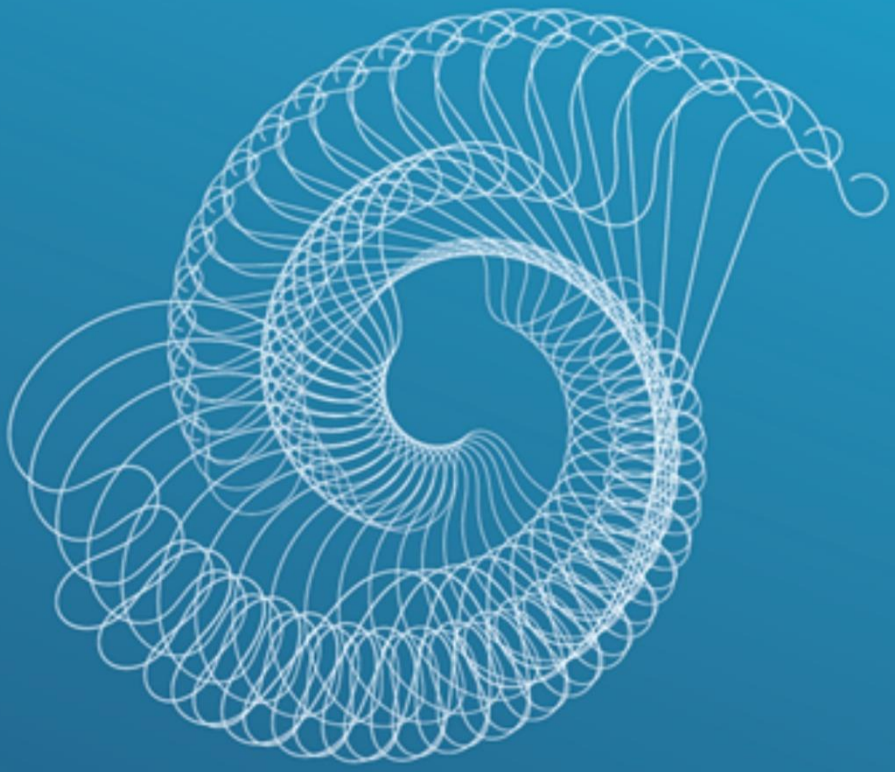
Journal of Soft Computing Exploration



JOSCEX Volume 2 Number 1 Pages 1-60 Semarang March 2021 e-ISSN 2746-0991 p-ISSN 2746-7686

Journal of Soft Computing Exploration

Publisher:
SHM Publisher, Jalan Karanglo 64, Pedurungan, Semarang City, Central Java, Indonesia

Journal of Soft Computing Exploration

Journal of Soft Computing Exploration is a journal that publishes scientific research papers related to soft computing and disseminates it widely to the public, especially soft computing enthusiasts. We hereby invite friends to post articles and citation articles in our journals. We appreciate if you would like to submit your paper for publication in JOSCEX.

The process of publishing articles in the Journal of Soft Computing Exploration will be carried out immediately after going through several stages, namely being examined by editors, reviewers, and final checking. Researchers who submit to JOSCEX are expected to be responsive to make revisions when contacted by JOSCEX so that they can produce qualified work, add knowledge, especially in the field of soft computing, and can be published on JOSCEX.

Publisher:
SHM Publisher, Jalan Karanglo 64, Pedurungan, Semarang City, Central Java, Indonesia

Indexing:
Journal of Soft Computing Exploration **is** indexed by google scholar and doaj

EDITORIAL TEAM

Contact:
Secretariat of the Journal of Soft Computing Exploration
SHM Publisher, Jalan Karanglo 64, Pedurungan, Semarang City, Central Java, Indonesia
Telp. (024) 6735842, email: shmpublisher@gmail.com
Website: https://shmpublisher.com/index.php/joscex

Table of Contents

Journal of Soft Computing Exploration

Analysis of twitter sentiment in COVID-19 era using fuzzy logic method

Devi Ajeng Efrilianda[1], Erika Noor Dianti[2], Oktaria Gina Khoirunnisa[3]
[12,3]Department of Computer Science, Universitas Negeri Semarang, Indonesia

Article Info

Article history:

Received Jan 8, 2021
Revised Feb 18, 2021
Accepted March 12, 2021

Keywords:

Soft computing
Fuzzy logic
Sentiment analysis
Social media

ABSTRACT

The sentiment is an assessment of attitudes towards certain events or things. Collecting opinion is known as a sentiment from existing data. This technique can also help analyze the opinions given by people in assessing certain objects. The best available source for gathering sentiment is the internet. In the era of the Covid-19 pandemic, many people access social media, especially Twitter to give their opinion on certain objects. Twitter is known as the social media that is accessed by users to post their opinions online. By using soft computing, especially fuzzy logic, it is possible to design, create and build bots that can analyze user opinions on Twitter. This model is used for data sentiment analysis on Twitter. The results showed that the sentiment analysis during the Covid-19 period was still dominant with positive tweets. As many as 48% of tweets are positive, 30% are negative tweets and 22% are neutral tweets. The use of applications to identify tweet sentiment during Covid-19 uses a combination of fuzzy logic methods with artificial intelligence. With the help of the Twitter API, you can get tweet data during the Covid-19 pandemic so you can find out the frequently used tweet sentiments.

Corresponding Author:

Devi Ajeng Efrilianda,
Department of Computer Science,
Universitas Negeri Semarang,
Sekaran, Gunungpati, Semarang, Indonesia.
Email: deviajeng@mail.unnes.ac.id

1. INTRODUCTION

This year there has been a change in lifestyle almost all over the world. This is due to the emergence of a new type of virus that attacks humans, namely Coronavirus Disease 2019 (Covid-19). Coronavirus itself is a group of viruses from the Ortho Coronavirus subfamily that can cause disease not only in humans but also in birds and mammals [1]. This virus then spread from the Chinese city of Wuhan to cause a large and prolonged pandemic all over the world until now. At the same time, the current condition of the Covid-19 pandemic affects all areas of life from the economy, society, politics, education, and so on.

Because the Covid-19 pandemic did not end, the government then implemented Large-Scale Social Restrictions (PSBB) which impacted the community to implement work from home (WFH) and distance learning for students and office workers. Therefore, to support these conditions, people are required to use gadgets. Over time, people are more active in socializing indirectly in cyberspace via the internet which makes social media users increase. Now more than 100 million active users open applications such as social media [2].

Social media is a persistent, internet-based, personal mass communication channel that facilitates the perception of interaction among users, derives value primarily from user-generated content, and is untrained [3]. The most commonly used social media networks such as Facebook, Youtube. Instagram, and Twitter [4]. Twitter itself has become one of the most popular social media in the world with more than 200 million active users and 10.6 billion tweets worldwide. Each user can easily provide their arguments, stories, expressions,

breaking news, or hot topics via Twitter. Indonesia, as one of the largest countries in the world with a population of more than 200 million people and has a large number of active Twitter users as well [5].

Social media, especially Twitter, have become increasingly popular with tweets discussing the Covid-19 pandemic in cyberspace with the emergence of hashtags about events that occurred during the Covid-19 pandemic making it a trending topic. That way, many responses or perceptions emerge from the community. The data obtained from Twitter will later be processed and analyzed to be useful for society and an organization. Sentiment analysis is used in this study to analyze arguments, stories, expressions of an individual on events that occurred during the Covid-19 pandemic through text mining data posts on Twitter. Sentiment analysis is a study that analyzes people's arguments and opinions towards entities such as services in text or products [6], [7].

Based on the description above, this research will apply a sentiment analysis based on an opinion by sorting the topics from those that are often discussed to those that are rarely discussed. This research uses fuzzy logic to design, create and build bots that can analyze user opinions on Twitter. Fuzzy logic is useful for processing and evaluating information [8]. The purpose of this study is to analyze public sentiment towards events during Covid-19, to find out the trending order of topics on Twitter during the pandemic, and to design, build, and build bots that can analyze user opinions on Twitter.

2. LITERATURE REVIEW

In a systematic literature review based on the reviewed papers, 7 papers are using the lexicon method, 10 papers using the machine learning method, and 7 papers mixing the two when applying sentiment analysis. The lexicon-based method is known as the unsupervised learning method. When conducting sentiment analysis most of the studies adopted the Sentiwordnet and TF-IDR methods. The Sentiwordnet method calculates based on positive or negative words. Meanwhile, the TF-IDR method converts numbers into words and is calculated using the frequency inversion document frequency method. Machine learning methods require training data to be processed. The methods that are often used in machine learning methods are the Naive Bayes and SVM, models. To improve results, combine the lexicon and machine learning methods. The data taken is mostly from the social media site Twitter [9].

The CorE Q9 bootstrap algorithm to find semantic lexicons that can be used to divide tweets into two categories: stressed and non-stressed. The Twitter data used in this article is collected via the Twitter API from January to April 2020 for the continental United States. One of the main innovations in our research is mapping the symptoms of stress that causes COVID-19 on a temporal scale. The algorithm takes a large, unnamed corpus from which it finds new related words and writes them into the wrong semantic category (for example, stress and non-stress in the case). Before the bootstrapping process, the pattern extraction was carried out on the unmarked corpus. It is used to extract all subject noun phrases, direct objects, or prepositional phrases. A universal sentence encoder is used to generate word embeddings. These text embeddings transform the tweet into a numeric vector, encoding the tweet text into the high-dimensional vector needed to find semantic similarities and perform classification tasks. The classifications used in the training process are SVM, logistic regression, naïve Bayes classification, and simple neural networks. By doing this, we were able to observe the spatiotemporal patterns of stress symptoms and answer questions about what are the main concerns regarding pandemics in different geographic areas on different time scales [10].

Sentiment analysis on Twitter users about the anti-LGBT campaign in Indonesia by seeking positive, negative, or neutral responses. Then use the Naive Bayes algorithm because it has a high degree of accuracy in analyzing sentiment. In this study, it can be concluded that on average giving neutral comments with an accuracy of 86.43% obtained through the Rapid Miner tool [11].

3. METHOD

This research is a study by analyzing the responses or perceptions of the community in a situation that occurs. Therefore, this study uses two sources of information, namely, the collection of data on public sentiment on Twitter social media through hashtags related to events during the Covid-19 pandemic. The population obtained from Twitter residents was 212 tweets. This data collection is using a random sampling method. Furthermore, this study also uses literature reviews such as journals or books related to the research being carried out.

After determining the selected data source, data collection is then carried out. As previously mentioned, data was obtained from the Twitter social media application using queries or special hashtags regarding events that occurred during Covid-19. The use of queries here is to make it easier to find various public sentiments about events that occurred during Covid-19.

After the data is collected, it performs calculations using the fuzzy method. Fuzzy itself is seen as being able to decipher an input into an output without neglecting the existing aspects [12], [13].

Fuzzy logic itself is considered very easy to apply because of its flexible nature and can be based on human logic with everyday language that is easy to understand [4], [14]. This fuzzy calculation begins with converting linguistic variables into functions that describe variables in fuzzy form. Each of these functions represents a linguistic variable and describes a fuzzy set combined with certain criteria. After the set is described, the next step is to determine the if-then which explains how the relationship between the fuzzy set, namely max (representing and) and min (representing or). In the connection completion step, all estimates are combined on the final fuzzy set.

The simplest form of data collected is text on the Twitter application. Searching for topics on social media Twitter will also be easier with the hashtags of topics or related events. So, as mentioned earlier, this study collects data through hashtags used by Twitter users which will later be analyzed through data mining methods. Data mining itself is a method of finding information or patterns in selected data [15], [16].

3.1 Text analysis

Anyone can manually analyze a sentimental tweet. However, over time, there are more and more users of the Twitter social media application. This of course causes the analysis process to be less concise and time-consuming. However, this can be made easy with information technology. This process uses the fuzzy method. This method has the ability in the process of language reasoning (linguistic reasoning) [17]–[19]. In this process, we get 200 data which consists of which we sort into negative tweets and positive tweets.

3.2 Twitter API

Application Programming Interface or what we often API is a program used to retrieve or modify data [20]. The program provided by Twitter is intended to make it easier for someone to obtain information available on Twitter. The Twitter API itself will help in selecting data so that later the data used will be more concise. Some of the data obtained will be deleted later. This process is often called the preprocessing process [21], [22]. After that, proceed to the Case Folding Process where we will improve the tweet data that we get. The fix here is intended to homogenize the text into lowercase. Then it will be continued with the data cleaning process. This process aims to remove the RT component (retweet), URL, and username. Next, the last one is Stopword Removal. Stopword itself is a word that is in the data but is less helpful in the process of tweeting analysis [23].

3.3 Fuzzy Application

The first step before we start applying fuzzy in this research is connecting the data with the existing database. Next, we enter the nature and hashtags text which will later be processed into text, and hashtags which will later be processed using fuzzy logic. At this stage, it will produce positive and negative scores for each tweet. After getting both of them, we do the calculation of each score which will then be subtracted from each other and produce the final score. After that, the data normalization stage is carried out. This data normalization aims to assist in data measurement and reduce redundancies [20]. After the weighting is done, we will begin to count and classify the data into several categories, namely very positive tweets, positive tweets, neutral tweets, negative tweets, and very negative tweets. The information about tweet analysis model can be seen at figure 1.

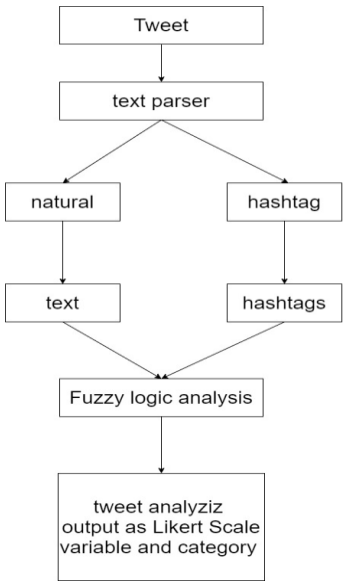


Figure 1. Tweet analysis model.

4. RESULT AND DISCUSSION

First, the tweet will be extracted or deciphered, using the fuzzy logic method, the sentiment value of the tweet will be calculated. If you have a tweet value, you can calculate the arithmetic average value. The results of this will know the percentage of positive, neutral, and negative values. Many of the conversations that occurred during Covid-19 discussed economic topics, case reports, and health. In [24] the authors analyzed Twitter data in the period 2 February-15 March 2020 from the identification of topics, 10 of which had positive sentiments and 2 were negative. Case [25] analyzed data from 01 January - 23 March found a greater number of positive and neutral tweets. Meanwhile, negative tweets were few. This is because people think the Covid-19 virus will end soon. However, as time went on negative tweets grew more and more.

Overall sentiment about Covid-19 is growing rapidly and the problem of the pandemic has become more complex as time goes on. In this study, the results showed that the sentiment analysis during the Covid-19 period was still dominant with positive tweets. As many as 48% of tweets are positive, 30% are negative tweets and 22% are neutral tweets. The use of applications to identify tweet sentiment during Covid-19 uses a combination of fuzzy logic methods with artificial intelligence. With the help of the Twitter API, you can get tweet data during the Covid-19 pandemic so you can find out the frequently used tweet sentiments.

5. CONCLUSION

Based on calculations using the fuzzy method, the results show that during the Covid-19 pandemic this study found 48% positive tweets, 30% negative tweets, and 22% neutral tweets. It can be concluded that there are more positive tweets than negative and neutral tweets. This is because people think that the pandemic will not last long, even though there are more and more negative tweets every day.

REFERENCES

[1] M. Mitra and R. K. Samanta, "A study on UCI hepatitis disease dataset using soft computing," *Model. Meas. Control C*, vol. 78, no. 4, pp. 467–477, 2017, doi: 10.18280/mmc_c.780405.

[2] Y. Hou, D. Xiong, T. Jiang, L. Song, and Q. Wang, "Social media addiction: Its impact, mediation, and intervention," *Cyberpsychology*, vol. 13, no. 1, 2019, doi: 10.5817/CP2019-1-4.

[3] R. A., E. S. -, A. A., and B. Ibrahim, "Modifications of particle swarm optimization techniques and its application on stock market: a survey," *Int. J. Adv. Comput. Sci. Appl.*, vol. 6, no. 3, 2015, doi: 10.14569/ijacsa.2015.060315.

[4] K. Howells and A. Ertugan, "Applying fuzzy logic for sentiment analysis of social media network data in marketing," *Procedia Comput. Sci.*, vol. 120, pp. 664–670, 2017, doi: 10.1016/j.procs.2017.11.293.

[5] M. Ibrahim, O. Abdillah, A. F. Wicaksono, and M. Adriani, "Buzzer detection and sentiment analysis for predicting presidential election results in a twitter nation," *Proc. - 15th IEEE Int. Conf. Data Min. Work. ICDMW 2015*, pp. 1348–1353, 2016, doi: 10.1109/ICDMW.2015.113.

[6] B. Agarwal, N. Mittal, P. Bansal, and S. Garg, "Sentiment analysis using common-sense and context information," *Comput. Intell. Neurosci.*, vol. 2015, 2015, doi: 10.1155/2015/715730.

[7] I.E. Tiffani, "Optimization of naïve bayes classifier by implemented unigram, bigram, trigram for sentiment analysis of hotel review," *J. Soft Comput. Explor.*, vol. 1. no. 1, pp. 1–7, 2020.
[8] R. Santhosh and M. Mohanapriya, "Generalized fuzzy logic based performance prediction in data mining," *Mater. Today Proc.*, 2020, doi: 10.1016/j.matpr.2020.08.626.
[9] Z. Drus and H. Khalid, "Sentiment analysis in social media and its application: Systematic literature review," *Procedia Comput. Sci.*, vol. 161, pp. 707–714, 2019, doi: 10.1016/j.procs.2019.11.174.
[10] D. Li, H. Chaudhary, and Z. Zhang, "Modeling spatiotemporal pattern of depressive symptoms caused by COVID-19 using social media data mining," *Int. J. Environ. Res. Public Health*, vol. 17, no. 14, pp. 1–23, 2020, doi: 10.3390/ijerph17144988.
[11] S. Kaur, P. Kaul, and P. M. Zadeh, "Monitoring the dynamics of emotions during Covid-19 using twitter data," *Procedia Comput. Sci.*, vol. 177, pp. 423–430, 2020, doi: 10.1016/j.procs.2020.10.056.
[12] Sunjana, "Prediction of production using the fuzzy mamdani inference method," *Int. J. Adv. Sci. Technol.*, vol. 28, no. 6, pp. 136–139, 2019.
[13] T. Yulianto, S. Komariyah, and N. Ulfaniyah, "Application of fuzzy inference system by Sugeno method on estimating of salt production," *AIP Conf. Proc.*, vol. 1867, 2017, doi: 10.1063/1.4994442.
[14] A. Chakraborty, A. Kolya, and S. Dutta, "A fuzzy logic inspired approach for social media sentiment analysis via deep neural network," *Adv. Intell. Syst. Comput.*, vol. 706, pp. 175–185, 2018, doi: 10.1007/978-981-10-8237-5_17.
[15] B. Çiğşar and D. Ünal, "Comparison of data mining classification algorithms determining the default risk," *Sci. Program.*, vol. 2019, 2019, doi: 10.1155/2019/8706505.
[16] S. Perveen, M. Shahbaz, A. Guergachi, and K. Keshavjee, "Performance Analysis of Data Mining Classification Techniques to Predict Diabetes," *Procedia Comput. Sci.*, vol. 82, pp. 115–121, 2016, doi: 10.1016/j.procs.2016.04.016.
[17] J. M. Sánchez, J. P. Rodríguez, and H. E. Espitia, "Review of artificial intelligence applied in decision-making processes in agricultural public policy," *Processes*, vol. 8, no. 11, pp. 1–23, 2020, doi: 10.3390/pr8111374.
[18] T. Tashtoush, A. Alazzam, and A. Rodan, "Utilizing fuzzy logic controller in manufacturing facilities design: Machine and operator allocation," *Cogent Eng.*, vol. 7, no. 1, 2020, doi: 10.1080/23311916.2020.1771820.
[19] Sulistiana and M. A. Muslim, "Support Vector Machine (SVM) optimization using grid search and unigram to improve e-commerce review accuracy," *J. Soft Comput. Explor.*, vol. 1, no. 1, pp. 8–15, 2020.
[20] A. Haque and T. Rahman, "Sentiment analysis by using fuzzy logic," *Int. J. Comput. Sci. Eng. Inf. Technol.*, vol. 4, no. 1, pp. 33–48, 2014, doi: 10.5121/ijcseit.2014.4104.
[21] I. M. Yulietha, S. A. Faraby, Adiwijaya, and W. C. Widyaningtyas, "An implementation of support vector machine on sentiment classification of movie reviews," *J. Phys. Conf. Ser.*, vol. 971, no. 1, 2018, doi: 10.1088/1742-6596/971/1/012056.
[22] A. Lestari and Alamsyah, "Increasing Accuracy of C4.5 algorithm using information gain ratio and adaboost for classification of chronic kidney disease," *J. Soft Comput. Explor.*, vol. 1, no. 1, pp. 32–38, 2020.
[23] D. Ramachandran and R. Parvathi, "Analysis of twitter specific preprocessing technique for tweets," *Procedia Comput. Sci.*, vol. 165, pp. 245–251, 2019, doi: 10.1016/j.procs.2020.01.083.
[24] A.-A. A., A. D., H. M., H. M., and S. Z., "Top Concerns of tweeters during the covid-19 pandemic: Infoveillance Study," *J. Med. Internet Res.*, vol. 22, no. 4, p. e19016, 2020.
[25] K. Chakraborty, S. Bhatia, S. Bhattacharyya, J. Platos, R. Bag, and A. E. Hassanien, "Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers—A study to show how popularity is affecting accuracy in social media," *Appl. Soft Comput. J.*, vol. 97, 2020, doi: 10.1016/j.asoc.2020.106754.

Journal of Soft Computing Exploration

Recommendation of Yogyakarta tourism based on simple additive weighting under fuzziness

Eko Yunanto Utomo
Postgraduate Student of Mathematics Department, Faculty of Science and Mathematics,
Universitas Diponegoro, Indonesia

Article Info

Article history:

Received Jan 14, 2021
Revised Feb 20, 2021
Accepted March 12, 2021

Keywords:

Tourism
Triangular fuzzy number
Multi-attribute decision making
Simple additive weighting

ABSTRACT

Tourists who do not understand the situation or the desired tourist attraction can choose tour and travel services. Tour and travel provide a choice of tour packages with various variations. Determining the right tour and travel package and agency can benefit tourists, both in terms of financial and vacation quality. The data used in this study were obtained from several Tour and Travel agents. There are several variables used, namely the price of the package, the number of participants, and the number of facilities obtained. The method used in this study combines the Triangular Fuzzy Number (TFN) and the Simple Additive Weighting (SAW) method. The purpose of this study is to help tourists determine the most profitable or best packages. The results of this study obtained the best 2 packages recommended for tourists to choose. In the SAW calculation, it can be seen that packages that have a preference value above 0.7 are highly recommended to be selected. Meanwhile, preference values above 0.6 to 0.7 are still considered to be selected, because they have an advantage in one of the criteria. For preference values below 0.6, it is not recommended to choose because the package is too expensive and not worth what you get.

Corresponding Author:

Eko Yunanto Utomo,
Postgraduate Student of Mathematics Department,
Faculty of Science and Mathematics, Universitas Diponegoro,
Jl. Prof. Sudarto No.13, Tembalang, Semarang, 50275, Indonesia.
Email: ekoyunanto1358@gmail.com

1. INTRODUCTION

Tourism is an important economic sector to support the progress of a country [1]. Both local and foreign tourists are business opportunities for several companies engaged in the tourism sector. So that a tour and travel agency emerged whose function was to assist tourists in accommodation, transportation, and famous tourist objects in certain areas [1]–[3]. It also [4] helps tourists identify popular tourist objects, especially in the Province of Yogyakarta Special Region.

The method used is Simple Additive Weighting (SAW) with fuzzy additions to help solve uncertain problems. The basic concept of the SAW method is to find the weighted sum of the performance ratings for each alternative in all attributes [5]–[7]. The advantage of the SAW method is to make a more precise assessment because it is based on predetermined criteria and weight values [8]–[11].

In [12], [13] use AHP and SAW methods. In [14]–[17], use the TOPSIS and SAW methods. In [18], [19] uses the SAW method to build a decision support system for selecting Banjar restaurants in Banjarmasin city. Based on several previous studies, it can be seen that the SAW method can be used to assist in decision-making. So that the Fuzzy SAW method can also be used to help make decisions on Tour and Travel problems.

2. METHOD

The method used to solve the problem of determining tourism is SAW using the TFN. The selected input variables are price, the number of participants, and the number of facilities available. The data is obtained from several tour and travel websites in Yogyakarta Special Region Province.

