

*Corresponding author: Muhammad Kasim Aidid, Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Makassar, Indonesia

E-mail: kasimaidid@unm.ac.id

RESEARCH / REVIEW ARTICLE

Application Of Ensemble K-Modes And SWFM For Grouping Sulawesi Tengah Regions By Underdeveloped Indicators

Zulkifli Rais, Muhammad Kasim Aidid*, Husnul Amira

Statistics Study Program: Faculty of Mathematics and Natural Sciences.
Makassar State University, Makassar City, South Sulawesi, Indonesia

Abstract: This research aims to determine the best final clustering results and clustering statistics for regencies/cities in Central Sulawesi based on underdeveloped region indicators. The study uses categorical and numerical data variables, consisting of 10 numerical variables and 3 categorical variables. The methods used in this research are the mixed data Ensemble K-Modes and the Similarity Weight and Filter Method (SWFM). The best mixed data clustering method shows that the Ensemble K-Modes method produces better clustering results than the SWFM method, as Ensemble K-Modes has a higher accuracy score of 0,8462.

Keywords: Ensemble K-Modes, Similarity Weight and Filter Method (SWFM), underdeveloped region indicators.

1. Introduction

Cluster analysis is a multivariate technique primarily aimed at grouping objects based on their statistical attributes. It classifies objects so that each object is most similar to others within the same cluster (Mara & Intisari, 2013). Objects within the same cluster share similar characteristics, distinguishing them from objects in other clusters (Yulianto & Hidayatullah, 2014). There are two main approaches in cluster analysis: hierarchical and non-hierarchical methods. The hierarchical approach identifies group structures through agglomerative (merging) or divisive (splitting) methods, whereas non-hierarchical approaches, like C-Means, are more suitable when the number of clusters is predetermined (Purnamasari, 2011). However, C-Means is less ideal if objects fall between two or more partitions, leading to the development of Fuzzy C-Means (FCM), a fuzzy approach allowing objects to belong to more than one cluster with a certain degree of membership (Anggara Cahya Putra & Dwi Hartomo, 2021).

For categorical data, the K-Modes method an adaptation of C-Means provides efficient clustering by updating attribute modes as cluster centers (Suguna & Arul Selvi, 2012). However, this method has limitations with mixed data types, including numerical and categorical data. In some cases, data transformation techniques can standardize variable scales, though these transformations often risk losing essential information from the original data.

Another approach to clustering mixed data is Ensemble Clustering methods, such as Ensemble K-Modes, which combine the clustering results of numerical and categorical data. The Similarity Weight and Filter Method (SWFM) is a mixed data clustering method that uses hierarchical partitioning for numerical data and K-Modes for categorical data (Science dkk., 2023). The strengths of SWFM include efficient multidimensional use, lower time complexity,



and effective distance measurement, even when cluster boundaries are irregular (Reddy & Kavitha, 2012). In the context of underdeveloped region indicators, which comprise both numerical and categorical data, the Ensemble K-Modes and SWFM approach can be used to cluster regencies and cities in Central Sulawesi according to the criteria for underdeveloped regions as defined by Presidential Regulation No. 63 of 2020.

2. Literature Review

2.1. Fuzzy C-Means

Fuzzy C-Means (FCM) is a data clustering technique where the presence of each data point in a cluster is determined by a degree of membership (H. L. Sari & Suranti, 2016). Fuzzy C-Means allows data to be a member of any class or cluster formed, with a membership degree between 0 and 1 (Kusuma & Sangadji, 2016).

The basic concept of Fuzzy C-Means is to find cluster centers that will represent the average position of each cluster. Each data point has a degree of membership for each cluster. By adjusting the cluster center values and the membership degree for each data point, the cluster centers will be positioned correctly (Muhardi, 2019). One method for determining membership values is through a function-based approach (Arifah dkk., 2017). Commonly used functions include :

Representation of an Increasing Linear Curve

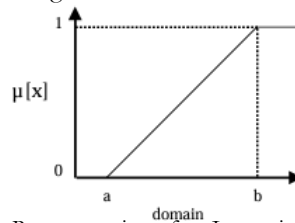


Figure 1. Representation of an Increasing Linear Curve

The formula for finding the membership function is as follows (Arifah dkk., 2017) :

$$\mu_{\text{linier naik}}(x, a, b, c) = \begin{cases} 0 & \text{if } x \leq a \\ \frac{(x-a)}{(b-a)} & \text{if } a \leq x \leq b \\ 1 & \text{if } x \geq b \end{cases} \quad (1)$$

