Fuzzy Subtractive C-Means for Teacher Distribution Analysis

Nirwana[1(✉)], Arie Vatresia[2], and Asahar Johar[3]

[1] Program Studi Pendidikan Fisika, Universitas Bengkulu, Jl. WR Supratman, Kota Bengkulu 38122, Indonesia
nirwana@unib.ac.id

[2] Program Studi Informatika, Universitas Bengkulu, Jl. WR Supratman, Kota Bengkulu 38122, Indonesia

[3] Program Studi Sistem Informasi, Universitas Bengkulu, Jl. WR Supratman, Kota Bengkulu 38122, Indonesia

Abstract. One of factors to maintain the quality of education is a good teacher distribution based on the school resources and requirements. The better balance of the teacher will improve the quality of management education in school. Although there is a data to record the number of teachers in the school, the adequacy value of teacher distribution over the school in Indonesia remain an open question. Here, we showed the implementation of fuzzy subtractive C-Means to see the cluster of teacher distribution over Bengkulu area. This research used new improved method of K-Means clustering that can show a better representation of teacher distribution in Bengkulu. The data used are numbers of the teachers teaching in each school, number of students, numbers of group study, number of teaching hours needed. It was implemented on to 25 secondary schools Bengkulu are to see the cluster developed in the province for 8 subjects. The research clustered the teachers on to 3 cluster with its unique characteristics to see the condition of the distribution in Bengkulu. The validation value of the system was 100% using the black box method. The result shows that 35.5% schools were categorized as cluster 1, 36.5% clustered as cluster 2, and only 28% were categorized in cluster 3.

Keywords: Fuzzy Subtractive C-Means · Teacher Distribution Analysis

1 Introduction

In order to ensure equal distribution of teachers among educational units, between levels, and between types of education, between districts, between cities, and between provinces and in an effort to realize the improvement and equity of the quality of formal education nationally and the achievement of national education goals [1], a Joint Regulation of the State Minister for the Empowerment of State Apparatus and Bureaucratic Reform, Minister of National Education, Minister of Home Affairs, Minister of Finance, and Minister of Religion Number 05/X/PB/2011, SPB/03/M.PAN-RB/10/2011, 48 of 2011, 158/PMK.01/2011, 11 of 2011 concerning the Arrangement and Equity of Civil Servant

M. L. Firdaus and A. Defianti (Eds.): MASEIS 2021, ASSEHR 718, pp. 233–244, 2023.
https://doi.org/10.2991/978-2-38476-012-1_31

Teachers, and also the teacher zoning system regulated in Permendikbud No 51 of 2018 but in its implementation there are still many problems such as the accumulation of teachers in favorite schools and the lack of teachers in remote areas. Improving the quality of education in Bengkulu Province is currently still constrained by the uneven placement of teachers, and the limited basic infrastructure in this area. These problems still in the stage of low maintain by the Ministry in Indonesia to know how the teachers are clustered in Indonesia [2–4].

Clustering is one of the most important techniques in data mining [2, 5, 6]. One of the most frequently used grouping methods is Fuzzy C-Means (FCM). The FCM method has several weaknesses, including the need for a large number of groups and a predetermined group membership matrix. Usually, the initial group membership matrix is initialized randomly which causes the FCM method to have inconsistency problems. Another alternative grouping method that can be used if the number of groups is not known beforehand is the Subtractive Clustering (SC) method. SC obtained more consistent results compared to FCM [7–9]. In addition, SC has a better speed than FCM, but SC has lower accuracy than FCM. To bridge the advantages and disadvantages of the two methods, a new statistical method is proposed, which is a hybrid of the two called Subtractive Fuzzy C-Means (SFCM). This method was used [9] in his research and concluded that SFCM in general provides a better solution than FCM and provides a higher level of speed in terms of objective function convergence.