2.1 Triangular Fuzzy Number (TFN)

TFN denotes the relative strength of each feature pair in the same hierarchy and can be indicated as $M = (l, m, u)$, where $l \leq m \leq u$. The membership function of the TFN can be defined as follows.

$$\mu_M(x) = \begin{cases} 0. \;\; x < l \; atau \; x > u \\ \dfrac{x-l}{m-l}, \;\; l \leq x \leq m \\ \dfrac{u-x}{u-m}, \;\; m \leq x \leq u \end{cases}$$

2.2 Fuzzy Simple Additive Weighting (FSAW)

FSAW is a method that is often used to solve problems in spatial decision analysis [20], [21]. This method requires the process of normalizing the decision matrix (X) to a scale that can be compared with all existing alternative ratings [22]. The Fuzzy SAW method is divided into 2 criteria, namely benefit (profitable) and cost (disadvantage) [23]. The steps of the Fuzzy SAW method are as follows.

First, changing all alternatives into a TFN form or called fuzzification. Second, classifying the benefit and cost criteria to determine the normalization formula. Third, normalize using the following formula.

$$r_{ij} = \begin{cases} \dfrac{X_{ij}}{\max X_{ij}}, \;\; if \; j \; is \; benefit \; criteria \\ \dfrac{\min X_{ij}}{X_{ij}}, \;\; if \; j \; is \; cost \; criteria \end{cases}$$

where r_{ij} is the normalized performance rating of the alternatives A_i on criteria C_j with $i = 1,2, \dots, m$ and $j = 1,2, \dots, n$. Preference value for each alternative (V_i) as follows.

$$V_i = \sum_{j=1}^{n} w_j r_{ij}$$

where V_i is the preference value and w_j is the ranking weight.The value of V_i the larger one indicates that the alternative is preferred.

3. RESULT AND DISCUSSION

In making decisions about determining tour and travel, researchers use 3 criteria, namely price, number of participants (NoP), and number of facilities. In Table 1, we can see the data that will be used in this study. The tour and travel data below is obtained from several tours and travel agencies and the data below has been grouped according to the existing variables

The step that needs to be done is to change the crisp numbers into fuzzy numbers (fuzzification), with the following groupings Low (1,1,3), Medium (1,3,5), and High (3,5,5). Fuzzification results can be seen in Table 2.

The next step is to determine the benefit and cost criteria for this research. The criteria included in the benefits are the number of participants and the number of facilities, while the cost criteria are the package price. Furthermore, normalization is carried out for each alternative, according to the formula for the benefit and cost criteria. After getting a normalized performance rating, we continue to look for the preferential value. The prevalence value is obtained by multiplying the normalized performance rating of each alternative with each criterion weight. The weights for each criterion are as follows, Price (0.2072; 0.4145; 0.5181), number of participants (0.1036; 0.2072; 0.4145), and number of facilities (0.2072; 0.4145 ; 0.5181). The preference value of each alternative can be seen in Table 3.

Table 1. Data on tour and travel vacation packages.

No.	Package Name	Price	NoP	Facilities
1.	Gua Pindul1	290.000	4	7
2.	Kalibiru1	220.000	4	7
3.	Merapi Lava Tour	310.000	4	7
4.	Rafting Sungai Elo	390.000	5	6
5.	Snorkeling	240.000	4	6
6.	Sunrise	220.000	3	7
7.	Mangunan	310.000	4	6
8.	Gua Pindul2	320.000	3	8
9.	Merapi	350.000	4	8
10.	Kalibiru2	320.000	3	8
11.	Indrayanti	300.000	4	7
12.	Candi Borobudur	270.000	3	6
13.	Gua Pindul3	290.000	3	6
14.	Hutan Pinus	240.000	4	6
15.	Kalibiru3	360.000	3	6
16.	Air Terjun Sri Gethuk	220.000	5	7
17.	Umbul Ponggok	320.000	4	6
18.	Candi Prambanan	230.000	4	6
19.	Indrayanti	270.000	3	6
20.	Rafting Sungai Elo	425.000	5	6
21.	Keraton Jogja	225.000	3	5
22.	Dieng Wonosobo	350.000	3	6
23	Keraton Solo	215.000	3	5

Table 2. Triangular Fuzzy Number from Data

No.	Package Name	Price	NoP	Facilities
1.	Gua Pindul1	(3,5,5)	(1,3,5)	(3,5,5)
2.	Kalibiru1	(1,3,5)	(1,3,5)	(3,5,5)
3.	Merapi Lava Tour	(3,5,5)	(1,3,5)	(3,5,5)
4.	Rafting Sungai Elo	(3,5,5)	(3,5,5)	(1,3,5)
5.	Snorkeling	(1,3,5)	(1,3,5)	(1,3,5)
6.	Sunrise	(1,3,5)	(1,1,3)	(3,5,5)
7.	Mangunan	(3,5,5)	(1,3,5)	(1,3,5)
8.	Gua Pindul2	(3,5,5)	(1,1,3)	(3,5,5)
9.	Merapi	(3,5,5)	(1,3,5)	(3,5,5)
10.	Kalibiru2	(3,5,5)	(1,1,3)	(3,5,5)
11.	Indrayanti	(3,5,5)	(1,3,5)	(3,5,5)
12.	Candi Borobudur	(1,3,5)	(1,1,3)	(1,3,5)
13.	Gua Pindul3	(3,5,5)	(1,1,3)	(1,3,5)
14.	Hutan Pinus	(1,3,5)	(1,3,5)	(1,3,5)
15.	Kalibiru3	(3,5,5)	(1,1,3)	(1,3,5)
16.	Air Terjun Sri Gethuk	(1,3,5)	(3,5,5)	(3,5,5)
17.	Umbul Ponggok	(3,5,5)	(1,3,5)	(1,3,5)
18.	Candi Prambanan	(1,3,5)	(1,3,5)	(1,3,5)
19.	Indrayanti	(1,3,5)	(1,1,3)	(1,3,5)
20.	Rafting Sungai Elo	(3,5,5)	(3,5,5)	(1,3,5)
21.	Keraton Jogja	(1,3,5)	(1,1,3)	(1,3,5)
22.	Dieng Wonosobo	(3,5,5)	(1,1,3)	(1,3,5)
23	Keraton Solo	(1,3,5)	(1,1,3)	(1,3,5)

Table 3. Preference value of fuzzy SAW

No.	Package Name	Preference Value
1.	Gua Pindul1	0,6372
2.	Kalibiru1	0,7708
3.	Merapi Lava Tour	0,6372
4.	Rafting Sungai Elo	0,5957
5.	Snorkeling	0,6879
6.	Sunrise	0,6879
7.	Mangunan	0,5543
8.	Gua Pindul2	0,5543
9.	Merapi	0,6372
10.	Kalibiru2	0,5543
11.	Indrayanti	0,6372
12.	Candi Borobudur	0,605
13.	Gua Pindul3	0,4714
14.	Hutan Pinus	0,6879
15.	Kalibiru3	0,4714
16.	Air Terjun Sri Gethuk	0,8123
17.	Umbul Ponggok	0,5543
18.	Candi Prambanan	0,6879
19.	Indrayanti	0,605
20.	Rafting Sungai Elo	0,5957
21.	Keraton Jogja	0,605
22.	Dieng Wonosobo	0,4714
23	Keraton Solo	0,605

From Table 3, it can be seen that the package with the highest preference value shows the best package for tourists to choose from. The Sri Gethuk Waterfall Package has the highest value because, with a cheap package price, it gets quite a lot of facilities and the largest number of participants, namely 5. Furthermore, several packages have low preference values, namely the Kalibiru3 Package, the Pindul3 Cave Package, and the Dieng Wonosobo Package. The three packages have relatively expensive prices for the number of participants as many as 3 people and fewer facilities too.

4. CONCLUSION

The Fuzzy SAW method can help tourists choose the best Tour and Travel package based on the three available criteria. From the SAW calculation, it can be seen that packages that have a preference value above 0.7 are highly recommended to be selected. Meanwhile, preference values above 0.6 to 0.7 are still considered to be selected, because they have an advantage in one of the criteria. For preference values below 0.6 it is not recommended to choose because the package is too expensive and not worth what you get.

REFERENCES

[1] O. I. Nima Farokhi, Mozhdeh Vahid, Mehrbakhsh Nilashi, "A multi-criteria recommender system for tourism using fuzzy approach," *J. Soft Comput. Decis. Support Syst.*, vol. 3, no. 4, pp. 19–29, 2016.

[2] M. O. Ilban and H. H. Yıldırım, "Determination of tourism activities of the world's best tourism destinations using the multi-criteria decision-making method," *Cogent Soc. Sci.*, vol. 3, no. 1, 2017, doi: 10.1080/23311886.2017.1301763.

[3] H. Alrasheed, A. Alzeer, A. Alhowimel, N. Shameri, and A. Althyabi, "A multi-level tourism destination recommender system," *Procedia Comput. Sci.*, vol. 170, pp. 333–340, 2020, doi: 10.1016/j.procs.2020.03.047.

[4] R. Schegg and B. Stangl, "Special section on recommendations and analytics in tourism," *Inf. Technol. Tour.*, vol. 18, no. 1–4, 2018, doi: 10.1007/s40558-018-0109-8.

[5] M. A. Maulana, M. Habib, A. Setyanto, and Oktavia, "Tourism Trend Mapping Based on Social Media Using SAW Algorithm," *J. Phys. Conf. Ser.*, vol. 1140, no. 1, 2018, doi: 10.1088/1742-6596/1140/1/012041.

[6] P. I. Ciptayani, N. G. A. P. H. Saptarini, P. A. W. Santiary, and I. N. G. A. Astawa, "Decision support system for tourist destination using the combination of ahp and saw," *Proc. - 2nd East Indones. Conf. Comput. Inf. Technol. Internet Things Ind. EIConCIT 2018*, pp. 271–275, 2018, doi: 10.1109/EIConCIT.2018.8878638.

[7] E. Charou, K. Kabassi, A. Martinis, and M. Stefouli, "Integrating multimedia GIS technologies in a recommendation system for geotourism," *Smart Innov. Syst. Technol.*, vol. 3, pp. 63–74, 2010, doi: 10.1007/978-3-642-13396-1_3.

[8] K. Piasecki, E. Roszkowska, and A. Łyczkowska-Hanćkowiak, "Simple additive weighting method equipped with fuzzy ranking of evaluated alternatives," *Symmetry (Basel).*, vol. 11, no. 4, 2019, doi: 10.3390/sym11040482.

[9] S. A. Pratiwi and H. Medyawati, "Design of supporting application for deciding the best mountain climbing 'Hiking-Yuk!,'" *Int. J. Eng. Technol. Manag. Res.*, vol. 7, no. 2, pp. 53–80, 2020, doi: 10.29121/ijetmr.v7.i2.2020.510.

[10] M. Z. Rohman, Irwansyah, and W. E. Sari, "The medical facilities selection based on location-based services application using SAW and TOPSIS Algorithm," *J. Phys. Conf. Ser.*, vol. 1577, no. 1, 2020, doi: 10.1088/1742-6596/1577/1/012012.

[11] H. A. Prihanditya, "The implementation of z-score normalization and boosting techniques to increase accuracy of c4.5 algorithm in diagnosing chronic kidney disease," *J. Soft Comput. Explor.*, vol. 5, no. 1, pp. 63–69, 2020.

[12] A. Cahyapratama and R. Sarno, "Application of Analytic Hierarchy Process (AHP) and Simple Additive Weighting (SAW) methods in singer selection process," *2018 Int. Conf. Inf. Commun. Technol. ICOIACT 2018*, vol. 2018-Janua, pp. 234–239, 2018, doi: 10.1109/ICOIACT.2018.8350707.

[13] M. Saputra, O. S. Sitompul, and P. Sihombing, "Comparison AHP and SAW to promotion of head major department SMK Muhammadiyah 04 Medan," *J. Phys. Conf. Ser.*, vol. 1007, no. 1, 2018, doi: 10.1088/1742-6596/1007/1/012034.

[14] M. M. D. Widianta, T. Rizaldi, D. P. S. Setyohadi, and H. Y. Riskiawan, "Comparison of multi-criteria decision support methods (AHP, TOPSIS, SAW & PROMENTHEE) for employee placement," *J. Phys. Conf. Ser.*, vol. 953, no. 1, 2018, doi: 10.1088/1742-6596/953/1/012116.

[15] I. Tahyudin, R. Rosyidi, A. S. Ahmar, and Haviluddin, "Comparison of the simple additive weighting (SAW) with the technique for others reference by similarity to ideal solution (TOPSIS) methods," *Int. J. Eng. Technol.*, vol. 7, no. 2, pp. 87–89, 2018, doi: 10.14419/ijet.v7i2.2.12740.

[16] R. Simanaviciene and L. Ustinovichius, "Sensitivity analysis for multiple criteria decision-making methods: TOPSIS and SAW," *Procedia - Soc. Behav. Sci.*, vol. 2, no. 6, pp. 7743–7744, 2010, doi: 10.1016/j.sbspro.2010.05.207.

[17] A. M. Yaakob and A. Gegov, "Interactive TOPSIS Based Group Decision Making Methodology Using Z-Numbers," *Int. J. Comput. Intell. Syst.*, vol. 9, no. 2, pp. 311–324, 2016, doi: 10.1080/18756891.2016.1150003.

[18] D. J, Windarsyah, and R. R, "Decision support system for selecting banjar restaurant in banjarmasin city using simple additive weighting method," *J. K6, Educ. Manag.*, vol. 1, no. 4, pp. 35–43, 2018, doi: 10.11594/jk6em.01.04.05.

[19] A. Göksu and Seniye Erdinç Kaya, "Ranking of tourist destinations with multi-criteria decision making methods in bosnia and herzegovina," *Econ. Rev. J. Econ. Bus.*, vol. 12, no. 2, pp. 91–103, 2014.

[20] J. Seyedmohammadi, F. Sarmadian, A. A. Jafarzadeh, M. A. Ghorbani, and F. Shahbazi, "Application of SAW, TOPSIS and fuzzy TOPSIS models in cultivation priority planning for maize, rapeseed and soybean crops," *Geoderma*, vol. 310, pp. 178–190, 2018, doi: 10.1016/j.geoderma.2017.09.012.

[21] D. Aprilianto, "SVM Optimization with correlation feature selection based binary particle swarm optimization for diagnosis of chronic kidney disease," *J. Soft Comput. Explor.*, vol. 1, no.1 pp. 24–31, 2020.

[22] A. R. Safitri and M. A. Muslim, "Improved accuracy of naive bayes classifier for determination of customer churn uses smote and genetic algorithms," *J. Soft Comput. Explor.*, vol. 1, no. 1, pp. 70–75, 2020.

[23] R. H. Saputra and B. Prasetyo, "Improve the accuracy of c.5 algorithm using Particle Swarm Optimization (PSO) feature selection and bagging technique in breast cancer diagnosis," *J. Soft Comput. Explor.*, vol. 1. no.1. pp. 47–55, 2020.

Journal of Soft Computing Exploration

The forecasting of palm oil based on fuzzy time series-two factor

Ratri Wulandari [1], Bayu Surarso[2], Bambang Irawanto[3], Farikhin[4]
[1]Postgraduate Program of Mathematics, Department of Mathematics, Faculty of Science and Mathematics, Universitas Diponegoro, Semarang, Indonesia
[2,3,4]Department of Mathematics, Faculty of Science and Mathematics, Universitas Diponegoro, Semarang, Indonesia

Article Info

Article history:

Received Jan 20, 2021
Revised Feb 25, 2021
Accepted March 14, 2021

Keywords:

Palm oil
Fuzzy forecasting
Fuzzy logical relationship
Two-factor
One-order

ABSTRACT

Palm oil is a vegetable oil obtained from the mesocarp fruit of the palm tree, generally, from the species, Elaeis guineensis, and slightly from the species Elaeis oleifera and Attalea maripa. Palm oil is naturally red due to its high alpha and beta-carotenoid content. Palm kernel oil is different from palm kernel oil produced from the same fruit core. Planning for palm oil production is necessary because it greatly affects to the level of the country's economy. Forecasting can reduce uncertainty in planning. Forecasting used in the palm oil problem is two-factor forecasting using the Kumar method with uama factors in the form of palm oil production and supporting factors in the form of land area. The forecasting is evaluated using AFER and MSE, from the acquisition of AFER value of 1.212% <10%, then the forecasting has very good criteria.

Corresponding Author:

Ratri Wulandari,
Postgraduate Program of Mathematics,
Faculty of Science and Mathematics, Universitas Diponegoro,
Jl. Prof. Sudarto No.13, Tembalang, Semarang, 50275, Indonesia.
Email: dariwulan711@gmail.com

1. INTRODUCTION

Palmoil Grown only in the tropics, the oil palm tree produces high-quality oil used primarily for cooking in developing countries [1]. It's an edible vegetable oil that comes from the fruit of oil palm trees, the scientific name is Elaeis guineensis [2]. Two types of oil can be produced; crude palm oil comes from squeezing the fleshy fruit, and palm kernel oil which comes from crushing the kernel, or the stone in the middle of the fruit. Palm oil is in nearly everything – it's in close to 50% of the packaged products we find in supermarkets, everything from pizza, doughnuts and chocolate, to deodorant, shampoo, toothpaste and lipstick. It's also used in animal feed and as a biofuel in many parts of the world [3].

Planning, sense of planning can also be defined as an activity or coordinated activity to achieve that particular goal within a certain period. Thus, in a plan there will be activities or activities testing several directions of achievement, assessing uncertainty, measuring capacity, determining the direction of achievement, and also determining steps to achieve it. According to [4] stated that the understanding of this planning is a process of determining what to achieve in the future and also sets the necessary steps to achieve it. In real life not all the sets we encounter in everyday life are firmly defined. In addressing the problem of set with an indecisive boundary Professor Lotfi A. Zadeh, an expert in computer science, University of California, Berkeley attributed the set to a function that expresses the degree of conformity of the elements in its universe to the concept that is a condition of membership of the function [5]. That function is called a membership function and the value of that function is called the degree of membership of an element in the set. A fuzzy set is a set expressed by a membership function, which maps each domain in the fuzzy set to exactly one number

at intervals of zero to one. Time series data forecasting predicts what will happen based on past data. Data consist of a set word and sentence [6]. A probabilistic feature model was generated as a reference for the calculation of the data test classification [7]. A time series is a collection of regular observations on a variable over the same period and successively. By learning how variable changes over time, a relationship between need and time can be formulated and used to predict the level of future needs.

Research using FST has been developed by various researchers, among others, is the use of FST method to solve the problem of forecasting of historical data in the form of Linguistic values [8], [9], research on the use of new methods in interval sharing that is frequency density-based partitioning that produces more accurate forecasting results than using the usual interval sharing method in the fuzzy time series method [10]–[14]. In addition, FST is also used for sales forecasting, stock price forecasting, inflation forecasting to electricity load forecasting. The application of fuzzy forecasting that has been carried out is time series forecasting with various methods implemented in the University of Alabama enrollment data from 1971-1992 to obtain a method that has a small error value [10], [15]–[17]. One-factor forecasting begins to develop into two or more factors forecasting, in contrast to forecasting that uses only one factor, the development of this forecast allows researchers to predict time series data using the main factors by considering the supporting factors that are then referred to as high order factors. Production of land area, the reason for the selection of variables in the form of the land area is a place or container used to cultivate palm oil in units per hectare (Ha). Land (land area) is one of the production factors [18].

2. METHOD

2.1 Fuzzy time series

Definition 1 [19]

Suppose $F(t)$ and $G(t)(t = \cdots,0,1,2,3, \dots)$are two fuzzy time series. If $F(t)$ caused $F(t-1), G(t-1)$, then the FLR is represented by $\big(F(t-1)G(t-1)\big) \rightarrow F(t)$, and called the two-factor FLR, where each $F(t)$ and $G(t)$ is called the main factor and the second factor on $(t = \cdots,0,1,2,3, \dots)$. For those factors k-faktor $(k > 2)$, the FLR can be built. When building a time series fuzzy forecasting model, the number of fuzzy sets $F(t)$ will always be the same for every moment $t(t = \cdots,0,1,2,3, \dots)$. For example, if $F(t-1) = A_t$, $G_1(t-1) = B_j^1$ dan A_r is a fuzzy set, $F(t-1), G_1(t-1) \rightarrow F(t)$ can be written as $A_t B_j^1 \rightarrow A_r$where $A_t B_j^1$ is called the premise and A_r called consequences. If $A_t B_j^1 \rightarrow A_r$ appear v times in the fuzzy time series, it will be written as $A_t B_j^1 \rightarrow A_r(v)$ where v is a positive integer.

This method is used the fuzzy time series model. In this approach, the value of the fuzzy time series is the fuzzy set, and there is a relationship between observation at t time and observation at the previous time. The model has been developed using two fuzzy time series techniques, the Chen-type arithmetic model, and the modified one.

Given fuzzy the main factor in the i year is the $A_i$and no fuzzy group relationships are available, then the forecast of the main factor in the year $i+1$is the middle value of the A_i. with $A_i \rightarrow A_i$, $A_i$have a maximum membership value at U_{Ai} intervals weighted average model. The model developed has been customized and

Definition 2 [20]

Given fuzzy the main factor in the i year is A_i, the relationship of the fuzzy group is expressed with the $A_i \rightarrow A_k$, then the forecast of the main factor in the i year is the average middle value of A_i and A_k. By A_k is a linguistic that has a maximum membership value at U_{An}intervals.

Definition 3.9.2

Given fuzzy the main factor in the i year is the A_i, the fuzzy group relationship is expressed by $A_i \rightarrow A_{k1}, A_i \rightarrow A_{k2}, A_i \rightarrow A_{kp}$, then the forecast value at $i+1$ is in the equation below:

$$\frac{\left(\frac{(m_1 + m_2 + \dots + m_p)}{p} + \frac{(l_1 + l_2 + \dots + l_q)}{q}\right)}{2} \quad (1)$$

with $A_{k1}, A_{k2}, \dots, A_{kp}$ have a maximum membership value at intervals of $U_{Ak1}, U_{Ak2}, \dots, U_{Akp}$and each has a middle interval value of $m_1, m_2, \dots, m_p$and fuzzy supporting factors in the i year are $B_m$and fuzzy group relationships its $B_m \rightarrow [A_{n1}, A_{n2}, \dots, A_{nq}]$ where $A_{n1}, A_{n2}, \dots, A_{nq}$have a maximum membership value at intervals of $U_{An1}, U_{An2}, \dots, U_{Anq}$and each has a middle interval value of $l_1, l_2, \dots, l_q$.

Definition 3 [20].

Given fuzzy the main factor in the i year is the $A_i$and no fuzzy group relationships are available, then the forecast of the main factor in the year $i+1$is the middle value of the A_i. with $A_i \rightarrow A_i$, $A_i$have a maximum membership value at U_{Ai} intervals, shown in Figure 1.

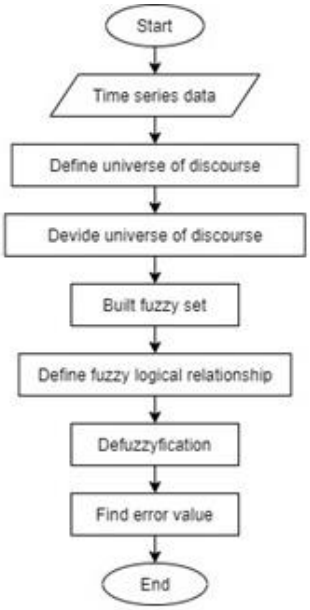


Figure 1. Kumar's method

3. RESULT AND DISCUSSION

The data used in this study relates to the production of palm oil in the form of palm oil production data in tonnes, and palm oil plantation area in hectares from 2000 to 2019.

The palm oil production variable is the dependent variable with a time horizon in the form of a time series which is then referred to as the main factor, namely X

The variable of oil palm plantation area is an independent variable with a time horizon in the form of a time series which is then referred to as a supporting factor, namely Y^i

Step 1: Defines the discussion universe for each factor, using the following definitions:

$$X_{min} = 2991.3 \quad X_{max} = 8688.9 \quad D_{X1} = 391.3 \quad D_{X2} = 211.1$$
$$Y^1{}_{min} = 5094.86 \quad Y^1{}_{max} = 29637.5 \quad D_{Y^1{}_1} = 94.86 \quad D_{Y^1{}_2} = 162.5$$

Suppose that the main factor universe (U) in equation below

$$U = [X_{min} - D_{X1}\ , X_{max} + D_{X2}\ ] = [2600{,}8900\ ] \tag{2}$$

Suppose the universe of supporting factors (W)

$$W = [Y^1{}_{min} - D_{Y^1{}_1}\ , Y^1{}_{max} + D_{Y^1{}_2}\ ] = [5000\ ,29800\ ] \tag{3}$$

Defining the length of the interval for the main factor

$$l_X = \frac{[\,(X_{max}+D_{X2}) - (X_{min}-D_{X1})\,]}{m} = 835 \tag{4}$$

Defining the length of the interval for a contributing factor

$$l_{Y^1} = \frac{\left[\left(Y^1{}_{max}+D_{Y^1{}_2}\right) - \left(Y^1{}_{min}-D_{Y^1{}_1}\right)\right]}{m} = 2480 \tag{5}$$

Steps 2 and 3: Division of the universe into intervals [21] shown in Table 1.

