



# Indonesian Document Text Summarization Based on Extractive Using Sentences Scoring and Fuzzy Logic

Basirudin Ansor<sup>1,\*</sup> Achmad Solichan<sup>2</sup> Muhammad Zainudin Al Amin<sup>3</sup> Aditya Putra Ramdani<sup>4</sup> Mulil Khaira<sup>5</sup> Nova Christina Sari<sup>6</sup>

<sup>1,2,3,4,5,6</sup>Universitas Muhamamdiyah Semarang, Semarang, Central Java 50273, Indonesia  
basirudinansor@unimus.ac.id

**Abstract.** The large number of text documents available on the internet has resulted in a demand for quick access to get the essence for making decisions based on the available information. One method to overcome this problem is to use text summarization. There are 2 ways to summarize, namely abstractive and extractive. In this study, the extractive method was used to find sentences that were considered important. By using sentence scoring as a sentence weighting feature, the sentences are given weight by paying attention to the features frequency, Uppercase, Proper Noun, Sentence to Sentence Similarity, Numerical Data, Sentence Length, Sentence Position, and Similarity to the Title. Then the fuzzy logic algorithm is used to select words based on the value of the sentence score which creates a total of 6561 rules. Evaluation in this research uses a confusion matrix, using precision, recall and f-measure measurements. In Indonesian CNN data, the evaluation results show an average precision value of 0.581, an average recall of 1.394 and an average f-measure of 0.796. Meanwhile, Indonesian media documents produce an average precision value of 0.481, recall 1.267 and f-measure 0.709.

Keywords: Text Summarization, Sentence Scoring, Fuzzy Logic, Indonesian News Document.

## 1. Introduction

Information is a collection of messages that cannot be separated from human life, both information given and obtained from anywhere at any time and from anyone. Information overload is a problem caused by the ease of manipulating information, storing and distributing information. Bawden and Robinson define information overload as a term to represent the state of efficiency of individuals when using information in their activities is hampered due to the large amount of relevant information available [1]. The number of electronic text documents stored worldwide is incalculable. The development of the Internet played a role in the spread of articles and text documents. APJII website survey in 2023 has 215 million and in 2024 has 221 million users and will continue to increase as time goes by [2].

© The Author(s) 2024

I. Yustar Afif and R. Nindyo Sumarno (eds.), *Proceedings of the 2nd Lawang Sewu International Symposium on Engineering and Applied Sciences (LEWIS-EAS 2023)*, Advances in Engineering Research 234,

[https://doi.org/10.2991/978-94-6463-480-8\\_11](https://doi.org/10.2991/978-94-6463-480-8_11)

The large number of text documents available on the internet has resulted in a demand for quick access to get the essence for making decisions based on the available information. One method to overcome this problem is to use text summarization. Text Summarization is the process of distilling the most important information from a source to produce a condensed version for a particular user and a particular task [3][4]. Feature selection in text summarization can be done using the sentences scoring method [5] to give a value to a sentence whose results will be used as a value that is processed in other methods to find the purpose of the research. There is research that uses feature extraction in text summarization in the form of four features, namely PositionScore, TitleScore, Similarity of semantic representation between sentences and document titles, Similarity of semantic representation between sentences and sentence clusters [6].

Indonesian texts are very suitable for using the sentence scoring method to produce high scores. By using the sentence extraction method, sentences are produced in short form, or sentence summaries. Sentences are ranked by assigning weights and ranked based on their weights. High ranking sentences are extracted from the input document thereby extracting important sentences leading to a high quality summary of the input document [7].

