Journal of Information Systems Engineering and Management
2025, 10(26s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Hybrid Clustering Using N-Soft Set and Artificial Bee Colony for Digital Literacy

Fatia Fatimah[1], Selly Anastassia Amellia Kharis[1], Andriyansah[2]
[1]*Department of Mathematics, Faculty of Science and Technology, Universitas Terbuka, Banten, Indonesia*
[2]*Department of Management, Faculty of Economics and Business, Universitas Terbuka, Banten, Indonesia*

ARTICLE INFO

Received: 22 Dec 2024

Revised: 14 Feb 2025

Accepted: 25 Feb 2025

ABSTRACT

Introduction: Open and distance education requires a good level of digital literacy. Students in open and distance educatiom come from various ages and backgrounds, resulting in differing levels of literacy. Clustering is needed to identify the digital literacy skills of new students so that digital literacy improvement can be tailored to their respective clusters. Traditional clustering methods like K-Means and Agglomerative Clustering often struggle with uncertainty in digital literacy data. This study introduces the NSS-ABC clustering method, which combines N-Soft Set (NSS) theory with the Artificial Bee Colony (ABC) algorithm, to improve the accuracy of clustering digital literacy profiles in open and distance education students.

Objectives: The primary objective of this research is to develop and evaluate the performance of the NSS-ABC clustering method for digital literacy data. The purpose of the study is to ascertain whether NSS-ABC performs better at managing uncertainty and enhancing cluster separation than traditional clustering techniques.

Methods: The population of this study was students of Universitas Terbuka, with a sample size of 3088. The samples were taken randomly with a research instrument in the form of a questionnaire distributed online. Based on data processing using the NSS-ABC algorithm, UT students' grouping, and digital behavior patterns were obtained based on generation profiles. The collected data was preprocessed before being subjected to clustering analysis using the NSS-ABC algorithm. The NSS-ABC method was implemented by integrating N-Soft Set decision making principles with the optimization capabilities of the Artificial Bee Colony algorithm to enhance cluster performance.

Results: The NSS-ABC, K-Means, and Agglomerative Clustering algorithms are used to analyze primary data on distance learning students' digital literacy. The results show that the NSS-ABC hybrid approach regularly outperforms K-Means and Agglomerative Clustering, particularly when three to six clusters are used. This shows that the combination of N-Soft Set and ABC can optimize clustering by handling data uncertainty better than conventional methods. The division of clusters in digital literacy can facilitate various parties in determining programs that suit the needs of each cluster.

Conclusions: The NSS-ABC method improves clustering by addressing data uncertainty and oprimizing group formations. The results suggest that the method can be applied to digital transformation policies in open and distance education.

Keywords: N-Softset, Artificial Bee Colony, Clustering, Digital Literacy, Decision-Making

INTRODUCTION

The rapid development of technology has made digital literacy an essential skill for individuals to access, understand, and utilize various digital platforms effectively. Not only that, society's increasing dependence on the internet and various digital tools also requires users to adapt quickly to using this technology. A deep understanding of user behavior, internet access patterns, and levels of digital competence is crucial for many parties, especially researchers, policymakers, and educational institutions. By understanding how users interact with technology, policymakers can design more targeted policies, while researchers can develop innovative approaches to address the challenges that

arise in the digital era. Good digital competence is also needed to ensure comprehensive digital inclusion so that all levels of society can participate actively and effectively in this increasingly digitalized world. Therefore, research on digital literacy patterns and internet usage behavior continues to develop to face these challenges.

Decisions can be generally simplified into agree, disagree, or neutral. However, decision-making produces various values by considering various data. [1], [2], [3]. Researchers combine several uncertainty theories to enrich science, find novelty, and provide decision-making solutions. One of the rapidly developing uncertainty theories is the N-Soft Set [4]. N-soft sets (NSS) can handle decision-making with binary judgment types, closed intervals between 0 and 1, and N-arrays. N-soft sets can also handle parameter reduction [5]. The combination of NSS with other theories produces several new approaches that are useful theoretically and practically, including fuzzy N-soft sets. [6], multi-fuzzy N-soft set [7], Pythagorean fuzzy set with NSS for the selection of scenic spots [8], complex fuzzy N-soft sets [9], Picture Fuzzy N-Soft Sets for Corona vaccine selection [10], and Complex neutrosophic N-soft sets for analysis for the performance of the Islamic banking industry [11]. N-Soft Set can also be an alternative solution to help handle the uncertainty and complexity of data mining caused by incomplete data, data noise, and the many features, such as implementing a hybrid Association Rule with NSS [12].