Research related to the title "Clustering Using the Subtractive Fuzzy C-Means (SFCM) Method, Case Study of Dengue Fever in East Java by [9] in this study used the Subtractive Fuzzy C-Means method on Dengue Fever data every year region, by grouping with 3 clusters. Other related research with the title "Application of Micro Business Classification of Home Dry Fish Industry with Fuzzy C-Means Clustering Algorithm in Coastal Areas of Bengkulu City" by. This study uses the Fuzzy C-Means Clustering method to classify fish industry businesses with 3 classifications, using 7 criteria, namely capital, profit, length of business, large assets, sales turnover, labor costs, and product capacity. Another related research is entitled Mapping the Distribution of Teachers in Banten Province Using the K-Means Spatial Clustering Method (Case Study: Banten Province Region). This study uses school data, student data and teacher data to find the best cluster to described how the condition of teacher distribution in Bengkulu area.

2 Method

The population in this study uses school data, namely the number of teachers in schools for each subject, the number of students in the school in each class, the number of study groups, the number of teaching hours needed for each subject each week, ratio data (number of teaching hours needed x total class) divided by the number of teachers, as well as the student to Class Ratio Data. Sampling of data in this research is using purposive sampling. Purposive sampling is a method or technique to determine samples with certain considerations [10]. Which means that the subject that will be taken by the population will be chosen deliberately based on certain considerations, namely School Data, Teachers, Students, Classes, and the need for teaching hours at SMP in Bengkulu City. In this study, the sample data used is junior high school data in the 2018/2019 academic year.

The system to be made requires input, output, and interface requirements [11, 12]. The purpose of the needs analysis is as a limitation of the system to be built, determining the capabilities and functions of the system according to user needs, and features that are added value. The analysis of system requirements to be built is as follows:

1. Need for input data
   The input data used in the study are number of teachers in school for each subject, number of students in school per class, number of study groups, number of teaching hours needed for each subject each week, ratio (number of teaching hours required x total class) divided by number of teachers, student to class ratio
2. Output data requirement
   The output data needed is the Clustering Distribution of Middle School Teachers in Bengkulu City and is divided into fit groups to determine three conditions. The first distribution of excess teachers where the number of teachers in the school or in the subject exceeds the required capacity. Second, the condition where the distribution of teachers who are understaffed where the number of teachers in the school or in the subject does not meet the required capacity. Third, even distribution of teachers in which the number of teachers in the school or in the subject is in accordance with the required capacity.
3. Interface requirements
   The need for the interface in the application is the ease and convenience of the user when accessing the application in accordance with the existing problems.

   The next stage in making this application is the implementation stage of the design that has been made into program codes. In this case, the distribution grouping of teachers has three distribution classifications, namely the Distribution of Teachers who are Excess, Even or Lack of Teachers. The database design used is using a MySQL database with the help of xampp software. The School Teacher Distribution Clustering Application using the Subtractive Fuzzy C-Means method is based on an online website using the programming language used is the Hypertext Pre-processor (PHP) programming language with the help of the CodeIgniter framework using sublime text software using the help of google chrome software so that it can be operated [2, 13–15].

   In this research, functional testing uses Blackbox testing system testing [16, 17]. Test data is generated, executed on the software and then the output of the software is tested whether it is as expected or not. Black box testing tries to find errors of incorrect or missing functions, interface error (interface), error in data structure or external database access, performance error, and initialization or termination error. The black box testing technique carried out in the School Teacher Distribution Clustering [18–20] is the equivalence partitioning technique, which is a testing technique that divides the input domain of a program into data classes, determines test cases by revealing error classes, so that it will reduce the number of test cases. Then the system functional testing and the subtractive fuzzy c-means algorithm will also be tested.

Fuzzy Subtractive C-Means

The input data in this study are text data for teacher distribution data and the initial value of Subtractive Fuzzy C-Means [7, 8, 21]. The system will process the data in the process of calculating the initial generation with the Subtractive algorithm and will display the cluster center generation value. The value of the cluster center generation is in the form of values U_1, U_2, and U_3. And will be continued with the next process using the Fuzzy C-Means algorithm to determine the cluster results and will be displayed in the form of data. The value obtained from the Fuzzy C-Means algorithm with the results obtained by the U partition matrix, namely new U_1, new U_2, and new U_3 and the three U partition matrices will be compared for each partition.

IF new U_1 > new U_2 and new U_1 > new U_3
Then the point is included in cluster 1,
Else
IF new U_2 > new U_1 and new U_2 > new U_3
Then that point belongs to cluster 2,
Else new U_3 > new U_1 and new U_3 > new U_2
Then that point belongs to cluster 3.