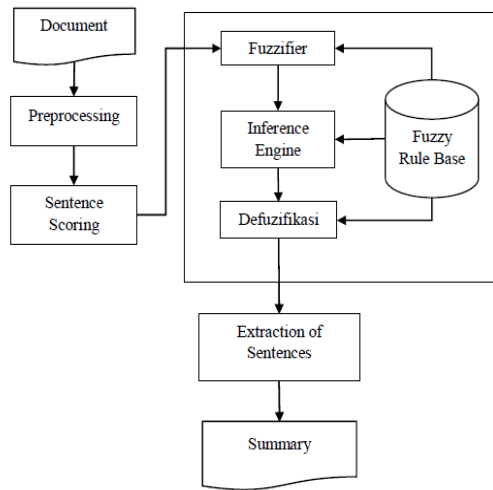
There are several methods for producing decision models or rules based on training data, such as decision trees. However, other algorithms that are more sensitive to rules should be used, such as the fuzzy logic algorithm [8]. Fuzzy logic is developed based on human language (natural language). The aim is to bridge precise machine language with human language which emphasizes meaning or meaning (significance). In determining whether a sentence is important or not, an algorithm is not immediately used which only produces binary values, because the word "important" does not have a value of 1 or 0, therefore the fuzzy algorithm is very suitable to be used to determine whether a sentence is important or not.

Sentences score is used to produce weights in each sentence based on 8 text features, namely TF / IDF, capital letters, nouns, signal phrases, numerical data, sentence length, sentence position, and similarity to the title [5]. Combined with the fuzzy logic method to determine the rules used, this will be a combination that is likely to be able to increase the F-Measure Score value.

## **2. Materials And Methods**

### **2.1. Research Flow**

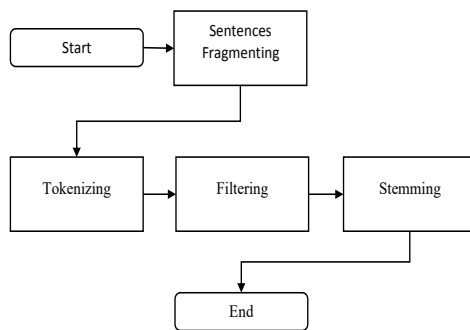
In this research, the methods used to summarize are sentence scoring and the fuzzy logic algorithm. The stages in this research are as follows:



**Fig 1** Text Summarization stages with sentence scoring and fuzzy logic.

In this research, the methods used to summarize are sentence scoring and the fuzzy logic algorithm. The stages in this research are as follows:

- a). Document: This is data used as complete news text from the data collection process carried out.
- b). Preprocessing: The stage of separating the document into sentences, from sentences into words per word, filtering the resulting words such as removing unimportant words, and turning each word into a basic word.



**Fig 2.** Preprocessing

- c). **Sentence Scoring:** Making every sentence have value. Sentence assessment consists of several methods, namely: tf-idf, uppercase, proper noun, sentence to sentence similarity, numerical data, sentence length, sentence position, similarity to title.
- d). **Fuzzy logic:** after obtaining the value for each sentence, fuzzy logic carries out a process to select sentences that are considered important, from the fuzzification, inference engine, and defuzzification stages.
- e). **Extraction of Sentence:** after the fuzzy logic process, a collection of sentences deemed important are produced which are then made into one paragraph.
- f). **Summary:** The results of extraction of sentences produce sentences which are the results of the summary.

## **2.2. Data collection**

The data that will be used is data in the form of news articles originating from the Media Indonesia and CNN news websites. Initially, data that has many paragraphs will become several paragraphs depending on the important sentences. The more important sentences there are, the more sentences you get as a summary.

## **2.3. Sentence Scoring**

Sentence scoring is a stage for giving a value to each sentence that has been carried out in the preprocessing stage, each sentence will be given a score or weight based on 8 weighting features [9]. There are TF / TDF sentences with capital letters, sentences with verbs, important phrases, numerical data, sentence length, sentence position, and similarities between sentences and titles.

- a). **Frequency (F1)**

The weight is based on the number of terms that appear or words in the sentence [10] (STF) and the number of terms that appear or words in all sentences in the text (ISF).

- b). **Uppercase (F2)**

Capital letters are a useful signal to readers. They have three main purposes: to tell the reader that a sentence is starting, to point out important words in a title, and to mark proper names and official titles[11]. This method gives higher weight to words that contain one or more capital letters, for example a person's name, city, country, and abbreviations.

