A Hybrid Evaluation Index Approach in Optimizing Single Tuition Fee Cluster Validity

by Wiyli Yustanti

2022 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE) | 979-8-3503-9961-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICITISEE57756.2022.10057653

A Hybrid Evaluation Index Approach in Optimizing Single Tuition Fee Cluster Validity

Wiyli Yustanti*, Nur Iriawan†, Irhamah†
†Department of Statistics,Faculty of Science and Data Analytics,Institut Teknologi Sepuluh Nopember
*Department of Informatics,Faculty of Engineering ,Universitas Negeri Surabaya
Surabaya, Indonesia
wiyliyustanti@unesa.ac.id, nur_i@statistika.its.ac.id , irhamah@statistika.its.ac.id

***Abstract*— The grouping of the socio-economic level of new students at the time of registration at public universities is a problem faced by all state universities. Identifying the right group will have an impact on the students and the university. The quality of the results of a valid grouping will give a sense of fairness to the parents of students in paying tuition fees. On the other hand, the university also expects that the results of a valid grouping will contribute to optimal revenue. This study aims to evaluate the cluster structure of a single tuition fee at the State University of Surabaya. The existing cluster structure is compared with the results of grouping using nine clustering methods, namely K-Mean, Hierarchical, BIRCH, DBSCAN, Mini Batch K-Mean, Mean Shift, OPTICS, Spectral Clustering, and Mixture Gaussian. The proposed evaluation method is a combination of three evaluation concepts, namely internal validity (Silhouette-Index), external validity (Rand Index), and the percentage conformity value to the expected income factor (Revenue-Index). These three indicators are then calculated as the average value for each clustering method as Hybrid-Index. The highest Hybrid-Index is shown by the Mini Batch K-Mean algorithm, with an average value of 0.6420, so the Mini Batch K-Mean algorithm can be recommended as a method for grouping single tuition fees.**

Keywords* — *clustering, clustering validity, hybrid evaluation, rand index, silhouette index

I. Introduction

The determination of a single tuition group for both students and universities is very important. An evaluation of the grouping structure that has been carried out by the State University of Surabaya will be evaluated in this study. By looking at the results of the internal validity of the existing clustering structure, further improvements will be made to the cluster data structure through machine learning methods. The machine learning method used is the unsupervised learning method. Research for single tuition fees (STF) based on unsupervised learning using Fuzzy C-Means and Simple Additive Weighting methods [1],[2]. Several unsupervised methods, especially for clustering problems, namely K-Mean, Agglomerative or Hierarchical, BIRCH, DBSCAN, Mini Batch K-Mean, Mean Shift, OPTICS, Spectral and Mixture Gaussian will be implemented. The use of this method is possible because the library is already available in the Scikit Learn module.

The results of grouping with these algorithms are then evaluated using an internal validity index using the Silhouette index and an external validity index using the Rand Index. Several researchers who have conducted studies on comparison of the validity of clustering in recent years include Khairul [3] presents an analysis between external and internal cluster validity indices with similar finite index ranges for ten datasets. Hasna [4] and Joonas[5] also conducted experiments to evaluate many clustering validity indices. In addition, this study proposes a new evaluation indicator, namely the revenue function, that considers the number of members in each group and the cost weight of each group. This indicator is called the Revenue Index (R-Index). Furthermore, the three validity index values are totaled, and the average value is computed for each algorithm. The average value in this research is called the Hybrid Evaluation Index (HE-Index). It is expected that the highest HE-Index value will produce the most optimal cluster model.