Table 1. Linguistic of main and second factor

Fuzzy			
Second factor	Linguistic	Main factor	Linguistic
[5000 ; 7480]	A1	[2600 ; 3500]	B1
[7480 ; 9960]	A2	[3500 ; 4400]	B2
[9960 ; 12440]	A3	[4400 ; 5300]	B3
[12440 ; 14920]	A4	[5300 ; 6200]	B4
[14920 ; 17400]	A5	[6200 ; 7100]	B5
[17400 ; 19880]	A6	[7100 ; 8000]	B6
[19880 ; 22360]	A7	[8000 ; 8900]	B7
[22360 ; 24840]	A8		
[24840 ; 27320]	A9		
[27320 ; 29800]	A10		

$B_1 \rightarrow B_1$
$B_1 \rightarrow B_1$
$B_1 \rightarrow B_1$
$B_1 \rightarrow B_1$
$B_1 \rightarrow B_1$
$B_1 \rightarrow B_2$
$B_2 \rightarrow B_2$

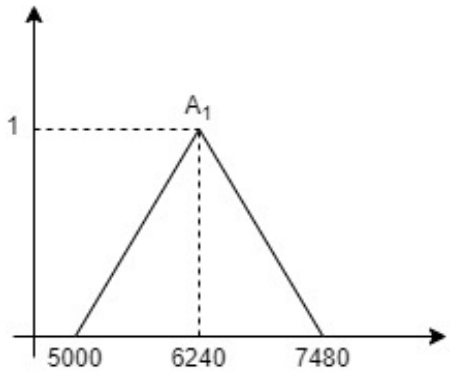
$B_2 \rightarrow B_3$
$B_3 \rightarrow B_3$
$B_3 \rightarrow B_3$
$B_3 \rightarrow B_4$
$B_4 \rightarrow B_4$
$B_4 \rightarrow B_5$
$B_5 \rightarrow B_5$

$B_5 \rightarrow B_5$
$B_5 \rightarrow B_5$
$B_5 \rightarrow B_7$
$B_7 \rightarrow B_7$
$B_5 \rightarrow \#$

Figure 2. Triangular membership function

$$\mu_{\tilde{a}}(x) = \begin{cases} \dfrac{x - 5000}{6240 - 5000} & , \quad 5000 \leq x \leq 6240 \\ \dfrac{7480 - x}{7480 - 6240} & , \quad 6240 \leq x \leq 7480 \\ 0 & , \quad yang\ lainnya \end{cases} \tag{6}$$

Steps 5: Set define a fuzzy set group for each factor, FLR from the main factor:



FLR from the second factor:
Step 6: Forecasting, using the following definition is shown in Table 2.

Table 2. Forecast value

Year	Production (Ton)	Forecast
2000	2991.3	
2001	3152.4	3125
2002	3258.6	3125
2003	3429.2	3125
2004	3496.7	3125
2005	3593.4	3125
2006	3748.5	3425
2007	4101.7	4025
2008	4451.8	4025
2009	4888	5000
2010	5161.6	5000
2011	5349.8	5000
2012	5995.7	5900
2013	6108.9	5900
2014	6332.4	6125
2015	6724.9	6650
2016	6462.1	6875
2017	6685.2	6875
2018	8507.4	6875
2019	8688.9	8450

Step 7: Count the error
Mean Squared Error (MSE)

$$MSE = \frac{\sum_{i=1}^{n}(X_i - F_i)^2}{n} = 197607.2 \quad (7)$$

Then AFER (Average Forecast Error Rate)

$$AFER = \frac{\frac{|X_i - F_i|}{X_i}}{n} \times 100\% = 1.212\% \quad (8)$$

by the following criteria be seen in Table 3.

Table 3. AFER criteria

AFER criteria	
AFER	Criteria
<10%	Very good
10% -20%	Good
20% -50%	Pretty good
> 50%	Not good (bad)

4. CONCLUSION

From the research conducted, In this section, we have fuzzy time series forecasting multi-factor one cross association as a forecasting method to forecast palm oil production influenced by plantation land area then using the frequency density partitioning modification method applied to the database that has been grouped. The results of the estimates show that the proposed method has a higher forecast performance. It is known that the AFER value is 1.212% according to the AFER criteria table 1.212%<10%, it can be concluded that the forecast has very good criteria.

REFERENCES

[1] I. E. Henson, "A brief history of the oil palm," *Palm Oil Prod. Process. Charact. Uses*, pp. 1–29, 2012, doi: 10.1016/B978-0-9818936-9-3.50004-6.

[2] H. Purnomo *et al.*, "Reconciling oil palm economic development and environmental conservation in Indonesia: A value chain dynamic approach," *For. Policy Econ.*, vol. 111, 2020, doi: 10.1016/j.forpol.2020.102089.

[3] H. Herdiansyah, H. A. Negoro, N. Rusdayanti, and S. Shara, "Palm oil plantation and cultivation: Prosperity and productivity of smallholders," *Open Agric.*, vol. 5, no. 1, pp. 617–630, 2020, doi: 10.1515/opag-2020-0063.

[4] E. Hambali and M. Rivai, "the potential of palm oil waste biomass in Indonesia in 2020 and 2030," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 65, no. 1, 2017, doi: 10.1088/1755-1315/65/1/012050.

[5] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, 1965, doi: 10.1016/S0019-9958(65)90241-X.

[6] N. Kurniawan, "Electrical energy monitoring system and Automatic Transfer Switch (ATS) Controller with the internet of things for solar power plants,". *J. Soft Comput. Explor.*, vol. 1, no. 1. pp. 16–23, 2020.

[7] I. E. Tiffani, "Optimization of naïve bayes classifier by implemented unigram , bigram , trigram for sentiment analysis of hotel review,". *J. Soft Comput. Explor.*, vol. 1, no. 1, pp. 1–7, 2020.

[8] Q. Song and B. S. Chissom, "Forecasting enrollments with fuzzy time series - Part I," *Fuzzy Sets Syst.*, vol. 54, no. 1, pp. 1–9, 1993, doi: 10.1016/0165-0114(93)90355-L.

[9] Q. Cai, D. Zhang, W. Zheng, and S. C. H. Leung, "A new fuzzy time series forecasting model combined with ant colony optimization and auto-regression," *Knowledge-Based Syst.*, vol. 74, pp. 61–68, 2015, doi: 10.1016/j.knosys.2014.11.003.

[10] T. A. Jilani, S. Muhammad, A. Burney, and C. Ardil, "Fuzzy metric approach for fuzzy time series forecasting based on frequency density based partitioning," *Internaltional J. Comput. Electr. Control Inf. Eng.*, vol. 4, no. 7, pp. 39–44, 2010.

[11] B. Garg, M. M. Sufyan Beg, A. Q. Ansari, and B. M. Imran, "Fuzzy time series prediction model," *Commun. Comput. Inf. Sci.*, vol. 141 CCIS, pp. 126–137, 2011, doi: 10.1007/978-3-642-19423-8_14.

[12] O. Cagcag Yolcu, U. Yolcu, E. Egrioglu, and C. H. Aladag, "High order fuzzy time series forecasting method based on an intersection operation," *Appl. Math. Model.*, vol. 40, no. 19–20, pp. 8750–8765, 2016, doi: 10.1016/j.apm.2016.05.012.

[13] R. Wulandari, Farikhin, B. Surarso, and B. Irawanto, "First-order fuzzy time series based on frequency density partitioning for forecasting production of petroleum," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 846, no. 1, 2020, doi: 10.1088/1757-899X/846/1/012063.

[14] B. Irawanto, R. W. Ningrum, B. Surarso, and Farikhin, "An improved forecasting method of frequency density partitioning (FDP) based on fuzzy time series (FTS)," *J. Phys. Conf. Ser.*, vol. 1321, no. 2, 2019, doi: 10.1088/1742-6596/1321/2/022082.

[15] M. Y. Chen, "A high-order fuzzy time series forecasting model for internet stock trading," *Futur. Gener. Comput. Syst.*, vol. 37, pp. 461–467, 2014, doi: 10.1016/j.future.2013.09.025.

[16] S. M. Chen, "Forecasting enrollments based on high-order fuzzy time series," *Cybern. Syst.*, vol. 33, no. 1, pp. 1–16, 2002, doi: 10.1080/019697202753306479.

[17] R. Zhang, B. Ashuri, and Y. Deng, "A novel method for forecasting time series based on fuzzy logic and visibility graph," *Adv. Data Anal. Classif.*, vol. 11, no. 4, pp. 759–783, 2017, doi: 10.1007/s11634-017-0300-3.

[18] M. Bose and K. Mali, "Forecasting with multivariate fuzzy time series: A statistical approach," *Adv. Intell. Syst. Comput.*, vol. 1085, pp. 247–257, 2020, doi: 10.1007/978-981-15-1366-4_20.

[19] F. Li and F. Yu, "Multi-factor one-order cross-association fuzzy logical relationships based forecasting models of time series," *Inf. Sci. (Ny).*, vol. 508, pp. 309–328, 2020, doi: 10.1016/j.ins.2019.08.058.

[20] S. Kumar and N. Kumar, "Two factor fuzzy time series model for rice forecasting," *Int. J. Comput. Math. Sci*, vol. 4, no. 1, pp. 56–61, 2015.

[21] C. Yasunobu, M. Kosaka, K. Yokomura, and K. Honda, "A decision support system building tool with fuzzy logic and its application to chart technical analysis," *Adapt. Intell. Syst.*, pp. 19–32, 1992, doi: 10.1016/b978-0-444-89838-8.50005-9.

Journal of Soft Computing Exploration

A combination of TDM and KSAM to determine initial feasible solution of transportation problems

Muhammad Sam'an[1], Yahya Nur Ifriza[2]
[1]Postgraduate Student, Faculty of Technology Management and Business, Universiti Tun Hussein Onn Malaysia, Malaysia
[2]Department of computer science, Universitas Negeri Semarang, Indonesia

Article Info

Article history:

Received Jan 22, 2021
Revised Feb 26, 2021
Accepted March 15, 2021

Keywords:

Transportation problem
Initial feasible solution
Approximation method
Total difference method

ABSTRACT

In case of the Transportation Problem (TP), it was found that TP had equal the smallest c_{ij} so that the existing methods will be generated two or more IFS values. The newly developed algorithm is generated through a combination of Total Difference Method (TDM) and Karagul-Sahin Approximation Method (KSAM) algorithm, is capable to determine the initial feasible solution of TP. Based on the numerical illustration of the TP example to evaluate the performance of the new proposed algorithm. The computational performances have been compared to the existing methods (TDM1 and KSAM) and the results shown this algorithm achieved better performance than the existing methods for TP example. The recommendation of the research is the new proposed algorithm integration with Stepping Stone and MODI methods in future research to evaluate the optimal solution of TP and determination of optimal solution in the event of information uncertainty about the parameters of the TP.

Corresponding Author:

Muhammad Sam'an,
Postgraduate Student,
Faculty of Technology Management and Business,
Universiti Tun Hussein Onn Malaysia (UTHM),
Batu Pahat, Johor, 86400, Malaysia.
Email: muhammad.92sam@gmail.com

1. INTRODUCTION

The method of finding Initial Feasible Solution (IFS) by assigning the smallest c_{ij} as base cell in Transportation Problem (TP) table has been widely used by several algorithms such as classical algorithm namely Least Cost Method (LCM) takes into account the smallest c_{ij} in the TP table and Vogel's Approximation Method (VAM) takes into account the highest penalty c_{ij} which is denotes the difference between the smallest c_{ij} and the next smallest c_{ij} in each row and each column of the TP table.

Subsequently, various types of finding IFS methods by determining the smallest c_{ij} of TP table were introduced by many researchers [1]–[3]. The pointer c_{ij} calculated by selecting the difference from the highest c_{ij} and second-highest c_{ij} for each row and each column, unlike VAM [4]. Total Opportunity Cost Table (TOCT) calculate the Distribution Indicators (DI) by the difference of the greatest unit c_{ij} and second gretest unit c_{ij}, then the highest two DI are selected as the base cell, and loads are imposed on the original TP corresponding to the base cells of the TOCT [5]. The lowest Allocation Method (LAM) was developed from LCM that also take into account the minimum of supply or demand for allocation in the smallest c_{ij} of TP table

[6]. Advanced Vogel's Approximation Method (AVAM) to overcome case of TP which has the equal smallest c_{ij} in the TP table [7]. TOCM-SUM Approach is calculated by the pointer c_{ij} for each row and column of the TOCM by selecting the sum of all entries in the respective row or column and made a maximum possible allocation to the smallest c_{ij} cell corresponding to the highest pointer c_{ij} [8]. Tuncay Cana's Approximation Method determined to the cell having the nearest c_{ij} to this average c_{ij} considering the demand and production constraints [9]. Incessant Allocation Method (IAM) [10] and Allocation Table Method (ATM) [11] are iterative methods based on the allocation table. Modification of TDM (Total Different Method) 1 considers penalty only for a row of TP table [12]. Global Minimum Method (GMM) [13]. Azad Hasan Method (AHM) is an effective algorithm for allocating the lowest number of demands and supplies to the lowest distribution cost of TP [14], [15]. The modification of TDM by adding the rules for selecting the highest penalty value and checking the lowest distribution cost followed by a combination of TOCM and TDM modification [16]. Total Opportunity Cost Matrix – Supreme Cell [17]. The novel approximation method is called Karagul-Sahin Approximation Method (KSAM) [18]. Zack algorithm [19] and Key Cell method (KCM) based avoiding maximum cost cells for making allocation [20]. An innovative strategy to obtain IBFS of TP based on penalty value [21]. The bottleneck transportation problem is based on pseudo cost to repair the IBFS that was obtained by classical transportation algorithm [22]. The revised ASM method [23]. The combination of TOCM and ASM (TOCM-ASM) [24], [25].

All proposed methods above by previous researchers illustrated the importance of taking the smallest c_{ij} of TP as a base cell for allocating the numbers of supplies and demands of TP because it will affect the value of IFS. But, these proposed methods will only run perfectly if the TP does not have equal smallest c_{ij} during the TP computations. Whenever there is equal smallest c_{ij} in TP computations, the outcomes of the methods will fail to determine the smallest c_{ij}. Several methods have proposed alternatives solution by choosing one of the equal smallest c_{ij} value available. Nevertheless, these alternative solutions for TP will result in more than one IFS value. Consequently, the IFS value generated will be different and it is depending on the smallest c_{ij} value is chosen which is used to compute the IFS.

Therefore, the objective of this research is to overcome the limitations and complications of having two or more IFS values generated through the modification of TDM and followed by integrating with KSAM algorithm. This new algorithm is proposed which is design to obtain optimal IFS value of TP.

2. MODELLING OF TRANSPORTATION PROBLEM

The modeling of transportation problems (TP) can be formulated the following equation to determine the approximation value of x_{ij} that minimizes the total distribution cost as follows,

$$\text{minT} = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \quad (1)$$

subject to

$$\sum_{j=1}^{n} x_{ij} \leq s_i \quad (2)$$

$$\sum_{i=1}^{m} x_{ij} \leq d_j \quad (3)$$

and

$$x_{ij} \geq 0 \forall i, j (i = 1,2, \dots, m; j = 1,2, \dots, n). \quad (4)$$

where m represents a total of supply, n represents a total of demand, s_i is i^{th} suppply, d_j is j^{th} demand, c_{ij} is distribution cost from i^{th} suppply to j^{th} demand, x_{ij} is the number of approximation unit to assign from i^{th} suppply to j^{th} demand, minT is minimal total distribution cost.

If total supply identically with the total demand is called a balance TP and given as,

$$\sum_{i=1}^{m} s_i = \sum_{j=1}^{n} d_j \quad (5)$$

3. THE NEW PROPOSED ALGORITHM

The setting of the smallest cij which has equal values in the base cells in TP table affects the location of x_{ij} allocated on the resulting IFS value depends on the setting of the smallest cij. The new proposed algorithm is based on the modification cation of the TDM algorithm and is integrated with the KSAM algorithm, it is advantageous that the original of TDM and KSAM algorithms are shown as is Algorithm 1 in shown Figure 1 and Figure 2, respectively.

Algorithm 1: TDM Algorithm

Data: Initilization: Number of rows is m, Number of column is n, s_i is i^{th} suppply, d_j is j^{th} demand, c_{ij} is distribution cost from i^{th} suppply to j^{th} demand, x_{ij} is the number of approximation unit to assign from i^{th} suppply to j^{th} demand, $i = 1,2,\ldots,m; j = 1,2,\ldots,n$.

Result: minT by Eq. (1)

repeat

{Produce a minimal total distribution cost} ;

for *i=1 to m* **do**

Find the penalty (F_i) for each i^{th} row by $F_i = \sum_{j=1}^{n} (c_{i,j} - \min(c_{i,j}))$;

Select highest of F_i (HF) by HF = $\max(F_i)$;

Select the least $c_{i,j}$ of HF;

Allocate the x_{ij} to it;

There may arise the following three cases;

if $\min(s_i, d_j) = s_i$ **then**

$x_{ij} = s_i$, $d_j = d_j - s_i$, $s_i = 0$, cross out of s_i

end

if $\min(s_i, d_j) = d_j$ **then**

$x_{ij} = d_j$, $s_i = s_i - d_j$, $d_j = 0$, cross out of d_j

end

if $s_i = d_j$ **then**

$s_i = 0$, $d_j = 0$, cross out of s_i and d_j

end

end

Recalculate the penalty without considering the cross out rows and columns

until $/\sum_{i=1}^{m} s_i = \sum_{j=1}^{n} d_j/$;

Figure 1. TDM1 algorithm

Algorithm 2: KSAM Algorithm

Data: Initilization: Number of rows is m, Number of column is n, s_i is i^{th} suppply, d_j is j^{th} demand, c_{ij} is distribution cost from i^{th} suppply to j^{th} demand, x_{ij} is the number of approximation unit to assign from i^{th} supply to j^{th} demand, $i = 1,2,\ldots,m; j = 1,2,\ldots,n$.

Result: minT by Eq. (1)

Calculate α_{ij} by Eq. (4) and α_{ji} by Eq. (5);

Calculate the weighted transportation cost matrix by multiplying $c_{i,j}$ with α_{ij} for demand (wcd) and $c_{i,j}$ with α_{ji} for supply (wcs);

repeat

{Produce a minimal total distribution cost} ;

Determine the least weighted costs in tables of wcd and wcs ;

Allocate the x_{ij} to it;

There may arise the following three cases;

if $\min(s_i, d_j) = s_i$ **then**

$x_{ij} = s_i$, $d_j = d_j - s_i$, $s_i = 0$, cross out of s_i

end

if $\min(s_i, d_j) = d_j$ **then**

$x_{ij} = d_j$, $s_i = s_i - d_j$, $d_j = 0$, cross out of d_j

end

if $s_i = d_j$ **then**

$s_i = 0$, $d_j = 0$, cross out of s_i and d_j

end

until $/\sum_{i=1}^{m} s_i = \sum_{j=1}^{n} d_j/$;

Compare the solution values of transportation tables. Set the smaller solution as the initial feasible solution

Figure 2. KSAM algorithm

The new proposed algorithm starts by checking either Eq. (5) is satisfied or not and followed by calculation of proportional demand matrix (α_{ij}) and proportional supply matrix (α_{ji}) by using Eq. (6) and (7), respectively [11].

$$\alpha_{ij} = \frac{d_j}{s_i}, i = 1,2,\ldots,m; j = 1,2,\ldots,n. \quad (6)$$

$$\alpha_{ji} = \frac{s_i}{d_j}, j = 1,2,\ldots,n; i = 1,2,\ldots,m. \quad (7)$$

$$\alpha_{ij}\alpha_{ji} = 1 \tag{8}$$

Determine the sum of α_{ij} and α_{ji}, the resulting value is called weighted distribution cost matrix (ω_{ij}). The new proposed algorithm in detail is Algorithm 3 in shown Figure 3.

Algorithm 3: The New Proposed Algorithm

Data: Initilization: Number of rows is m, Number of column is n, s_i is i^{th} suppply, d_j is j^{th} demand, c_{ij} is distribution cost from i^{th} suppply to j^{th} demand, x_{ij} is the number of approximation unit to assign from i^{th} supply to j^{th} demand, $i = 1,2,\ldots,m; j = 1,2,\ldots,n$.
Result: min T by Eq. (1)
Calculate α_{ij} by Eq. (4) and α_{ji} by Eq. (5);
Calculate ω_{ij} by $\omega_{ij} = (\alpha_{ij} + \alpha_{ji})c_{ij}$;
repeat
  {Produce a minimal total distribution cost} ;
  for $i=1$ *to* n **do**
    Find the penalty (F_j) for each j^{th} column by $F_j = \sum_{i=1}^{m} (\omega_{ij} - \min(\omega_{ij}))$;
    Select highest of F_j (HF) by HF = $\max(F_j)$;
    Select the least ω_{ij} of HF;
    Allocate the x_{ij} to it;
    There may arise the following three cases;
    if $\min(s_i, d_j) = s_i$ **then**
      $x_{ij} = s_i$, $d_j = d_j - s_i$, $s_i = 0$, cross out of s_i
    end
    if $\min(s_i, d_j) = d_j$ **then**
      $x_{ij} = d_j$, $s_i = s_i - d_j$, $d_j = 0$, cross out of d_j
    end
    if $s_i = d_j$ **then**
      $s_i = 0$, $d_j = 0$, cross out of s_i and d_j
    end
  end
  Recalculate the penalty without considering the cross out rows and columns
until $/\sum_{i=1}^{m} s_i = \sum_{j=1}^{n} d_j/$;

Figure 3. The new proposed algorithm

4. NUMERICAL ILLUSTRATIONS

The numerical illustration used an example of a transportation problem to explain the new proposed algorithm is shown Table 1 ([16]).

Table 1. An original data of transportation problem in example .

Sources	Destination						s_i
	D_1	D_2	D_3	D_4	D_5	D_6	
S_1	9	12	9	6	9	10	2
S_2	7	3	7	7	5	5	5
S_3	6	5	9	11	3	11	6
S_4	6	8	11	2	2	10	9
d_j	2	2	2	4	4	6	

The solution of the transportation problem example 1 in Table 1 is solved by using Algorithm 3 as follows:

Step 1 : Calculating α_{ij} and α_{ji}. For i=1,2,3,4;j=1,2,3,4,5,6 are obtained as bellows

$$\alpha_{46} = \begin{pmatrix} 1 & 1 & 2 & 2 & 2 & 3 \\ 0.4 & 0.4 & 0.8 & 0.8 & 0.8 & 1.2 \\ 0.33 & 0.33 & 0.67 & 0.67 & 0.67 & 1 \\ 0.22 & 0.22 & 0.44 & 0.44 & 0.44 & 0.67 \end{pmatrix} \text{ and } \alpha_{64} = \begin{pmatrix} 1 & 1 & 0.5 & 0.5 & 0.5 & 0.33 \\ 2.5 & 2.5 & 1.25 & 1.25 & 1.25 & 0.83 \\ 3 & 3 & 1.5 & 1.5 & 1.5 & 1 \\ 4.5 & 4.5 & 2.25 & 2.25 & 2.25 & 1.5 \end{pmatrix}^T$$

Step 2: Calculating ω_{ij}. For i=1,2,3,4;j=1,2,3,4,5,6 is produced as bellow

$$\boldsymbol{\omega_{46}} = \begin{pmatrix} 18 & 24 & 22.5 & 15 & 22.5 & 33.3 \\ 23.2 & 8.7 & 14.35 & 14.35 & 10.25 & 10.17 \\ 20 & 16.67 & 19.5 & 23.8 & 6.5 & 22 \\ 28.33 & 37.78 & 29.69 & 5.39 & 5.39 & 21.67 \end{pmatrix}$$

Step 3: Calculating F_j. For j=1 so F_1 is obtained as bellows

$$\begin{aligned} F_1 \text{ of column-1} &= 14.63 \\ F_1 \text{ of column-2} &= 52.34 \\ F_1 \text{ of column-3} &= 28.59 \\ F_1 \text{ of column-4} &= 37.01 \\ F_1 \text{ of column-5} &= 23.03 \\ F_1 \text{ of column-6} &= 46.5 \end{aligned}$$

Step 4: Selecting the highest of F_1 is F_1 of column-2
Step 5: Selecting the least ω_{ij} in F_1 of column-2 is ω_{22}.
Step 6: Allocating the x_{22} to ω_{22} with $x_{32} = \min(s_2, d_2) = \min(2,5) = 2$ such that $s_2 = s_2 - d_2 = 5 - 2 = 3$ and $d_2 = d_2 - d_2 = 2 - 2 = 0$. Since $s_2 \neq 0$ and $d_2 = 0$ then cross out of d_2.
Step 7: Re-calculating the penalty without considering d_2 such that for j= 2 is obtained as bellows.

$$\begin{aligned} F_2 \text{ of column-1} &= 14.63 \\ F_2 \text{ of column-3} &= 28.59 \\ F_2 \text{ of column-4} &= 37.01 \\ F_2 \text{ of column-5} &= 23.03 \\ F_2 \text{ of column-6} &= 46.5 \end{aligned}$$

Step 8: Repeat Step 3 until Step 6 such that $\sum_{i=1}^{m} s_i = \sum_{j=1}^{n} d_j$. The final result is shown in Table 2.