School teacher distribution data and initial value parameters are inputted for the Subtractive Fuzzy C-Means algorithm. For the next stage, carry out the data normalization stage using the formula:

$$X_{ij} = \frac{X_{ij} - X_{minj}}{X_{maxjj} - X_{minj}} \tag{1}$$

That is, with X_{ij} as the data normalization value, X_{maxj} as the maximum data value and X_{minj} as the minimum data value. The process is continued by calculating the stages of finding potential points by using the formula:

$$P_i = \sum_n^{j=1} e^{-\frac{4}{r_a^2}|x_i - x_j|^2} \tag{2}$$

With Pi as the potential point r_a as a constant value and $|x_i - x_j||x_i - x_j|$ is the Euclidean point distance between x_i and x_j. Next, look for the point value with the highest potential value to be the centre of the point potential. After getting the highest point potential value, then subtract the existing potential with the surrounding potential with the formula:

$$P_i = P_i - P_i^* e^{-\frac{4}{r_b}^2 \left|x_i - x_j^*\right|^2} \tag{3}$$

After reducing the surrounding potential, it will decide or conditioning with if the ratio is greater than the accepted ratio then the potential point will become the centre of the cluster, otherwise conditioning will be carried out by comparing if the ratio is less than the rejected ratio, the cluster centre will not be updated (Fig. 1).

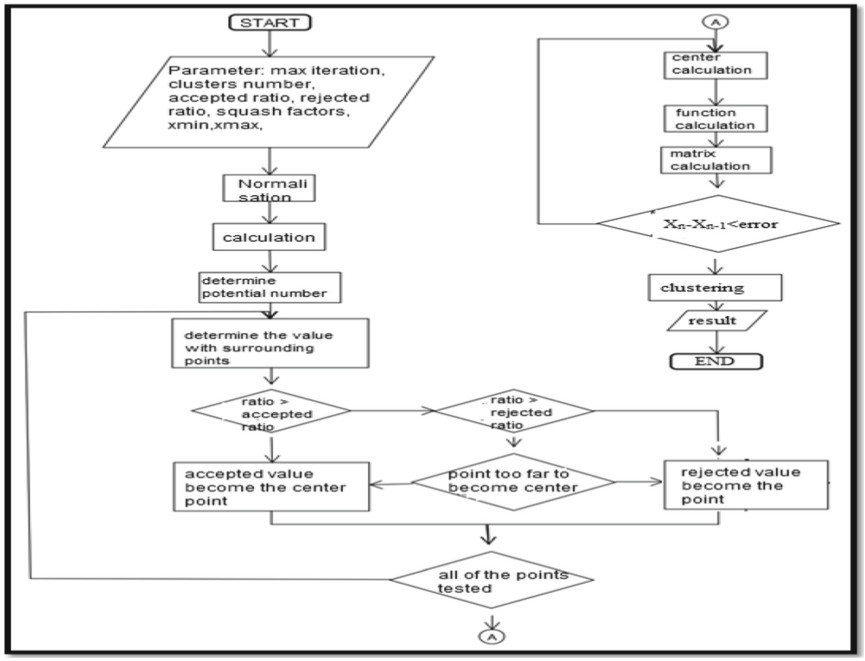

Fig. 1. Flowchart Fuzzy

After getting the initial generation value, it will calculate the center of the cluster with the formula:

$$V_{kj} = \frac{\sum_{n}^{i=1} (\mu_{ik})^{w} \cdot x_i}{\sum_{n}^{i=1} (\mu_{ik})^{w}} \tag{4}$$

After calculating the center of the cluster, it will be continued by calculating the value of the objective function using the formula:

$$\mathrm{P}_t = \sum_{n}^{i=1} \sum_{C}^{k=1} \left(\left[\sum_{m}^{j=1} \left(X_{ij} - V_{kj} \right)^2 \right] \mu_{ik}^{w} \right) \tag{5}$$

After performing calculations using the objective function value, it will continue to calculate the value of the partition matrix by using the formula:

$$\mu_{ik} = \frac{\left(\sum_{m}^{j=1} \left(X_{ij} - V_{kj} \right)^2 \right)^{\frac{-1}{w-1}}}{\sum_{c}^{k} 1 \left(\sum_{m}^{j=1} \left(X_{ij} - V_{kj} \right)^2 \right)^{\frac{-1}{w-1}}} \tag{6}$$

Next, perform conditioning where if the value of the cluster centre minus the previous cluster centre is smaller than the error, the system will stop looping. If not, repeat starting from the initial stage in the Fuzzy C-Means algorithm using the current cluster centre value as the initial generation value.