Clustering has been widely used to understand digital interaction patterns and usage more effectively. Clustering is an analytical technique that aims to group data into groups that share similar characteristics. Its use allows researchers to identify user segments based on digital behavior and competencies. Clustering methods such as K-Means and Agglomerative Clustering often have limitations in handling complex and uncertain data, which often appear in real-world data, including digital literacy data. The Artificial Bee Colony method must also be modified to obtain optimal results. [13], [14].

To overcome these limitations, a more advanced and flexible approach is needed. One proposed solution combines the N-Soft Set (NSS) theory and the Artificial Bee Colony (ABC) algorithm. The N-Soft Set theory excels in handling uncertainty in data. [4]. Meanwhile, the ABC algorithm, inspired by the behavior of bees in searching for food, functions to optimize the grouping process based on similar characteristics. [15], [16]. The combination of these two approaches is called the NSS-ABC method, which can produce more accurate groupings, even when the data contains uncertainty or incomplete information.

This study focuses on user segmentation based on their digital literacy level and internet usage patterns. By applying the NSS-ABC clustering method to digital literacy data, this study aims to identify user profiles that reflect varying levels of digital competence and Internet access behavior. The contribution of this study is the proposal of a new clustering method that integrates the N-Soft Set theory with Artificial Bee Colony optimization to handle uncertain data more effectively.

The rest of this article is structured as follows: Section 2 presents the methodology and experimental setup, including data pre-processing, NSS-ABC clustering algorithm, and evaluation metrics. Section 3 details the results and analysis and compares the performance of the proposed method with other clustering approaches. Finally, Section 4 concludes with the implications of this study and suggestions for future research.

OBJECTIVES

This study's primary goal is to present and assess the N-Softset Artificial Bee Colony (NSS-ABC) clustering algorithm as a novel approach for clustering digital literacy profiles among open and distance education students. The specific goal of this research is to handle data uncertainty and increase clustrering analytical decision making by developing an improved clustering method that integrates N-Softset theory with the Artificial Bee Colony optimization algorithm. Using silhouette score as the evaluation metric, compare NSS-ABC's performance to those of more conventional clustering techniques like K-Means and Agglomerative Clustering.

METHODS

This study uses theoretical and practical studies. Theoretical studies are needed because they will define the form of NSS-ABC, so it is necessary to discuss the underlying definitions, namely N-Soft Sets and Artificial Bee Colony. In the next stage, the NSS-ABC decision-making algorithm is created. In the practical study, the algorithm's findings are applied to actual data, in this case, primary data, namely the digital literacy of distance education students.

In simple terms, the state of the art of the proposed NSS-ABC decision-making is presented in Figure 1.