Table 2. The new proposed algorithm with a penalty.

Sources	Destination					
	D_1	D_2	D_3	D_4	D_5	D_6
S_1	18	24	22.5	15	22.5	33.3
S_2	23.2	8.7	14.35	14.35	10.25	10.17
S_3	20	16.67	19.5	23.8	6.5	22
S_4	28.33	37.78	29.69	5.39	5.39	21.67
F_1	14.63	52.34	28.59	37.01	23.08	46.5
F_2	14.63	0	28.59	37.01	23.08	46.5
F_3	12.33	0	13.13	28.05	18.22	12
F_4	12.33	0	13.13	0	18.22	12
F_5	12.33	0	13.13	0	0	12
F_6	2	0	3	0	0	11.33
F_7	2	0	3	0	0	0
F_8	2	0	0	0	0	0
F_9	0	0	0	0	0	0

Step 9: Finally, calculating minimal total distribution cost by using Eq. (1). The result of minT is 115. The feasible solution table of example 1 can be seen Table 3.

Table 3. The feasible solution table of example.

Sources	Destination						s_i
	D_1	D_2	D_3	D_4	D_5	D_6	
S_1	0	0	0	0	0	2	2
S_2	0	2	0	0	0	3	5
S_3	2	0	4	0	0	0	6
S_4	0	0	0	4	4	1	9
d_j	2	2	2	4	4	6	

5. DISCUSSION

The application of the new proposed has been described with a numerical illustration of transportation problem example which has not only one or more the smallest c_{ij} of the TP table. The solving of TP example used the Algorithm 3 did not raise the potential to produce more than one IFS. It can be shown from the highest penalty used to determine which is base cell has not the equal value.

Meanwhile, the solving of TP example used Algorithm 3 (TDM 1[12]) raised the potential to produce more than on IFS. It is because the highest penalty has equal value. The highest penalty of TP example was produced by Algorithm 1 is shown in Table 4. If we compare the highest penalty obtained by Algorithm 1 with Algorithm 3, then Algorithm 3 is better at producing the highest penalty than Algorithm 1. Besides that, Algorithm 1 does not provide additional conditions if there is the highest penalty with the equal value such that it can potentially be Algorithm 1 cannot be made to solve TP example.

Table 4. The penalties of transportation problem in example generated TDM 1.

Sources	Destination						F_i
	D_1	D_2	D_3	D_4	D_5	D_6	
S_1	9	12	9	6	9	10	19
S_2	7	3	7	7	5	5	16
S_3	6	5	9	11	3	11	27
S_4	6	8	11	2	2	10	27

Furthermore, the solving of TP example used Algorithm 2 (KSAM [18]) showed also the potential to generate not only one or more on IFS. It is caused by weighted transportation cost matrix by demand (WCD) or weighted transportation cost matrix by supply (WCS) which is assigned to select the smallest c_{ij} as a base cell has equal value. The value of WCD and WCS on TP example as follows,

$$\text{WCD} = \begin{pmatrix} 9 & 12 & 18 & 12 & 18 & 30 \\ 2.8 & 1.2 & 5.6 & 5.6 & 4 & 6 \\ 2 & 1.67 & 6 & 7.3 & 2 & 11 \\ 1.3 & 1.78 & 4.89 & 0.89 & 0.89 & 6.67 \end{pmatrix} \text{ and}$$

$$\text{WCS} = \begin{pmatrix} 9 & 12 & 4.5 & 3 & 4.5 & 3.3 \\ 17.5 & 7.5 & 8.75 & 8.75 & 6.25 & 4.17 \\ 18 & 15 & 13.5 & 16.5 & 4.5 & 11 \\ 27 & 36 & 24.75 & 4.5 & 4.5 & 15 \end{pmatrix}^T$$

If we also compare WCS and WCD obtained by Algorithm 2 with the penalty of Algorithm 3, then Algorithm 3 is better than Algorithm 2. Similarly, with the TDM 1, Algorithm 2 does not provide additional conditions if there is the WCD or WCS with the equal value such that it can potentially be Algorithm 2 cannot be made to solve TP example. TDM 1 and KSAM add criteria by freely selected the equal highest penalty (Algorithm 1) and equal value of WCD or WCS (Algorithm 2) so that IFS will be obtained. Therefore, the result comparison Of IFS between the new proposed method (Algorithm 3) with the existing methods (Algorithm 1 [12] and Algorithm 2 [18]) is shown Table 5. In this table can be seen the new proposed algorithm solved the TP example better than TDM 1 and KSAM. The algorithm of TDM and KSAM produced two IFS. The first IFS produced by TDM 1 algorithm is better than the new proposed algorithm and otherwise, the second IFS produced by TDM algorithm fewer results compared to the new proposed algorithm. Meanwhile, both IFS (WCS) generated by KSAM algorithm is better than the new proposed algorithm.

Table 5. The result comparison of IFS.

The existing method	Initial Feasible Solution (IFS)
TDM1	117 (1)
	110 (2)
	123 (WCD)
KSAM	121 (WCS-1)
	131 (WCS-2)
The new proposed algorithm	115

6. CONCLUSION

The importance of setting the smallest c_{ij} as the base cell for finding IFS of TP to determine the optimal solution which is the minimum total cost. The new proposed algorithm is developed based on the KSAM characteristics improvisation plus integration of modified TDM algorithm. It also considers the supply and demand coverage ratio (weights) as well as the smallest c_{ij} and included the penalties calculations for each column. This new proposed algorithm is capable to find the IFS effectively and efficiently. The numerical illustration of TP example indicated that the new proposed algorithm numerical results is comparable to the existing methods. The new proposed algorithm achieved better performance than TDM 1 and KSAM algorithms for TP example. In addition, it is also recommended that the new proposed algorithm integration with Stepping Stone and MODI methods in future research to evaluate the optimal solution of TP and determination of optimal solution in the event of information uncertainty about the parameters of TP.

REFERENCES

[1] D. Aprilianto, "SVM optimization with correlation feature selection based binary particle swarm optimization for diagnosis of chronic kidney disease," *J. Soft Comput. Explor.*, vol. 1, no. 1, pp. 24–31, 2020.

[2] R. H. Saputra and B. Prasetyo, "Improve the accuracy of c4.5 algorithm using particle swarm optimization (pso) feature selection and bagging technique in breast cancer diagnosis," *J. Soft Comput. Explor.*, vol. 1, no. 1, pp. 47–55, 2020.

[3] I. E.Tiffani, "Optimization of naïve bayes classifier by implemented unigram, bigram, trigram for sentiment analysis of hotel review," *J. Soft Comput. Explor.*, vol. 1, no. 1, pp. 1–7, 2020.

[4] A. Khan, "A Re-solution of the Transportation Problem: An Algorithmic Approach," *Jahangirnagar Univ. J. Sci.*, vol. 34, pp. 49–62, 2011.

[5] M. A. Islam, M. M. Haque, and M. S. Uddin, "Extremum difference formula on total opportunity cost: a transportation cost minimization technique," *Prime Univ. Journa*, vol. 6, no. 2, pp. 125–130, 2012.

[6] M. Ashraful Babu, M. Abu Helal, M. S. Hasan, and U. K. Das, "Lowest Allocation Method (LAM): a new approach to obtain feasible solution of transporta-tion model," *Int. J. Sci. Eng. Res.*, vol. 4, no. 11, 2013.

[7] U. Kanti Das, M. Ashraful Babu, A. Rahman Khan, and D. Sharif Uddin, "Advanced Vogel's Approximation Method (AVAM): a new approach to determine penalty cost for better feasible solution of transportation problem," *Int. J. Eng. Res. Technol.*, vol. 3, p. 182, 2014.

[8] A. R. Khan, V. Adrian, N. Sultana, and S. S. Ahmed, "Determination of initial basic feasible solution of a transportation problem: a TOCM-SUM approach," *Bul. Institutului Politeh. Din Iasi*, vol. 61, pp. 39–49, 2015.

[9] T. Can, "Tuncay Can's Approximation method to obtain initial basic feasible solution to transport problem," *Appl. Comput. Math.*, vol. 5, no. 2, p. 78, 2016, doi: 10.11648/j.acm.20160502.17.

[10] M. M. Ahmed, A. R. Khan, F. Ahmed, and M. S. Uddin, "Incessant Allocation Method for Solving Transportation Problems," *Am. J. Oper. Res.*, vol. 06, no. 03, pp. 236–244, 2016, doi: 10.4236/ajor.2016.63024.

[11] M. M. Ahmed, A. R. Khan, M. S. Uddin, and F. Ahmed, "A new approach to solve transportation problems," *Open J. Optim.*, vol. 05, no. 01, pp. 22–30, 2016, doi: 10.4236/ojop.2016.51003.

[12] E. Hosseini, "Three new methods to find initial basic feasible solution of transportation problems," *Appl. Math. Sci.*, vol. 11, pp. 1803–1814, 2017, doi: 10.12988/ams.2017.75178.

[13] Y. Harrath and J. Kaabi, "New heuristic to generate an initial basic feasible solution for the balanced transportation problem," *Int. J. Ind. Syst. Eng.*, vol. 30, no. 2, p. 193, 2018, doi: 10.1504/ijise.2018.10016216.

[14] S. M. Abul Kalam Azad and M. Kamrul Hasan, "An effective algorithm to solve cost minimising transportation problem," *Int. J. Math. Oper. Res.*, vol. 15, no. 4, pp. 434–445, 2019, doi: 10.1504/IJMOR.2019.103005.

[15] Z. H. Radthy, F. H. Maghool, and A. H. Khaleel, "Application the linear programming according to transportation problem on real data," *Int. J. Sci. Technol. Res.*, vol. 8, no. 1, pp. 100–102, 2019.

[16] B. Amaliah, C. Fatichah, and E. Suryani, "Total opportunity cost matrix – Minimal total: A new approach to determine initial basic feasible solution of a transportation problem," *Egypt. Informatics J.*, vol. 20, no. 2, pp. 131–141, 2019, doi: 10.1016/j.eij.2019.01.002.

[17] B. Amaliah, C. Fatichah, E. Suryani, and A. Muswar, "Total opportunity cost matrix - Supreme cell: A new method to obtain initial basic feasible solution of transportation problems," in *ACM International Conference Proceeding Series*, 2020, pp. 151–156, doi: 10.1145/3411174.3411198.

[18] K. Karagul and Y. Sahin, "A novel approximation method to obtain initial basic feasible solution of transportation problem," *J. King Saud Univ. - Eng. Sci.*, vol. 32, no. 3, pp. 211–218, 2020, doi: 10.1016/j.jksues.2019.03.003.

[19] Z. U. Rizqi, "Zack algorithm: A heuristic approach to solve transportation problem," in *Proceedings of the International Conference on Industrial Engineering and Operations Management*, 2019, pp. 1127–1131.
[20] L. Kaur, M. Rakshit, and S. Singh, "An alternate approach for finding the initial basic feasible solution of transportation problem," *Int. J. Sci. Technol. Res.*, vol. 8, no. 9, pp. 442–447, 2019.
[21] P. Sumathi and C. V. Sathiya Bama, "An innovative route to acquire least cost in transportation problems," *Int. J. Eng. Adv. Technol.*, vol. 9, no. 1, pp. 5368–5369, 2019, doi: 10.35940/ijeat.A3070.109119.
[22] B. Mallia, M. Das, and C. Das, "A New Algorithm for Bottleneck Transportation Problem," *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 8, pp. 419–423, 2019.
[23] R. Murugesan and T. Esakkiammal, "Determination of best initial basic feasible solution of a transportation problem: A TOCM-ASM approach," *Adv. Math. Sci. J.*, vol. 9, no. 7, pp. 4563–4577, 2020, doi: 10.37418/amsj.9.7.25.
[24] R. Murugesan and T. Esakkiammal, "An improved ASM method for the transportation problem," *Adv. Math. Sci. J.*, vol. 9, no. 10, pp. 8259–8271, 2020, doi: 10.37418/amsj.9.10.55.
[25] R. Murugesan and T. Esakkiammal, "TOCM-VAM method versus asm method in transportation problems," *Adv. Math. Sci. J.*, vol. 9, no. 6, pp. 3549–3566, 2020, doi: 10.37418/amsj.9.6.34.

Journal of Soft Computing Exploration

Three stages algorithm for finding optimal solution of balanced triangular fuzzy transportation problems

Muhammad Sam'an
Postgradute Student, Faculty of Technology Management and Business, Universiti Tunn Hussein Onn, Malaysia

Article Info

Article history:

Received Jan 25, 2021
Revised Feb 28, 2021
Accepted March 18, 2021

Keywords:

Fuzzy transportation problem
Segregated method
Balanced triangular fuzzy number
Total difference method
Modified distribution method
Fuzzy optimal solution

ABSTRACT

In the literature, the fuzzy optimal solution of balanced triangular fuzzy transportation problem is negative fuzzy number. This is contrary to the constraints that must be non-negative. Therefore, the three stages algorithm is proposed to overcome this problem. The proposed algorithm consists of segregated method with segregating triangular fuzzy parameters into three crisp parameters. This method avoids the ranking technique. Next, total difference method is used to get the initial basic feasible solution (IBFS) value based on segregating triangular fuzzy parameters. While modified distribution algorithm is used to determine optimal solution based on IBFS value. In order to illustrate the proposed algorithm is given the numerical example and based on the result comparison, the proposed algorithm equality to the two existing algorithms and better than the one existing algorithm. The proposed algorithm can solve the fuzzy decision-making problems and can also be extended to an unbalanced fuzzy transportation problem.

Corresponding Author:

Muhammad Sam'an,
Postgraduate Student, Faculty of Technology Management and Business,
Universiti Tunn Hussein Onn Malaysia (UTHM),
Batu Pahat, Johor, 86400, Malaysia.
Email: muhammad.92sam@gmail.com

1. INTRODUCTION

In real life, parameters of transportation problems (supply, demand, and transportation cost) sometimes are uncertain in values. The transportation cost, depends on fuel prices, congested routes and weather, while supply is caused by reduced quantities of raw materials, machine breakdowns and production failures. In addition, volatile market situations create uncertainty in demand. However, Zadeh [1] introduced transportation problem with the numbers of supply, demand, and transportation cost are represented by fuzzy number that is called Fuzzy Transportation Problem (FTP). The solving of FTP, many researchers changed the fuzzy parameters to crisp sets and solved by crisp transportation algorithm [2]–[12].

[13] used Yanger's ranking function to change L-R flat fuzzy parameters to crisp numbers and fuzzy classical transportation algorithm i.e. VAM (Vogel's Approximation Method), LCM (Least Cost Method) and NWC (North West Corner) to obtain L-R flat fuzzy optimal solution. [14] proposed the new ranking function-based integral value to rank triangular fuzzy parameters and generalized fuzzy classical algorithm i.e. GVAM, GLCM and GNWC to determine a generalized triangular fuzzy optimal solution. [6] simplified the new ranking function that was proposed by [3] to rank trapezoidal fuzzy parameters and GLCM to obtain generalized trapezoidal fuzzy optimal solution. Chandran and Kandaswamy presented the score ranking method to change

the triangular and trapezoidal fuzzy parameters and the modification of fuzzy LCM to find fuzzy optimal solution. [7] discussed the ranking method based mean of triangular fuzzy numbers and minimum demand-supply method to obtain crisp optimal solution. [9] proposed the classical ranking method and VAM-based harmonic mean method to obtain penalty value of each column and row. [15] introduced a score and accuracy function based on the Pythagorean fuzzy number and applied it on the fuzzy VAM to obtain fuzzy optimal solution. [16] presented the classical ranking function and the generalized minimum supply-demand to obtain triangular fuzzy optimal solution. [11] presented a segregated advancement scheme based minimum demand-supply method and stepping stone to obtain triangular fuzzy optimal solution.

Furthermore, the direct approach of solving FTP such as the Zero point method to obtain fuzzy optimal solution [17]. Robust ranking to rank fuzzy number and Zero suffix method to obtain the fuzzy optimal solution Fuzzy dual matrix to obtain a fuzzy optimal solution [18], [19], Improved and revised Zero point [20]–[22]. The modification of Zero-point [23]. Particle Swarm Optimization algorithm (PSO) with fuzzy constraint and conjugate constraint [24]–[27].

The use of the ranking function in the solving of FTP has a very significant impact on the resulting fuzzy optimal solution. As the results of [13] which produce a negative fuzzy optimal solution, this is contrary to the constraints that must be non-negative in the FTP model. In addition, the ranking process takes a long time to compute so that it will affect the computational performance that is not good. Therefore, in this article, we use the segregated advancement scheme approach based on the total difference method and modified distribution method to produce a fuzzy optimal solution without ranking function.

2. THE MODELING OF BALANCED TRIANGULAR FUZZY TRANSPORTATION PROBLEM

Based on the triangular fuzzy transportation problem with a delivers and b receivers. Given $\tilde{s}_i = s_i^l, s_i^m, s_i^u$ be the triangular fuzzy supply at the i^{th} deliverer, $\tilde{d}_j = d_j^l, d_j^m, d_j^u$ be the triangular fuzzy demand at the j^{th} receiver where $i = 1,2,\dots,a; j = i = 1,2,\dots,b$. Let $\tilde{c}_{ij} = (c_{ij}^l, c_{ij}^m, c_{ij}^u)$ be the per unit triangular fuzzy transportation cost from the i^{th} deliverer to the j^{th} receiver and $\tilde{x}_{ij} = (x_{ij}^l, x_{ij}^m, x_{ij}^u)$ be the number of triangular fuzzy approximation unit to assign from the i^{th} deliverer to the j^{th} receiver. The modeling of balanced triangular FTP can be formulated as follows,

$$\min Z = \sum_{i=1}^{a} \sum_{j=1}^{b} \tilde{c}_{ij}\tilde{x}_{ij} \tag{1}$$

subject to

$$\begin{aligned} &\sum_{j=1}^{b} \tilde{x}_{ij} = \tilde{s}_i \quad i = 1,2,\dots,a, \\ &\sum_{i=1}^{a} \tilde{x}_{ij} = \tilde{d}_j \quad j = i = 1,2,\dots,b \\ &\tilde{x}_{ij} \geq 0 \qquad \forall i,j \end{aligned} \tag{2}$$

3. THREE STAGES ALGORITHM FOR FINDING TRIANGULAR FUZZY OPTIMAL SOLUTION

In this section, the three stages algorithm is proposed to find fuzzy optimal solution of a balanced triangular fuzzy transportation problem. The proposed algorithm consists of a segregated scheme to partition triangular fuzzy numbers, Total Difference Method (TDM)1 [12] to find Initial Basic Feasible Solution (IBFS) and Modified Distribution method (MODI) to obtain optimal solution based IBFS.

4. A SEGREGATED SCHEME

The segregated method is based on the possibility of optimized output concept of balanced triangular fuzzy transportation problem when the triangular fuzzy parameters with the corresponding demand and supply be partitioned one by one. This scheme consists of pointwise segregation of each triangular fuzzy parameters such that the first element of each triangular fuzzy parameters will be defined first segregated transportation problem is denoted STP^1. Similarly, the second and third elements of each triangular fuzzy parameters are denoted STP^3 and STP^3 respectively.

Using the segregated method, Eq. (1) and (2) can be transformed into the followings three STPs as follows,

1) STP^1

$$\min Z_1^l = \sum_{i=1}^{a} \sum_{j=1}^{b} c_{ij}^l x_{ij}^l \quad (3)$$

$$\begin{aligned} &\sum_{j=1}^{b} x_{ij}^l = s_i^l \quad i = 1,2,\dots,a, \\ &\sum_{i=1}^{a} x_{ij}^l = d_j^l \quad j = i = 1,2,\dots,b \\ &x_{ij}^l \geq 0 \qquad \forall i,j \end{aligned} \quad (4)$$

2) STP^2

$$\min Z_2^m = \sum_{i=1}^{a} \sum_{j=1}^{b} c_{ij}^m x_{ij}^m \quad (5)$$

$$\begin{aligned} &\sum_{j=1}^{b} x_{ij}^m = s_i^m \quad i = 1,2,\dots,a, \\ &\sum_{i=1}^{a} x_{ij}^m = d_j^m \quad j = i = 1,2,\dots,b \\ &x_{ij}^m \geq 0 \qquad \forall i,j \end{aligned} \quad (6)$$

3) STP^3

$$\min Z_3^u = \sum_{i=1}^{a} \sum_{j=1}^{b} c_{ij}^u x_{ij}^u \quad (7)$$

$$\begin{aligned} &\sum_{j=1}^{b} x_{ij}^u = s_i^u \quad i = 1,2,\dots,a, \\ &\sum_{i=1}^{a} x_{ij}^u = d_j^u \quad j = i = 1,2,\dots,b \\ &x_{ij}^u \geq 0 \qquad \forall i,j \end{aligned} \quad (8)$$

4.1 Total Difference Method 1

For a crisp balanced transportation problem with transportation cost matrix of order (a, b) having supply $s_i^q, i = 1,2,\dots,a$; demand $d_j^q, i = j = 1,2,\dots,b$ and the corresponding transportation cost $c_{ij}^q, q = l,m,u$. In detail, TDM 1 algorithm shown in Algorithm 1 shown in Figure 1.

4.2 Modified Distribution Method

MODI to obtain the optimal solution of balanced triangular fuzzy transportation problem Algorithm 2 shown in Figure 2.

4.3 Three Stages Algorithm

Three stages algorithm to obtain the optimal solution of balanced triangular fuzzy transportation problem shown Algorithm 3 shown in Figure 3.

Algorithm 1: TDM Algorithm

Data: Initialization: Number of rows is a, Number of column is b, s_i^q is i^{th} supply, d_j^q is j^{th} demand, c_{ij}^q is distribution cost from i^{th} supply to j^{th} demand, x_{ij}^q is the number of approximation unit to assign from i^{th} supply to j^{th} demand, $i = 1, 2, \ldots, a; j = 1, 2, \ldots, b; q = l, m, u$.

Result: $\min z_q$

repeat

{Produce a minimal total distribution cost} ;

for *i=1 to m+n-1* **do**

Find the penalty (F_i^q) for each i^{th} row by $F_i = \sum_{j=1}^{n} (c_{i,j}^q - \min(c_{i,j}^q))$;

Select highest of F_i^q (HF) by HF = $\max(F_i^q)$;

Select the least $c_{i,j}^q$ of HF;

Allocate the x_{ij}^q to it;

There may arise the following three cases;

if $\min(s_i^q, d_j^q) = s_i^q$ **then**

$x_{ij}^q = s_i^q$, $d_j^q = d_j^q - s_i^q$, $s_i^q = 0$, cross out of s_i^q

end

if $\min(s_i^q, d_j^q) = d_j^q$ **then**

$x_{ij}^q = d_j^q$, $s_i^q = s_i^q - d_j^q$, $d_j^q = 0$, cross out of d_j^q

end

if $s_i^q = d_j^q$ **then**

$s_i^q = 0$, $d_j^q = 0$, cross out of s_i^q and d_j^q

end

end

Recalculate the penalty without considering the cross out rows and columns

until $\{\sum_{i=1}^{a} s_i^q = \sum_{j=1}^{b} d_j^q\}$;

Figure 1. TDM 1 algorithm

Algorithm 2: Modified-Distribution Method

Step 1: The IBFS obtained of FTP by using Algorithm 1

Step 2: Introduce m_i^q and n_j^q as variable convenient for every i^th and j^th, respectively. In front of i^th write m_i^q in row and at n_j^q the under ofj^th in column. Let $m_i^q = 0$ is maximum number of allocations row;

Step 3: Determine $\lambda_{i,j}^q$ and n_j^q by using $c_{ij}^q = m_i^q + n_j^q$ for base of cell, then determine $\lambda_{i,j}^q = \tilde{c}_{ij}^q - (u_i^q + v_j^q)$ of non-base of cells. Next, two possibilities as follow;

(a) If $\lambda_{i,j}^q \geq 0, \forall i, j$, then the resulted of IBFS is done. In other words, fuzzy optimal solution has been satisfied;

(b) Otherwise, $\exists \lambda_{i,j}^q$, then the resulted of IBFS do not finished yet. In other words, fuzzy optimal solution is not optimal. Therefore, fuzzy optimal solution is chosen a cell of $(i, j)^th$ in which $\lambda_{i,j}^q$ is smallest negative. Next, make a horizontal and vertical closed path that starts from unchosen base of cell of $(i, j)^{th}$. The path can only replace to angle on base of cell $(i, j)^th$ and the path is chosen must pass through base and non-base cell of $(i, j)^{th}$;

Step 4: Give sign $(+)$ and $(-)$ for closed loop started with $(+)$ for chosen non-base cells. After that, determine fuzzy quantity on cells with signs $(+)$ and $(-)$. Consequently, new TP table is obtained.