Representation of an Decreasing Linear Curve

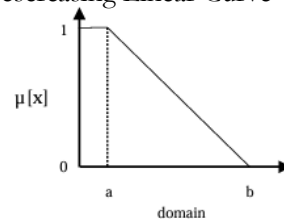


Figure 2. Representation of a Decreasing Linear Curve

The formula for finding the membership function is as follows (Arifah dkk., 2017) :

$$\mu_{\text{segitiga}}(x, a, b, c) = \begin{cases} \frac{x-a}{b-a} & \text{if } a \leq x \leq b \\ \frac{c-x}{c-b} & \text{if } b \leq x \leq c \\ 0 & \text{For other values of } x \end{cases} \quad (2)$$

Representation of a Triangular Curve

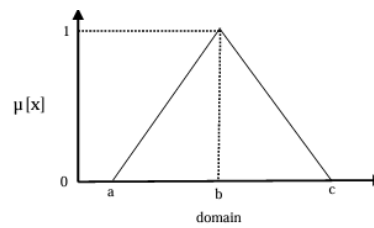


Figure 3. Representation of a Triangular Curvae

The formula for finding the membership function is as follows (Tiro & Rasyid, 2018) :

$$\mu_{\text{segitiga}}(x, a, b, c) = \begin{cases} \frac{x-a}{b-a} & \text{if } a \leq x \leq b \\ \frac{c-x}{c-b} & \text{if } b \leq x \leq c \\ 0 & \text{or other values of } x \end{cases} \quad (3)$$

One method for determining the ideal number of clusters in numerical data clustering is the **Pseudo-F statistic**, which compares the between-cluster and within-cluster variability to determine the optimal number of clusters (Orpin & Kostylev, 2006). The highest value of the Pseudo F-statistic indicates that the number of clusters used to partition the data is optimal (Solikin dkk., 2021). Below is the formula to calculate the Pseudo F-statistic value (Solikin dkk., 2021):

$$\text{pseudo } F - \text{statistics} = \frac{\left(\frac{R^2}{k-1}\right)}{\left(\frac{1-R^2}{n-k}\right)} \quad (4)$$

Where

$$R^2 = \frac{(SST-SSW)}{SST} \quad (5)$$

$$SST = \sum_{i=1}^n \sum_{j=1}^c \sum_{k=1}^p (x_{ijk} - \bar{x}_j)^2 \quad (6)$$

$$SSW = \sum_{i=1}^n \sum_{j=1}^c \sum_{k=1}^p (x_{ijk} - \bar{x}_{jk})^2 \quad (7)$$

After calculating the optimal number of clusters for each membership function, the optimum membership function is determined based on the Internal Cluster Dispersion (ICD) Rate.

$$ICD\text{rate} = 1 - \frac{(SST-SSW)}{SST} \quad (8)$$

2.2. Similarity Weight and Filter Method

Clustering in the Similarity Weight and Filter Method (SWFM) is similar to the ensemble concept but differs in the final cluster formation stage. In ensemble clustering, the final clustering uses an algorithm for categorical data. SWFM, on the other hand, applies an algorithm based on Similarity Weight and Filter Method. The SWFM stage involves a similarity measure that includes weighting factors based on the number of clusters formed, as formulated by Reddy & Kavitha

$$\text{sim}(X_i, X_j) = \sum_{i \leq n_i, j \leq n_j} \frac{S_{ij}}{\max(n_i, n_j)}, i \neq j \quad (9)$$

The Similarity Weight-based filter method algorithm can be used to cluster mixed data. The first step is to divide the data into sub-data for categorical and numerical data. Each sub-data is clustered based on its data type to produce appropriate clusters. The clustering results for numerical and categorical data are then combined to obtain the final clustering using the filter method (Reddy & Kavitha, 2012).

$$(X_i, X_j) = \sum_{i=1}^{n_i} \sum_{j=1}^{n_j} w_{ij} d(X_i, X_j) \quad (10)$$

2.3. Underdeveloped Areas

An underdeveloped area refers to a district whose region and community are less developed compared to other areas at the national level (R. Sari, 2014). As outlined in Presidential Regulation (Perpres) Number 63 of 2020 regarding the determination of underdeveloped areas for 2020-2024, article 1, paragraph (1), “An underdeveloped area refers to a district whose region and community are less developed compared to other areas on a national scale,” and article 2, paragraph (1), “An area is designated as underdeveloped based on criteria such as: the economy of the community, human resources, infrastructure, local financial capacity, accessibility, and regional characteristics.”

3. Research Method and Materials

The type of research to be used is quantitative research. This study will collect the necessary data, which will later employ the ensemble K-Modes and Similarity Weight and Filter Method (SWFM) clustering methods to identify the clustering of areas and the statistical results of the clustering of districts/cities in Central Sulawesi.