II. Related Works

A. *Clustering Algorithm*

Clustering is an unsupervised learning method that is widely used in various fields. This algorithm aims to group data points based on their equations. In theory, data points that are in the same group should have similar properties or characteristics, while points that are in different groups should have very different properties or characteristics. Some algorithms for clustering problems that are commonly used can be explained as follows [6]:

1. K-Mean
   The advantage of this algorithm is that it is easy to learn and simple because basically this algorithm only calculates the distance of each data to the center point of the cluster. However, the K-Means clustering algorithm still has a weakness, it needs to determine the number of clusters first. K-Means algorithm also starts with a random selection of cluster centers and can generate different clusters when running data.
2. Hierarchical
   This method forms a hierarchy based on a certain level to resemble a tree structure. Thus, the grouping process is carried out in stages or stages. Usually, this method is used on data that is not too large, and the number of clusters to be formed is unknown. In the hierarchical method, there are two grouping strategies: agglomerative and divisive.
3. BIRCH
   BIRCH stands for Balanced Iterative Reducing and Clustering using Hierarchy is an integrated hierarchical grouping algorithm. BIRCH introduces two concepts, namely feature clustering (CF) and clustering feature tree (CF tree) which are used to describe cluster summaries [7].
4. DBSCAN
   Density-Based Spatial Clustering of Applications with Noise or abbreviated as DBSCAN is a density-based clustering algorithm similar to mean-shift only slightly better than the previous algorithm. DBSCAN has several advantages, including not needing to determine the

number of clusters at the beginning, being able to identify outliers as noise, and forming clusters of various sizes and shapes. The disadvantage of this algorithm is that it does not work well when the clusters have varying densities due to the setting of distance thresholds and minimum points [8],[9].

5. Mini Batch K-Mean
   When clustering very large data sets, this algorithm, a variant of the K-means algorithm, may be used instead of the K-means algorithm. Because it does not cycle over the complete data set, it occasionally outperforms the traditional K-means algorithm when working with huge data sets. The key benefit of adopting the mini-batch K-means technique is that it makes locating clusters less computationally expensive. Although you might choose to use the K-means technique, you should employ the mini-batch method when working with huge data sets.
6. Mean Shift
   The means shift clustering algorithm is a sliding-window-based algorithm that will identify dense point areas. This clustering algorithm is a centroid-based algorithm, so the purpose of this algorithm is to find the center point of each cluster. Unlike the previous algorithm, this algorithm does not need to select the number of clusters because this algorithm can automatically find the optimal number of clusters.
7. OPTICS
   In general, the way the OPTICS algorithm works is the same as the DBSCAN algorithm. The parameters that are owned are the same, namely the epsilon parameter (eps) and the minimum points parameter. However, in the OPTICS algorithm, there are two new terms that were not previously available in the DBSCAN algorithm, namely core distance and reachability distance.
8. Spectral
   The Spectral Clustering algorithm uses Laplacian Matrix calculations. The calculation of the Eigen Vector is obtained from the Laplacian Matrix (L), and the data grouping is done based on the threshold process on the Eigen Vector with the second largest Eigen Value [10],[11], [12],[13].
9. Mixture Gaussian
   This algorithm has two main advantages, first is that it is much more flexible in terms of covariance. Second, using probability so that it can have several clusters per data point. So, if a data point is in the middle of two overlapping clusters, you can easily identify its class by comparing the percentages.

B. Evaluation Performance

In general, there are three basic criteria to investigate the validity of the results of clustering, namely external criteria, internal criteria, and relative criteria [14]. The first two approaches involve statistical and computational testing, while the third, namely relative criteria, does not involve statistical testing. The basic concept of measuring clustering validity is to find out whether the data comes from a random distribution or not [15].

1. Internal Validity
   The purpose of the internal criteria is to evaluate the clustering structure generated by a clustering algorithm through the number and features inherited from the data set. To apply internal criteria, there are two situations: (a) hierarchical clustering scheme (hierarchical clustering method) and (b) non-hierarchical clustering scheme (partition-based/central point-based method). The idea for validating the hierarchical clustering scheme is to use the so-called cophenetic matrix $\mathbf{T}_c$ and then use the cophenetic correlation coefficient to measure the degree of similarity between $\mathbf{T}_c$ and the proximity matrix $\mathbf{T}$. The cophenetic matrix $\mathbf{T}_c$ is defined such that the elements $\mathbf{T}_c(i,j)$ represent the degree of closeness between two points X_i and X_j are found around the same cluster for the first time [16].
2. External Validity
   External validity criteria aim to measure how well the grouping results match previous knowledge about the data. It is assumed that prior information cannot be computed from $\mathbf{X}$. Perhaps the most used forms of external information are classes (categories) and class labels for objects associated with $\mathbf{X}$. This information is usually obtained through manual classification. So, in principle an external criterion is an index designed to measure the similarity between two partitions which only considers the distribution of points in different groups and is not used to measure the quality of this distribution[17] There are two approaches that can be taken, the first is to evaluate the resulting clustering structure of S, by comparing it with an independent partition of the $\mathbf{T}$ data that is constructed according to intuition or previous information about the clustering structure of the data. Then, the second approach is to compare the proximity matrix $\mathbf{T}$ with the partition matrix $\mathbf{T}$.

