Clustering ECE and NFE Accredited Statuses with Unsupervised Possibilistic Fuzzy C-Means

Agung Prihantoro[1]*; Kartianom[2]; Begimbetova Guldana Atymtaevna[3]
[1] Universitas Cokroaminoto Yogyakarta Indonesia; [2] Institut Agama Islam Negeri Bone Indonesia; [3] Abai Kazakh National Pedagogical University Kazakhstan
*Corresponding Author, email: aprihantoro@ucy.ac.id

Abstract

The research aims to have clusters of accredited statuses of early childhood education (ECE) and non-formal education (NFE) institutions in Yogyakarta Special Province in Indonesia, which are created by unsupervised possibilistic fuzzy c-means (UPFC) and to organize the institutions into the clusters created. The Board of National Accreditation for ECE and NFE determined four accredited statuses of A, B, C, and TT. The research employs a method of machine learning, especially UPFC. The dataset is a data of accreditation 2022 from the Board of National Accreditation for ECE and NFE of Yogyakarta Special Province. The data consists of 760 institutions composed of 749 (98.55%) ECE institutions and 11 (1.45%) NFE institutions. The analysis of UPFC created two clusters of accredited statuses of the institutions, thar are Accredited A that consists 437 (57.5%) institutions and Accredited B consisting of 323 (42.5%) institutions. The names of the clusters have political impact.

Keywords: UPFC, accredited status, early childhood education, non-formal education

Introduction

There are four accredited statuses of early childhood education (ECE) and non-formal education (NFE) in Indonesia. They are A (>850), B (700—850), C (551—700) and unaccredited (TT, <550) (Table 1) that are decided by the Board of National Accreditation for ECE and NFE (BAN PAUD dan PNF 2019). These are manually clustered before empirical data are collected, not by machine after collecting data.

Researches on accreditation of ECE focus on three topics and do not discuss clustering ECE accredited statuses. Three topics are (1) evaluation of Ece accreditation and quality assurance (Tout 2013), (Karoly 2014), (Cannon dkk. 2017), (2) validation of accreditation (Elicker dan Ruprecht 2019), and (3) experimentation of alternative accreditation systems (Elicker dan Ruprecht

2019). In Indonesia, there is no research on clustering ECE accredited statuses by machine learning (Fithri dan Anggraini 2018; Liana dan Sjamsir 2023; Nugroho dkk. 2024; Oktaria dkk. 2019:2023, 2019; Prihantoro 2025; Prihantoro dan Setiawati 2023; Rukhiyah 2020; Sari 2019).

Table 1
Scores and Accredited Statuses of ECE and NFE in Indonesia

Accredited statuses	A	B	C	TT
Score	>850	701—850	551—700	<551

Based on the scores and accredited statuses, ECE and NFE instiusions in 34 provinces in Indonesia including Yogyakarta Special Province are accredited and clustered into 4 statuses. In Yogyakarta Special Province in 2022 there are 760 institutions of ECE and NFE accredited and clustered into A, B, C, and TT (Table 2) (BAN PAUD dan PNF Provinsi DIY 2022). Twu hundred and seventy one (35.66%) institutions are accredited A, 224 (29.47%) B, 262 (34.47%) C, and 3 (0.40%) TT. Then two institutions accredited TT went to appeal and their appeal was agreed by the Board of National Accreditation for ECE and NFE and their statuses became C. But the research employed the earlier data before the two institutions went to appeal because the researchers do not get the scores after appealing.

Table 2
Accredited Statuses of ECE and NFE Institutions in Yogyakarta Special Province, Indonesia in 2022

Accredited Statuses	A	B	C	TT	Total
Number of Institutions	271	224	262	3	760
Percentage	35.66%	29.47%	34.47%	0.40%	100%

The manually clustering of the 760 ECE and NFE institutions in Yogyakarta Special Province in 2022 into 4 accredited statuses is going to be tested by machine learning, that is unsupervised possibilistic fuzzy C-means (UPFC). Therefore, two research questions are (1) How many clusters of accredited statuses of ECE and NFE institutions in Yogyakarta Special Province in 2022 are created by UPFC? and (2) How many ECE and NFE institutions of 760 ones in Yogyakarta Special Province in 2022 are organized into the clusters created?

Methods

The research employs a method of machine learning, especially UPFC.

Machine learning is employed to analyze data in descriptive and predictive ways (Yağcı 2022). The characteristics of machine learning are proper to achieve the objectives of the research, that are to have clusters which are created by UPFC and to organize ECE and NFE institutions of 760 ones in Yogyakarta Special Province in 2022 into the clusters created.