Step 5: Repeat of steps 2, 3 and 4 for TP table until $\lambda_{i,j}^q \geq 0, \forall i, j$

Step 6: Obtain a new improved solution by allocating units to the unfilled cell according step 5 and calculate the new TP.

Step 7:Determine the value of fuzzy optimal solution or objective function $\min z_q$

Figure 2. MODI algorithm

Algorithm 3: Three stages algorithm

Step 1 Imply the segregated scheme discussed in Section 3.1. to determine the corresponding three $STP_q, q = l, m, n.$

Step 2 Imply the TDM 1 in Algorithm 1 to each STP_q to find IBFS.

Step 3 Imply the MODI method in Algorithm 2 to each of the three STP_q to obtain the optimal solutions.

Step 4 Determine a fuzzy optimal solution of balanced triangular fuzzy transportation problem as a combination of the optimal solutions obtained in Step 3.

Figure 3. Three stages algorithm

5. NUMERICAL ILLUSTRATION

In this section , we illustrate the proposed method by using numerical example is adapted from Ebrahimnejad [28] in which a trading company wants to optimize the transportation cost of products.

Example 1 A leading trading factory wants to obtain the fuzzy number of the commodity that should be distributed from each warehouse to each object such that the total fuzzy transportation cost is at a minimum. The factory has two production houses and three distribution centers. The parameters of transportation are fuzzy numbers that are represented by triangular fuzzy numbers because of disinformation real-life conditions. The summary of fuzzy parameters i.e. supply, demand and transportation cost (dollar) are shown in Table 1 with l as the least amount value, m as the most possible value and u as the greatest amount value.

Table 1. Data of example 1 (in U.S. dollar)

Parameters		Triangular fuzzy numbers		
		Least amount value (l)	Most possible value (m)	Greatest amount value (n)
Transportation cost ($)	c_{11}	15	25	35
	c_{12}	55	65	85
	c_{13}	85	95	105
	c_{21}	65	75	85
	c_{22}	80	90	110
	c_{23}	30	40	50
Supply in unit	s_1	75	95	125
	s_2	45	65	95
Demand in unit	d_1	35	45	65
	d_2	25	35	45
	d_3	60	80	110

Solution: After implementing the segregated scheme discussed in Section 5.1, three crisp transportation problems i.e. STP^1, STP^2, STP^3 are resulted as follows:

1) for STP^1 by using Eq. (3) and (4)

$$\min Z_1^l = 15x_{11}^l + 55x_{12}^l + 85x_{13}^l + 65x_{21}^l + 80x_{22}^l + 30x_{23}^l \tag{9}$$

subject to

$$\begin{aligned} x_{11}^l + x_{12}^l + x_{13}^l &= 75 \\ x_{21}^l + x_{22}^l + x_{23}^l &= 45 \\ x_{11}^l + x_{21}^l &= 35 \\ x_{12}^l + x_{22}^l &= 25 \\ x_{13}^l + x_{23}^l &= 25 \end{aligned} \tag{10}$$

2) for STP^2 by using Eq. (5) and (6)

$$\min Z_2^m = 25x_{11}^m + 65x_{12}^m + 95x_{13}^m + 75x_{21}^m + 90x_{22}^m + 40x_{23}^m \tag{11}$$

subject to

$$\begin{aligned} x_{11}^m + x_{12}^m + x_{13}^m &= 95 \\ x_{21}^m + x_{22}^m + x_{23}^m &= 65 \\ x_{11}^m + x_{21}^m &= 45 \\ x_{12}^m + x_{22}^m &= 35 \\ x_{13}^m + x_{23}^m &= 805 \end{aligned} \tag{12}$$

3) for STP^3 by using Eq. (7) and (8)

$$\min Z_3^u = 35x_{11}^u + 85x_{12}^u + 105x_{13}^u + 85x_{21}^u + 110x_{22}^u + 50x_{23}^u \tag{13}$$

subject to

$$\begin{aligned} x_{11}^u + x_{12}^u + x_{13}^u &= 125 \\ x_{21}^u + x_{22}^u + x_{23}^u &= 95 \\ x_{11}^u + x_{21}^u &= 65 \\ x_{12}^u + x_{22}^u &= 45 \\ x_{13}^u + x_{23}^u &= 110 \end{aligned} \tag{14}$$

After implementing the TDM 1 in Algorithm 5.2, the penalty value resulted

(a) for STP^1 is $F_1 = 110$ and $F_2 = 85$ such that the IBFS obtained $\begin{matrix} x_{11}^l = 35 & x_{12}^l = 25 & x_{13}^l = 15 \\ x_{21}^l = 0 & x_{22}^l = 0 & x_{23}^l = 45 \end{matrix}$

(b) for STP^2 is $F_1 = 110$ and $F_2 = 85$ such that the IBFS obtained $\begin{matrix} x_{11}^m = 45 & x_{12}^m = 35 & x_{13}^m = 15 \\ x_{21}^m = 0 & x_{22}^m = 0 & x_{23}^m = 65 \end{matrix}$

(c) for STP^3 is $F_1 = 120$ and $F_2 = 95$ such that the IBFS obtained $\begin{matrix} x_{11}^u = 65 & x_{12}^u = 45 & x_{13}^u = 15 \\ x_{21}^u = 0 & x_{22}^u = 0 & x_{23}^u = 95 \end{matrix}$

After implementing the MODI method shown in Algorithm 2(`)@ is used to obtain optimal solution based on the IBFS value of all three $STP_q, q = l, m, n$ and resulted in minimal transportation cost for $STP_l = 4535, STP_m = 7425$ and $STP_u = 12425$.

After combining the minimal transportation cost of all three STP_q, so that the fuzzy optimal solution of balanced triangular fuzzy optimal solution is (4525, 7425, 1245) where represents that the minimal transportation cost most likely will be \$7,425 but certainly not less than \$4,525. Meanwhile, if things are not in favor of decision-maker, it could be as high as \$12,425.

6. CONCLUSION

In this article, the three stages algorithm consist of a segregated method, total difference method 1 and modified distribution method is proposed to optimize of balanced triangular fuzzy transportation problem. A segregated method scheme and TDM 1 are used to get IBFS value, meanwhile, the MODI method to obtain optimal solution based IBFS value. The numerical example is given to illustrate the justification of proposed algorithm. The Comparative study of results with the literature journal shows that the fuzzy optimal solution that is resulted by proposed algorithm equivalent to those by [13] and [11] and better than those by [2]. All the acquired allocations by proposed algorithm there are non-negative triangular fuzzy numbers whereas one allocation by [2] is negative and violates the basic rule of trading.

REFERENCES

[1] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, 1965, doi: 10.1016/S0019-9958(65)90241-X.

[2] A. Kumar, J. Kaur, and P. Singh, "A new method for solving fully fuzzy linear programming problems," *Appl. Math. Model.*, vol. 35, no. 2, pp. 817–823, 2011, doi: 10.1016/j.apm.2010.07.037.

[3] S. Chandran and G. Kandaswamy, "A fuzzy approach to transport optimization problem," *Optim. Eng.*, vol. 17, no. 4, pp. 965–980, 2016, doi: 10.1007/s11081-012-9202-6.

[4] G. De França Aguiar, B. De Cássia Xavier Cassins Aguiar, and V. E. Wilhelm, "New methodology to find initial solution for transportation problems: A case study with fuzzy parameters," *Appl. Math. Sci.*,

vol. 9, no. 17–20, pp. 915–927, 2015, doi: 10.12988/ams.2015.4121018.

[5] A. Kumar and P. Singh, "A new method for solving fully fuzzy linear programming problems," *Ann. Fuzzy Math. Inf.*, vol. 3, pp. 103–118, 2012.

[6] A. Ebrahimnejad, "A simplified new approach for solving fuzzy transportation problems with generalized trapezoidal fuzzy numbers," *Appl. Soft Comput. J.*, vol. 19, pp. 171–176, 2014, doi: 10.1016/j.asoc.2014.01.041.

[7] N. Mathur, P. K. Srivastava, and A. Paul, "Trapezoidal fuzzy model to optimize transportation problem," *Int. J. Model. Simulation, Sci. Comput.*, vol. 7, no. 3, 2016, doi: 10.1142/S1793962316500288.

[8] D. Hunwisai and P. Kumam, "A method for solving a fuzzy transportation problem via Robust ranking technique and ATM," *Cogent Math.*, vol. 4, no. 1, 2017, doi: 10.1080/23311835.2017.1283730.

[9] K. Balasubramanian and S. Subramanian, "Optimal solution of fuzzy transportation problems using ranking function," *Int. J. Mech. Prod. Eng. Res. Dev.*, vol. 8, no. 4, pp. 551–558, 2018, doi: 10.24247/ijmperdaug201856.

[10] D. Chakraborty, D. K. Jana, and T. K. Roy, "A new approach to solve fully fuzzy transportation problem using triangular fuzzy number," *Int. J. Oper. Res.*, vol. 26, no. 2, pp. 153–179, 2016, doi: 10.1504/IJOR.2016.076299.

[11] R. K. Saini, A. Sangal, and O. Prakash, "Fuzzy transportation problem with generalized triangular-trapezoidal fuzzy number," in *Advances in Intelligent Systems and Computing*, 2018, vol. 583, pp. 723–734, doi: 10.1007/978-981-10-5687-1_64.

[12] D. Rani and T. R. Gulati, "A new approach to solve unbalanced transportation problems in imprecise environment," *J. Transp. Secur.*, vol. 7, no. 3, pp. 277–287, 2014, doi: 10.1007/s12198-014-0143-5.

[13] A. Kumar and A. Kaur, "Application of Classical Transportation Methods to Find the Fuzzy Optimal Solution of Fuzzy Transportation Problems," *Fuzzy Inf. Eng.*, vol. 3, no. 1, pp. 81–99, 2011, doi: 10.1007/s12543-011-0068-7.

[14] A. Kaur and A. Kumar, "A new approach for solving fuzzy transportation problems using generalized trapezoidal fuzzy numbers," *Appl. Soft Comput. J.*, vol. 12, no. 3, pp. 1201–1213, 2012, doi: 10.1016/j.asoc.2011.10.014.

[15] R. Kumar, S. A. Edalatpanah, S. Jha, and R. Singh, "A Pythagorean fuzzy approach to the transportation problem," *Complex Intell. Syst.*, vol. 5, no. 2, pp. 255–263, 2019, doi: 10.1007/s40747-019-0108-1.

[16] P. K. Srivastava and D. C. S. Bisht, "A Segregated Advancement in the Solution of Triangular Fuzzy Transportation Problems," *Am. J. Math. Manag. Sci.*, pp. 1–11, 2020, doi: 10.1080/01966324.2020.1854137.

[17] P. Pandian and G. Natarajan, "A new algorithm for finding a fuzzy optimal solution for fuzzy transportation problems," *Appl. Math. Sci.*, vol. 4, no. 1–4, pp. 79–90, 2010, doi: 10.1504/IJMMNO.2021.111715.

[18] A. Edward Samuel and M. Venkatachalapathy, "IZPM for unbalanced fuzzy transportation problems," *Int. J. Pure Appl. Math.*, 2013, doi: 10.12732/ijpam.v86i4.8.

[19] K. Selvakumari and S. Sathya Geetha, "A new approach for solving intuitionistic fuzzy transportation problem," *J. Adv. Res. Dyn. Control Syst.*, vol. 12, no. 5 Special Issue, pp. 956–963, 2020, doi: 10.5373/JARDCS/V12SP5/20201841.

[20] A. Edward Samuel and M. Venkatachalapathy, "Improved zero point method for solving fuzzy transportation problems using ranking function," *Far East J. Math. Sci.*, vol. 75, no. 1, pp. 85–100, 2013.

[21] A. Edward Samuel, "Improved zero point method (IZPM) for the transportation problems," *Appl. Math. Sci.*, vol. 6, no. 109–112, pp. 5421–5426, 2012.

[22] T. Karthy and K. Ganesan, "Improved zero point method for the fuzzy optimal solution to fuzzy transportation problems," *Glob. J. Pure Appl. Math.*, vol. 12, no. 1, pp. 255–260, 2016.

[23] A. Akilbasha, G. Natarajan, and P. Pandian, "A new approach for solving transportation problems in fuzzy nature," *Int. J. Appl. Eng. Res.*, vol. 11, no. 1, pp. 498–502, 2016.

[24] A. Baykasoğlu and K. Subulan, "A direct solution approach based on constrained fuzzy arithmetic and metaheuristic for fuzzy transportation problems," *Soft Comput.*, vol. 23, no. 5, pp. 1667–1698, 2019, doi: 10.1007/s00500-017-2890-2.

[25] M. A. Imron and B. Prasetyo, "Improving algorithm accuracy K-Nearest Neighbor using Z-Score Normalization and Particle Swarm Optimization to predict customer churn," *J. Soft Comput. Explor.*, vol. 1, no.1, pp. 56–62, 2020.

[26] R. H. Saputra and B. Prasetyo, "Improve the accuracy of c4.5 algorithm using particle swarm optimization (pso) feature selection and bagging technique in breast cancer diagnosis," *J. Soft Comput. Explor.*, vol. 1, no.1, pp. 47–55, 2020.

[27] D. Aprilianto, "SVM optimization with correlation feature selection based binary particle swarm optimization for diagnosis of chronic kidney disease," *J. Soft Comput. Explor.*, vol. 1, no.1, pp. 24–31, 2020.
[28] A. Ebrahimnejad, "An improved approach for solving fuzzy transportation problem with triangular fuzzy numbers," *J. Intell. Fuzzy Syst.*, vol. 29, no. 2, pp. 963–974, 2015, doi: 10.3233/IFS-151625.

Journal of Soft Computing Exploration

Improved logistics service quality for goods quality delivery services of companies using analytical hierarchy process

Popy Riliandini[1], Erika Noor Dianti[2], Sayidah Rohmatul Hidayah[3], Dwika Ananda Agustina Pertiwi[4]
[1] Informatics Department, Universitas Sebelas Maret
[2,3,4] Computer Science Department, Universitas Negeri Semarang

Article Info

Article history:

Received Jan 28, 2021
Revised Feb 28, 2021
Accepted March 17, 2021

Keywords:

Logistic of company
Logistic service quality
Decision making
AHP

ABSTRACT

Logistics plays a role in the smooth transaction between companies because it is a facilitator of buying and selling goods and services to fulfill the supply orders of consumer companies. This study aims to analyze how the impact of improved Logistic Service Quality (LSQ) for quality of goods delivery services by using LSQ dimensions from previous research. Sample data is obtained through the dissemination of questionnaires which are then processed quantitatively with convergent validity and reliability tests. Data processing with a sample count of 61 respondents. The results of this study show that there is the main dimension of logistic service quality in improving the quality of service, namely ordering condition, time, and information quality. Each comparison factor is tested for consistency using the Analytical Hierarchy Process (AHP), each of the main criteria has a consistency value of less than 0.1 so that the main criteria tested have a consistent comparison matrix and can be the basis of decision making for companies in choosing alternative criteria priorities.

Corresponding Author:

Popy Riliandini,
Informatics Department,
Universitas Sebelas Maret,
Jl. Ir. Sutami No.36, Kentingan, Kec. Jebres, Surakarta, Indonesia.
Email: popy.riliandini28@student.uns.ac.id

1. INTRODUCTION

The development of industrial companies in Indonesia is now increasing rapidly with the influence of advances in information technology. Especially in the field of goods delivery services which is the process framework of a raw material supplier company industrial company [1], [2]. Judging from the rapid development of the industry is one of the factors that the logistics industry will be on the verge of growth in manufacturing, retail, and real estate that began to emerge to the surface [3].

Logistics plays a role in the smooth transaction between companies because it is a facilitator of buying and selling goods and services to fulfill the supply orders of consumer companies. Customer satisfaction is assessed from the quality of service, where a good delivery service is determined by the concept of logistics owned [4], [5]. So, in the logistics of delivery of goods punctuality and place affects the satisfaction of consumers.

In this study, the author tried to apply the AHP method in the case of company problems, namely to know the influence of LSQ on goods delivery service companies. Previous research [6]–[9] that on the relationship between the quality of logistics services with a special emphasis on determining factors as well as customer satisfaction which also contains expectations from customers towards improving the quality of

logistics services. By applying the AHP and LSQ methods will be tried to know the sectors that contribute predominantly to the company so that the strategy to determine the quality of services in accordance with the conditions and capabilities of the company can be done using the concept of supporting the decision.

LSQ is an interconnected quality construction process, where the construction is reliable and valid in all customer segments. The emphasis is placed on each of the different constructions in several customer segments. Based on the scale of LSQ, supplier companies can understand the desire and expectations of customers in improving the quality of service [10], [11].

In improving the quality of goods delivery services, several alternative criteria focus on how the condition of supply orders, the speed of delivery of supply goods, and the quality of information delivery services when responding to consumer companies, so that good communication is established. The criteria of improving the quality of services will be weighted using the information of a company [12], [13].

Service quality measurement intends to evaluate the performance of supply goods delivery services with selected dimensions. Measurement dimensions can be taken from how the level of customer satisfaction. In a previous study [14]–[16] from the observation process it was obtained that the most significant criterion in improving customer satisfaction and comfort was the accuracy of order conditions.

Judging from the importance of improving the quality of goods delivery services, therefore we then conduct this research, intending to gain the main criteria weight based on the dimensions of LSQ in improving logistics services delivery of goods.

2. METHOD

The quantitative approach used in this study, quantitative method is an approach used for the research of certain populations or samples and based on the philosophy of positivism [16], after the data is collected then the data is analyzed which is quantitatively used as an instrument in this study.

This study using data retrieval techniques through the dissemination of questionnaires will then be quantitatively tested with *convergent validity and reliability tests* using SPSS applications. The data collection process uses questionnaires as instruments that are distributed to companies that use logistics services. The questionnaire is a technique of collecting sample data from respondents through the instrument of questions provided by researchers.

In order to obtain information there are steps in collecting data that is by making a questionnaire consisting of question instruments that correspond to the specified research variables, next determine the respondent to fill out a questionnaire that meets the criteria that have been determined, before filling out the questionnaire respondents are given the procedure of filling out the questionnaire and the purpose of this study, after filling out the questionnaire then the completeness of the questionnaire data, if the data is incomplete then the data will not be used as research sample data. The questionnaire data is tailored to the dimensions of LSQ from previous research by [15], [16]. Weighting criteria by using Multi-Criteria Decision Making (MCDM) namely AHP.

The object of this research sample is the company that owns the goods and users of logistics services delivery of goods. Characteristics of the company that we choose in this study include: (a) companies that often utilize logistics services at least one year; (b) the existence of delivery of goods in a year occurs twice delivery; (c) never get shipments of goods that are not appropriate; (d) through the website, the company utilizes logistics services. As for the framework of research can be seen in Figure 1.

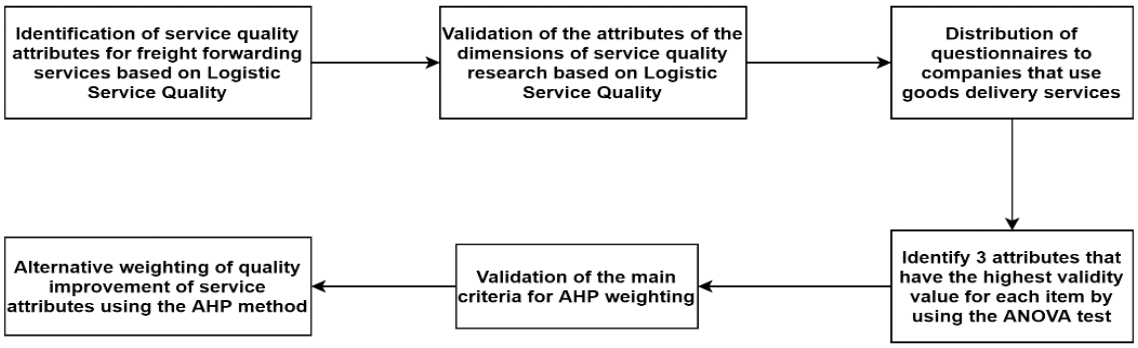


Figure 1. Framework of research

The next step is to analyze the factors that make consumers dissatisfied with the quality of services provided, the choice of criteria as an alternative weighting sample. Identify criteria tailored to Logistic Service Quality to review the main criteria to be taken. Then weighting is carried out on each alternative based on the

main criteria that have been identified [17]–[19]. This weighting is done by supporting the AHP decision method, where the alternative and comparison of each criterion are inputted in the AHP and it is considered that it has not been able to meet customer satisfaction is the output attribute of LSQ [16].

3. RESULT AND DISCUSSION

Sample data was obtained through the dissemination of questionnaires at a company that utilizes the logistics system as the fulfillment of the supply of raw materials production, where the sample data obtained by 61 respondents who are then processed through validity and reliability tests to find out the average of each attribute questionnaire using the method of statistical quantity analysis ANOVA and the results can be seen Table 1.

Table 1. Convergent validity test

Construct	Items	Mean	Std. Deviation
Information	Ease of accessing service information	3,97	0,73
Quality	Accuracy of information from the provider of goods delivery services (logistics)	3,54	0,721
	Clarity of information from the provider of goods delivery services (logistics)	3,7	0,715
Order Procedure	Provided website of the provider of goods delivery services (delivery of goods)	3,77	0,864
	Delivery service providers do not limit the volume of	3,44	1,025
	Procedure clarity	3,77	0,920
Time	Time to arrive fast	4,00	0,753
	The promised time is very appropriate	3,54	0,867
	Packing setup speed	3,56	0,786
Order Condition	Goods up to good condition	4,13	0,806
	Items not damaged	3,57	0,753
	Items not lost	3,62	0,730
Order Discrepancies Handling	Responsiveness in responding to problems	3,51	0,977
	Be agile in dealing with problems that arise	3,43	0,939
	Provide proper compensation	3,61	0,954

From the results of the convergent validity test, there is an average value of each criterion, taken the highest mean value based on the results among others, order condition with a mean value of 4.13, then time with a mean value of 4.00, and information quality with a mean value of 3.78. Where the three ratings mean value is used as the main criteria for weighting calculation on the method of supporting the decision.

Next, a correlation test is conducted to determine the strength and direction of the linear relationship of the correlation between variables by comparing the calculated R and R values of table. The result is said to be valid if the $_{\text{calculated}}$ R-value is $>R_{\text{of the table,}}$ and the value is significantly less than 0.05. The test variable correlations can be seen Table 2.

From the reliability test in Table 2, getting the result of Cronbach's Alpha value of each criterion greater than 0.6 then we can conclude that the level of stability of measurable data is high and reliable. The reliability test can be seen Table 3. From this table, got a value of Cronbach's Alpha greater than 0.6 then it can be concluded that the level of stability of measurable data is high and reliable. The next step is data processing for the determination of priority weights of each attribute using the AHP method, this determination aims to know the consistency of each of the main criteria as an attribute in improving the quality of logistics services delivery of goods for companies seen from the point of view of the main criteria.

Table 2. Test variable correlations

Items	Pearson Correlation	Sig.	N
IQ1	1	0,00	61
IQ2	0,587	0,00	61
IQ3	0,482	0,00	61
OP1	0,469	0,00	61
OP2	0,407	0,01	61
OP3	0,599	0,00	61
T1	0,472	0,00	61
T2	0,460	0,00	61
T3	0,519	0,00	61
OC1	0,474	0,00	61
OC2	0,516	0,00	61
OC3	0,512	0,00	61
ODH1	0,539	0,00	61
ODH2	0,482	0,00	61
ODH3	0,671	0,00	61

Table 3. Reliability test

Items	Cronbach's Alpha	Description
IQ1	0,915	Reliabel
IQ2	0,915	Reliabel
IQ3	0,919	Reliabel
OP1	0,921	Reliabel
OP2	0,920	Reliabel
OP3	0,915	Reliabel
T1	0,917	Reliabel
T2	0,918	Reliabel
T3	0,917	Reliabel
OC1	0,917	Reliabel
OC2	0,918	Reliabel
OC3	0,917	Reliabel
ODH1	0,912	Reliabel
ODH2	0,912	Reliabel

Consistency of assessment on AHP method using Consistency Ratio (CR), formulated:

$$CR = = \frac{CI}{RI}$$

where:

$$\text{CI} = \frac{(\lambda_{maks} - n)}{n-1}$$

The following are the results of the calculation of the Consistency Ratio (CR) of each alternative can be seen in Table 4 and 5,

Table 4. Eigenvector normalization criteria of order conditions

Criteria	C1	C2	C3	Total	EVN
C1	2.99	7.65	19	29.64	0.6403262985
C2	1.26	2.98	7.65	11.89	0.2568650368
C3	0.5089	1.26	2.99	4.7589	0.1028086647
	Overall			46.2889	

Table 5. Consistency ratio of order condition criteria

Emacs	Ci	Cr
3.017202828	0.008601414162	0.01483002442

From Table 5, the criteria ratio of the order condition is $CR < 0.1$. So, it can be drawn results in Tables 6 and 7 that the comparison matrix is consistent.