III. Method

The proposed methodological framework for this research can be described in Fig 1.

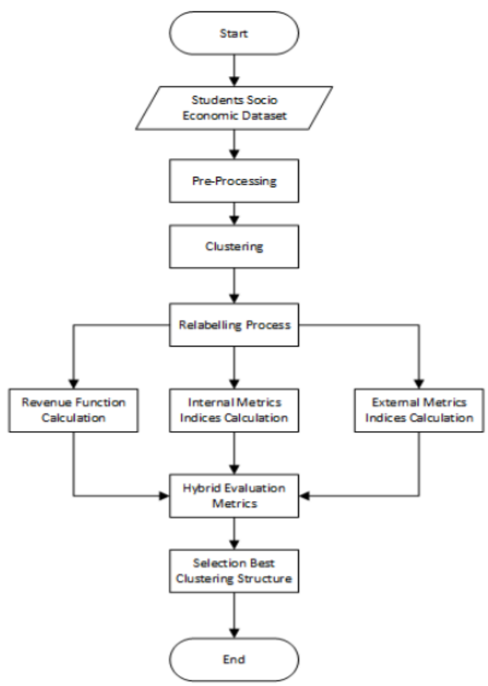

Fig. 1. The Proposed Method

The phase in this study is started with the preprocessing phase, modelling with clustering algorithms, relabeling process, and performance evaluation.

A. Dataset

The data used in this study is the socio-economic data of new students from 2017 to 2021 at the State University of Surabaya with a total of 15,875 rows. The number of groups of single tuition fees determined is 8 classes. The descriptive statistics of tuition fee at the State University of Surabaya according to Government Regulation of the Ministry of Research, Technology and Higher Education No. 22 of 2015 (in million) can be seen in Table I.

TABLE I. DESCRIPTIF STATISTICS OF SINGLE TUITION FEE (IN MILLION)

Statistics	K1	K2	K3	K4	K5	K6	K7	K8
MEAN	0.5	1	2.4	3.48	4.55	5.62	6.68	7.73
MINIMUM	0.5	1	2.4	3.12	3.84	4.56	5.28	6.00
MAKSIMUM	0.5	1	2.4	4.14	5.88	7.62	9.36	11.10
STDEV	0	0	0	3.08	6.16	9.33	12.46	15.70

B. Pre-Processing

Prior to the clustering process, the conversion process from categorical data to numeric data is carried out. This is because socio-economic data comes from student entry data through a registration system that has a categorical type. After becoming numerical data, then the process of normalization and dimension reduction is carried out using the principle component analysis (PCA) method.

C. Clustering

At this stage, the clustering process is carried out from the data that has been normalized and in the component variables. To facilitate visualization, clustering was carried out based on the first and second components of the PCA results. In this process, cluster modeling is carried out using 9 algorithms, namely the K-Mean, Agglomerative or Hierarchical, BIRCH, DBSCAN, Mini Batch K-Mean, Mean Shift, OPTICS, Spectral and Mixture Gaussian algorithms.

Fig 2. The Pattern Matching Between Reference Label and Random Label

D. Relabelling Cluster

The relabeling process is the stage of re-labeling the labeling results through a clustering algorithm. This stage is done because the label given by the cluster algorithm is random. Meanwhile, the determination of the label in the STF case study has an impact on income for the university and the burden of costs that must be incurred from the student side. Thus, this labeling process becomes important to give meaning to the labels generated from the algorithm. The method used in this study is to use the dominant variable in the first PCA component. The dominant variable in this component is the father's income, which shows an upward trend pattern as shown in Fig.2. This upward trend indicates that the higher the STF value, the greater the parent's income.