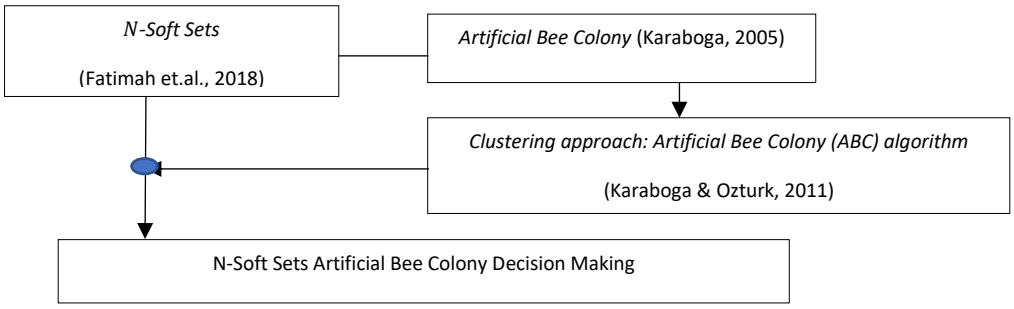

Figure 1. State of the Art NSS-ABC

Fatimah et al. define the -soft set as follows [4]. N-Soft Set (NSS) above U is denoted by (F, A, N) defined as mapping $F: A \rightarrow 2\text{^}(U \times R)$ where for each $a \in A$ here is a singly ordered pair $(u, r_a) \in U \times R$ so that$(u, r_a) \in F(a) \in U, r_a \in R$. The tabular form of the N-soft set is presented in Table 1.

Table 1. *N-Soft Set* (F, A, N)

(F, A, N)	a_1	a_2	…	a_q
u_1	$\{r_{11}\}$	$\{r_{12}\}$	…	$\{r_{1q}\}$
u_2	$\{r_{21}\}$	$\{r_{22}\}$	…	$\{r_{2q}\}$
…	…	…	…	…
u_p	$\{r_{p1}\}$	$\{r_{p2}\}$	…	$\{r_{pq}\}$

The ABC algorithm works by iterating between three types of bees. [15]. Employed Bee is used to generate new solutions from existing solutions. Onlooker Bee helps select solutions based on probability. Meanwhile, Scout Beegeneratesg new solutions randomly if the existing solutions are saturated.

Regarding the NSS-ABC decision-making, decision-making is the creation of a representative algorithm according to the definition made where different Employed Bees check the object u_i on the NSS. The NSS-ABC decision-making modification of the Artificial Bee Colony clustering approach [16] is done with a combination of N-soft sets.

The NSS-ABC Algorithm.

1. Let $U = \{u_i\}$ be a set of objects, E is a set of parameters where$A = \{a_j\} \subseteq E$ and $R = \{0, 1, \dots, N-1\}$ are ranks with $N = \{2,3, \dots\}$. Input multi N-soft set (F, A, N) so that $\forall u_i \in U, a_j \in A, \exists! r_{ij} \in R$.
2. Determine the iteration value M & threshold T where $T \in R$ s the minimum limit.
3. Calculate the value $f(x)$ using extended choice values (ECVs) or a specified function.
4. Calculate the fit(fit) Value of each object.

$$fit = \begin{cases} \frac{1}{1+f}; f \geq \\ 1 + |f|; f < 0 \end{cases}$$

5. Iteration repetition
   a) For each employed Bee, find a new solution by:
      i. The first Employed bee phase for checking the first object
         1) Select the parameter r_{1j} to focus on changing
         2) Select another object parameter in line with column no. 1) denoted b r_{pj}

3) Randomly select a weigh $\phi_1 = [-1,1]$

4) Calculate the value $r_{new} = r_{1j} + \phi_1(r_{1j} - r_{pj})$. If $r_{new} \in R$, then use the value r_{new}. If $r_{new} \notin R$ then change the value r_{new} =0 if negative and vice versa $r_{new} = N - 1$.

5) Change r_{1j} with r_{new}.

6) Calculate the value $f(x)$ & fit first object based on r_{new}.

7) If fit in Step 4 > fit Step 6) then use the change parameter.

8) If fit in Step 4 < fit Step 6) then use the previous parameter value in other words, there is no change. Give the iteration number to 1.

ii. Continue the second Employed bee phase to check the second object, and so on, until the last Employed bee phase for the remaining objects.

b) Calculate the probability valuep_i Based on the fit value in Step 7).

c) For every onlooker, Bee finds new solutions in a way:

i. Select a random number t. If $t < p_i$ Then, continue to find new solutions in the same way as Step a).i.

ii. If $t > p_i$ Then, the parameter value data does not change based on the previous data. Do the next iteration.

d) The Scout Bee phase can be carried out if there is an abandoned solution (trial>threshold), then replace the parameter value on the related object with a randomly generated value.