The dataset is a data of accreditation 2022 from the Board of National Accreditation for ECE and NFE of Yogyakarta Special Province. The data consists of 760 institutions composed of 749 (98.55%) ECE institutions and 11 (1.45%) NFE institutions in Yogyakarta Special Province.

The data have four measures or indicators, that are (1) document assessment, (2) & (3) individual field assessments by two assessors, and (4) group field assessment. Document assessment (DA) is done by an assessor to score document requirement of ECE and NFE institutions to be assessed in their empirical condition. Minimum score of document assessment to be able to be assessed in ECE and NFE empirical condition is 600. The second and third measures are individual field assessments by two assessors. The two assessors visit an ECE/NFE institution to analyze and then score individually its empirical condition (Visit A/VA and Visit B/VB). After scoring it individually, they score it in group by discussing (Visit in G/VG). These are four measures of the data to analyze with UPFC.

UPFC is one of clustering techniques to arrange data. UPFC is to arrange a dataset into c subsets based on similarities among the individual data items (MCA dan Baruah 2013). Initially, UPFC minimizes distance between prototype c (v_i) and feature vector n (x_k) in feature space R^p (Formula 1).

$$J_{UPFC}(X; U, V) = \sum_{k=1}^{n} \sum_{i=1}^{c} (au_{ik}^m + bt_{ik}^\eta) d^2(x_k, v_i) + \frac{\beta}{n^2 \sqrt{c}} \sum_{k=1}^{n} \sum_{i=1}^{c} (t_{ik}^\eta log t_{ik}^\eta - t_{ik}^\eta)$$

(1)

In Formula (1), u_{ik} and t_{ik} are fuzzy membership value and possibility of feature vector k^{th} (x_k) against cluster i^{th} consecutively. Objective function of UPFC has rules in Formula (2).

$$\sum_{i=1}^{c} u_{ik} = 1; \forall k; 0 \le u_{ik} \le 1; a > 0; b > 0; m > 1; \eta > 1$$

(2)

In PCA, parameter m is fuzziness exponent while parameter η is typicality

exponent. Both exponents are usually two things. But in objective function in Formula (1), parameters a and b are weight coefficients to determine relativeness of fuzziness and typicality. Both coefficients are generally one thing. To minimize J_{UPFC}, unique representativeness degree and membership degree u_{ik} are recalculated in each iteration step with Formula (3) and Formula (4).

$$u_{ik} = \left(\sum_{j=1}^{c} \left(\frac{d(x_k, v_i)}{d(x_k, v_j)} \right)^{2/(m-1)} \right)^{-1} \forall i, k$$

(3)

$$t_{ik} = exp\left(\frac{bn\sqrt{cd^2}(x_k, v_i)}{\beta} \right) \forall i$$

(4)

As seen in Formula (4), typicality degree is a possibilistic degree indicating membership degree of a data point to its cluster. Value β is a distance variance Formula (5).

$$\beta = \frac{1}{n} \sum_{k=1}^{n} d^2(x_k, \bar{x}); \bar{x} = \frac{1}{n} \sum_{k=1}^{n} x_k$$

(5)

Formula (6) is used to renew cluster center in each iteration. UPFC algorithm stops when it accomplishes a determined convergence point.

$$v_i = \frac{\sum_{i=1}^{c} (au_{ik}^m + bt_{ik}^\eta) x_k}{\sum_{i=1}^{c} (au_{ik}^m + bt_{ik}^\eta)} \forall i$$

(6)

UPFC in the research is analyzed with R package "ppclust" (Cebeci 2020). First, it initializes to construct matrix of earlier cluster prototype and matrix of fuzzy membership degree. The initialization is done by determining number of cluster (centers) in integer. Value of the centers is determined based on optimal number of cluster with 5 multiple starts to find the best solution. The best solution is made sure by calculating index of cluster internal validity. There are four indices of cluster internal validity, namely fuzzy *silhouette* (SI), *partitiony entropy* (PE), *partition coefficient* (PC), *modified partition coefficient* (MPC) and Xie-Beni (XB) (Cebeci 2020).

Result And Discussion

At first, the data was analyzed in descriptive statistic way and then was correlated among the 4 indicators. The result of the analysis is showed in Table 3 indicating that values of skewness and kurtosis of all indicators range between -2 to + 2 and this is accepted (Demir 2022). It means that the data is normally distributed. Meanwhile Table 4 shows that correlation coefficients of three indicators (VA and DA, VB and DA, VG and DA) are <0.5 and their other threes (VB and VA, VG and VA, VG and VB) are >0.5. There is multicollinearity, but the problem is solved by UPFC (Bezdek, Ehrlich, dan Full 1984) (Wu dkk. 2010).