E. Proposed Performance Evaluation

This research proposes a hybrid method of evaluating clustering performance. The internal validity used is the silhouette index bye eq (1) [18], [19] and the external validity uses the Rand Index y eq (2).

$$Sill-index = \max_{k} \frac{1}{N}\sum_{i=1} \frac{b(i)-a(i)}{\max\{b(i),a(i)\}} \qquad (1)$$

where,

$a(i) = d(i, \mu C_i)$ and $b(i) = \min_{C_j \neq C_i} d(i, C_j)$

While the equation of Rand-Index [20], [21] can be formulated as eq (2). It is a measure of the percentage of correct decisions.

$$Rand-index = \frac{TP+TN}{TP+FP+FN+TN} \qquad (2)$$

In addition to the two validity indices, this study proposes a new measure called the Revenue (R) Index, where this revenue index measures the suitability of the revenue obtained from the clustering results with the existing revenue. Suppose the existing revenue is expressed by R, then the Revenue Index can be formulated by eq (3).

$$R-index = \frac{\sum_{i=1}^{k} in_i}{R} \times 100\% \qquad (3)$$

Where i is STF group level and n_i is the number of members of the i-th STF group. Next, the Hybrid Evaluation HE) Index value will be calculated with the formulation as in equation (4)

$$HE-Index = \frac{Sill\text{-}Index + Rand\text{-}Index + R\text{-}Index}{3} \qquad (4)$$

The interpretation of all the index values is that the higher the index value, the more valid the result of grouping the data. This means that each data point has been grouped into the right cluster.

IV. RESULTS AND DISCUSSION

In this section, the results of exploring the existing cluster structure will be explained through the results of internal validity checks using the Silhouette index value. The results of the internal validity measurement produce a value of -0.0274, which means that many data points fall into the wrong group. Visually, it can be seen in Fig. 3 how the to scatter data is based on the STF group.

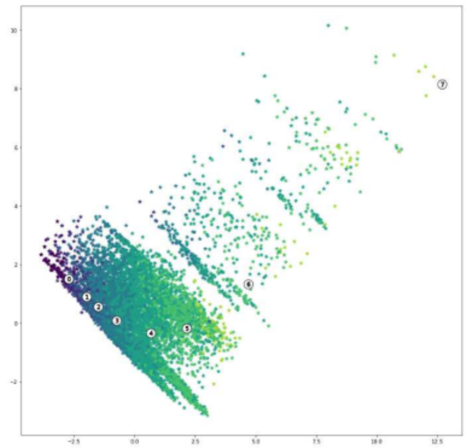

Fig. 3. Scatter of Existing Single Tuition Fee Cluster

In Fig.3, there is much overlap between groups, so the boundaries between one group and another are not visible. The existing data structure will be used as a baseline to find STF data groupings with a better structure. In this instance, the group's point density is higher, and the boundaries between groups are more precise. In Fig. 4, the highest Silhouette index value was obtained by clustering with the spectral algorithm. The next rank is followed by the results of Mean Shift and K-Mean, while the results of the BIRCH and Mini Batch K-Mean algorithms show that the Silhouette index values are not much different.

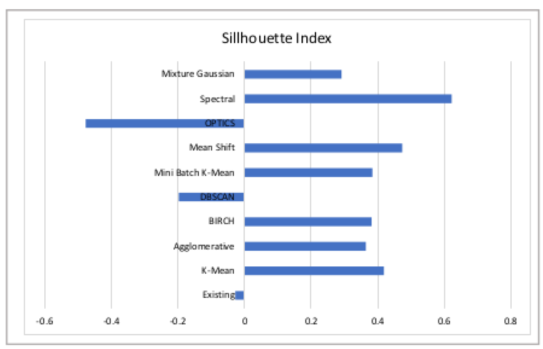

Fig. 4. Silhouette Index Existing Cluster vs Cluster based Algorithms

The silhouette index value is in the range of -1 to 1. Consider the case where the silhouette coefficient value is near to 1, suggesting improved data clustering. In contrast, if the silhouette coefficient is close to 0, the data clustering is deteriorating. The Rand-Index measurement in Table II shows that the results of clusters with centroids (K-Mean and Mini Batch) and hierarchies (Agglomerative and BIRCH) have a higher degree of similarity with the reference cluster.