Table 6. Eigenvector normalization criteria delivery speed

Criteria	C1	C2	C3	Total	EVN
C1	2.99	0.5089	1.26	4.7589	0.1028086647
C2	19	2.99	7.65	29.64	0.6403262985
C3	7.65	1.26	2.98	11.89	0.2568650368
Overall				46.2889	

Table 7. Delivery speed criteria consistency ratio

Emacs	Ci	Cr
3.017202828	0.008601414162	0.01483002442

From the consistency table, the ratio of the delivery speed criteria is $CR < 0.1$. So it can be drawn results in Tables 8 and 9 that the comparison matrix is consistent.

Table 8. Eigenvector normalization of information quality

Criteria	C1	C2	C3	Total	EVN
C1	3	11.4	1.1	15.5	0.2115174672
C2	1.1	2.98	0.32	4.4	0.06004366812
C3	11.4	39	2.98	53.38	0.7284388646
Overall				73.28	

Table 9. Consistency ratio of information quality criteria

Emacs	Ci	Cr
3.068084061	0.03404203057	0.05869315615

From the consistency table, the ratio of quality criteria is information that the value $CR < 0.1$. So it can be drawn results that the comparison matrix is consistent. Next, the final result of the selection of criteria using multiplication of the results of eigenvector normalization of each criterion tested is C1 = 0.5008846749, C2 = 0.3261559147 and C3 = 0.1729594104.

In criteria 1, alternative 1 ranks first which means excelling in maintaining supply order conditions. Then in criteria 2, alternative 2 is superior to other alternatives in the delivery time/speed criteria. And alternative 3 becomes the first order on criteria 3 that is superior in considering consumers of the company through good quality information. The alternative priority of criterion can be seen in Figure 2.

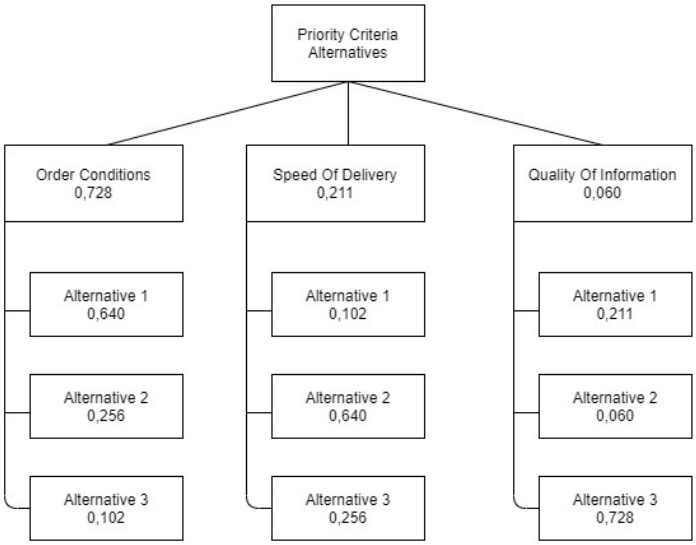


Figure 2. Alternative priority criteria

On the alternative priority chart of criteria, each dimension of the main criteria can be used as a decision-supporting construction in choosing an alternative delivery service company, where the weighting construction decision is tailored to the needs of the consumer company. The alternative priority chart of criteria can be seen Figure 3.

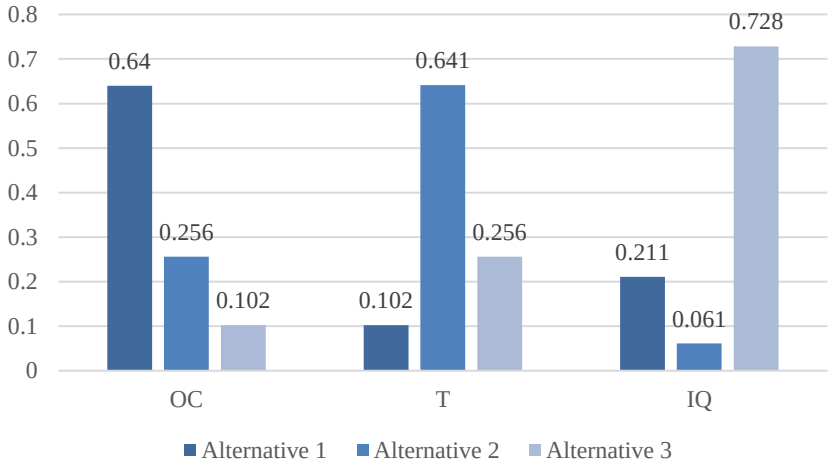


Figure 3. Alternative priority chart of criteria

4. CONCLUSION

Research on improving the quality of goods delivery services based on Logistic Service Quality by using sample data obtained through the dissemination of questionnaires in service user companies, obtained from 5 dimensions of LSQ namely Information Quality (IQ), Order Procedure (OP), Time (T), Order Condition (OC), Order Discrepancies Handling (ODH). From sample data obtained processed through AHP decision support method, resulting in three main criteria, namely good condition of goods, speed of delivery of goods, quality of information services. The main criteria of weighting results are factors that can influence in improving the quality of the company's services in the field of logistics of delivery of supply goods.

REFERENCES

[1] M. C. Cooper and J. T. Gardner, “Building good business relationships: more than just partnering or strategic alliances?,” *Int. J. Phys. Distrib. Logist. Manag.*, vol. 23, no. 6, pp. 14–26, 1993, doi: 10.1108/09600039310044876.

[2] Z. Lemma, K. Sripruetkiat, and I. Nitithanprapas, "Factors affecting sustainability of business relationships in ethiopia banana value chain," *J. Austrian Soc. Agric. Econ.*, vol. 16, no. 05, pp. 83–95, 2020.

[3] R. A. Kahnali and A. Esmaeili, "An integration of SERVQUAL dimensions and logistics service quality indicators (A case study)," *Int. J. Serv. Oper. Manag.*, vol. 21, no. 3, pp. 289–309, 2015, doi: 10.1504/IJSOM.2015.069650.

[4] R. Bouzaabia, O. Bouzaabia, and A. Capatina, "Retail logistics service quality: A cross-cultural survey on customer perceptions," *Int. J. Retail Distrib. Manag.*, vol. 41, no. 8, pp. 627–647, 2013, doi: 10.1108/IJRDM-02-2012-0012.

[5] A. N. Giovanis, P. Tomaras, and D. Zondiros, "Suppliers Logistics Service Quality Performance and its effect on retailers' behavioral intentions," *Procedia - Soc. Behav. Sci.*, vol. 73, pp. 302–309, 2013, doi: 10.1016/j.sbspro.2013.02.056.

[6] I. G. Saura, D. S. Francés, G. B. Contrí, and M. F. Blasco, "Logistics service quality: A new way to loyalty," *Ind. Manag. Data Syst.*, vol. 108, no. 5, pp. 650–668, 2008, doi: 10.1108/02635570810876778.

[7] X.-Y. Zhou, D.-M. Wang, and Y. Liu, "The empirical study of logistics service quality evaluation within electronic retail," 2018, doi: 10.2991/icmesd-18.2018.32.

[8] S. Nugroho, S. Kempa, and T. Wiliater Soaloon Panjaitan, "Logistic service quality and customer satisfaction to customer retention on rice producer industry," *SHS Web Conf.*, vol. 76, p. 01048, 2020, doi: 10.1051/shsconf/20207601048.

[9] M. Er, T. Wasusri, H. M. Astuti, and A. Herdiyanti, "The service quality of Indonesia's logistics service provider in preparation for ASEAN economic community," *Lect. Notes Electr. Eng.*, vol. 349, pp. 647–656, 2015, doi: 10.1007/978-3-662-47200-2_68.

[10] J. T. Mentzer, D. J. Flint, and G. T. M. Hult, "Logistics service quality as a segment-customized process," *J. Mark.*, vol. 65, no. 4, pp. 82–104, 2001, doi: 10.1509/jmkg.65.4.82.18390.

[11] A. Sutrisno, E. Andajani, and F. N. Widjaja, "The Effects of Service Quality on Customer Satisfaction and Loyalty in a Logistics Company," *KnE Soc. Sci.*, 2019, doi: 10.18502/kss.v3i26.5360.

[12] W. H. Yu, S. K. Chiu, and C. M. Tung, "The study of evolution among logistic service quality, service compensation and long-term cooperation commitment," *Procedia Manuf.*, vol. 39, pp. 1493–1500, 2019, doi: 10.1016/j.promfg.2020.01.299.

[13] H. M. Jang, P. B. Marlow, and K. Mitroussi, "The effect of logistics service quality on customer loyalty through relationship quality in the container shipping context," *Transp. J.*, vol. 52, no. 4, pp. 493–521, 2013, doi: 10.5325/transportationj.52.4.0493.

[14] A. K. Rahmat and N. Faisol, "Manufacturers satisfaction on logistics service quality: operational, relational and national culture," *Procedia - Soc. Behav. Sci.*, vol. 224, pp. 339–346, 2016, doi: 10.1016/j.sbspro.2016.05.385.

[15] J. Sze, D. O. L. Teik, F. Tiffany, L. F. Kok, and T.Y. Teh, "Logistic service quality among courier services in malaysia," *Int. Conf. Econ. Bus. Innov.*, vol. 38, pp. 113–117, 2012.

[16] J. S. Y. Ho, D. O. L. Teik, F. Tiffany, L. F. Kok, and T. Y. Teh, "The Moderating Effect of Local VS. Foreign CourierService Providers on Logistic Service Quality (LSQ)," *Int. J. Trade, Econ. Financ.*, pp. 257–261, 2012, doi: 10.7763/ijtef.2012.v3.210.

[17] M. A. Imron and B. Prasetyo, "Improving algorithm accuracy k-nearest neighbor using z-score normalization and particle swarm optimization to predict customer churn," *J. Soft Comput. Explor.*, vol. 1, pp. 56–62, 2020.

[18] A. R. Safitri and M. A. Muslim, "Improved accuracy of naive bayes classifier for determination of customer churn uses smote and genetic algorithms," *J. Soft Comput. Explor.*, vol. 1, no. 1, pp. 70–75, 2020.

[19] H. A. Prihanditya, "The implementation of z-score normalization and boosting techniques to increase accuracy of c4 . 5 algorithm in diagnosing chronic kidney disease," *J. Soft Comput. Explor.*, vol. 5, no. 1, pp. 63–69, 2020.

Journal of Soft Computing Exploration

Irrigation management of agricultural reservoir with correlation feature selection based binary particle swarm optimization

Yahya Nur Ifriza[1], Muhammad Sam'an[2]
[1]Department of Computer Science, Universitas Negeri Semarang, Indonesia
[2]Postgradute Student, Faculty of Technology Management and Business, Universiti Tunn Husein Onn, Malaysia

Article Info

Article history:

Received Feb 8, 2021
Revised March 7, 2021
Accepted March 20, 2021

Keywords:

Support vector machine
Correlation feature
Binary particle swarm
Irrigation management
Agricultural reservoir

ABSTRACT

The requirement for the applied innovation to farming water system is especially required for supplies, as rural water system focuses. Supplies as one of horticulture water system asset focus that are regularly constraints identified with the conveyance of repository water stream, this brought about lopsided dissemination of rural water system and the term of control of agrarian water system that streams from water system asset focuses. At the point when ranchers need to change the water system way, it will take a long effort to make another water system way. From these troubles to convey rural water systems simpler, it is important to plan a specialist framework to decide rural water system choices. A few researchers focused on improved quality of plant. There have been limited studies concerned with irrigation management Therefore, this research intends to design The objectives of this research are optimization irrigation management of agricultural reservoirs with CFS-BPSO. The consequences of this investigation demonstrate that the exactness of the utilization of the SVM calculation is 62.32%, while after utilizing the CFS calculation precision of 84.12% is acquired and exactness of ten SVM calculations by applying a blend of CFS highlight choice. also, BPSO 91.84%.

Corresponding Author:

Yahya Nur Ifriza,
Department of Computer Science,
Universitas Negeri Semarang,
Sekaran, Gunungpati, Semarang 50229, Indonesia.
Email: yahyanurifriza@mail.unnes.ac.id

1. INTRODUCTION

The value and practical period of farming water system improvement or recovery will decrease if there isn't sufficient legitimate upkeep and operational administration to look after them, feasible and suitable from all invested individuals in water system water use exercises. A stream is a level or land 75 where surface water comes from downpour, snow melts, or ice unites to a point at low height [1]. Supplies of good-quality water system water are required to diminish in a few districts because of expanded metropolitan mechanical rural rivalry, accessible freshwater supplies should be utilized all the more effectively [2]. A day-by-day soil dampness balance was utilized to foresee dissipation and happening from which harvest yields were assessed [3]. Improvement of profits to water assets could emerge out of creating better caliber and higher value crops for homegrown and unfamiliar business sectors utilizing improved water system procedures [4]. To foresee a choice in information mining can utilize arrangement methods [5]. Support Vector Machine [6]–[8], K-Nearest

Neighbor [9]–[11], Decision Tree [12], [13], and Artificial Neural Network [12], [14] are instances of arrangement calculations.

The requirement for the applied innovation to farming water system is especially required for supplies, as rural water system focuses. Supplies as one of horticulture water system asset focus that is regularly constrained identified with the conveyance of repository water stream, this brought about lopsided dissemination of rural water system and the term of control of agrarian water system that streams from water system asset focuses, thusly the developments are required for the dispersion of rural water system should be possible dependent on the necessities Water system of farmland by ranchers [15]. A water-saving water system can be utilized rather than nonstop flood water system to utilize the groundwater [16]. Water system stretch, or soil water accessibility for crop use, soil type, evapotranspiration request, dispersion of plant root framework and saltiness, and harmfulness of explicit particles on crop development [17].

There are three stages to create accuracy cultivating innovation. The initial step depends on ordinary agrarian innovation, by escalating machines to lessen works. The subsequent advance includes the improvement of planning procedures, variable-level of machine innovation, and presentation of the fundamental framework choice emotionally supportive network. The third step applies innovation well, the model applied in past research was to coordinate the utilization of fitting innovation in farming [18].

At the point when ranchers need to change the water system way, it will take a long effort to make another water system way. From these troubles to convey rural water systems simpler, it is important to plan a specialist framework to decide rural water system choices. A few researchers focused on improved quality of plant. There have been limited studies concerned with irrigation management Therefore, this research intends to design the objectives of this research are optimization irrigation management of agricultural reservoir with CFS-BPSO.

2. METHOD

The interaction utilized by an agronomist (master) to decide the measure of water system is showed in Figure 1.

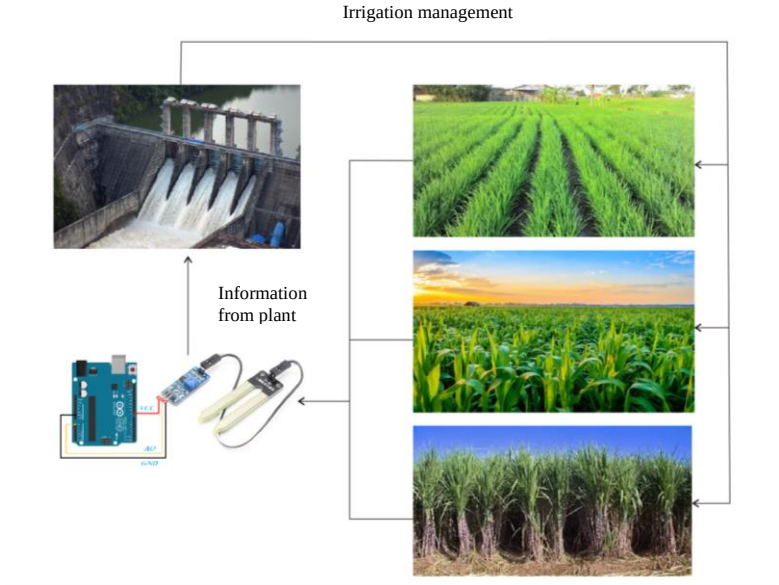


Figure 1. Cycle utilized design to decide irrigation management.

This article uses sensor data for all data models and is compiled from the conditions of each plant. includes agricultural humidity sensor data, reservoir water level data, data on types of agricultural commodities. Data may vary based on the data set available from the source. This dataset has 11 attributes that is (1) water level of reservoir (kaw); (2) moisture of land 1 (kel1); (3) moisture of land 2 (kel2); (4) land humidity 3 (kel3); (5) moisture of land 4 (kel4); (6) land humidity 5 (kel5); (7) type of commodity land 1 (koml1); (8) type of commodity land 2 (koml2); (9) kinds of land commodity 3 (koml3); (10) type of commodity land 4 (koml4); (11) commodity land type 5 (koml5).

In this examination, the blend of CFS and BPSO was completed as an element choice. CFS is utilized to decrease the elements of the dataset dependent on the connection among's highlights and target class yet doesn't associate with different highlights. BPSO is utilized to locate the best blend of highlights. The characterization strategy utilized is the Support Vector Machine calculation. From the arrangement results, we

will get an expansion in precision from Support Vector Machine when the blend of CFS and BPSO is applied [8]. The flowchart of optimization CFS-BPSO is depicted in Figure 2.

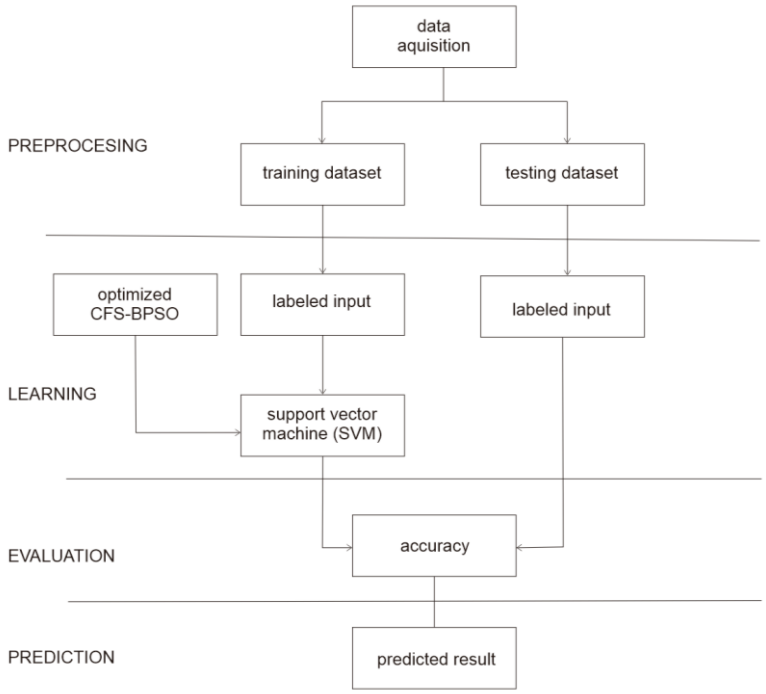


Figure 2. Flowchart of Optimization CFS-BPSO

Relationship-based Component Choice (CFS) is a channel calculation that positions subset ascribes as indicated by heuristic assessment capacities dependent on connection [18]. CFS will assess includes by thinking about the prescient abilities of each component and the degree of repetition between them. If the relationship among's traits and class is known, and the connection between's each character is given, at that point the connection can be anticipated.

PSO was presented in 1995 by Kennedy and Eberhart dependent on a friendly reproduction model known as a stochastic enhancement calculation [19], [20]. Examination and applications on Particle Swarm Optimization (PSO) have expanded quickly since its development and this has brought about many improved PSO calculations in different kinds of advancement issues. In PSO, the hyper boundary is streamlined by two highlights; calculation and its capacity [21]. This has been applied in PSO calculations to take care of the streamlining issues or to improve the first PSO [22]. Heaps of work and investigation of the adequacy of PSO contrasted with other AI and multitude knowledge calculation for designing and software engineering issue has been finished by specialists to assess its exhibitions [23]. It can be joined PSO-SVM to figure horticulture water utilization [24]. The idea of PSO is that every molecule is flown in search space to locate the best arrangement (wellness) called pbest. At that point, the best generally esteem (worldwide worth) is called gbest. Every molecule has two vectors in particular position vectors and speed vectors to move around in search space. Every molecule has memory and every molecule will follow the best position already [24].

Support Vector Machine (SVM) was first proposed by Vladimir Vapnik. Proposed in the field of measurable learning hypothesis and primary danger minimization [25], [26]. SVM has been utilized in an assortment of issues like information characterization, picture grouping, text order, tone acknowledgment, digit acknowledgment of penmanship [27], [28].

3. RESULTS AND DISCUSSIONS

In this investigation, the proposed calculation testing utilizes the Python programming language by using the scikit-learn, sk-highlight and pyswarms libraries. The information utilized is a dataset taken from UCI Machine Learning. This dataset has 11 attributes. CFS will choose attributes that have the highest correlation weighting value. From the CFS process, 11 selected attributes were obtained. The list of attributes and weights of the CFS process is shown in Table 1.

Table 1. List of attributes and results of CFS weight

No	Attributes	CFS Weight
1	kaw	0,53212231
2	kel1	0,54326545
3	kel2	0,54356437
4	kel3	0,53478954
5	kel4	0,54587690
6	kel5	0,53976512
7	koml1	0,53987652
8	koml2	0,54543217
9	koml3	0,54667584
10	koml4	0,53485769
11	koml5	0,53212231

The attributes, picked by the CFS calculation, don't generally create the best blend of characteristics. Thusly, the BPSO calculation is utilized to decide the best element mix of the credits picked by the CFS. At this stage, 11 tests are hurried to decide the best element mix. BPSO boundaries utilized in this examination have appeared in Table 2.

Table 2. BPSO parameters

Parameters	Value
Number of computation	50
Max iteration	100
Number of particles	50
Cognitif c1	2,00
Social c2	2,00

At this stage, 3 tests were done, specifically the independent SVM calculation, the SVM calculation with the execution of the CFS calculation and the SVM by actualizing a blend of CFS and BPSO. In the main application, the SVM calculation will handle the CKD dataset with 11 credits. The utilization of the SVM calculation has a precision of 62.32%. The aftereffects of this precision express that the SVM calculation can characterize the dataset well because the exactness results are more noteworthy than the mistake rate. Be that as it may, the consequences of this precision can be improved by applying a few preprocessing techniques. In the subsequent application, the SVM calculation will be joined with the CFS calculation. So SVM will handle the dataset with 11 attributes and 1 class. The exactness of this grouping model is 84.12%. The precision of applying this model can expand the exactness of the SVM calculation by 21.8%. Notwithstanding, these outcomes can in any case be improved by picking the best element mix utilizing the BPSO calculation. The third application, the SVM algorithm will be joined with the CFS and BPSO algorithm. In this usage, 6 tests are rushed to decide the best component mix. The exactness of this arrangement model can be found in Table 3.

Table 3. SVM Accuracy Results with CFS and BPSO

Execution	Number of attributes	Value
1	11	91,32%
2	10	90,87%
3	10	90,58%
4	11	92,31%
5	11	93,62%
6	11	92,37%

The precision of applying this model can expand the exactness of the SVM + CFS calculation by 7.72% and can build the SVM calculation by 29.52%. The correlation of the precision of every calculation application can be found in Figure 3.

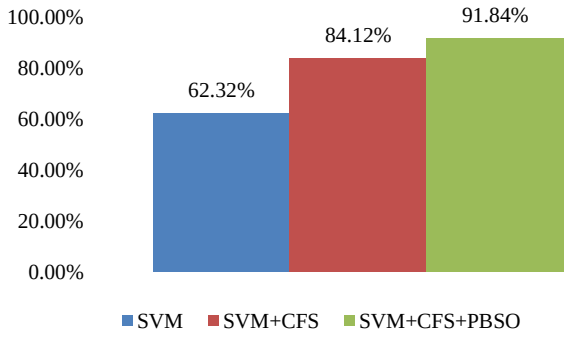


Figure 3. The comparison result SVM, CFS and BPSO

4. CONCLUSION

The SVM algorithm was tried by applying the CFS algorithm and the BPSO calculation utilizing a dataset. The CFS calculation is utilized to get credits with a great connection, while BPSO is utilized to acquire the best characteristic blends. The consequences of this investigation demonstrate that the exactness of the utilization of the SVM calculation is 62.32%, while after utilizing the CFS calculation precision of 84.12% is acquired and exactness of ten SVM calculations by applying a blend of CFS highlight choice. also, BPSO 91.84%. Subsequently, it very well may be presumed that the use of a mix of CFS and BPSO is a choice component in the SVM calculation can build the precision in diagnosing by 29.52%.