Table 3
Descriptive statistic of accreditation of ECE and NFE institutions in Yogyakarta Special Province, Indonesia in 2022

Variabel	M	SD	Min	Max	Skewness	Kurtosis
DA	748.65	67.89	605	938	0.00	-0.66
VA	730.91	77.86	529	929	-0.02	-0.25
VB	727.20	78.55	541	970	0.02	-0.30
VG	732.76	77.16	538	953	0.00	-0.29

Table 4
Correlation among indicators of accreditation of ECE and NFE institutions in Yogyakarta Special Province, Indonesia in 2022

Variabel	DA	VA	VB	VG
DA	1.00			
VA	0.29	1.00		
VB	0.28	0.96	1.00	
VG	0.30	0.98	0.97	1.00

Afterward, it analyzed the data to create optimally clusters of accredited statuses of ECE and NFE institutions in Yogyakarta Special Province in 2022. To do so is made sure by calculating index of cluster internal validity. There are four indices of cluster internal validity, namely fuzzy silhouette (SI), partitiony entropy (PE), partition coefficient (PC), modified partition coefficient (MPC) and Xie-Beni (XB). Table 5 shows that the highest SI index, the lowest PE index, and the highest PC index are in the 3 clusters and that the highest MPC index and the highest XB index are in 5 clusters. Therefore, the optimum number of cluster is 2 clusters.

Later, it analyzed the data in 5 multiple starts to decide final number of cluster. The analysis shows that the first cluster (Profile-1) is categorized in low

cluster (accredited statuses < average = 0) and that the second cluster (Profile-2) is categorized in high cluster (accredited statuses ≥ average = 0) (Figure 1). These two clusters explain 98.38% data (Figure 2). Profile-1 is named with Accredited B and Profile-2 is Accredited A.

Table 5
Result of cluster internal validity

n Cluster	Fuzzy Silhouette Index (SI)	Partitiony Entropy (PE)	Partition Coefficient (PC)	Modified Partition Coefficient (MPC)	Xie-Beni (XB)
2	**0.68**	**0.00**	**0.48**	0.05	0.11
3	0.60	0.44	0.45	0.32	0.19
4	0.58	0.92	0.44	0.43	0.33
5	0.56	1.34	0.43	**0.48**	**0.34**

Figure 1.
Final clusters based on the best solution

Analysis of profile membership of ECE and NFE institutions in Yogyakarta Special Province in 2022 indicates that the 437 (57.5%) institutions are members of Profile-2 or Accredited A and the 323 (42.5%) institutions are members of Profile-1 or Accredited B. Figure 3 presents the profile membership of the institutions.

The result of UPFC analysis tell us that the optimum number of clusters is two clusters of accredited statuses. The result is different from what the Board of National Accreditation for ECE and NFE manually decided accredited statuses into four. UPFC offers the optimum number of clusters of accredited statuses as an alternative to decide the institutions' accredited statuses.

The names of the clusters have political impact in Indonesia. The names are Accredited A and Accredited B, not Accredited and Unaccredited, because of Minister of Education, Culture, Research and Technology Regulation Number 38 of 2023. The regulation (Article 14) recommends District of Province Offices of Education to close or merge ECE and NFE institutions, whose second accredited status is Unaccredited. The regulation (Article 8) says the ECE and NFE institutions having been built in minimum two years are obliged to be accredited. If the institutions get Unaccredited status, they have two years to be accredited again. If their second accredited status is still Unaccredited, they are recommended to be merged or closed. Therefore, the names of the two clusters are not Accredited and Unaccredited, but Accredited A and Accredited B.

If the names of the clusters are Accredited and Unaccredited, according to the Regulation Number 38 of 2023, there will be many ECE and NFE institutions that should be closed or merged. Closing or merging many ECE and NFE institutions has problems. Finally, the names of the clusters are Accredited A and Accredited B, not Accredited and Unaccredited, because the names have political impact.