TABLE II. VALIDITY SCORE FOR CLUSTERING ALGORITHMS

Clustering Algorithm	Silhouette Index	Rand Index	Revenue Index	Hybrid Evaluation Index
K-Mean	0.4191	0.6930	0.7318	0.6146*
Agglomerative	0.3646	0.6835	0.6488	0.5656
BIRCH	0.3815	0.6659	0.6430	0.5635
DBSCAN	-0.1970	0.5386	0.8265	0.3894
Mini Batch K-Mean	0.3842	0.6939	0.8478	0.6420*
Mean Shift	0.4737	0.3642	0.2593	0.3657
OPTIC	-0.4771	0.3414	0.9214	0.2619
Spectral	0.6223	0.3225	0.2360	0.3936
Mixture Gaussian	0.2911	0.6771	0.7796	0.5826

The Rand-Index value ranges from 0 to 1, where if the value is close to 1, the cluster results are more like the existing cluster. Scatter of clustering result can be shown in Fig.5.

Furthermore, when viewed from the distribution of the number of members of each clustering result multiplied by the weight of the cost factor to produce a revenue index, the OPTICS algorithm has the highest R-Index. By seeing that there is a considerable enough variation in the measurement results of each validity metric, a calculation is carried out that combines the three validities using equation (2). The proposed cluster performance evaluation formulation has found that the clustering algorithm using Mini Batch K-Mean gets the highest HE-Index value with a value of 0.6420. As with other cluster validity measures, the closer to a value of 1 the better the cluster structure. This value means that the clustering results have a reasonably homogeneous structure within the cluster, have a pretty good resemblance to the existing cluster, and are not too much different from the revenue value expected by the university.

