Design and Implementation of the Culinary Recommendation System Using Sentiment Analysis and Simple Adaptive Weighting in Bengkulu, Indonesia

Yudi Setiawan[1], Boko Susilo[1], Aan Erlanshari[1], Sumitra J Firdaus[1], Evi Maryanti[2]

[1]*Department of Informatics, Faculty of Engineering*

[2]*Department of Chemistry, Faculty of Mathematics and Natural Sciences*

Universitas Bengkulu

Jl. W.R Supratman Kandang Limun Bengkulu 38371, Indonesia

ysetiawan@unib.ac.id, boko.susilo@unib.ac.id, aan_erlanshari@unib.ac.id, sumitra.firdaus77@gmail.com, *evimaryanti82@yahoo.com*

Abstract— **In 2017, the minister of Indonesia tourism stated that everyone who travels spends his time for culinary about 30-40%. The key point in increasing the tourism income; especially from the culinary sector is about how to inform and promote the wealth of Indonesia culinary to all travellers. The information system of Bengkulu tourism has been developed in the previous study. However, that system has not been able to provide the best culinary recommendations to the travellers. This study focuses on designing and implementing the recommendation system of Bengkulu culinary by using sentiment analysis and simple adaptive weighting (SAW). The recommendation offered is based on the user review and criteria as well. The user review will be classified into positive, negative and neutral reviews by the sentiment analysis method. If the user needs culinary information based on criteria, the system will provide a recommendation and rank of culinary by using simple adaptive weighting. These criteria used are the average price, opening hours, facilities, distance from a central city, and transportation as well. Sentiment analysis method obtains the accuracy of recommendation classification at 79% while the recommendation rank obtained by the SAW method is 90.83%. These results show that the proposed method has a potential for assisting the travellers to gain the best culinary recommendation, especially in the Bengkulu area.**

Keywords— **Culinary, Tourism, Recommendation System, Sentiment Analysis, Simple Adaptive Weighting**

I. INTRODUCTION

Tourism is an activity directly involves and helps the community in managing some benefits especially in the local economy aspect for local communities and governments However, the Central bank of Indonesia (BI) stated that the economic growth of Bengkulu province is the lowest compared the other provinces in Sumatra Island. Whereas in the economy sector, Bengkulu has a great potential to be more developed. Until 2016, there was not seen yet the supporting from the tourism sector to the regional economy [1].

Bengkulu province has the lowest economic growth of tourism, whereas it has many tourist destinations with the stunning natural attractions that have great potential to develop and expose [2]. Moreover, the culinary has a strong magnet that can attract tourists for coming in Bengkulu, Indonesia. However, some of the culinary places haven't been well-known by Bengkulu people itself, even by the tourists who are coming to Bengkulu. Therefore, an information system is needed to provide the explanation and optional recommendation what the best culinary place which can be reached by tourists and Bengkulu people as well.

Sentiment analysis (SA) known as opinion mining is a method for analysing public opinion, sentiments, evaluation, or emotion on an entity such as product, service, organisation, individual, issue, topic and others [3]. Having travelled, tourists oftentimes will share their experiences by uploading captured moments and information into many social media. Of course, this big data will be hard to understand by people looking for the best recommendation for tourism. These experiences can be analysed by using sentiment analysis for obtaining the summary information and a recommended destination [4]. Naïve Bayes is one of the sentiment analysis algorithms. Naïve Bayes process involves items describing that may be recommended; comparing items to the other (items or users); and automatically in response to feedback on the desirability of items that have been presented to the user [5].

Convolution Neural Network (CNN) has been quite reliable proved and fast in classifying complex and detail objects to obtain the information about Indonesian food for tourists. By using CNN method, the classification process runs accurately. So that, the detailed information such as the forms, names and food ingredients are generated appropriately [6]. It is proved by the accuracy rate of 70% with 500-times iteration.

In other studies, semantic analysis algorithm is used to measure the words in two texts. The rate of tourism destination is created by comparing the dataset of the travel blog to make the review quality is better [7]. The semantic algorithm is used as knowledge in Tourism Mobile Recommender System (RSs) for providing information about the tourism destination [8]. The tourists can make travel planning based on some criteria such as distance and variety places [9].

Recommendation system for choosing the hotels in Yogyakarta based on simple adaptive weighting achieved more objectively of weighting value than that of direct weighting [10]. An information sharing system such as social media, blogs and website can support the government to promote all about tourism destination, heritage and culinary as well. Photos and videos shared via Apps will be a trending topic [11].