In this study, mixed data is used, consisting of both numerical and categorical scales. There are 13 variables employed, which are based on the criteria for underdeveloped areas. These variables include the percentage of the poor population, the Human Development Index (HDI), life expectancy at birth, the percentage of open unemployment, the percentage of villages with shops, the ratio of villages with the widest asphalt roads, the percentage of villages with hospitals, population density ratio, the percentage of villages not affected by disasters, the percentage of households with access to proper sanitation services, the presence of industrial areas, local fiscal capacity, and drinking water sources. These variables will be analyzed to identify the clustering of districts and cities in Central Sulawesi based on the underdeveloped area criteria.

3.1. Research Steps

The data analysis techniques in this study are as follows:

- (1). Describing statistics for districts/cities based on underdeveloped area indicators.
- (2). Clustering districts/cities based on underdeveloped area indicators using the Ensemble K-Modes mixed data method:
 - (a). Separation of numerical and categorical data.
 - (b). Reduction of numerical variables using factor analysis, selecting the highest variables from each factor formed. These selected variables will be used in the numerical data clustering analysis.
 - (c). Clustering of numerical data using Fuzzy C-Means, where the initial membership degrees are determined by calculating the distance between the objects and the average of each variable. The optimum number of clusters is determined by looking at the largest Pseudo-F value and ICD rate, which will produce an output with categorical scale.
 - (d). Clustering of categorical data using the K-Modes method, with the optimum number of clusters determined by the accuracy of the categorical data
- (3). Clustering districts/cities based on underdeveloped area indicators using the Similarity Weight and Filter Method (SWFM):
 - (a). Preparation of data from the merging results.
 - (b). Clustering using Similarity Weight and its filter algorithm.
 - (c). Distance matrices generated by the filter algorithm are clustered using the hierarchical method, specifically single linkage.
 - (d). Determining the optimum number of clusters by evaluating its accuracy.
- (4). Comparing the best clustering results from the Ensemble K-Modes method and the Similarity Weight and Filter Method (SWFM) using the highest accuracy.
- (5). Conclusion and recommendations

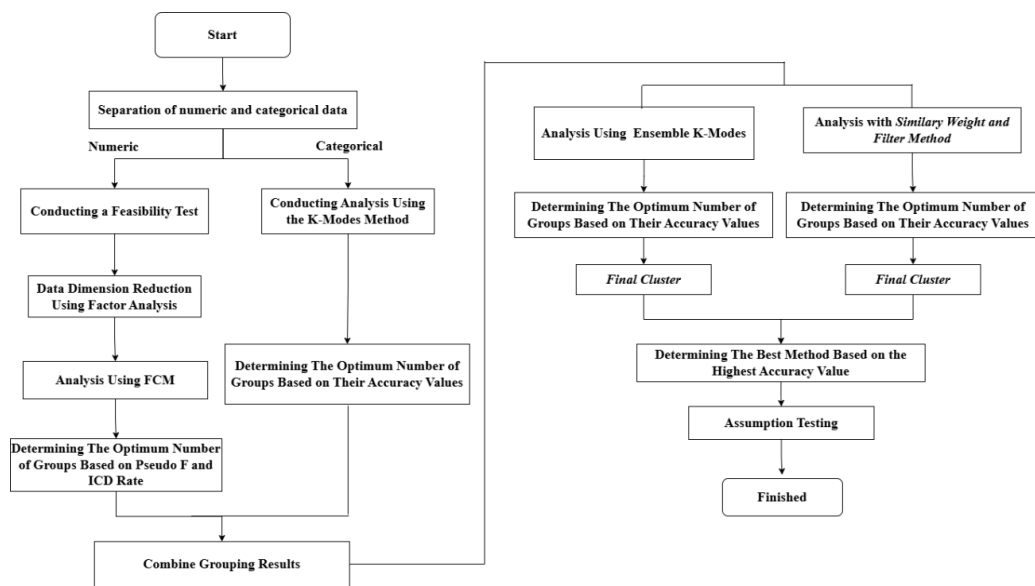


Figure 4. Data Analysis Techniques

4. Results and Discussion

4.1. Factor Analysis

In factor analysis, it is necessary to meet the assumptions of correlation adequacy and variable dependency. The assumption of correlation adequacy can be tested with the Kaiser Meyer Olkin (KMO) measure to determine whether the variables are sufficiently correlated for factor analysis. The KMO value should be greater than 0.50 for the analysis to proceed (Verdian, 2019).. In this study, the KMO value is 0.717, which is greater than 0.50, indicating that all variables in this study can be analyzed further.

Next, the assumption of variable dependency is tested. Variable dependency is examined using Bartlett's Test to determine whether the variables are correlated with each other. If the Bartlett's Test result is less than the 5% significance level ($\alpha = 0.05$), the variables are considered to be correlated.