- c). **Proper Noun (F3)**

Sentences containing more proper nouns receive higher weighting and have a tendency to be selected into the document summary.

d). Sentence to Sentence Similarity (F4)

The similarity between one sentence and another [12] gives higher weight if the sentence has similarities with other sentences.

e). Numerical Data (F5)

Numbers guarantee objectivity because they provide definite facts. The numbers themselves do not change based on interpretation, and different statistical models answer questions using the same data [13]. Summarize text considering the numerical data in the document. that's because usually sentences with it provide important information.

f). Sentence Length (F6)

A long sentence has higher weight. Length is calculated based on the total words in the sentence times the average length in the document.

g). Sentence Position (F7)

Sentence position is the position of the sentence in the paragraph. An assumption states that the last sentence is the conclusion so it is called the most important sentence [9].

h). Similarity to The Title

Sentences that are similar to the document title are the same words appearing in the sentence and title.

#### 2.4. Fuzzy Logic

After obtaining the value for each sentence feature, this value is used to create a rule for deciding which sentence will be used as a summary. The following are the stages of the fuzzy logic algorithm:

a). Variables used: Sentence Scoring(F1-F8), Importance

b). Variable Input: Sentence Scoring

c). Determine the variable value

- Sentence Scoring: High, Medium, Low
- Importance: Important, Average, Not Important

d). Determining Rules:

The rules generated from the sentence scoring variables F1, F2, F3, F4, F5, F6, F7 and F8 with 3 memberships namely Important, Average and Not Important above are  $3^8 = 6561$  rules.

The following is an example of a rule that produces the output value for each output variable:

- a). IF((TF-IDF is High) and (Uppercase is High) and (Proper Noun is High) and (Sentence to sentence similarity is High) and (Numerical Data is High) and (Sentence Length is High) and (Sentence Position is High) and (Similarity to the title is High) then Important)
- b). IF((TF-IDF is Medium) and (Uppercase is Medium) and (Proper Noun is Medium) and (Sentence to sentence similarity is Medium) and (Numerical Data is Medium) and (Sentence Length is Medium) and (Sentence Position is Medium) and (Similarity to the title is Medium) then Average)
- c). IF((TF-IDF is Low) and (Uppercase is Low) and (Proper Noun is Low) and (Sentence to sentence similarity is Low) and (Numerical Data is Low) and (Sentence Length is Low) and (Sentence Position is Low) and (Similarity to the title is Low) then Unimportant)

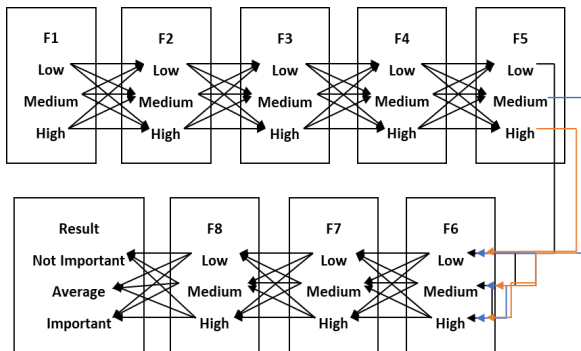
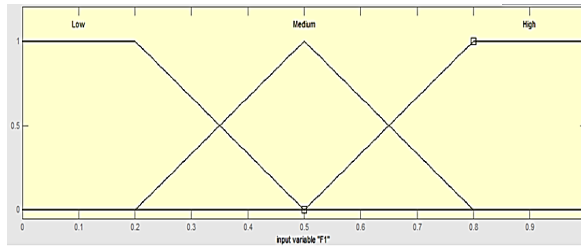


Fig 3. Fuzzy Logic Rules

Next, in the fuzzification stage, the text features given as input are converted into linguistic values using membership functions. Fuzzification is the process of mapping crisp values into fuzzy values using membership functions to fuzzy sets. In the fuzzification process, the criteria that must be met are that all members of the crisp set must be in the fuzzy set [14]. The proposed model uses Triangular Member Function (TriMF) and Trapezoidal Member Function (TrapMF) for each feature and is divided into three fuzzy sets: Low, Medium, dan High.



