the Department of Informatics at the UIN Sunan Gunung Djati Bandung, Indonesia. His current research interests focus on information system, expert system, decision support system, strategic management, and research methodology. He can be contacted at email: m_ali_ramdhani@uinsgd.ac.id.