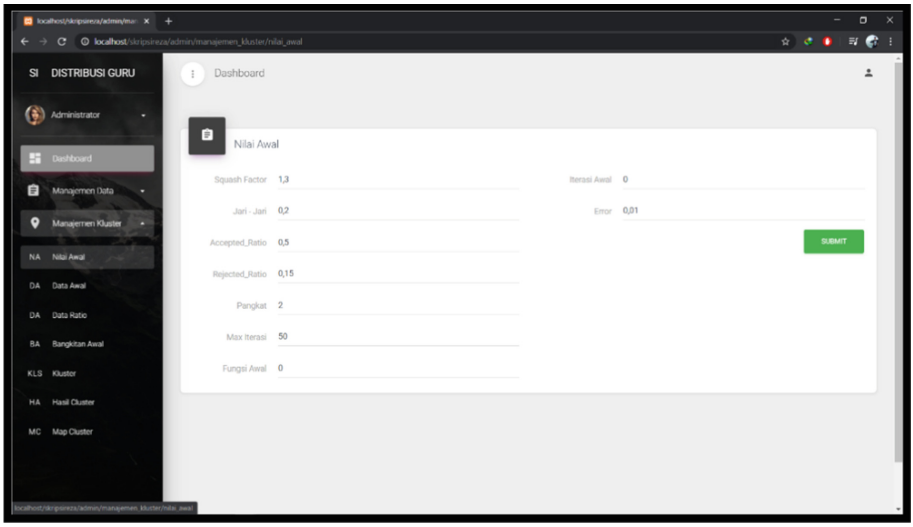

Fig. 2. Main Menu

3 Result and Discussion

3.1 System Implementation

Below is the core of the system implementation. In this system, there are pages that can be accessed by users, namely the home page, data management page, initial value page, initial generation page, cluster page and cluster results page.

Main Menu
The Initial Value page in this system is a page that contains the values used in subtractive fuzzy c-means, namely Squash Factor, Radius Accept Ratio, Reject Ratio, Rank, Max Iteration, Initial Function, Initial Iteration, and the smallest error value. The following is a display of the main system page shown in Fig. 2.

Initialized Data
The Initial Data page in this system is a page that contains the data used in subtractive fuzzy c-means, namely school data, student data, classroom data, data on the number of teachers and teaching hours for each lesson. The following is a display of the main system page shown in Fig. 3.

Initial Ratio
The initial generation page is a submenu of the cluster management page. This page performs the Subtractive algorithm process and displays the calculation results in tabular form. The following is a display of the main system page shown in Fig. 4.

Fuzzy C-Means Result
On this cluster page, the Fuzzy C-Means algorithm will process and the calculation