**Fig 4** Input Variable Membership Function.

The sentence scoring membership function has 8 input variables, namely F1-F8, the image above is the F1 membership function. The range of membership function numbers F1 to F8 is described in the Table 1.

**Table 1.** Fuzzy Input Variable Range

Variable	Range		
	Low	Medium	High
F1	0-0.5	0.2-0.8	0.5-1
F2	0-0.2	0.1-0.3	0.2-1
F3	0-0.2	0.1-0.3	0.2-1
F4	0-0.5	0.2-0.8	0.5-1
F5	0-0.1	0.05-0.2	0.1-1
F6	0-0.5	0.2-0.8	0.5-1
F7	0-0.5	0.2-0.8	0.5-1
F8	0-0.2	0.1-0.3	0.2-1

### 3. Results

Testing is carried out from the preprocessing process to the summary process which produces an extractive summary of the documents to be summarized, the Table 2 is an example of one of the data from Media Indonesia [15] that is processed.

**Table 2.** Sentence Score & Fuzzy Score Overall Media Indonesia data.

F1	F2	F3	F4	F5	F6	F7	F8	Fuzzy score
0.348	0.584	0.584	1	0	0.632	0.034	0.077	0.497
0.476	0.79	0.79	0.127	0	1	0.067	0	0.188
0.351	0.778	0.778	0.934	0	0.474	0.1	0.1	0.501

F1	F2	F3	F4	F5	F6	F7	F8	Fuzzy score
0.422	0.358	0.358	0.629	0	0.737	0.134	0.067	0.543
0.46	0.4	0.4	0.16	0	0.264	0.167	0	0.195
0.45	0.75	0.75	0.2	0	0.211	0.2	0	0.193
0.404	0.637	0.637	0.655	0.091	0.579	0.234	0.084	0.549
0.406	0.308	0.308	0.585	0	0.685	0.267	0.072	0.474
0.413	0.334	0.334	0.4	0	0.158	0.3	0	0.202
0.366	0.5	0.5	0.95	0	0.843	0.334	0.059	0.655
0.372	0.77	0.77	0.924	0	0.685	0.367	0.072	0.543
0.433	0.334	0.334	0.334	0.167	0.316	0.4	0	0.47
0.443	0.75	0.75	0.35	0	0.843	0.434	0	0.362
0.461	0.2	0.2	0.24	0.2	0.527	0.467	0	0.464
0.399	0.5	0.5	0.4	0	0.106	0.5	0	0.202
0.45	0.334	0.334	0.2	0	0.316	0.534	0	0.205
0.396	0.143	0.143	0.6	0.143	0.737	0.567	0	0.34
0.403	0.2	0.2	0.64	0	0.264	0.6	0	0.372
0.5	0.334	0.334	0	0	0.158	0.634	0	0.209
0.48	0.8	0.8	0.08	0	0.527	0.667	0	0.244
0.443	0.334	0.334	0.4	0	0.632	0.7	0	0.403
0.423	0.2	0.2	0.56	0	0.264	0.734	0	0.43
0.427	0.375	0.375	0.45	0	0.422	0.767	0	0.429
0.417	0.467	0.467	0.507	0	0.79	0.8	0	0.431
0.419	0.429	0.429	0.343	0.143	0.369	0.834	0	0.521
0.464	0.527	0.527	0.295	0	1	0.867	0	0.304
0.438	0.706	0.706	0.4	0	0.895	0.9	0	0.41
0.384	0.25	0.25	0.6	0	0.211	0.934	0	0.372
0.447	0.334	0.334	0.267	0	0.158	0.967	0	0.195
0.46	0.2	0.2	0.16	0	0.264	1	0	0.268

**Table 3.** Sentence Score & Fuzzy Score Indonesian Media Summary.

F1	F2	F3	F4	F5	F6	F7	F8	id	Fuzzy score
0.348	0.584	0.584	1	0	0.632	0.034	0.077	0	0.497