In this study, Sentiment Analysis and Simple Adaptive Weighting algorithms are implemented to design the

recommendation system of Bengkulu culinary. Thus, the information is gained easily for assisting the tourists and Bengkulu people itself. The structure of this paper is organised as follows. Section II describes the experimental method followed by implementation and discussion in Section III. The last section presents the conclusion of this study.

II. METHOD

A. Sentiment Analysis

Text pre-processing is the initial process to prepare the text to be ready data for further process. The text should be separated on several different levels. A document can be broken down into chapters, sub-chapters, paragraphs, sentences and ultimately into pieces of words/ tokens. Furthermore, the presence of digit numbers, capital letters, or other characters are removed and changed in this stage [12]. Sentiment analysis process including pre-processing, calculation of probability, classification, gain the alternative result and the data rank as a recommendation result as shown in Fig.1.

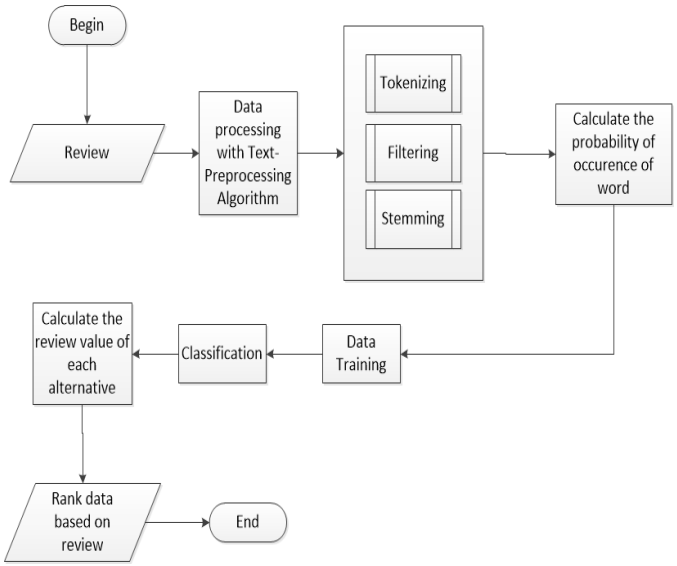

Fig. 1. The flowchart of sentiment analysis in culinary recommendation system

The text pre-processing in sentiment analysis consists of three stages as following described.

1) Case Folding and Tokenizing

Case folding changes all the letters in the document into lowercase, only the letter 'a' up to the letter 'z' allowed. Characterisation letters are omitted and considered as a delimiter. The tokenizing is the stage to enter the string based on the compiled word [12].

2) Filtering

Filtering is the stage of taking important words from tokenising stage results by using a stop-word algorithm (removing the less important words) [12]. Stop-words are non-descriptive words that can be removed in the bag-of-words approach.

3) Stemming

Stemming is the stages of finding the basic word of each word results from the filtering process [12]. The process of stemming in Indonesian text is more complicated since there are variations of affixes that should be removed to get the root word of a word. This algorithm refers to the rules of KBBI (Indonesian language dictionary) which grouped allowed affixes or unauthorized affixes [13].

B. Naïve Bayes Algorithm

Naïve Bayes is a machine learning method that uses probability calculations. The algorithm takes advantage of the probability and statistical methods proposed by British scientist Thomas Bayes, predicting future probabilities based on past experience. There are two stages in classifying the documents. The first stage is training documents that have known its category. While the second one is that of documents that unknown its category.

C. Simple Adaptive Weighting (SAW)

Simple Additive Weighting (SAW) is also known as terms weighted summation approach. The basic concept of SAW is to find the weighted sum of performance ratings on each alternative on all attributes. The SAW method requires the process of normalising the X results of the sciences that can be compared with all the alternate ratings available [14]. Simple adaptive weighting process is depicted in Fig. 2.