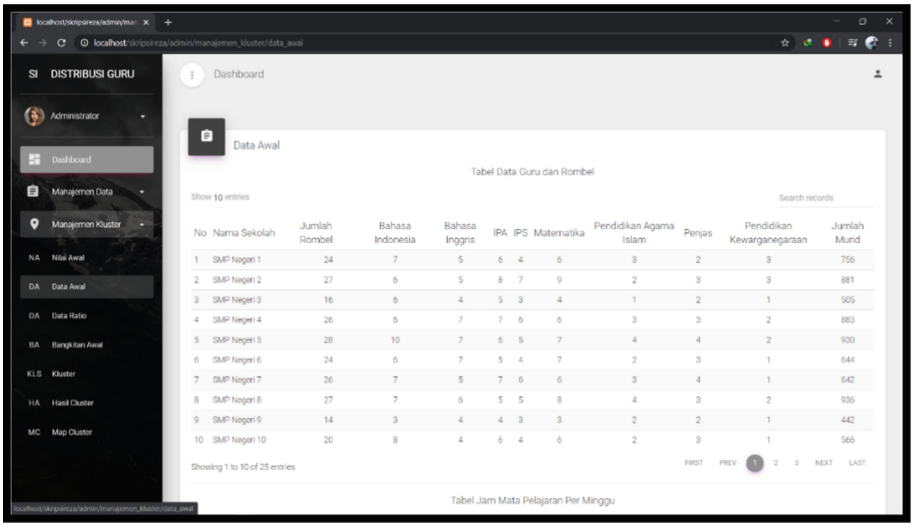

Fig. 3. Initial Data Page

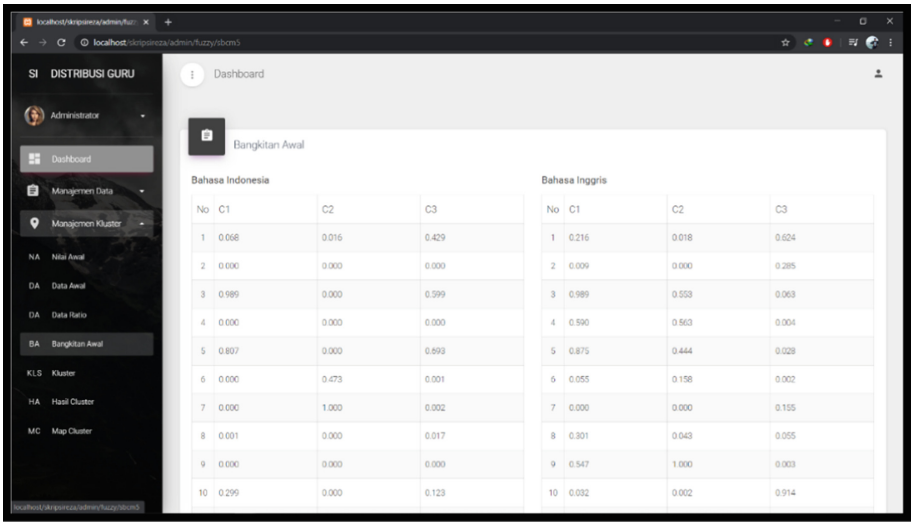

Fig. 4. Initial Ratio Page

results will be saved to the database. The data saved to the database is the result of the new U_{i1}, new U_{i2}, and new U_{i3} calculations and will immediately cluster the data into which cluster (Fig. 5).

Page of Cluster Result

Calculations that have been done on the previous pages using the Subtractive Fuzzy C-Means algorithm, the data will be displayed in the form of a map (Fig. 6).