REFERENCES

[1] S. A. Rahaman, S. A. Ajeez, S. Aruchamy, and R. Jegankumar, "Prioritization of sub watershed based on morphometric characteristics using fuzzy analytical hierarchy process and geographical information system – a study of kallar watershed, tamil nadu," *Aquat. Procedia*, vol. 4, pp. 1322–1330, 2015, doi: 10.1016/j.aqpro.2015.02.172.

[2] M. Qadir, "Crop and irrigation management strategies for saline-sodic soils and waters aimed at environmentally sustainable agriculture," *Sci. Total Environ.*, 2003, doi: 10.1016/s0048-9697(03)00596-5.

[3] D. L. Martin, D. G. Watts, and J. R. Gilley, "Model and production function for irrigation management," *J. Irrig. Drain. Eng.*, vol. 110, no. 2, pp. 149–164, 1984, doi: 10.1061/(asce)0733-9437(1984)110:2(149).

[4] M. Valipour, "A comprehensive study on irrigation management in Asia and Oceania," *Arch. Agron. Soil Sci.*, vol. 61, no. 9, pp. 1247–1271, 2015, doi: 10.1080/03650340.2014.986471.

[5] N. Deb, V. Jha, A. K. Panjiyar, and R. K. Gupta, "A comparative analysis of news categorization using machine learning approaches," *Int. J. Sci. Technol. Res.*, vol. 9, no. 1, pp. 2469–2472, 2020.

[6] S. Mayor and B. Pant, "Document classification using Support Vector Machine," *Int. J. Eng. Sci. Technol.*, vol. 4, p. 5, 2012.

[7] A. Gupta and A. Kumar, "A new method for solving linear multi-objective transportation problems with fuzzy parameters," *Appl. Math. Model.*, vol. 36, no. 4, pp. 1421–1430, 2012, doi: 10.1016/j.apm.2011.08.044.

[8] Sulistiana and M. A. Muslim, "Support Vector Machine (SVM) optimization using grid search and unigram to improve e-commerce review accuracy," *J. Soft Comput. Explor.*, vol. 1, no. 1, pp. 8–15, 2020.

[9] D. A. Adeniyi, Z. Wei, and Y. Yongquan, "Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method," *Appl. Comput. Informatics*, vol. 12, no. 1, pp. 90–108, 2016, doi: 10.1016/j.aci.2014.10.001.

[10] H. Shahabi, A. Shirzadi. A. Ghaderi and E. Omidvar, "Flood detection and susceptibility mapping using Sentinel-1 remote sensing data and a machine learning approach: Hybrid intelligence of bagging ensemble based on K-Nearest Neighbor classifier," *Remote Sens.*, vol. 12, no. 2, 2020, doi: 10.3390/rs12020266.

[11] Z. Pan, Y. Wang, and Y. Pan, "A new locally adaptive k-nearest neighbor algorithm based on discrimination class," *Knowledge-Based Syst.*, vol. 204, 2020, doi: 10.1016/j.knosys.2020.106185.

[12] X. Sun, S. Su, Z. Zuo, X. Guo, and X. Tan, "Modulation classification using compressed sensing and decision tree–support vector machine in cognitive radio system," *Sensors (Switzerland)*, vol. 20, no. 5, 2020, doi: 10.3390/s20051438.

[13] T. Lajnef *et al.*, "Learning machines and sleeping brains: Automatic sleep stage classification using decision-tree multi-class support vector machines," *J. Neurosci. Methods*, vol. 250, pp. 94–105, 2015, doi: 10.1016/j.jneumeth.2015.01.022.

[14] M. Ghiassi, M. Olschimke, B. Moon, and P. Arnaudo, "Automated text classification using a dynamic artificial neural network model," *Expert Syst. Appl.*, vol. 39, no. 12, pp. 10967–10976, 2012, doi: 10.1016/j.eswa.2012.03.027.

[15] Y. N. Ifriza, C. E. Edi, and J. E. Suseno, "Expert system irrigation management of agricultural reservoir system using analytical hierarchy process (AHP) and forward chaining method," *Int. Conf. on Math. Scien. and Edu.*, vol. 4. no.1. pp. 74–83, 2017.

[16] R. Talebnejad and A. R. Sepaskhah, "Effects of water-saving irrigation and groundwater depth on direct seeding rice growth, yield, and water use in a semi-arid region," *Arch. Agron. Soil Sci.*, vol. 60, no. 1, pp. 15–31, 2014, doi: 10.1080/03650340.2013.770598.

[17] N. Pirmoradian and A. R. Sepaskhah, "A very simple model for yield prediction of rice under different water and nitrogen applications," *Biosyst. Eng.*, vol. 93, no. 1, pp. 25–34, 2006, doi: 10.1016/j.biosystemseng.2005.09.004.

[18] S. Tohidyan Far and K. Rezaei-Moghaddam, "Determinants of Iranian agricultural consultants' intentions toward precision agriculture: Integrating innovativeness to the technology acceptance model," *J. Saudi Soc. Agric. Sci.*, vol. 16, no. 3, pp. 280–286, 2017, doi: 10.1016/j.jssas.2015.09.003.

[19] N. Gopika and A. E. A. Meena Kowshalaya, "Correlation based feature selection algorithm for machine learning," *Proc. 3rd Int. Conf. Commun. Electron. Syst. ICCES 2018*, pp. 692–695, 2018, doi: 10.1109/CESYS.2018.8723980.

[20] D. Aprilianto, "SVM Optimization with Correlation feature selection based binary particle swarm optimization for diagnosis of chronic kidney disease," *J. Soft Comput. Explor.*, vol. 1, no.1, pp. 24–31, 2020.

[21] R. A., E. S. -, A. A., and B. Ibrahim, "Modifications of Particle Swarm Optimization techniques and its application on stock market: a survey," *Int. J. Adv. Comput. Sci. Appl.*, vol. 6, no. 3, 2015, doi: 10.14569/ijacsa.2015.060315.

[22] R. H. Saputra and B. Prasetyo, "Improve the accuracy of c4.5 algorithm using particle swarm optimization (pso) feature selection and bagging technique in breast cancer diagnosis," *J. Soft Comput. Explor.*, vol. 1, no.1, pp. 47–55, 2020.

[23] N. A. Hitam and A. R. Ismail, "Comparative performance of machine learning algorithms for cryptocurrency forecasting," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 11, no. 3, pp. 1121–1128, 2018, doi: 10.11591/ijeecs.v11.i3.pp1121-1128.

[24] Y. Tang and J. Zhou, "The performance of PSO-SVM in inflation forecasting," *2015 12th Int. Conf. Serv. Syst. Serv. Manag. ICSSSM 2015*, 2015, doi: 10.1109/ICSSSM.2015.7170251.

[25] J. Nayak, B. Naik, and H. S. Behera, "A Comprehensive survey on support vector machine in data mining tasks: applications & challenges," *Int. J. Database Theory Appl.*, vol. 8, no. 1, pp. 169–186, 2015, doi: 10.14257/ijdta.2015.8.1.18.

[26] B. Ghaddar and J. Naoum-Sawaya, "High dimensional data classification and feature selection using support vector machines," *Eur.J. Oper. Res.*, vol. 265, no. 3, pp. 993–1004, 2018, doi: 10.1016/j.ejor.2017.08.040.

[27] D. K. Srivastava and L. Bhambhu, "Data classification using support vector machine," *J. Theor. Appl. Inf. Technol.*, vol. 12, no. 1, pp. 1–7, 2010.

[28] J. Dou and A.P. Yunus., "Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed, Japan," *Landslides*, vol. 17, no. 3, pp. 641–658, 2020, doi: 10.1007/s10346-019-01286-5.

Journal of Soft Computing Exploration

Comparation analysis of naïve bayes and decision tree C4.5 for caesarean section prediction

I Gusti Ayu Suciningsih[1], Muhammad Arif Hidayat [2], Renita Arianti Hapsari[3]
[12,3]Department of Computer Science, Universitas Negeri Semarang, Indonesia

Article Info

Article history:

Received Jan 8, 2021
Revised Feb 18, 2021
Accepted March 12, 2021

Keywords:

Caesar section
Pregnancy
C4.5
Naïve Bayes

ABSTRACT

The development of technology can be used to facilitate many matters. One of them is childbirth in the medical fields. Maternal mortality rate (MMR) is the number of maternal deaths during pregnancy to postpartum caused by pregnancy, childbirth or its management. There are several methods of labors that can be done. The determination of the labor is based on many factors and must be in accordance with the conditions of pregnant patient. Caesarean birth is the last alternative in labor, due to high risk factors. The objective of this research is to predicte and analyse caesarean section using C4.5 and Naïve Bayes classifier models. For experimentation the dataset is collected from UCI Machine Learning Repository and the main attributes represented in this dataset are age, delivery number, delivery time, blood of pressure, and heart problem. The accuracy using C4.5 by 80 training cases is 45% And the accuracy using Naïve Bayes is 50%.

Corresponding Author:

I Gusti Ayu Suciningsih,
Department of Computer Science,
Universitas Negeri Semarang,
Sekaran, Gunungpati, Semarang, Indonesia.
Email: suciningsihayu@gmail.com

1. INTRODUCTION

Maternal mortality rate (MMR) in Indonesia is still high. MMR represents number of maternal deaths during lifetime pregnancy until the post-term childbirth caused by pregnancy, childbirth and the puerperium or the management and not caused by accident or fell on every 100,000 live births [1]. Caesarean section is the last alternative in action labor. This is due to high risk factors, both risk for mother and babies [2]. Despite the high risk, numbers of caesarean birth experienced increase significantly, particularly in Indonesia. World Health Organization (WHO) set the standard for caesar section delivery in a country about 5-15 percent per thousand births in the world. Based on WHO data, in 2004-2008 in three continents (Latin America, Africa, and Asia) the lowest Caesarean birth rate was in Angola (2.3%) and the highest in China (46.2%). Caesar births data in Indonesia has increased sharply, especially in big cities. Lowest rate in Southeast Sulawesi (5.5%) and the highest in DKI Jakarta (27.2%) [3].

Information technology will continue to develop and needed to meet the needs of fast and accurate information for life [4]. Technology has been used in various fields, for example in the health sector [5]. At this time the health sector has been supported by technology that is able to visualize and predict a patient's condition. From existing patient data, it can be used as material to classify a patient's condition using technology. One area that requires classification of a patient's condition is a childbirth [6]. Based on the explanation above, it is necessaary to have an algorithm that can support the work of medical personnel in determining the type of labor [1]. Classification is one of the methods contained in data mining [7]. Classification is necessary to find patterns in order to be able to produce correct predictions even in critical conditions [8]. To perform the classification process, there are several algorithm that can be used including

Support Vector Machine (SVM), Naïve Bayes, K-Nearest Neighbor (KNN), Decission Tree, and Artificial Neural Network (ANN). These methods have their own level of accuracy for each object to be classified. The methods that will be used in this project are comparation of the Decision Tree C4.5 and Naïve Bayes methods to classify the caesarean section.

2. METHOD

2.1 Application of Naïve Bayes Algorithm

This classifier is based on the Naïve Bayes Theorem, which gives a way to estimate the posterior probability. Posterior probability of a class gives the estimation of an item belonging to that class based on the given attributes. Naïve bayes is the simplest calculation of the Bayes theorem, because it is able to reduce computational complexity to simple multiplication of probability [9]. Apart from that, the Naïve Bayes algorithm is also capable of handling data sets which has many attributes.

The application of Caesarean Section data set on Naïve Bayes algorithm process as follows:

Prepare caesarean section data set.
Classifying using Naïve Bayes algorithm.
Count the number of classes or labels in the data set.
Count the number of cases on each class.
Multiply all the class variables.
Compare the results of each classes.
The following is the equation of the Naïve Bayes:

$P(H|X) = (P(X \mid H)P(H))/(P(X))$

In wich:

X : data or tuple object (class C)
H: : hypothesis
P(H|X) : probability that hypothesis H is in condition
P(H) : prior probability that the H hypothesis is valid (true)
P(X) : prior probability of tuple X.

2.2 Application of Decision Tree C4.5 Algorithm

The C4.5 algorithm [10] is used in Data Mining as a Decision Tree Classifier [11] which can be employed to generate a decision, based on a certain sample of data (univariate or multivariate predictors). The following is the application of the research Decision Tree C4.5 algorithm [12].

1. Determine the root of the tree.
2. Calculate entropy for the classes
3. Calculate entropy after split each attribute
4. Calculate information gain for each split
5. Perform the split
6. Perform further splits
7. Complete the decision tree

For choosing attribute as a root, based on the highest gain value of the exsting attributes. To calculate gain, a formula is used as shown in the equation:

Gain(S,A) = Entropy(S) 〖- Σ〗 _(i=1)^n (|S_i |)/(|S|) x Entropy(S_i)

In which:

S : case set
A : attribute
n : number of partitions attribute A
|S_i| : number of cases of i partitions
|S| : number of cases in S

2.3 Dataset Attribute Information

The dataset attribute on information can be seen at Table 1.

Table 1. Dataset attribute information

Attributes	Type	Description
Age	Integer	Age in years
Delivery Number	Integer	Birth stage
Delivery Time	Integer (0,1,2)	0 = Timely, 1= Premature, or 2 = Latecomer
Blood of Pressure	Integer (0,1,2)	0 = low, 1 = normal, or 2 = high
Heart Problem	Integer (0,1)	0 = apt, 1 = inept
Caesarean	Integer (0,1)	Whether patient is allowed to caesarean delivery. 0 = No or 1 = Yes

3. RESULT AND DISCUSSION

3.1 Sampel Data

The data isconsidered in ARFF format. The following gives the name of relation, name of attributes and sample instances in the given data set.
@attribute 'Age' {22,26,28,27,32,36,33,23,20,29,25,37,24,18,30,40,31,19,21,35,17, 38}
@attribute 'Age' {22, 26, 28, 27, 32, 36, 33, 23, 20, 29, 25, 37, 24, 18, 30, 40, 31, 19, 21, 35, 17, 38}, has 22 distinct values with a maximum value 40 and minimum value 17.
@attribute 'Delivery number' {1, 2, 3, 4}, considered up to the first four deliveries.
@attribute 'Delivery time' {0, 1, 2}, premature and late deliveries are taken into consideration.
@attribute 'Blood of Pressure' {2, 1, 0}, various blood pressure moods are noted at the timeof delivery.
@attribute 'Heart Problem' {1, 0}, heart response is apt or inapt.
@attribute Cesarean {0, 1}, a class attribute whether cesarean section delivery or not.

3.2 Cleaning Data

Cleaning data is checked on the dataset, if there is a missing value in the dataset, treatment must be given to the data [13]. In the dataset used for this study, there are no missing values as shown in Figure 1 which show a dataset of caesarean section, because there are no missing value then we can go to the next step .

	id	Age	Dev_number	Dev_time	Blood_preassure	Heart_problem	Caesarian
0	1	22	1	0	2	0	0
1	2	26	2	0	1	0	1
2	3	26	2	1	1	0	0
3	4	28	1	0	2	0	0
4	5	22	2	0	1	0	1

Figure 1. The dataset that shown in google colab

3.3 Determining Independet Variables and Dependent Variables

The dependent variable used here is the caesarian variable, because we want to see whether the patient is classified as caesar labor or normal labor. The other varible that are age, delivery number, delivery time, blood pressure, and heart problem became an independent variable, can be seen at figure 2.

```
[ ] # Variabel independen
    x = caesarian.drop(["Caesarian","id"], axis = 1)
    x.head()
```

	Age	Dev_number	Dev_time	Blood_preassure	Heart_problem
0	22	1	0	2	0
1	26	2	0	1	0
2	26	2	1	1	0
3	28	1	0	2	0
4	22	2	0	1	0

```
[ ] # Variabel dependen
    y = caesarian["Caesarian"]
    y.head()

0    0
1    1
2    0
3    0
4    1
Name: Caesarian, dtype: int64
```

Figure 2. The table of independent and dependent variable

3.4 Normalization

Normalization is rescaling real numeric attributes into range 0 and 1. That in the dataset there is data with values other than 0 and 2, then the normalization stage will be carried out so the data becomes values in the range 0 and 1.

```
[ ] from sklearn import preprocessing
    x1 = preprocessing.normalize(x)
    print (x1)

    [[0.99487439 0.04522156 0.         0.09044313 0.        ]
     [0.99632216 0.07664017 0.         0.03832008 0.        ]
     [0.99559146 0.07658396 0.03829198 0.03829198 0.        ]
     [0.9968264  0.03560094 0.         0.07120189 0.        ]
     [0.99487439 0.09044313 0.         0.04522156 0.        ]
     [0.99852398 0.03840477 0.03840477 0.         0.        ]
     [0.99658819 0.07382135 0.         0.03691067 0.        ]
     [0.99515266 0.09329556 0.         0.03109852 0.        ]
     [0.9968264  0.07120189 0.         0.03560094 0.        ]
     [0.99794872 0.03696106 0.03696106 0.03696106 0.        ]
     [0.99922929 0.02775637 0.         0.02775637 0.        ]
     [0.99908299 0.03027524 0.03027524 0.         0.        ]
     [0.99717646 0.0433555  0.0433555  0.0433555  0.        ]
     [0.99627096 0.04981355 0.         0.04981355 0.04981355]
     [0.99645179 0.03436041 0.06872081 0.         0.03436041]
     [0.99602384 0.03984095 0.07968191 0.         0.        ]
     [0.99840383 0.03993615 0.         0.03993615 0.        ]
     [0.98893635 0.04944682 0.09889364 0.09889364 0.        ]
     [0.99600652 0.08075729 0.         0.0269191  0.0269191 ]
     [0.99483201 0.04145133 0.08290267 0.         0.04145133]
     [0.99778842 0.03837648 0.03837648 0.03837648 0.        ]
     [0.99771219 0.06046741 0.         0.         0.0302337 ]
     [0.99523429 0.03980937 0.03980937 0.07961874 0.        ]
     [0.99863107 0.03698634 0.         0.         0.03698634]
     [0.99258333 0.04962917 0.         0.09925833 0.04962917]
     [0.9969278  0.05538488 0.         0.05538488 0.        ]
     [0.98936948 0.05496497 0.05496497 0.10992994 0.05496497]
     [0.99889074 0.03329636 0.         0.03329636 0.        ]
     [0.99708312 0.03115885 0.         0.0623177  0.03115885]
     [0.99486235 0.07652787 0.03826394 0.03826394 0.03826394]
     [0.99920096 0.03996804 0.         0.         0.        ]
     [0.99906382 0.0249766  0.         0.0249766  0.0249766 ]
     [0.99563423 0.06222714 0.         0.06222714 0.03111357]
     [0.99591    0.07377111 0.         0.03688556 0.03688556]
     [0.99340894 0.07641607 0.07641607 0.03820804 0.        ]
```

Figure 3. The stage of normalization

3.5 Data Testing and Data Training

The classification using naïve bayes is contained in the sklearn package [6]. In this classification, testing data and training data are needed. Dividing the data set into Data Testing and Data training aims to adjust the data set into the Algorithm model. Divided by the ratio of Data Training 75% and Data Testing 25%. Training data with random state is 123. The random state value is independent, the random state shows how many times the data is randomized. However, this time using 123 so that the random results we get are the same.

3.6 Calculate the Probability Value and the Predicted Results

```
[ ] # Menentukan probabilitas hasil prediksi
    nbtrain.predict_proba(x_test)

    array([[0.28134695, 0.71865305],
           [0.73095067, 0.26904933],
           [0.15167084, 0.84832916],
           [0.46259235, 0.53740765],
           [0.39529401, 0.60470599],
           [0.70104384, 0.29895616],
           [0.49926902, 0.50073098],
           [0.62370872, 0.37629128],
           [0.66441541, 0.33558459],
           [0.80417774, 0.19582226],
           [0.83222515, 0.16777485],
           [0.68946998, 0.31053002],
           [0.65620501, 0.34379499],
           [0.07948917, 0.92051083],
           [0.66441541, 0.33558459],
           [0.4562007 , 0.5437993 ],
           [0.53810831, 0.46189169],
           [0.61264038, 0.38735962],
           [0.03816595, 0.96183405],
           [0.10885903, 0.89114097]])
```

Figure 4. The result of probability values

The results that seen at Figure 4. For the example, the first data is 0.71 is rounded to 1, the second data is 0.26 is rounded to 0, and so on.

3.7 Confusion of Matrix

In figure 2. We can know that there are 5 pregnant women who are predicted to have normal labor and in actual circumstances do deliver normal. Meanwhile, the number of pregnant women who are predicted to have normal labor but in actual fact give birth by caesarean section is also 6. Then, there were 5 pregnant women who were predicted to give birth by caesarean section and in actual fact they gave birth by caesarean section. Meanwhile, there were 4 pregnant women who were predicted to give birth by caesarean section but in actual circumstances gave birth normally, the result can be seen at Figure 5.

prediction	0	1
actual		
0	5	4
1	6	5

Figure 5. The result confusion of matrix

3.8 Memory Usage

In the decision tree models memory that been used is 111,57 MB, shown in Figure 6.

```
[ ]  import os, psutil
     process = psutil.Process(os.getpid())
     print(process.memory_info().rss)  # in bytes

     111575040
```

Figure 6. The result of usage memory by Decision Tree C4.5

And in the naïve bayes algorithm, memory that been used is 111,70 MB, shown in Figure 7.

```
[ ]  import os, psutil
     process = psutil.Process(os.getpid())
     print(process.memory_info().rss)  # in bytes

     111706112
```

Figure 7. The result of usage memory by Naïve Bayes

3.9 Measure the Program Execution Time

In measuring the program execution time using a decision tree algorithm, the results are displayed for 0.013 seconds, shown in Figure 8.

```
[ ]  import timeit
     code_to_test = """
     a = range(100000)
     b = []
     for i in a:
         b.append(i*2)
     """
     elapsed_time = timeit.timeit(code_to_test, number=100)/100
     print(elapsed_time)

     0.013129985089999536
```

Figure 8. The result of execution time by Decision Tree C4.5

And in the execution time using naïve bayes algorithm the results are displayed for 0.012 seconds which means by using naïve bayes algorithm the execution time is faster, shown in Figure 9.

```
[ ]  a = range(100000)
     b = []
     for i in a:
         b.append(i*2)
```

```
[ ]  import timeit
     code_to_test = """
     a = range(100000)
     b = []
     for i in a:
         b.append(i*2)
     """
     elapsed_time = timeit.timeit(code_to_test, number=100)/100
     print(elapsed_time)

     0.01271887993999826
```

Figure 9. The result of execution time by Naïve Bayes Algorithm

3.10 Level of Accuracy

Decision tree models are created using 2 steps: Induction and Pruning. Induction is where we actually build the tree i.e set all of the hierarchial decision boundaries based on our data. Because of the nature of training decision tree they can be prone to mjor overfitting. Pruning is the process of removing the unnecessary structure from a decision tree, effectively reducing the complexity to combat overfitting with the added bonus of making it even easier to interpet. By using this method, the result of accuracy are shown in figure 10.

```
[ ]  from sklearn.tree import DecisionTreeClassifier
     from sklearn.metrics import accuracy_score
     from sklearn import tree
     clf_gini = DecisionTreeClassifier(criterion = "gini", random_state= 123, max_depth=3,min_sampl
     clf_gini.fit(x_train,y_train)
     y_pred = clf_gini.predict(x_test)
     from sklearn import metrics
     metrics.accuracy_score(y_test, y_pred)*100

     45.0
```

Figure 10. The result of accuracy in implementation by Decision Tree C4.5

After getting the Naïve Bayes algorithm classification model, calculate the accuracy using a confusion matrix. Naïve Bayes classification algorithm will produce better results if using more training data. The results of the accuracy in the Naïve Bayes classification are shown in Figure 11.

	precision	recall	f1-score	support
0	0.45	0.56	0.50	9
1	0.56	0.45	0.50	11
accuracy			0.50	20
macro avg	0.51	0.51	0.50	20
weighted avg	0.51	0.50	0.50	20

Figure 11. The result of accuracy in implementation by Naïve Bayes

4. CONCLUSION

Using C4.5 and Naïve Bayes classifier models , the result of accuracy are 45% and after getting the Naïve Bayes algorithm classification model, calculate the accuracy using a confusion matrix. Naïve Bayes classification algorithm will produce better results if using more training data. The results of the accuracy in the Naïve Bayes classification are 50%. So the level of accuracy using the Naïve Bayes method is greater or more accurate than the decision tree method.