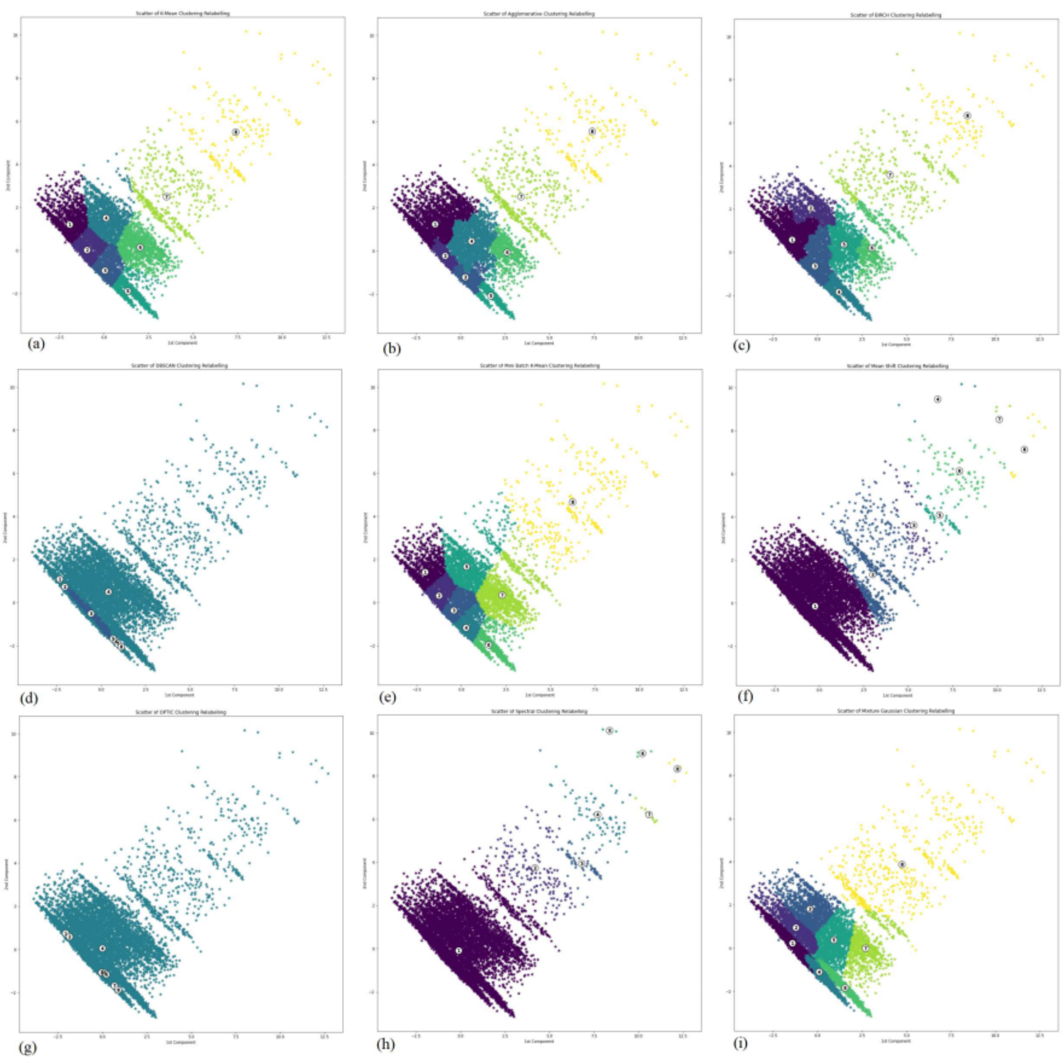

Fig. 5. Scatter Data of Clustering Algorithm

V. CONCLUSION

This study concludes that the proposed hybrid evaluation index can help provide recommendations in choosing a clustering algorithm method that can accommodate three indicators, namely the homogeneity indicator in the cluster (Silhouette Index), the cluster similarity indicator with the reference cluster (Rand Index) and the suitability indicator with the expected income (Revenue Index). Based on implementation of nine types of algorithms, it is found that the Mini Batch K-Mean algorithm is the most recommended for grouping single tuition fees on the socio-economic dataset of new students at the State University of Surabaya. The clustering results with the Mini Batch K-Mean algorithm get a Hybrid Evaluation Index value of 0.6420, which means that the three aspects of clustering validity indicators have an average value of more than 0.5 (close to 1).

REFERENCES

[1] Indrawati, Anwar, and N. Amalia, "Determination System of Single Tuition Group Using a Combination of Fuzzy C-Means Clustering and Simple Additive Weighting Methods," in *IOP Conference Series: Materials Science and Engineering*, 2019, vol. 536, no. 1. doi: 10.1088/1757-899X/536/1/012148.

[2] H. Syahputra, Sutrisno, and S. Gultom, "Decision Support System for Determining the Single Tuition Group (UKT) in State University of Medan Using Fuzzy C-Means," in *Journal of Physics: Conference Series*, Mar. 2020, vol. 1462, no. 1. doi: 10. 1088/1742-6596/1462/1/012071.

[3] K. N. Ismail, A. Seman, and K. A. F. Abu Samah, "A Comparison between External and Internal Cluster Validity Indices," in *2021 IEEE 11th International Conference on System Engineering and Technology, ICSET 2021 - Proceedings*, 2021, pp. 229–233. doi: 10.1109/ICSET53708.2021.9612525.

[4] H. Chouikhi, M. Charrad, and N. Ghazzali, "A comparison study of clustering validity indices," in *GSCIT 2015 - Global Summit on Computer and Information Technology - Proceedings*, Dec. 2015. doi: 10.1109/GSCIT.2015.7353330.

[5] J. Hämäläinen, S. Jauhiainen, and T. Kärkkäinen, "Comparison of internal clustering validation indices for prototype-based clustering," *Algorithms*, vol. 10, no. 3, 2017, doi: 10.3390 /a100 30105.

[6] F. Murtagh, "A survey of recent advances in hierarchical clustering algorithms," *Computer Journal*, vol. 26, no. 4, pp. 354–359, 1983, doi: 10.1093/comjnl/26.4.354.