F1	F2	F3	F4	F5	F6	F7	F8	id	Fuzzy score
0.351	0.778	0.778	0.934	0	0.474	0.1	0.1	2	0.501
0.422	0.358	0.358	0.629	0	0.737	0.134	0.067	3	0.543
0.404	0.637	0.637	0.655	0.091	0.579	0.234	0.084	6	0.549
0.406	0.308	0.308	0.585	0	0.685	0.267	0.072	7	0.474
0.366	0.5	0.5	0.95	0	0.843	0.334	0.059	9	0.655
0.372	0.77	0.77	0.924	0	0.685	0.367	0.072	10	0.543
0.433	0.334	0.334	0.334	0.167	0.316	0.4	0	11	0.47
0.461	0.2	0.2	0.24	0.2	0.527	0.467	0	13	0.464
0.443	0.334	0.334	0.4	0	0.632	0.7	0	20	0.403
0.423	0.2	0.2	0.56	0	0.264	0.734	0	21	0.43
0.427	0.375	0.375	0.45	0	0.422	0.767	0	22	0.429
0.417	0.467	0.467	0.507	0	0.79	0.8	0	23	0.431
0.419	0.429	0.429	0.343	0.143	0.369	0.834	0	24	0.521
0.438	0.706	0.706	0.4	0	0.895	0.9	0	26	0.41

**Table 4.** Sentence Score & Fuzzy Score Overall CNN Indonesia data.

F1	F2	F3	F4	F5	F6	F7	F8	id	Fuzzy score
0.388	0.762	0.762	0.381	0	0.914	0.04	0.116	0	0.392
0.357	0.75	0.75	0.389	0.125	0.348	0.08	0.067	1	0.506
0.287	0.534	0.534	0.815	0	0.653	0.12	0.046	2	0.449
0.403	0.445	0.445	0.494	0	0.392	0.16	0.063	3	0.204
0.411	0.273	0.273	0.526	0.091	0.479	0.2	0.056	4	0.443
0.391	0.353	0.353	0.367	0.059	0.74	0.24	0.042	5	0.377
0.304	0.445	0.445	0.889	0	0.392	0.28	0.063	6	0.419
0.3	0.2	0.2	0.623	0	0.218	0.32	0	7	0.32
0.353	0.358	0.358	0.635	0.072	0.609	0.36	0.048	8	0.486
0.419	0.445	0.445	0.42	0	0.392	0.4	0.063	9	0.204
0.369	0.445	0.445	0.544	0	0.392	0.44	0.134	10	0.375
0.358	0.3	0.3	0.445	0	0.435	0.48	0.059	11	0.211
0.419	0.4	0.4	0.578	0	0.218	0.52	0.084	12	0.253
0.353	0.5	0.5	0.612	0	0.348	0.56	0.143	13	0.465

F1	F2	F3	F4	F5	F6	F7	F8	id	Fuzzy score
0.3	0.53	0.53	0.746	0.059	0.74	0.6	0.042	14	0.646
0.406	0.625	0.625	0.278	0.125	0.348	0.64	0	15	0.499
0.636	0.305	0.305	0.107	0.348	1	0.68	0	16	0.792
0.5	0.334	0.334	0	0	0.131	0.72	0	17	0.198
0.396	0.616	0.616	0.342	0.077	0.566	0.76	0	18	0.372
0.371	0.5	0.5	0.778	0	0.174	0.8	0.091	19	0.341
0.357	0.5	0.5	0.612	0	0.174	0.84	0	20	0.339
0.406	0.8	0.8	0.489	0	0.435	0.88	0.059	21	0.414
0.276	0.5	0.5	1	0	0.174	0.92	0	22	0.432
0.523	0.286	0.286	0.508	0	0.305	0.96	0.072	23	0.339
0.418	0.75	0.75	0.26	0.084	0.522	1	0	24	0.41