6. The selected object is based on the f(x) value obtained in the last iteration.

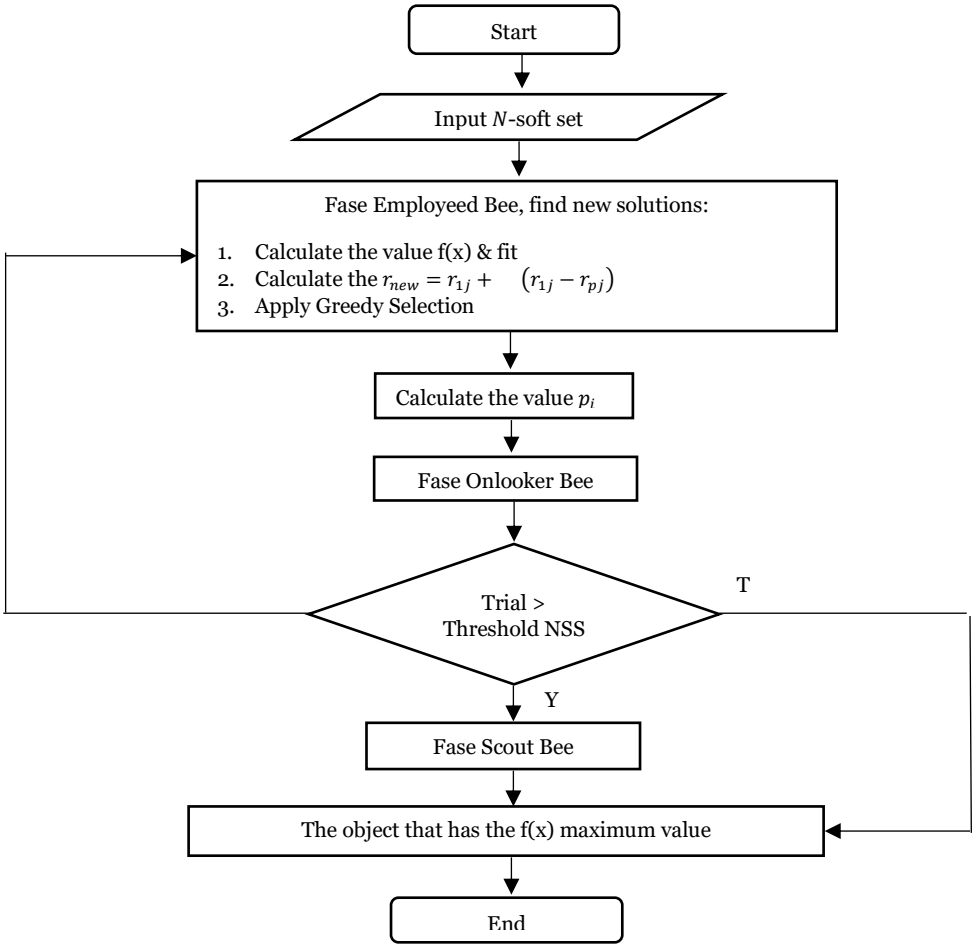

Figure 2. Flowchart the Algorithm of NSS-ABC

The stages of implementing the NSS-ABC algorithm in this study are shown in Figure 2. The study population was Universitas Terbuka (UT) studends in Odd Semester 2024/2025. The sampling technique used was probability sampling. The research instrument used a questionnaire related to the digital literacy profile of UT students, for example, with the notation U. The questionnaire questions consisted of eight parameters to explore the digital behavior of UT students. These parameters are the generation profile (a_1), ob profile (a_2), the most frequently performed internet activities (a_3), the most frequently performed internet activities (a_4), the most frequently accessed source for information (a_5), most frequently used social media (a_6), duration of hours using the internet other than social media in a day (a_7), and duration of accessing social media in a day (a_8). The results of the respondents' choices are given a score $R = \{r_{ij}\}$ according to the number of preference items.