[7] T. Zhang, R. Ramakrishnan, and M. Livny, "BIRCH: An Efficient Data Clustering Method for Very Large Databases," *SIGMOD Record (ACM Special Interest Group on Management of Data)*, vol. 25, no. 2, pp. 103–114, 1996, doi: 10.1145/ 235968.233324.

[8] A. C. Benabdellah, A. Benghabrit, and I. Bouhaddou, "A survey of clustering algorithms for an industrial context," *Procedia Comput Sci*, vol. 148, pp. 291–302, 2019, doi: 10.1016/ j.procs. 2019.01.022.

[9] J. Oyelade *et al.*, "Data Clustering: Algorithms and Its Applications," *Proceedings - 2019 19th International Conference on Computational Science and Its Applications, ICCSA 2019*, no. ii, pp. 71–81, 2019, doi: 10.1109/ICCSA.2019.000-1.

[10] M. Z. Rodriguez *et al.*, *Clustering algorithms: A comparative approach*,vol.14,no.1.2019.doi:10.1371/ journal. pone. 0210 236.

[11] M. C. V. Nascimento and A. C. P. L. F. De Carvalho, "Spectral methods for graph clustering - A survey," *Eur J Oper Res*, vol. 211, no. 2, pp. 221–231, 2011, doi: 10.1016/j.ejor.2010.08.012.

[12] Y. Yang and H. Wang, "Multi-view clustering: A survey," *Big Data Mining and Analytics*, vol. 1, no. 2, pp. 83–107, 2018, doi: 10.26599/bdma.2018.9020003.

[13] H. Zha, X. He, C. Ding, H. Simon, and M. Gu, "Spectral Relaxation for K-means Clustering."

[14] G. Gan, C. Ma, and J. Wu, *Data Clustering : Theory, Algorithms and Applications*. American Statistical Association and the Society for Industrial and Applied Mathematics. 10, 2007. doi: 10.1017/CBO9781107415324.004.

[15] M. Halkidi, Y. Batistakis, and M. Vazirgiannis, "Cluster Validity Methods: Part I," *SIGMOD Rec.*, vol. 31, no. 2, pp. 40–45, Jun. 2002, doi: 10.1145/565117.565124.

[16] S. Guan and M. Loew, "An Internal Cluster Validity Index Using a Distance-based Separability Measure," in *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*, Nov. 2020, vol. 2020-November, pp. 827–834. doi: 10.1109/ICTAI50040.2020.00131.

[17] B. Desgraupes, "Clustering Indices," *Quest - Lab Modal'X*, vol. 1, no. November, p. 34, 2017.

[18] L. Lenseen and E. Schubert,"Similarity Search and Applications," 2022, vol. 13590. doi: 10.1007/978-3-031-17849-8.

[19] F. Batool and C. Hennig, "Clustering with the Average Silhouette Width," *Comput Stat Data Anal*, vol. 158, Jun. 2021, doi: 10.1016/j.csda.2021.107190.

[20] N. Xuan Vinh, J. Epps, and J. Bailey, "Information Theoretic Measures for Clusterings Comparison: Variants, Properties, Normalization and Correction for Chance," 2010.

[21] A. J. Gates and Y.-Y. Ahn, "The Impact of Random Models on Clustering Similarity," 2017. [Online]. Available: http://jmlr.org/ papers/ v18/17-039.html.

A Hybrid Evaluation Index Approach in Optimizing Single Tuition Fee Cluster Validity

ORIGINALITY REPORT

17%	12%	10%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	apps.dtic.mil Internet Source	1%
2	Submitted to University of Southampton Student Paper	1%
3	Submitted to Liverpool John Moores University Student Paper	1%
4	cs.uef.fi Internet Source	1%
5	Submitted to IUBH - Internationale Hochschule Bad Honnef-Bonn Student Paper	1%
6	digilib.unimed.ac.id Internet Source	1%
7	medium.com Internet Source	1%
8	rc.library.uta.edu Internet Source	1%