**Table 5.** Sentence Score & Fuzzy Score CNN Indonesia Summary.

F1	F2	F3	F4	F5	F6	F7	F8	id	Fuzzy score
0.357	0.75	0.75	0.389	0.125	0.348	0.08	0.067	1	0.506
0.287	0.534	0.534	0.815	0	0.653	0.12	0.046	2	0.449
0.411	0.273	0.273	0.526	0.091	0.479	0.2	0.056	4	0.443
0.304	0.445	0.445	0.889	0	0.392	0.28	0.063	6	0.419
0.353	0.358	0.358	0.635	0.072	0.609	0.36	0.048	8	0.486
0.353	0.5	0.5	0.612	0	0.348	0.56	0.143	13	0.465
0.3	0.53	0.53	0.746	0.059	0.74	0.6	0.042	14	0.646
0.406	0.625	0.625	0.278	0.125	0.348	0.64	0	15	0.499
0.636	0.305	0.305	0.107	0.348	1	0.68	0	16	0.792
0.406	0.8	0.8	0.489	0	0.435	0.88	0.059	21	0.414
0.276	0.5	0.5	1	0	0.174	0.92	0	22	0.432
0.418	0.75	0.75	0.26	0.084	0.522	1	0	24	0.41

The Table 2 is the result of the sentence scoring process for Indonesian media data which produces F1 to F8, then the process is carried out again, namely the values from F1 to F8 to produce a Fuzzy Score in Table 3.

Sort based on the largest to smallest fuzzy score, then take 50% of the sentences from all the sentences in the document to be like the Table 3.

Next, testing is carried out from the preprocessing process to the summary process which produces an extractive summary of the documents to be summarized. The Table 4 is an example of one of the data from CNN Indonesia that is processed and the result of the sentence scoring process for Indonesian media data which produces F1 to F8. Then, the process is carried out again, namely the values from F1 to F8 to produce a Fuzzy Score in the Table 5.

Summarization system testing uses precision, recall, and f-measure. The precision method evaluates the accurate proportion of sentences in the summary, while the recall method evaluates the relevance of the sentence proportion of the summary.

The following are the evaluation results from CNN Indonesia data:

**Table 6.** Fuzzy Input Variable Range

No	Precision	Recall	F1-Score
1	0.5	1.4	0.73684
2	0.75	1.1875	0.91935
3	0.61538	1.38462	0.85207
4	0.54545	1.35714	0.77816
5	0.58333	1.41667	0.82639
6	0.46154	1.58333	0.71473
7	0.875	1.07143	0.9633
8	0.6	1.3871	0.83766
9	0.36364	1.53846	0.58824
10	0.44444	1.5	0.68571
11	0.6	1.38095	0.83654
12	0.5	1.42857	0.74074
13	0.36842	1.92308	0.61837
14	0.81818	1.13333	0.95031
15	0.7	1.23077	0.89243

If illustrated graphically, it can be seen as follows:

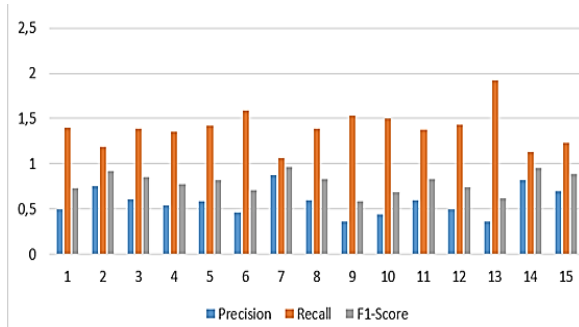


Fig 5. CNN Indonesia Document Evaluation Score Graph.

The image above is the precision, recall and f-measure values from CNN Indonesia documents produced by the sentence scoring and fuzzy logic algorithms.