RESULTS AND DISCUSSION

The NSS-ABC algorithm is applied to primary data on the digital literacy of distance education students. The research sample was obtained from 3088 students who filled out the questionnaire. The evaluation metric used in this research is the silhoette score. The silhouette score is used to assess how well a clustering has been performed. The silhouette score is used to assess how ell a clustering has been performed. The silhouette score ranges from -1 to 1. The higher the silhouette score, the better the clustering results. A value of 0 indicates the occurence of overlap between clusters. This study compares the performance of three clustering methods, namely K-Means [17], Agglomerative Clustering [18], and the hybrid method of NSS and ABC that optimizes K-Means.

The experiment was conducted by running three programs for clusters ranging from 3 to 7 clusters. The calculation results are shown in Table 2.

Table 2. The performance of k-means, agglomerative, NSS-ABC Silhouette Score

Number of Clusters	K-Means Silhouette Score	Agglomerative Silhouette Score	NSS-ABC Silhouette Score
3	0.5127	0.4980	0.5683
4	0.4557	0.5009	0.5542
5	0.4733	0.4611	0.5455
6	0.5158	0.4861	0.5417
7	0.5390	0.5297	0.5390

The NSS-ABC method gives the highest silhouette score (0.568) compared to k-means (0.513) and Agglomerative Clustering (0.498) for the three-cluster configuration. This shows that the hybrid method produces better clusters than K-Means and Agglomerative Clustering, with more apparent separation between clusters and more substantial similarities within clusters. The performance of the NSS and ABC methods on three clusters shows that these techniques are superior than K-Means and agglomerative method in handling uncertainty and complex data.

Agglomerative Clustering slightly outperforms K-Means in the four-cluster configuration in the Silhouette Score. The Silhoutte score for NSS-ABC for the four-clustering configuration is 0.5542. The silhouette score obtained is higher compared to the silhouette score using the K-Means method, which is 0.4557, and agglomerative clustering, which is 0.5009. This is in line with relevant research [19]. However, NSS-ABC again gave the best result with a value of 0.554, indicating that even though the number of clusters increased, this method was still able to maintain good separation between clusters and cohesion within the clusters. In five clusters, K-means produced a slightly higher value than Agglomerative Clustering (0.473 vs 0.461), but both were still behind NSS-ABC, which had a silhouette score of 0.546. This result shows that NSS-ABC maintained its superiority in producing more separated and coherent clusters. In seven clusters, the silhouette score values of K-Means and NSS-ABC were recorded at 0.539, with Agglomerative Clustering slightly behind at 0.530. Other relevant studies also found that K-Means was superior to Agglomerative Clustering [20]. In this case, K-Means and NSS-ABC show almost comparable performance.

Based on the findings, the NSS-ABC hybrid method consistently provides better results than K-Means and Agglomerative Clustering, especially in configurations of 3 to 6 clusters. This shows that the combination of N-Soft Set and ABC can optimize clustering by handling data uncertainty better than conventional methods. NSS-ABC maintains a higher silhouette score, indicating that the resulting clusters are more defined and well-separated. These clustering results show that the NSS-ABC method performs superior in various cluster configurations, especially in

scenarios where data uncertainty is challenging. This method can produce more well-defined clusters, even when the data becomes more complex.

Based on the clustering results, it was found that the highest silhouette score in in 3 clusters, using the NSS-ABC method that optimizes K-Means. This study identified three main groups with different data amounts and characteristics. Cluster 1 contains 1559 data, which is the cluster with the most significant number, Cluster 2 contains 1219 data, which is the second largest cluster, and Cluster 3 contains 310 data, which is the smallest cluster. From this distribution, researchers can conclude that Cluster 1 and 2 cover most users, while Cluster 3 is a smaller segment. The number of members in each cluster is shown in Tables 3.