Figure 2.
Explanation of the data diversity
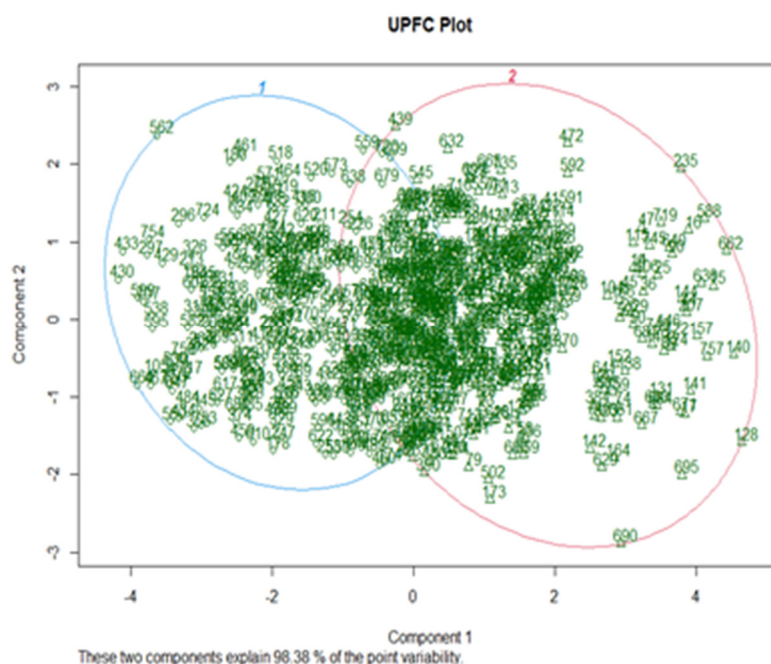
The political impact of result of UPFC is not discussed by other researchers of machine learning primarily UPFC. They discus comparison of machine learning techniques (Hogo 2010) and detail of machine learning techniques (Goel dan Tushir 2020) (Cebeci 2020) (Demir 2022). They do not talk about the political impact that is the novelty of the research.

Conclusion

UPFC created two clusters of accredited statuses of ECE and NFE institutions in Yogyakarta Special Province in 2022. The clusters of accredited statuses are Accredited A that consists 437 (57.5%) institutions and Accredited B consisting of 323 (42.5%) institutions. The number of clusters (2) is different from the number (4) decided by the Board of National Accreditation for ECE and NFE. The names of two clusters have political impact in Indonesia context because Minister of Education, Culture, Research and Technology Regulation Number 38 of 2023 (Article 14) recommends District of Province Offices of Education to close or merge ECE and NFE institutions, whose second accredited status is Unaccredited. If the names are Accredited and Unaccredited, 323 ECE and NFE institutions in Yogyakarta Special Province may be merged or closed.

The limitation of the research is incomplete data analyzed. The data has four indicators but does not have variables of accreditation of ECE and NFE institutions. If the data is complete, UPFC will give more information of clustering ECE and NFE institutions.

References

BAN PAUD dan PNF. 2019. "Panduan Penilaian Akreditasi Satuan Pendidikan Tahun 2019." 26.

BAN PAUD dan PNF Provinsi DIY. 2022. *Status Akreditasi dan Skor KPA Visitasi dan Validasi PAUD dan PNF Provinsi DIY Tahun 2022*. Yogyakarta.

Bezdek, James C., Robert Ehrlich, dan William Full. 1984. "FCM: The Fuzzy c-Means Clustering Algorithm." *Computers & Geosciences* 10(2–3):191–203. doi: https://doi.org/10.1016/0098-3004(84)90020-7.

Cannon, Jill S., Gail L. Zellman, Lynn A. Karoly, dan Heather L. Schwartz. 2017. "Quality Rating and Improvement Systems for Early Care and Education Programs: Making the Second Generation Better." *RAND Corporation* 15. Diambil 16 Januari 2025 (File:///E:/Agungp Pri Ucy/010-Research References/Evaluation Of Quality Assurance/

Cannon 2017 Quality Rating and Improvement Systems for Early Care and Education Programs.pdf).

Cebeci, Zeynel. 2020. "fcvalid: An R Package for Internal Validation of Probabilistic and Possibilistic Clustering." *Sakarya University Journal of Computer And Information Sciences* 3(1):11–27. doi: DOI: 10.35377/ saucis.03.01.664560.

Demir, Suleyman. 2022. "Comparison of Normality Tests in Terms of Sample Sizes under Different Skewness and Kurtosis Coefficients." *International Journal of Assessment Tools in Education* 9(2):397–409. doi: https://doi.org/ 10.21449/ijate.1101295.

Elicker, James, dan Karen Ruprecht. 2019. "Child Care Quality Rating and Improvement Systems (QRIS): National Experiment for Improving Early Childhood Education Quality." Hlm. 515–36 dalam *The Wiley Handbook of Early Childhood Care and Education*, disunting oleh C. P. Brown, M. B. McMullen, dan N. File. Hoboken: John Wiley & Sons, Inc.