REFERENCES

[1] H. Manik, M. F. G. Siregar, R. Kintoko Rochadi, E. Sudaryati, I. Yustina, and R. S. Triyoga, "Maternal mortality classification for health promotive in Dairi using machine learning approach," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 851, no. 1, 2020, doi: 10.1088/1757-899X/851/1/012055.

[2] Anggorowati and N. Sudiharjani, "Mobilisasi Dini dan Penyembuhan Luka Operasi Pada Ibu Post Sectio Caesarea (SC) di Ruang Dahlia Rumah Sakit Umum Daerah Kota Salatiga," *Pros. Semin. Nas.*

dan Int. Univ. Muhammadiyah Semarang, pp. 30–35, 2010, [Online]. Available: https://jurnal.unimus.ac.id/index.php/psn12012010/article/viewFile/1281/1334.

[3] L. Andayasari *et al.*, "Proporsi seksio sesarea dan faktor yang berhubungan dengan seksio sesarea di Proporsi Seksio Sesarea dan Faktor yang Berhubungan dengan Seksio Sesarea di Jakarta THE PROPORTION OF CAESAREAN SECTION AND ASSOCIATED FACTORS IN HOSPITAL OF JAKARTA," pp. 6–16, 2014.

[4] N. R. Indraswari and Y. I. Kurniawan, "Aplikasi Prediksi Usia Kelahiran Dengan Metode Naive Bayes," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 9, no. 1, pp. 129–138, 2018, doi: 10.24176/simet.v9i1.1827.

[5] P. Radha and B. Srinivasan, "Predicting Diabetes by cosequencing the various Data Mining Classification Techniques," vol. 1, no. 6, pp. 334–339, 2014.

[6] A. Kamat, V. Oswal, and M. Datar, "Implementation of Classification Algorithms to Predict Mode of Delivery," *Int. J. Comput. Sci. Inf. Technol.*, vol. 6, no. 5, pp. 4531–4534, 2015.

[7] R. H. Saputra and B. Prasetyo, "Improve the Accuracy of C4 . 5 Algorithm Using Particle Swarm Optimization (PSO) Feature Selection and Bagging Technique in Breast Cancer Diagnosis," pp. 47–55, 2020.

[8] M. W. L. Moreira, J. J. P. C. Rodrigues, A. M. B. Oliveira, K. Saleem, and A. V. Neto, "An inference mechanism using Bayes-based classifiers in pregnancy care," *2016 IEEE 18th Int. Conf. e-Health Networking, Appl. Serv. Heal. 2016*, no. Dm, pp. 0–4, 2016, doi: 10.1109/HealthCom.2016.7749475.

[9] I. E. Tiffani, "Optimization of Naïve Bayes Classifier By Implemented Unigram , Bigram , Trigram for Sentiment Analysis of Hotel Review," pp. 1–7, 2020.

[10] A. Wibowo, D. Manongga, and H. D. Purnomo, "The Utilization of Naive Bayes and C.45 in Predicting The Timeliness of Students' Graduation," *Sci. J. Informatics*, vol. 7, no. 1, pp. 99–112, 2020, doi: 10.15294/sji.v7i1.24241.

[11] A. De Ramón Fernández, D. Ruiz Fernández, and M. T. Prieto Sánchez, "A decision support system for predicting the treatment of ectopic pregnancies," *Int. J. Med. Inform.*, vol. 129, pp. 198–204, 2019, doi: 10.1016/j.ijmedinf.2019.06.002.

[12] E. Fitriani, "Perbandingan Algoritma C4.5 Dan Naïve Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan," *Sistemasi*, vol. 9, no. 1, p. 103, 2020, doi: 10.32520/stmsi.v9i1.596.

[13] X. Chu, I. F. Ilyas, S. Krishnan, and J. Wang, "Data cleaning: Overview and emerging challenges," *Proc. ACM SIGMOD Int. Conf. Manag. Data*, vol. 26-June-20, pp. 2201–2206, 2016, doi: 10.1145/2882903.2912574.

Journal of Soft Computing Exploration

Accuracy of classification poisonous or edible of mushroom using naïve bayes and K-nearest neighbors

Roni Hamonangan[1], Meidika Bagus Saputro[2], Cecep Bagus Surya Dinata Karta Atmaja[3]
[12,3]Department of Computer Science, Universitas Negeri Semarang, Indonesia

Article Info

Article history:

Received Jan 8, 2021
Revised Feb 18, 2021
Accepted March 12, 2021

Keywords:

Accuracy
Mushroom Classification
Naïve Bayes
K-Nearest Neighbors

ABSTRACT

Mushrooms are plants that are widely consumed by the general public, but not all mushrooms can be consumed directly, because the types of mushrooms are feasible and it is still too difficult to distinguish, then there are several ways to identify fungi, namely by means of morphology. The morphology referred to in this paper is the morphology of fungi which includes color, habitat, class, and others. We got the morphology of this mushroom from a datasets we get from UCI Machine Learning with the 23 atribut that we use in the program. In determining the classification of this fungus we use the Naive Bayes algorithm which produces an accuracy of around 90,2% which we then improve again so that it reaches 100% accuracy using the K-Nearest Neighbors algorithm. Furthermore, in this case to prove accuracy that we had before, we use calculation accuracy with confusion matrix to show it the accuracy of classification poisonous or edible mushroom.

Corresponding Author:

Roni Hamonangan,
Department of Computer Science,
Universitas Negeri Semarang,
Sekaran, Gunungpati, Semarang, Indonesia.
Email: ronihamonangan93@students.unnes.ac.id

1. INTRODUCTION

Fungi belong to the fungal kingdom, therefore mushrooms do not have true leaves and roots, and do not have chlorophyll so they cannot carry out photosynthesis like plants in general. Fungi are classified or classified separately because they cannot be classified in plants or animals. There are fungi that can be seen directly or are macroscopic and some must be observed using a microscope or microscopic shape. In general, fungi have many cells (multicellular) such as edible mushrooms and tempeh mushrooms, but some are single-celled (unicellular) such as yeast or yeast (Saccharomyces). Multicellular fungi are composed of threads called hyphae. When viewed with a microscope, hyphae have a separating form (septa) and some are not partitioned [1].

A mushroom is one of the fungi types' food that has the most potent nutrients on the plant. Mushrooms have major advantages such as kill cancer cells, viruses and enhancing the human immune system. Currently, the mushroom refers to the process that performed by robot in food industry. This technique used to limit the features such as color. Recently, mushroom system used specific characteristics that improve the selection process of mushrooms. Such system depends on analyzing and investigating the features in order to get better classification based on the well-known features [2].

To identify which mushrooms are edible and poisonous, there are several ways that can be used. One of the aspects that can be used as benchmarks in identifying a fungus is its morphological characteristics. The morphological features referred to are the shape of the umbrella, color, habitat, and other features visible to our eyes. We obtained these morphological characteristics from the datasets we took from UCI Machine Learning [3].

Datasets is a collection of data. In the case of tabular data, a data set corresponds to one or more database tables, where every column of a table represents a particular variable, and each row corresponds to a given record of the data set in question. The data set lists values for each of the variables, such as height and weight of an object, for each member of the data set [4].

In this case we used two methods to determine the classification of mushrooms, namely the Naive Bayes method and also the K-Nearest Neighbor method as the classifier [5]. We use these two methods because they have different accuracy and we can compare it with the method which we get better accuracy of the two methods that have been tested. Extraction of morphological features is used to help identify fungi, so that later it will be known including the types of edible or poisonous mushrooms.

The accurate accuracy of the model can classify correctly [6]. Thus, the accuracy of the ratio of the predictions is correct (positive and negative) to the total data [7]. In other words, accuracy is the level of closeness of the predicted value to the actual (actual) value.

2. METHOD

In this study case the analysis will be carried out to find the best accuracy in determining the classification of fungi using two classification algorithms. The proposed algorithm is Naive Bayes algorithm [8] and K-Nearest Neighbor (KNN) algorithm [9], then evaluates and validates the results by looking for the best accuracy results of these two algorithms. The next stage is compare the results of the accuracy of each algorithm, to get a model classification algorithm that obtains the highest accuracy and time complexity [10]. The highest accuracy results from this calculation can be said to be the best algorithm in determining the classification of poisonous or edible mushrooms. And then we use calculation accuracy to test again the result of the accuracy that we get before.

For application in this case we will using Naive Bayes algorithm and K-Nearest Neighbor first [11]. Next we calculate using the Naive Bayes algorithm to get its accuracy, then we continue using the K-Nearest Neighbor algorithm to get better accuracy than the previous algorithm. The following is the flowchart that we produce according to the classification results of mushrooms using the Naive Bayes algorithm and the K-Nearest Neighbors algorithm, shown in Figure 1.

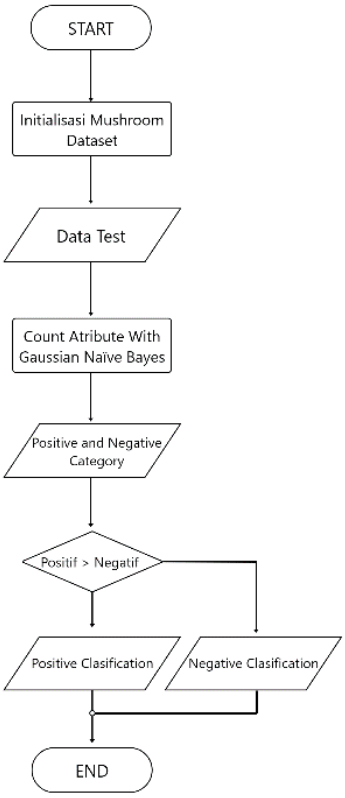


Figure 1. Flowchart of naïve bayes algorithm

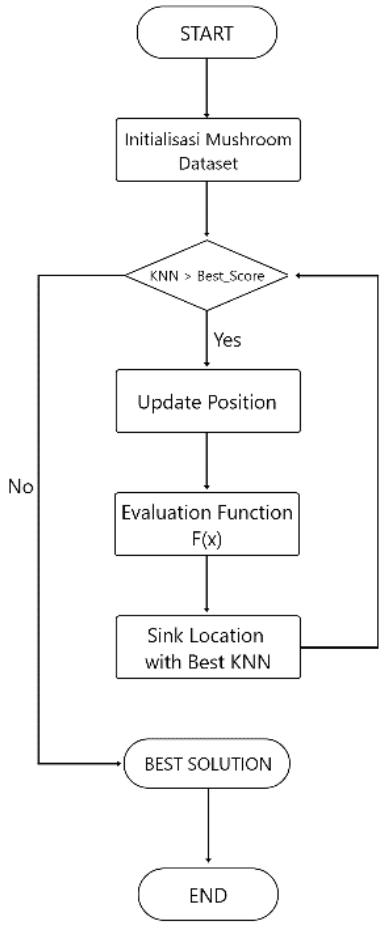


Figure 2. Flowchart of KNN algorithm

2.1 Data Collection

The data used is mushroom data obtained from the site archive.ics.uci.edu (UCI Machine Learning Repository) [12]. This data was donated by Jeffrey Schlimmer in 1987. In this study there are 2 classes, namely food mushrooms and poisonous mushrooms. For the number of each class, consisting of 4208 data included in the food mushroom category and 3916 data included in the poisonous mushroom category, so that the total number of data used was 8124 data. Each initial of each attribute and class is a representation of the type of attribute concerned. The feature extraction used in the training data is morphological features. The morphologicals that we use to calculate the accuracy of this classifocation poisonous or edible mushroom can show in the dataset that we have form UCI Machine Learning shown by the following this figure 3.

	class	cap-shape	cap-surface	cap-color	bruises	odor	gill-attachment	gill-spacing	gill-size	gill-color	stalk-shape	stalk-root
0	p	x	s	n	t	p	f	c	n	k	e	e
1	e	x	s	y	t	a	f	c	b	k	e	c
2	e	b	s	w	t	l	f	c	b	n	e	c
3	p	x	y	w	t	p	f	c	n	n	e	e
4	e	x	s	g	f	n	f	w	b	k	t	e

stalk-surface-above-ring	stalk-surface-below-ring	stalk-color-above-ring	stalk-color-below-ring	veil-type	veil-color	ring-number	ring-type	spore-print-color	population	habitat
s	s	w	w	p	w	o	p	k	s	u
s	s	w	w	p	w	o	p	n	n	g
s	s	w	w	p	w	o	p	n	n	m
s	s	w	w	p	w	o	p	k	s	u
s	s	w	w	p	w	o	e	n	a	g

Figure 3. Datasets mushroom classification

2.2 Data Processing

In this mushroom classification process data, we use two algorithms, namely naive bayes and k-nearest neighbors to get the best accuracy in this classification. Here we use 23 attributes according to the datasets we use [13]. The attributes that we used in datasets there are class, cap shape, cap surface, cap color, bruises, odor, gill attachment, gill spacing, gill size, gill color, stalk root, stalk surface above ring, stalk surface below ring, stalk surface above ring, stalk surface below ring, veil type, veil color, ring number, ring type, spore print color, population, and last habitat of mushroom.

2.3 The Algorithm Used

2.3.1. Naïve Bayes

Bayes' theorem is a statistical calculation by calculating the probability of the similarity of an existing old case on a case basis with a new case [5]. Bayes' theorem has a high degree of accuracy and good speed when applied to large databases. Naive Bayes is included in supervised learning, so that at the learning stage, initial data is needed in the form of training data to be able to make decisions. At the classification stage, the probability value of each class label that is available for the input will be calculated. The class label that has the greatest probability value will be used as the label for the input data class. Naive Bayes is the simplest calculation of the Bayes theorem, because it is able to reduce computational complexity to a simple multiplication of probability. In addition, the Naive Bayes algorithm is also able to handle data sets that have many attributes.

The application of Caesarean Section data set on Naïve Bayes algorithm process as follows:

1. Prepare mushroom classification data set.
2. Classifying using Naïve Bayes algorithm.
3. Count the number of classes or labels in the data set.
4. Count the number of cases on each class.
5. Multiply all the class variables.
6. Compare the results of each classes.
7. The following is the equation of the Naïve Bayes:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

In wich:

X : data or tuple object (class C)

H: : hypothesis
P(H|X) : probability that hypothesis H is in condition
P(H) : prior probability that the H hypothesis is valid (true)
P(X) : prior probability of tuple X

2.3.2. K-Nearest Neighbors

K-Nearest Neighbor or often abbreviated as KNN is one of the algorithms used to classify objects based on learning data (training data) which is the closest distance to the object [14]. The purpose of the KNN algorithm is to classify new objects based on attributes and samples from training data.

KNN is a supervised learning algorithm, which means that this algorithm uses existing data and the output is known. KNN is widely used in data mining, pattern recognition, image processing, etc.

1. Specifies the parameter K as the number of neighbors closest to the new object.
2. Calculate the distance between new objects / data against all objects / data that have been trained.
3. Sort the results of these calculations.
4. Determine the closest neighbor based on the minimum distance to K.
5. Determine the category of the closest neighbor to the object / data.
6. Use majority category as new object / data classification.

2.4 Calculation Accuracy

The method used to determine the final accuracy oftests performed is the confusion matrix for the multi-class method. This method is used to perform system calculations with many prediction classes [7]. The difference from the multi-class confusion matrix with the ordinary confusion matrix is that the final results are calculated cumulative accuracy of the overall accuracy of all test data. Parameters of the accuracy are presented in Table 1.

Table 1.Confusion matrix

Classification	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

where:
True Negative(TN) : if the prediction and actual results are negative
False Negative (FN) : if the positive prediction results, and the actual results negative
False Positive (FP) : if the negative prediction results, and the actual results positive
True Positive (TP) : if the predictive and actual results are positive.Calculation of the total accuracy of the tests performed using the following formula.

$$\text{Total Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (2)$$

3. RESULT AND DISCUSSION

To make this mushroom classification we use a software that is Google Colabs using python language. In determining this accuracy, we used the dataset we took from UCI Machine Learning, which was donated by Jeff Schlimmer in 1987. In this dataset, there are 23 attributes to calculate the accuracy of mushroom classification and with a lot of data, 8124 data.

3.1 Classification Mushroom with Naïve Bayes Algorithm

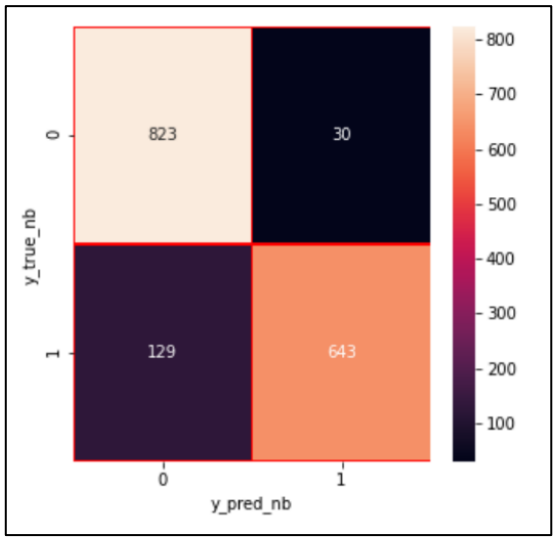


Figure 4. Confusion matrix of Accuracy Mushroom Classification with Naïve bayes

In Figure 3. It can be seen that the accuracy obtained using the Naive Bayes method is 90.21%. This accuracy is obtained from the import library in the form of GaussianNB to calculate the classification of this fungus. However, the calculation of the Naive Bayes method does not produce great accuracy results.

From the previous calculations that we have obtained before, we can prove it with the results we get the accuracy is 90,2%. To get the prove, one way to prove it is by using confusion matrix table. to show the confusion matrix table we can call the Naïve bayes values to the program and show the result to be output of this confusion matrix. The following is a calculation according to the confusion matrix that we can prove:

Accuracy =(TP+TN)/(TP+TN+FP+FN) × 100%
= (823+643)/(823+643+129+30) × 100%
=1466/1625 × 100%
= 0,9021 × 100%
= 90,2%

Calculation of accuracy in this study using a confusion matrix has been proven because we can see the result from true positive is 823 and true negative 643. At same time we can see if the value of false positive is 129 and false negative is 30. So, with that calculation we get the accuracy in this method using naïve bayes algorithm is 90,2%.

3.2 Classification Mushroom with K-Nearest Neighbors (KNN)

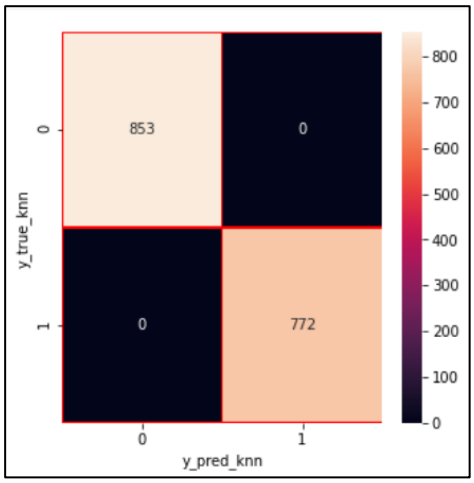


Figure 5. Confusion matrix of accuracy mushroom classification with KNN algorithm

Based on Figure 4, which calculates using the Naive Bayes method with an accuracy of 90.2% we think it has not reached the highest accuracy. Therefore, we tried to re-calculate the accuracy using the K-Nearest Neighbors algorithm and it is true that we found a very good accuracy of 100%. With using this KNN we need an import library, namely K-Nearest Neighbors Classifier. Not only that, we also calculated the best value from this classification of mushrooms by getting the best value, that is k = 1. So it can be concluded that by using the K-Nearest Neighbors algorithm we can produce a classification accuracy of this fungus reaching 100%.

From the previous calculations that we have obtained before, we can prove it with the results we get the accuracy is 100%. To get the prove, one way to prove it is by using confusion matrix table. to show the confusion matrix table we can call the KNN values to the program and show the result to be output of this confusion matrix. The following is a calculation according to the confusion matrix that we can prove:

$$\begin{aligned}\text{Accuracy} &= (TP+TN)/(TP+TN+FP+FN) \times 100\% \\ &= (853+772)/(853+772+0+0) \times 100\% \\ &= 1625/1625 \times 100\% \\ &= 1 \times 100\% \\ &= 100\%\end{aligned}$$

Calculation of accuracy in this study using a confusion matrix has been proven because we can see the result from true positive is 853 and true negative 772. At same time we can see if the value of false positive and false negative is 0. Therefore, to check the accuracy of mushroom classification with K-Nearest Neighbor algorithm is the best way to used that proven with 100% accuracy if we compare with Naïve Bayes algorithm.

4. CONCLUSION

The application of the two algorithms Naive Bayes algorithm and K-Nearest Neighbors produced two different accuracy as well. by using naive bayes algorithm based on our calculation, the results yield 90,2% accuracy. At the same time if we use K-Nearest Neihgbors based on the results of our calculations we produce an accuracy of up to 100%. So the best algorithm to determine the accuracy of the mushroom classification is the K-Nearest Neighbors algorithm.

REFERENCES

[1] H. G. Lewis and M. Brown, "A generalized confusion matrix for assessing area estimates from remotely sensed data," *Int. J. Remote Sens.*, vol. 22, no. 16, pp. 3223–3235, 2001, doi: 10.1080/01431160152558332.

[2] U. Lindequist, T. H. J. Niedermeyer, and W. D. Jülich, "The pharmacological potential of mushrooms," *Evidence-based Complement. Altern. Med.*, vol. 2, no. 3, pp. 285–299, 2005, doi: 10.1093/ecam/neh107.

[3] A. Srivastava, V. Singh, and G. S. Drall, "Sentiment analysis of twitter data: A hybrid approach," *Int. J. Healthc. Inf. Syst. Informatics*, vol. 14, no. 2, pp. 1–16, 2019, doi: 10.4018/IJHISI.2019040101.

[4] Kautsarina, A. N. Hidayanto, B. Anggorojati, Z. Abidin, and K. Phusavat, "Data modeling positive security behavior implementation among smart device users in Indonesia: A partial least squares structural equation modeling approach (PLS-SEM)," *Data Br.*, vol. 30, p. 105588, 2020, doi: 10.1016/j.dib.2020.105588.

[5] R. N. Devita, H. W. Herwanto, and A. P. Wibawa, "Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa indonesia," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, p. 427, 2018, doi: 10.25126/jtiik.201854773.

[6] G. Abdurrahman and J. T. Wijaya, "Analisis Klasifikasi Kelahiran Caesar Menggunakan Algoritma Naive Bayes," *JUSTINDO (Jurnal Sist. dan Teknol. Inf. Indones.*, vol. 4, no. 2, p. 46, 2019, doi: 10.32528/justindo.v4i2.2616.

[7] A. De Ramón Fernández, D. Ruiz Fernández, and M. T. Prieto Sánchez, "A decision support system for predicting the treatment of ectopic pregnancies," *Int. J. Med. Inform.*, vol. 129, pp. 198–204, 2019, doi: 10.1016/j.ijmedinf.2019.06.002.

[8] H. Manik, M. F. G. Siregar, R. Kintoko Rochadi, E. Sudaryati, I. Yustina, and R. S. Triyoga, "Maternal mortality classification for health promotive in Dairi using machine learning approach," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 851, no. 1, 2020, doi: 10.1088/1757-899X/851/1/012055.

[9] S. Sutarti, A. T. Putra, and E. Sugiharti, "Comparison of PCA and 2DPCA Accuracy with K-Nearest Neighbor Classification in Face Image Recognition," *Sci. J. Informatics*, vol. 6, no. 1, pp. 64–72, 2019, doi: 10.15294/sji.v6i1.18553.

[10] P. Radha and B. Srinivasan, "Predicting Diabetes by cosequencing the various Data Mining Classification Techniques," *.IJISET.* vol. 1, no. 6, pp. 334–339, 2014.

[11] N. R. Indraswari and Y. I. Kurniawan, "Aplikasi Prediksi Usia Kelahiran Dengan Metode Naive Bayes," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 9, no. 1, pp. 129–138, 2018, doi: 10.24176/simet.v9i1.1827.

[12] P. Ray and A. Chakrabarti, "Twitter sentiment analysis for product review using lexicon method," *2017 Int. Conf. Data Manag. Anal. Innov. ICDMAI 2017*, pp. 211–216, 2017, doi: 10.1109/ICDMAI.2017.8073512.

[13] A. Al-Thubaity, Q. Alqahtani, and A. Aljandal, "Sentiment lexicon for sentiment analysis of Saudi dialect tweets," *Procedia Comput. Sci.*, vol. 142, pp. 301–307, 2018, doi: 10.1016/j.procs.2018.10.494.

[14] K. Fukunaga and P. M. Narendra, "A Branch and Bound Algorithm for Computing k-Nearest Neighbors," *IEEE Trans. Comput.*, vol. C–24, no. 7, pp. 750–753, 1975, doi: 10.1109/T-C.1975.224297.