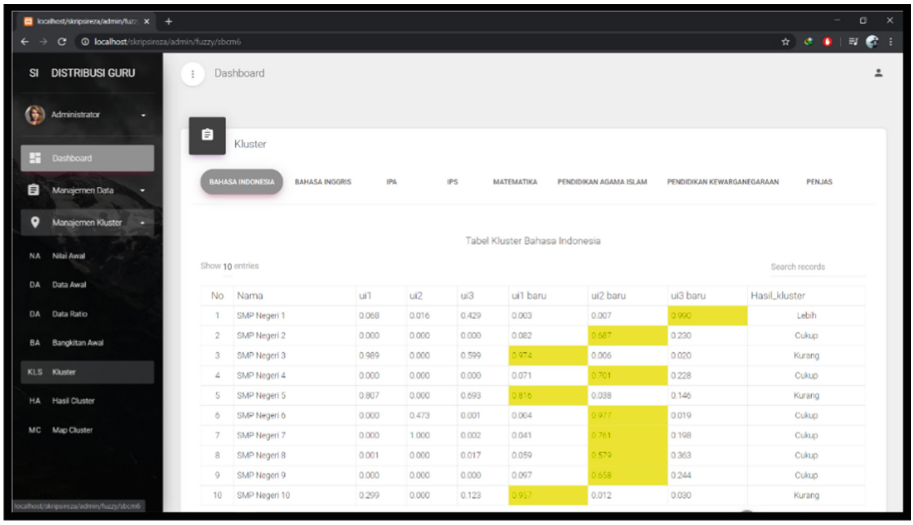

Fig. 5. Fuzzy C-Means Result

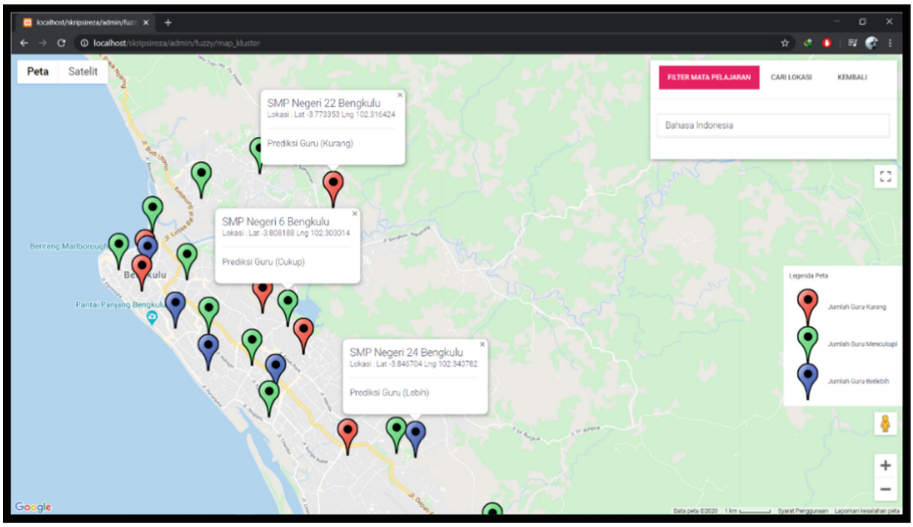

Fig. 6. Cluster Result

Black Box Testing

Black box is a test that is carried out to observe the results of the execution of the interface through the data being tested and check the functionality of the application that has been made. Black box testing is a test data design method based on software specifications. The following is a case for testing software that has been built using the black box method (Table 1).

Table 1. Functionality Testeing

No	Activity	Realization	Result	Note
1.	Main Menu	Main menus succeed to be processed and show school data and subject	Succeeded	
2.	Data Management	The data management page appears and displays 3 sub menus (sub-district management, subjects, and schools)	Succeeded	
3.	Add Data	Data will be added and stored in the database	Succeeded	
4.	Editing and show initial value	The data will be changed and will be saved to the database	Succeeded	
5.	Managing algorithm	The data will process the Subtractive algorithm and will display the cluster centre obtained.	Succeeded	Long-time processing
6.	Cluster menu	The data will be processed with the Fuzzy C-Means algorithm and the data will be saved to the database and will display the cluster centre, as well as the data cluster.	Succeeded	Long-time processing
7.	Cluster result	The data will be displayed in the form of a map or WEBGIS	Succeeded	
Succeeded scenario				7
Total Scenario				7
Functionality Result				100%

Data Validation

In the initial generation menu, the Subtractive algorithm process will be carried out in this case it will perform calculations. The system will display the results of these calculations. The page in the image below is a test with the Subtractive algorithm. The results of the Subtractive algorithm calculation process successfully display the searched cluster center data. In this case the cluster center will be used in the next algorithm. In the cluster menu there is a Fuzzy C-Means algorithm process, in this case the Fuzzy C-Means algorithm calculation process will be carried out in this menu. The data will be displayed, and the data has been grouped into which cluster. The data that has been