Table 3. Number of Data in Each Cluster

Cluster	Number of Data
1	1559
2	1219
3	310

Each cluster has different characteristics and uniqueness. These differences consist of various aspects, namely generation, type of work, internet access time, online activities, social media platforms used, and digital literacy levels. Cluster 1 consists of the majority of users who come from the younger generation. Users in this group are primarily employees or staff, indicating that they are involved in more operational or administrative job positions. Internet access by users in Cluster 1 occurs more often during the day to the afternoon, most likely related to work breaks or when work activities are not too busy. Users in this cluster tend to use the internet for work-related activities or information, such as reading news or searching for work-related data. They do not focus too much on entertainment or social media as their primary activity on the internet. WhatsApp is the most frequently used social media platform for personal needs and work communications. Regarding digital literacy, users in Cluster 1 have a moderate or low literacy level, indicating that although they are familiar with the internet, their ability to utilize digital technology in depth is still limited. They may use the internet only for basic needs, such as communication or searching for information, without mastering more sophisticated technology.

Cluster 2 consists of users from the younger generation, but there are significant differences compared to Cluster 1 in terms of job type. Users in Cluster 2 primarily work in higher positions, namely as professionals or executives who usually have greater responsibilities in their organizations. Internet access by users in this cluster occurs more often in the morning to afternoon, indicating that they start digital activities earlier on weekdays. This may be done to complete essential tasks before noon. Similar to Cluster 1, users in Cluster 2 also use the internet for work and to get information, but because they work in more strategic positions, they may be more often involved in searching for information that is more related to business decisions or data analysis. WhatsApp remains the leading social media platform for both internal and external communication. Although users in Cluster 2 have higher job positions, their digital literacy levels remain medium or low, indicating that they may only use digital technology for primary purposes and have not fully utilized digital tools optimally to improve work efficiency or productivity.

Cluster 3 is very different from Clusters 1 and 2 because most users in this group are from the older generation. Users in Cluster 3 also work in professional or executive positions but with a higher focus on jobs that require more in-depth skills and knowledge. Internet access by users in this cluster also occurs more often in the morning to afternoon, indicating a disciplined and structured work pattern, where they use the morning to complete necessary work. Their online activities are very focused on work and information, indicating that users in Cluster 3 use the internet productively to support their professional needs, such as research, decision-making, and business communication. WhatsApp is also a central communication platform in this group, but what is most striking about Cluster 3 is its high level of digital literacy. Users in this cluster have high digital literacy, indicating that they deeply understand how to use digital technology to support their work. They are likely to be able to utilize various digital tools and technologies better than users in other clusters, and this may contribute to their success in higher job positions.

The division of clusters in digital literacy can facilitate various parties in determining programs that suit the needs of each cluster. Suppose in Cluster 1 it is known that the majority of users are employees or staff who have a moderate or low literacy level. Policymakers can hold basic training sessions that support their administration in learning and

working, thereby enhancing their digital literacy. Meanwhile, for Cluster 2 and 3, which consist of professionals or executives, they can conduct training in business decision-making or data analysis. Especially in cluster 3, due to their high digital literacy, the training provided is advanced training. The application of NSS-ABC on digital literacy data has shown positive results in enhancing classification models. With the combination of these two methods, the division of clusters becomes better and more effective, which can be utilized in various other aspects.

REFERENCES

[1] F. Fatimah and J. C. R. Alcantud, "Expanded Dual Hesitant Fuzzy Sets," in *9th International Conference on Intelligent Systems 2018: Theory, Research and Innovation in Applications, IS 2018 - Proceedings*, 2018. doi: 10.1109/IS.2018.8710539.

[2] Andriyansah and F. Fatimah, "Developing the Concept of E-Customer Relationship Management Model to Improve Marketing Performance," in *ACM International Conference Proceeding Series*, Association for Computing Machinery, Jun. 2020, pp. 22–26. doi: 10.1145/3409929.3414746.

[3] Q. Q. Aini, I. Mukhlash, K. Fahim, Jasmir, and F. Fatimah, "Neutrosophic Soft Set for Forecasting Indonesian Bond Yields," in *Lecture Notes in Networks and Systems*, Springer Science and Business Media Deutschland GmbH, 2024, pp. 690–698. doi: 10.1007/978-3-031-67192-0_77.

[4] F. Fatimah, D. Rosadi, R. B. F. Hakim, and J. C. R. Alcantud, "N-soft sets and their decision making algorithms," *Soft comput*, vol. 22, no. 12, 2018, doi: 10.1007/s00500-017-2838-6.