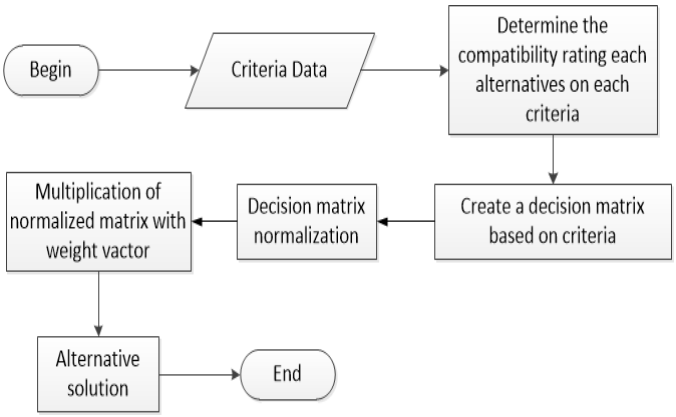

Fig. 2 Flowchart of simple adaptive weighting

The process of simple adaptive weighting is started by inserting criteria data which is subsequently undergo processed to determine the compatibility rating each alternative on each criterion for making the decision of normalisation matrix and then giving an alternative solution.

III. IMPLEMENTATION AND DISCUSSION

Bengkulu culinary recommendation system is implemented for desktop and mobile using internet browser. Sentiment Analysis and Simple Adaptive Weighting algorithm are embedded in the system as knowledge in providing the culinary recommendation. The user interface is coded in Bahasa, and there are two actors that can access this system namely end users (visitor) and administrator.

A. System User Interface

The homepage shows the variety of culinary places in alphabetical order. There is also informing the most popular destination and culinary place most opened by visitors. Moreover, a search box is available used to search the culinary place by inputting the keywords. The homepage of end-user interface is presented in Fig. 3.

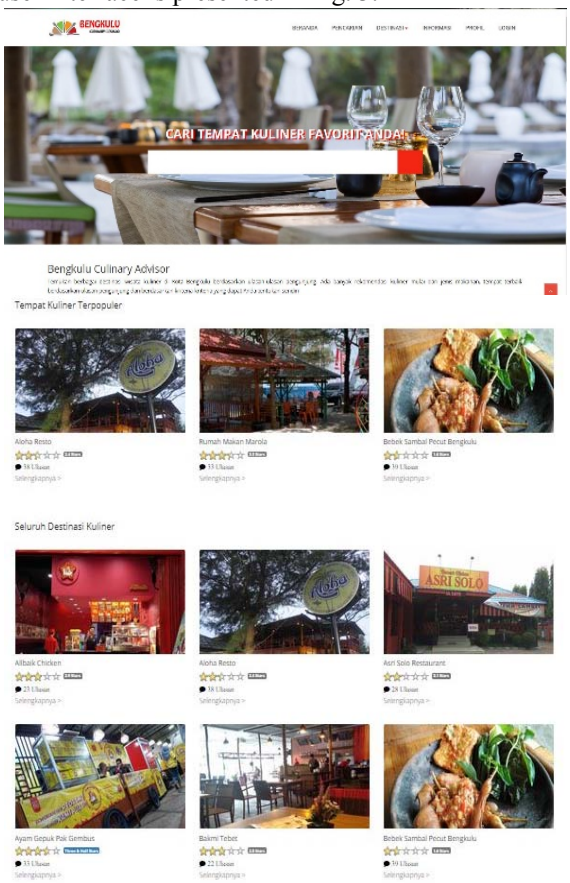

Fig. 3 Home Page Interface of Bengkulu Culinary Recommendation System

Search results are sorted from the highest rating obtained by measuring the ratio of positive reviews to the number of reviews on that culinary place.

The searching using data filtered will show the culinary place based on the selected filters and sorted by SAW value as shown in Fig. 4. These filters including prices, open hour, distance from the central city, and transportation availability.

On the administrator home page, there are several menus consists of users data menu, location data, password data, set criteria data, alternative table data, data, tokenising data, data filtering and stemmed data as displayed in Fig. 5.