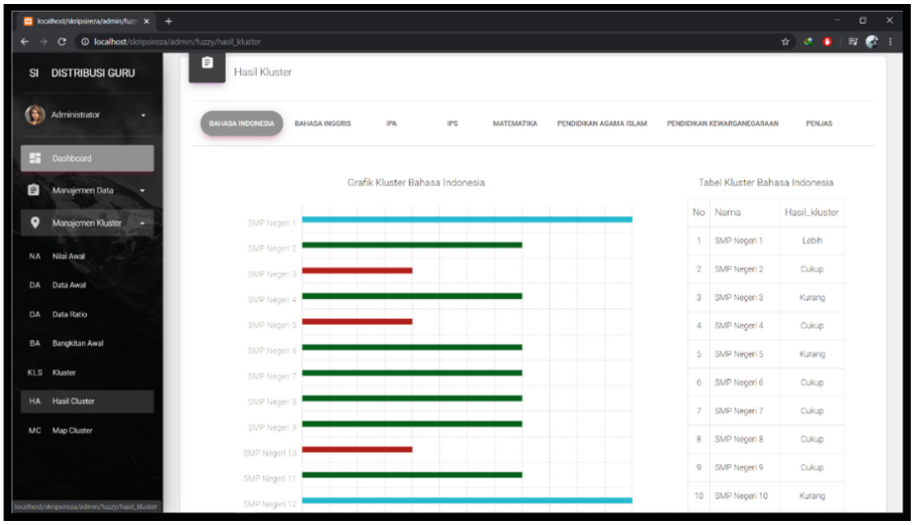

Fig. 7. Cluster result for each school

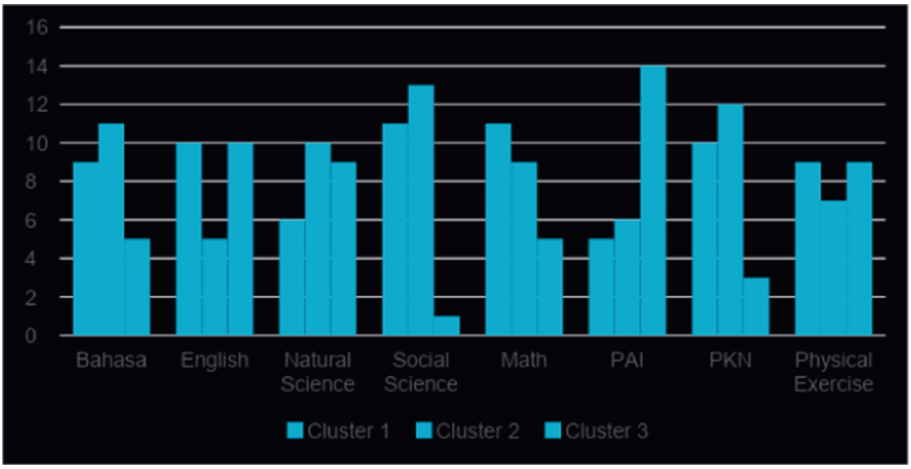

Fig. 8. School Cluster Distribution

clustered. Where in 8 subjects each school has been grouped into 3 clusters, the best fit of the cluster from the data, namely excess teachers (new U_{i3}) teacher adequacy (new U_{i2}) and teacher shortages (new U_{i1}). In Fig. 7 after the subtractive fuzzy c-means process was carried out on the distribution application of teachers in the city of Bengkulu using the subtractive fuzzy c-means method, the data obtained from the cluster results.

From clustering process before, the result shows that 35.5% schools were categorized as cluster 1, 36.5% clustered as cluster 2, and only 28% were categorized in cluster 3. The distribution of each subject can be seen in Fig. 8.

4 Conclusion

This research has produced a School Teacher Clustering application using a website-based Subtractive Fuzzy C-Means algorithm. This system has succeeded in displaying School Teacher Clustering using a website-based Subtractive Fuzzy C-Means algorithm. In the Subtractive Fuzzy C-Means algorithm, the clustering is good, but to get the best method for Clustering, it is recommended to compare the Subtractive Fuzzy C-Means method with other methods, Hard Fuzzy C-Means, Possibility Fuzzy C-Means, and others. In this system it can be applied using other programming languages to get better test results such as R, Matlab, or C.