Fithri, Radhiyatul, dan Wewen Anggraini. 2018. "Peningkatan Mutu Sekolah TK Melalui Akreditasi Se-Kecamatan Pucuk Rantau Kabupaten Kuantan Singingi." *PAUD Lectura: Jurnal Pendidikan Anak Usia* 2(1):1–15. doi: 10.31849/paudlectura.v2i01.1982.

Goel, Sonia, dan Meena Tushir. 2020. "A new iterative fuzzy clustering approach for incomplete data." *Journal of Statistics and Management Systems* 23(1):91–102. doi: DOI: 10.1080/09720510.2020.1714150.

Hogo, Mofreh A. 2010. "Evaluation of E-Learners Behaviour using Different Fuzzy Clustering Models: A Comparative Study." *International Journal of Computer Science and Information Security* 7(2):131–40.

Karoly, Lynn A. 2014. *Validation Studies for Early Learning and Care Quality Rating and Improvment Systems A Review of the Literature*.

Liana, Heppy, dan Hasbi Sjamsir. 2023. "Evaluasi Hasil Outcome dan Meningkatkan Minat Satuan PAUD Mengikuti Akreditasi Berbasis SISPENA 3.1 di BAN PAUD dan PNF Provinsi Kalimantan Timur." *Jurnal Bunga Rampai Usia Emas (BRUE)* 9(2):294–310. doi: https://doi.org/10.24114/jbrue.v9i2.52777.

MCA, Mr. Kaushik K. Phukon, dan Prof. Hemanta K. Baruah. 2013. "Extension of the Fuzzy C Means Clustering Algorithm to Fit With the Composite Graph Model for Web Document Representation." *International Journal of Cognitive Research in science, engineering and education* 1(2):1–8.

Nugroho, Taufik, Muhammad Nasrudin, Fadilah, Ahmad Reihan Syavicky, dan Yurniati. 2024. "Pendampingan Pengembangan Materi Ajar PAI Moderasi Beragama Di TPA Taman Bocah Permata Hati Sleman." *NUANSA: Jurnal Penelitian Ilmu Sosial dan Keagamaan Islam* 9(1):235–46. doi: https://doi.org/10.47200/jnajpm.v9i1.2352.

Oktaria, Renti, Lukman Hamid, Yuningsih Yuningsih, dan Suparti Suparti. 2019. "Evaluasi Hasil Akreditasi Lembaga Paud Se-Kota Depok." *Journal of Early Childhood Education (JECE)* 1(2):16–28. doi: 10.15408/ jece.v1i2.12888.

Prihantoro, Agung. 2025. "Kritik dan Dukungan Media Massa terhadap Kebijakan dan Implementasi Akreditasi Sekolah." *Asas wa Tandhim* 4(1):57–70. doi: https://doi.org/10.47200/awtjhpsa.v4i1.2782.

Prihantoro, Agung, dan Farida Agus Setiawati. 2023. "Keberhasilan Pendampingan Akreditasi Satuan PAUD Sejenis: Penelitian Kasus Tunggal." *Jurnal Obsesi : Jurnal Pendidikan Anak Usia Dini* 7(6):6895–6906. doi: DOI: 10.31004/obsesi.v7i6.4007.

Rukhiyah, Yayah. 2020. "Evaluasi Program Akreditasi PAUD di Kota Serang." Universitas Pakuan Bogor.

Sari, Ratih Permata. 2019. "Dampak Akreditasi Terhadap Mutu PAUD di KB Al Amin Desa Sumber Bendo Kecamatan Bantur Kabupaten Malang." *Jurnal Tinta* 1(1):117–33. doi: https://doi.org/10.35897/jurnaltinta.v1i1.159.

Tout, Kathryn. 2013. "Look to the Stars: Future Directions for the Evaluation of Quality Rating and Improvement Systems." *Early Education and Development* 24(1):71–78. doi: https://doi.org/10.1080/10409289.2013.741912.

Wu, X., B. Wu, J. Sun, dan H. Fu. 2010. "Unsupervised possibilistic fuzzy clustering." *Journal of Information & Computational Science* 7(5):1075–80.

Yağcı, Mustafa. 2022. "Educational data mining: prediction of students' academic performance using machine learning algorithms." *Smart Learning Environments* 9(11):1–19. doi: https://doi.org/10.1186/s40561-022-00192-z.