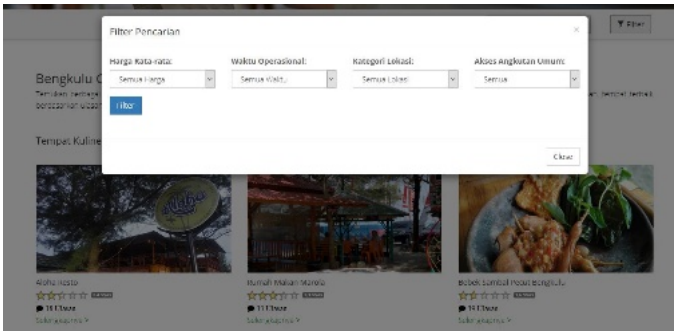

Fig. 4 The Search Filter of Culinary Recommendation System

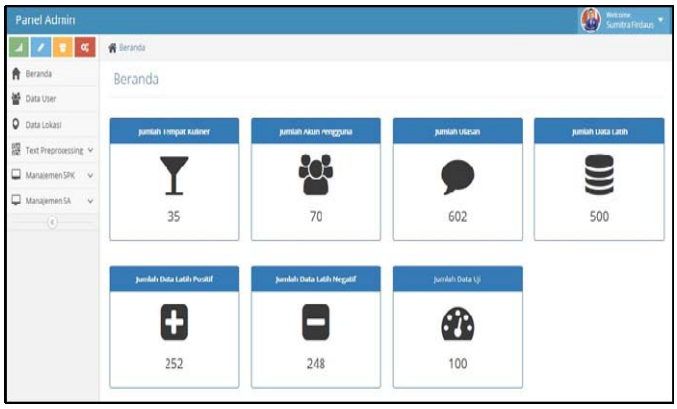

Fig. 5 Administrator interface

B. *Sentiment Analysis for Review Classification*

The dataset page displays the entire number of reviews, consisting of training data and test data. This page is used to determine whether training data as positive training data or negative ones. The frequency term page presents a dictionary of words derived from all reviews which training data that has been broken down into words per word. On this page each term (word) has a number of occurrences in each positive training or negative training.

The sentiment classification page displays testing data that has been classified as positive or negative. Furthermore, on this page is also available the button to test sentiment against unclassified reviews as shown in Fig. 6.

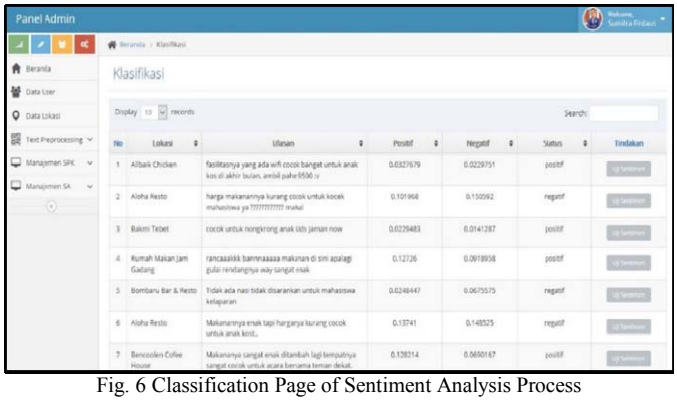

Fig. 6 Classification Page of Sentiment Analysis Process

C. *Simple Adaptive Weighting Process*

The SAW table page displays SAW tables consisting of alternate tables, yield matrices, normalisation matrices and culinary ranking tables which have been calculated by SAW method. Criterion is the basis assessment in determining the priorities order of culinary selection. There are two types of criteria in SAW method namely benefit and cost. A criterion is classified as a benefit since it provides benefits to decision makers, while criteria as a cost since criteria raise the cost for decision makers. The Table 1 shows the weighting of data criteria used such as price, open hour, facilities, distance from the central city, available transportation.

TABLE I
THE WEIGHTING OF DATA CRITERIA

No	Code	Criteria	Attribute	Weight (wj)
1.	C1	Open hour	Benefit	0.2
2.	C2	Average prices	Cost	0.4
3.	C3	Facilities	Benefit	0.2
4.	C4	Distance from central city	Cost	0.1
5.	C5	Availability of transportation	Benefit	0.1
Total				1

The complete ranking result is shown in Fig. 7. The first culinary priority is occupied by Kedai Bandung with the preference value of 0.733. The second priority is KFC Bencoleen Mall with the preference value of 0.693 while the third priority is Waroenk R.A with the preference value of 0.68.