References

1. Sondak, C. F. A., Kaligis, E. Y., and Bara, R. A., "Economic valuation of Lansa Mangrove forest, north Sulawesi, Indonesia," Biodiversitas, vol. 20, no. 4, (2019).
2. Mahbub, M., Purnamawati, D., Maslamah, Sopakua, S., and Fauziddin, M., "Educational data mining with clustering technique on the distribution of civil servant teachers in Indonesia," J. Adv. Res. Dyn. Control Syst., vol. 12, no. 6, (2020).
3. Nirwana, N., Vatresia, A., and Utama, F. P., "Mapping Teacher Distribution Analysis with Digitation Technology Implementation to Improve Education Management in Bengkulu City," in International Conference on Educational Sciences and Teacher Profession (ICETeP 2018), (2019).
4. Salas-Rodríguez, F. and Lara, S., "Systematic mapping of literature on collective teacher efficacy," Rev. Interuniv. Form. del Profr., vol. 34, no. 2, (2020).
5. Science, C. and Studies, M., "Data Mining Algorithms and their applications in Education Data Mining," Int. J. Adv. Res. Comput. Sci. Manag. Stud., vol. 2, no. 7, (2014).
6. Anonym, "Applications of Data Mining in Higher Education," Int. J. Comput. Sci. Issues, vol. 9, no. 2, (2012).
7. Abdeljabar, B., Rahmani, L., Mebrouk, B., and Kada, B., "Estimating daily diffuse solar radiation using neuro-fuzzy, fuzzy C-means and subtractive clustering: Case study: Adrar, Algeria," in PervasiveHealth: Pervasive Computing Technologies for Healthcare, (2020).
8. Qaiyum, S., Aziz, I., Jaafar, J., and Wong, A. K. L., "Ant colony optimization of interval Type-2 Fuzzy C-Means with subtractive clustering and multi-round sampling for large data," Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 1, (2019).
9. Haqiqi, B. N. and Kurniawan, R., "ANALISIS PERBANDINGAN METODE FUZZY C-MEANS DAN SUBTRACTIVE FUZZY C-MEANS," MEDIA Stat., vol. 8, no. 2, (2015).
10. Maione, C., Batista, B. L., Campiglia, A. D., Barbosa, F., and Barbosa, R. M., "Classification of geographic origin of rice by data mining and inductively coupled plasma mass spectrometry," Comput. Electron. Agric., vol. 121, pp. 101–107, (2016).
11. Hevner, A. R., March, S. T., Park, J. and Ram, S., "Design Science in Information Systems Research 1," Des. Sci. IS Res. MIS Q., vol. 28, no. 1, pp. 75–105, (2004).
12. Gerber, A., Van der Merwe, A., and Alberts, R., "Practical Implications of Rapid Development Methodologies," in Proceedings of the 2007 Computer Science + Information Technology Education Conference, 2007, pp. 233–245.
13. Romero, C. and Ventura, S., "Data mining in education," Wiley Interdiscip. Rev. Data Min. Knowl. Discov., vol. 3, no. 1, (2013).
14. Fischer, C. et al., "Mining Big Data in Education: Affordances and Challenges," Rev. Res. Educ., vol. 44, no. 1, (2020).

15. Su, G., “Analysis of optimisation method for online education data mining based on big data assessment technology,” Int. J. Contin. Eng. Educ. Life-Long Learn., vol. 29, no. 4, (2019).
16. Vatresia, A., Sadler, J. P., Rais, R. R., and Imandeka, E., “The development of mobile application for conservation activity and wildlife in Indonesia,” in Proceeding - 2016 International Conference on Computer, Control, Informatics and its Applications: Recent Progress in Computer, Control, and Informatics for Data Science, IC3INA 2016, 2017, vol. 2016-Septe, pp. 203–208.
17. Lamy, J. B., Sekar, B., Guezennec, G., Bouaud, J., and Séroussi, B., “Explainable artificial intelligence for breast cancer: A visual case-based reasoning approach,” Artif. Intell. Med., vol. 94, (2019).
18. Mamcenko, J., Kurilovas, E., Kurilovas, E., and Krikun, I., “On application of case-based reasoning to personalise learning,” Informatics Educ., vol. 18, no. 2, (2019).
19. Wotaifi, T. A., and Al-Shamery, E. S., “Modified random forest based graduates earning of higher education mining,” Int. J. Comput. Inf. Syst. Ind. Manag. Appl., vol. 12, (2020).
20. Shin, D. and Shim, J., “A Systematic Review on Data Mining for Mathematics and Science Education,” Int. J. Sci. Math. Educ., vol. 19, no. 4, (2021).
21. Benmouiza, K. and Cheknane, A., “Clustered ANFIS network using fuzzy c-means, subtractive clustering, and grid partitioning for hourly solar radiation forecasting,” Theor. Appl. Climatol., vol. 137, no. 1–2, 2019.Author, F.: Article title. Journal 2(5), 99–110 (2016).