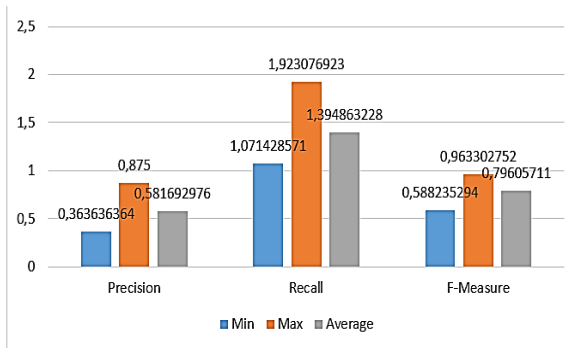


Fig 6. Precision, Recall and F-Measure CNN Indonesia.

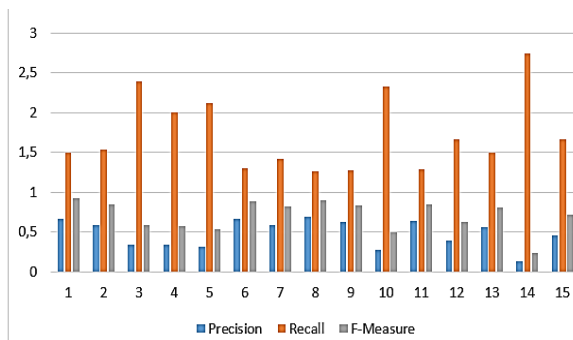
The following are the evaluation results from Media Indonesia data:

Table 7. Evaluation of Indonesian Media Data

No	Precision	Recall	F1-Score
1	0.66667	1.5	0.92308
2	0.58824	1.53846	0.85106
3	0.33333	2.4	0.58537
4	0.33333	2	0.57143
5	0.30769	2.125	0.53755
6	0.66667	1.30769	0.88312
7	0.58333	1.41667	0.82639
8	0.69231	1.26667	0.89529

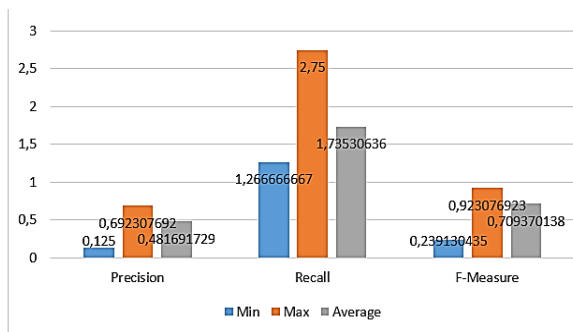
No	Precision	Recall	F1-Score
9	0.625	1.27273	0.83832
10	0.27273	2.33333	0.48837
11	0.63636	1.28571	0.85135
12	0.38462	1.66667	0.625
13	0.55556	1.5	0.81081
14	0.125	2.75	0.23913
15	0.45455	1.66667	0.71429

If illustrated graphically, it can be seen as follows:



**Fig 7.** Indonesian Media Document Evaluation Score Graph.

The image above is the precision, recall and f-measure values from Media Indonesia documents produced by the sentence scoring and fuzzy logic algorithms.



**Fig 8.** Precision, Recall and F-Measure Media Indonesia.

## 4. Conclusion

The conclusion that can be drawn from the results and discussion above is text summarization using sentence scoring as a feature and fuzzy logic as a rule for extractive word selection, with 8 features produced and the rules used to summarize Indonesian language document text from Indonesian CNN documents resulting in the average Precision value is 0.581, the average Recall is 1.394 and the average F-Measure is 0.796. Meanwhile, Indonesian media documents produce an average Precision value of 0.481, Recall 1.267 and F-Measure 0.709. From the results above, it shows that text summarization of Indonesian documents using sentence scoring and fuzzy logic is able to produce an F-Score value with a maximum value of 0.963, with an average F-Score of 0.796.

## Authors' Contributions

**Author 1:** Conceptualization, Methodology, & Investigation. **Author 2:** Supervision & Validation. **Author 3:** Writing – review & editing. **Author 4:** Formal analysis. **Author 5:** Writing – original draft. **Author 6:** Visualization.

## Conflicts of Interest

The authors declare they have no conflicts of interest.

## Acknowledgments

This work was supported by research and community service institutions (LPPM) Universitas Muhammadiyah Semarang.