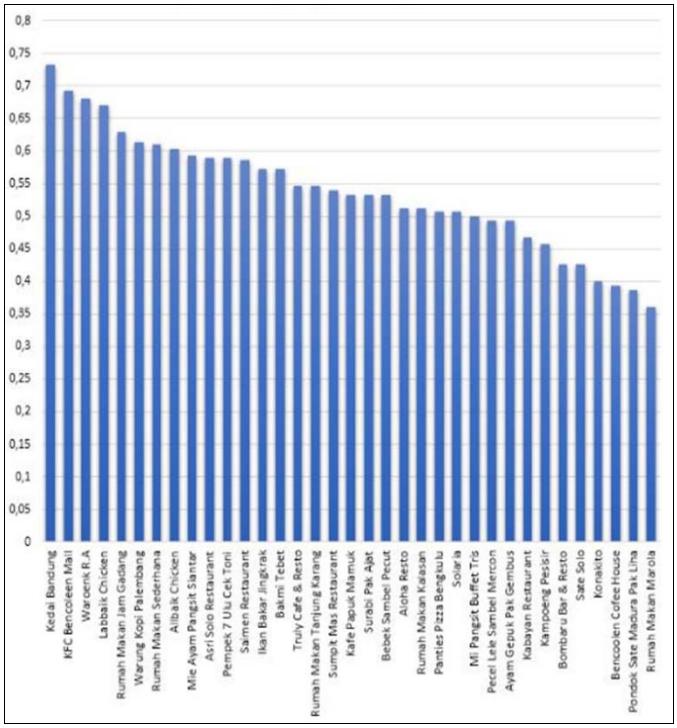

Fig. 7. Culinary recommendation system result ranking

D. System Evaluation

In this study, a user-based evaluation is conducted in order to uncover usability problems. Then, users participate in a satisfaction questionnaire to observe user-based evaluation. The questionnaires are then distributed to the respondents. The technique of selecting respondents is done randomly so that 30 samples are obtained from the general public. Before performing the calculations by using Likert Scale, the interval should be looking for firstly.

The rating category of the culinary recommendation system is valued in the range of 1-5. The highest value is 5 while the lowest is 1 with the number of classes and students are 5. The interval value can be calculated using (1). The calculation obtains interval values of 0.8 so that it can be generated the rating category as shown in Table 2. Table 3 presents the calculation results of questionnaire answer for 30 samples.

$$i = \frac{highest\ value - lowest\ value}{classes} = \frac{5-1}{5} = 0.8 \quad (1)$$

TABLE II
INTERVAL RANGE OF CATEGORIES

Interval	Category
4.24 – 5.04	Excellent
3.43 – 4.23	very good
2.62 – 3.42	Good
1.81 – 2.61	Fair
1.00 – 1.80	Poor

TABLE III
THE RESULTS OF USER CONVENIENCE VARIABLES

No	Assessment	Frequency of Answer					
		M	Excellent	Very Good	Good	Fair	Poor
1	Ease of operating the system.	4.4	13	16	1	0	0
2	Ease of searching for a culinary place using search queries	4.6	20	8	2	0	0
3	Ease of getting culinary recommendations based on rating reviews	4.67	20	10	0	0	0
4	Ease in getting the desired culinary information	4.5	15	15	0	0	0
Number of answer frequencies			68	49	3	0	0
Average percentage			56.67%	40.83%	2.5%	0%	0%
Average category total			4.54				
Category			Excellent				

From the number of answers, the score is calculated by multiplying each number of answers with each weight of the answer.

Excellent	13 x 5	=	65
Very Good	16 x 4	=	64
Good	1 x 3	=	3
Fair	0 x 2	=	0
Poor	0 x 1	=	0
Total			132

The above calculation obtained the number of scores for component 2 is 138, for component 3 is 140, and component 4 is 135. The total score is divided by the total number of answers. In this study, the number of respondents in the general public is 30, so the value of the score will be divided by 30. M value for each component variable is presented below.

$$m_1 = \frac{132}{30} = 4.4 \qquad m_2 = \frac{138}{30} = 4.6 \qquad m_3 = \frac{140}{30} = 4.67$$

$$m_4 = \frac{135}{30} = 4.5 \qquad m_{average} = \frac{4.4+4.6+4.67+4.5}{4} = 4.54$$

Based on Table II, the value of 4.54 falls within the interval 4.24 - 5.04 categorised as "excellent". The percentage for approval level of displayed variable can be done by comparing the total score of each component with the expected maximum score (assuming the respondent chooses the answer of SB, so 5 x 30 = 150).

$$\text{Component 1 :} \quad \frac{132}{150} \text{ x } 100\ \% \quad = \quad 88\%$$

Component 2 :	$\frac{138}{150}$ x 100 %	=	92%
Component 3 :	$\frac{140}{150}$ x 100 %	=	93.33%
Component 4 :	$\frac{135}{150}$ x 100 %	=	90%
	Average		90.83%

From the above calculation, it can be concluded that the percentage of respondent's approval to ease of operating the system is 88%. Ease of searching culinary place using search query obtains the percentage of 92%. Ease of getting culinary recommendations based on rating reviews obtains the percentage of 93.33%. The percentage of ease in getting the desired culinary information is 90%. The average percentage of this system is 90.83%.