9	espace.curtin.edu.au Internet Source	1 %
10	Hadi Santoso, Hilyah Magdalena. "Improved K-Means Algorithm on Home Industry Data Clustering in the Province of Bangka Belitung", 2020 International Conference on Smart Technology and Applications (ICoSTA), 2020 Publication	1 %
11	eprints.kfupm.edu.sa Internet Source	1 %
12	Atif Rizwan, Naeem Iqbal, Anam Nawaz Khan, Rashid Ahmad, Do Hyeun Kim. "Towards Effective Pattern Recognition Based on Enhanced Weighted K-mean Clustering Algorithm for Groundwater Resource Planning in Point Cloud", IEEE Access, 2021 Publication	<1 %
13	Mauladi, P E P Utomo, B F Hutabarat, R A Putra. "Decision Support System to Determine Uang Kuliah Tunggal (UKT) by Combining Naïve Bayes Classifier and Fuzzy-TOPSIS", Journal of Physics: Conference Series, 2020 Publication	<1 %
14	Submitted to Hellenic American University Student Paper	<1 %

15	Submitted to Universitas Airlangga Student Paper	<1%
16	Ika Wirawati, Nur Iriawan, and Irhamah. "Bayesian Hierarchical Random Intercept Model Based on Three Parameter Gamma Distribution", Journal of Physics: Conference Series, 2017 Publication	<1%
17	Submitted to University of Technology, Sydney Student Paper	<1%
18	Submitted to University of Lancaster Student Paper	<1%
19	repository.paramadina.ac.id Internet Source	<1%
20	www.ioinformatic.org Internet Source	<1%
21	Ika Marta Sari, Saskya Mary Soemartojo, Titin Siswantining, Devvi Sarwinda. "Mining Biological Information from 3D Medulloblastoma Cancerous Gene Expression Data Using TimesVector Triclustering Method", 2020 4th International Conference on Informatics and Computational Sciences (ICICoS), 2020 Publication	<1%

22	Indrawati, Anwar, Nova Amalia. "Determination System of Single Tuition Group Using a Combination of Fuzzy C-Means Clustering and Simple Additive Weighting Methods", IOP Conference Series: Materials Science and Engineering, 2019 Publication	<1%
23	Michael Botyarov, Erika E. Miller. "Partitioning around medoids as a systematic approach to generative design solution space reduction", Results in Engineering, 2022 Publication	<1%
24	doras.dcu.ie Internet Source	<1%
25	idus.us.es Internet Source	<1%
26	www.duo.uio.no Internet Source	<1%
27	Absalom E. Ezugwu, Abiodun M. Ikotun, Olaide O. Oyelade, Laith Abualigah et al. "A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects", Engineering Applications of Artificial Intelligence, 2022 Publication	<1%

28	Jie Wen, Ke Yan, Zheng Zhang, Yong Xu, Junqian Wang, Lunke Fei, Bob Zhang. "Adaptive Graph Completion Based Incomplete Multi-view Clustering", IEEE Transactions on Multimedia, 2020 Publication	<1 %
29	Siti Mariana Che Mat Nor, Shazlyn Milleana Shaharudin, Shuhaida Ismail, Sumayyah Aimi Mohd Najib, Mou Leong Tan, Norhaiza Ahmad. "Statistical Modeling of RPCA-FCM in Spatiotemporal Rainfall Patterns Recognition", Atmosphere, 2022 Publication	<1 %
30	edoc.hu-berlin.de Internet Source	<1 %
31	iopscience.iop.org Internet Source	<1 %
32	technodocbox.com Internet Source	<1 %
33	vtechworks.lib.vt.edu Internet Source	<1 %
34	www.egr.msu.edu Internet Source	<1 %
35	www.hindawi.com Internet Source	<1 %

36	Alp Ozdemir, Marcos Bolanos, Edward Bernat, Selin Aviyente. "Hierarchical Spectral Consensus Clustering for Group Analysis of Functional Brain Networks", IEEE Transactions on Biomedical Engineering, 2015 Publication	<1 %
37	K R Kavitha, S Sandeep, P R Praveen. "Improved spectral clustering using PCA based similarity measure on different Laplacian graphs", 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2016 Publication	<1 %

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		