## References

1. M. Arnold, M. Goldschmitt, and T. Rigotti, "Dealing with information overload: a comprehensive review," *Front. Psychol.*, vol. 14, no. June, 2023, doi: 10.3389/fpsyg.2023.1122200.
2. APJII, "Hasil Survei Internet APJII," Apjii. Accessed: Jan. 01, 2024. [Online]. Available: <https://survei.apjii.or.id/>
3. K. Chauhan, "Unsupervised Text Summarization using Sentence Embeddings," *medium.com*. Accessed: Dec. 22, 2024. [Online]. Available: <https://medium.com/jatana/unsupervised-text-summarization-using-sentence-embeddings-adb15ce83db1>
4. W. S. El-Kassas, C. R. Salama, A. A. Rafea, and H. K. Mohamed, "Automatic text summarization: A comprehensive survey," *Expert Syst. Appl.*, vol. 165, p. 113679, 2021, doi: 10.1016/j.eswa.2020.113679.

5. T. Sri, R. Raju, and B. Allarpu, "Text Summarization using Sentence Scoring Method," *Int. Res. J. Eng. Technol.*, pp. 2395–56, 2017, [Online]. Available: <https://irjet.net/archives/V4/i5/IRJET-V4I5493.pdf>
6. N. Lin, J. Li, and S. Jiang, "A simple but effective method for Indonesian automatic text summarisation," *Conn. Sci.*, vol. 34, no. 1, pp. 29–43, 2022, doi: 10.1080/09540091.2021.1937942.
7. J. N. Madhuri and R. Ganesh Kumar, "Extractive Text Summarization Using Sentence Ranking," 2019 *Int. Conf. Data Sci. Commun. IconDSC 2019*, pp. 1–3, 2019, doi: 10.1109/IconDSC.2019.8817040.
8. Y. Du and H. Huo, "News Text Summarization Based on Multi-Feature and Fuzzy Logic," *IEEE Access*, vol. 8, pp. 140261–140272, 2020, doi: 10.1109/ACCESS.2020.3007763.
9. P. M. Sabuna and D. B. Setyohadi, "Summarizing Indonesian text automatically by using sentence scoring and decision tree," *Proc. - 2017 2nd Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng. ICITISEE 2017*, vol. 2018-Janua, pp. 1–6, 2017, doi: 10.1109/ICITISEE.2017.8285473.
10. S. G. Jindal and A. Kaur, "Automatic Keyword and Sentence-Based Text Summarization for Software Bug Reports," *IEEE Access*, vol. 8, pp. 65352–65370, 2020, doi: 10.1109/ACCESS.2020.2985222.
11. Empire State University, "CAPITAL LETTERS," Empire State University. Accessed: Dec. 24, 2023. [Online]. Available: <https://www.sunyempire.edu/online-writing-support/resources/punctuation/capital-letters/#:~:text=Capital letters are useful signals,proper names and official titles>.
12. L. SERRANO, "What is Similarity Between Sentences?" COHERE. Accessed: Dec. 20, 2023. [Online]. Available: <https://txt.cohere.com/what-is-similarity-between-sentences/>
13. Indeed Editorial Team, "Defining Numerical Data (Importance and Comparison)," Indeed. Accessed: Dec. 25, 2023. [Online]. Available: <https://ca.indeed.com/career-advice/career-development/numerical-data>
14. A. Helsaputra, A. Luhur Prasasti, and R. R. Septiawan, "Implementasi Deep Learning Untuk Prediksi Tingkat Kematangan Dan Bobot Buah Pepaya Deep Learning Implementation for Weight and Ripeness Prediction of Papaya," *E-Proceeding Engineering Vo.8, No.6 Desember 2021*, vol. 8, no. 6, p. 11993, 2021.
15. D. K. Winata, "WNI Dievakuasi." Accessed: Jan. 28, 2020. [Online]. Available: <https://mediaindonesia.com/read/detail/286901-wni-dievakuasi>

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