In this study, a total of 500 data which consists of 250 positive and 250 negative data are used for classification training process based on Naïve Bayes classifier. Then, the classification testing process by involving 100 data is conducted and obtains the accuracy rate of 79%.

IV. Conclusion

Bengkulu culinary recommendation system has been developed to assist traveller/tourists and Bengkulu people itself in finding culinary information in easy and efficient based on user review and criteria. Sentiment analysis classifies the user review into positive, negative, and neutral classes which obtain the accuracy rate of 79%. While simple adaptive weighted provides the culinary recommendation based on some criteria and obtain the approval average percentage of 90.83% from 30 respondents. These results indicate that the developed system has a great potential to be implemented for reducing time in finding the culinary place, especially in Bengkulu area. Some observed issues during the testing process need to be addressed in future work. Moreover, this system will be designed in many languages to globalise the proposed application.

Acknowledgement

The authors would like to thank the Directorate General of Higher Education, Ministry of Research, Technology and Higher Education, Republic of Indonesia for financial support of this research.

References

[1] E. Poerwanto, "BI: Pertumbuhan Ekonomi Wisata Bengkulu Terendah se Sumatera | Portal Berita Bisnis Wisata." [Online]. Available: http://bisniswisata.co.id/bi-pertumbuhan-ekonomi-wisata-bengkulu-paling-rendah-di-sumatera/. [Accessed: 11-Apr-2018].

[2] Suwarni and A. Soleh, "Membangun Bengkulu Melalui Peningkatan Sektor Pariwisata," presented at the Seminar Nasional Riset Inovatif Ii, 2014.

[3] B. Liu, *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publisher, 2012.

[4] Y. H. Gu *et al.*, "Sentiment analysis and visualization of Chinese tourism blogs and reviews," in *2018 International Conference on Electronics, Information, and Communication (ICEIC)*, 2018, pp. 1–4.

[5] M. J. Pazzani and D. Billsus, "The Adaptive Web," 10th ed., Springer.

[6] R. P. Prasetya and F. A. Bachtiar, "Indonesian food items labeling for tourism information using Convolution Neural Network," in *2017 International Conference on Sustainable Information Engineering and Technology (SIET)*, 2017, pp. 327–331.

[7] D. M. Ramadhani, C. Rahmad, and F. Rahutomo, "Tourism destination rating system based on social media analysis (proposal and dataset development in Indonesian language)," in *2017 International Conference on Sustainable Information Engineering and Technology (SIET)*, 2017, pp. 41–46.

[8] D. Gavalasab, C. Konstantopoulosbc, K. Mastakasbd, and G. Pantziou, "Mobile recommender systems in tourism," vol. 39, no. March 2014, pp. 319–333.

[9] H. Kurdi and N. Alnashwan, "Design and implementation of mobile cloud tourism application," in *2017 Computing Conference*, 2017, pp. 681–687.

[10] A. Syafrianto, "Penerapan Algoritma AHP dan SAW Dalam Pemilihan Penginapan Di Yogyakarta," vol. 17, no. Desember 2016, pp. 7–12.

[11] Y. Setiawan, B. Susilo, A. Erlanshari, and D. Puspitaningrum, "Design Dan Implementasi Sistem Informasi Pariwisata Berbasis Konten Sebagai Startup Lokal Bengkulu," presented at the Seminar Nasional Teknologi Informasi, Univeritas Tarumanagara, 2017.

[12] R. Feldman and J. Sanger, *The Text Mining Handbook*. Cambridge: University Press.

[13] L. Agusta, "Perbandingan Algoritma Stemming Porter Dengan Algoritma Nazief & Adriani Untu K Stemming Dokumen Teks Bahasa Indonesia," in *Konferensi Nasional Sistem dan Informatika 2009*, Bali, 2009, pp. 196–201.

[14] S. Kusumadewi, S. Hartati, A. Harjoko, and R. Wardoyo, *Fuzzy Multi-Attribute Decision Making (FUZZY MADM)*. Yogyakarta: Penerbit Graha Ilmu, 2006.