[5] M. Akram, G. Ali, J. C. R. Alcantud, and F. Fatimah, "Parameter reductions in N-soft sets and their applications in decision-making," *Expert Syst*, vol. 38, no. 1, 2021, doi: 10.1111/exsy.12601.

[6] M. Akram, A. Adeel, and J. C. R. Alcantud, "Fuzzy N -soft sets: A novel model with applications," *Journal of Intelligent and Fuzzy Systems*, vol. 35, no. 4, pp. 4757–4771, 2018, doi: 10.3233/JIFS-18244.

[7] F. Fatimah and J. C. R. Alcantud, "The multi-fuzzy N-soft set and its applications to decision-making," *Neural Comput Appl*, vol. 33, no. 17, 2021, doi: 10.1007/s00521-020-05647-3.

[8] H. Zhang, D. Jia-Hua, and C. Yan, "Multi-attribute group decision-making methods based on pythagorean fuzzy N-soft sets," *IEEE Access*, vol. 8, pp. 62298–62309, 2020, doi: 10.1109/ACCESS.2020.2984583.

[9] T. Mahmood, U. ur Rehman, and Z. Ali, "A novel complex fuzzy N-soft sets and their decision-making algorithm," *Complex and Intelligent Systems*, vol. 7, no. 5, pp. 2255–2280, Oct. 2021, doi: 10.1007/s40747-021-00373-2.

[10] U. U. Rehman and T. Mahmood, "Picture Fuzzy N-Soft Sets and Their Applications in Decision-Making Problems," *Fuzzy Information and Engineering*, vol. 13, no. 3, pp. 335–367, 2021, doi: 10.1080/16168658.2021.1943187.

[11] M. Akram, M. Shabir, and A. Ashraf, "Neutrosophic Sets and Systems Neutrosophic Sets and Systems Complex neutrosophic N-soft sets: A new model with applications Complex neutrosophic N-soft sets: A new model with applications Complex neutrosophic N-soft sets: A new model with applications," 2021.

[12] F. Fatimah, S. A. A. Kharis, and F. I. Fajar, "N-Soft Sets Association Rule and its Application for Promotion Strategy in Distance Education," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 18, no. 3, pp. 1865–1878, Jul. 2024, doi: 10.30598/barekengvol18iss3pp1865-1878.

[13] P. C. Saibabu, H. Sai, S. Yadav, and C. R. Srinivasan, "Synthesis of model predictive controller for an identified model of MIMO process," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 17, no. 2, pp. 941–949, 2019, doi: 10.11591/ijeecs.

[14] S. Salhi, D. Naimi, A. Salhi, S. Abujarad, and A. Necira, "A novel hybrid approach based artificial bee colony and salp swarm algorithms for solving ORPD problem," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 23, no. 3, pp. 1825–1837, Sep. 2021, doi: 10.11591/ijeecs.v23.i3.pp1825-1837.

[15] D. Karaboga, "An idea based on honey bee swarm for numerical optimization," 2005.

[16] D. Karaboga and C. Ozturk, "A novel clustering approach: Artificial Bee Colony (ABC) algorithm," *Applied Soft Computing Journal*, vol. 11, no. 1, pp. 652–657, Jan. 2011, doi: 10.1016/j.asoc.2009.12.025.

[17] G. Hamerly and C. Elkan, "Learning the k in k-means," 2003.

[18] M. R. Ackermann, J. Blömer, D. Kuntze, and C. Sohler, "Analysis of agglomerative clustering," in *Algorithmica*, May 2014, pp. 184–215. doi: 10.1007/s00453-012-9717-4.

[19] K. B, "A Comparative Study on K-Means Clustering and Agglomerative Hierarchical Clustering," *International Journal of Emerging Trends in Engineering Research*, vol. 8, no. 5, pp. 1600–1604, May 2020, doi: 10.30534/ijeter/2020/20852020.

[20] N. Singh and D. Singh, “Performance Evaluation of K-Means and Heirarichal Clustering in Terms of Accuracy and Running Time.”