Table 1. Bartlett's Test

Chi-Square	148.3935
P-Value	0.0001
decision	Correlated

From Table 1, Bartlett's Test shows that the P-Value (0.000) < α (0.05), indicating that the variables are correlated and can proceed to factor analysis.

Once all assumptions are met, factor analysis is conducted to identify factors that represent all the variables in the study. Factors with eigenvalues ≥ 1 are considered to adequately explain the variables and are included in forming the variables (Paruntu dkk., 2021).. The number of factors formed can be seen from the eigenvalue table.

Table 2. Eigenvalue

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	6.528	65.275	65.275
2	1.250	12.503	77.778
3	0.878	8.784	86.563
4	0.733	7.330	93.893
5	0.296	2.965	96.857



6	0.189	1.887	98.744
7	0.105	1.045	99.790
8	0.018	0.176	99.965
9	0.002	0.021	99.986
10	0.001	0.014	100

Based on Table 2, there are two components with eigenvalues greater than 1, indicating that two factors are formed. These two components account for a cumulative total variance of 77.78% for the variables. After identifying the components, the next step is to review the component matrix to identify the members of each component. This study uses variables with the highest loadings on the formed factors (Paruntu dkk., 2021).

Table 3. Component Matrix

	Component	
	1	2
X_1	-0.760	0.62
X_2	0.960*	0.156
X_3	0.620	0.630
X_4	0.828	0.268
X_5	0.951	-0.013
X_6	0.504	-0.808*
X_7	0.940	0.002
X_8	0.919	-0.037
X_9	-0.614	0.303
X_{10}	0.833	-0.093

Table 3 shows the component matrix, where the members of each component are determined based on the highest component loadings. Component 1 consists of the Percentage of Poor Population (X_1), Human Development Index (X_2), Open Unemployment Rate (X_4), Percentage of Villages with Shops (X_5), Percentage of Villages with Hospitals (X_7), Population Density Ratio (X_8), Percentage of Villages Without Disasters (X_9), and Percentage of Households with Access to Proper Sanitation (X_{10}). Component 2 consists of Life Expectancy at Birth (X_3) and the Ratio of Villages with the Widest Asphalt Roads (X_6).

From the component matrix, one variable with the highest component loading in each component is selected for clustering numerical variables. The variable with the highest loading in Component 1 is the Human Development Index (X_2), and in Component 2, it is the Ratio of Villages with the Widest Asphalt Roads (X_6).

4.2. Fuzzy C-Means

The next step is to cluster the regencies/cities using the Fuzzy C-Means method with two tests. The first test uses 10 numerical variables, and the second test uses the selected variables from the factor analysis. The goal is to determine the best clustering from these two analyses

Table 4. Pseudo F Statistics



	Number of Clusters	Membership Function		
		Linear Up	Linear Down	Triangular Curve
Without Factor Analysis	2	14.0749837391	14.0749837404*	14.0749837382
	3	11.0747819876	11.0747817106	11.0747819733
	4	8.4680872991	8.4680811502	9.2471975226
	5	9.7297312964	8.9597395313	9.0000967773
Factor Analysis	2	6.3815012664	6.3815014013	6.3814997394
	3	16.9421108148	16.9421108715	16.9421108317
	4	16.8080928048	16.8080915181	16.8080927573
	5	18.0533858792	18.0558219043	18.5689784334*

Table 4 shows the Pseudo F Statistics values, which are used to determine the optimal number of clusters; the higher the Pseudo F value, the more optimal the clusters. For the FCM method without factor analysis, the optimal cluster count is 2, whereas with factor analysis, the optimal cluster count is 5.

Table 5. ICD Rate Fuzzy C-Means

	Number of Clusters	ICD Rate
Without Factor Analysis	2	0.4386842
Factor Analysis	5	0.0972338

Table 5 shows the ICD Rate values, which provide information on the clustering method's performance in group analysis; a smaller ICD Rate indicates a better method, as it reflects smaller within-group differences. The analysis shows that the ICD Rate for the Fuzzy C-Means (FCM) method with factor analysis is lower than without factor analysis. Therefore, FCM clustering with factor analysis is chosen to reduce variables, apply the triangular membership function, and form five clusters.

Table 6. Group Members 5 Clusters Fuzzy C-Means

Cluster	Member
1	Kab. Donggala, Kab. Tojo Una-Una
2	Kab. Banggai Kepulauan, Kab. Toli-Toli, Kab. Buol, Kab. Parigi Moutong, Kab. Banggai Laut
3	Kota Palu
4	Kab. Morowali, Kab. Sigi, Kab. Morowali Utara
5	Kab. Banggai, Kab. Poso

4.3. K-Modes

The K-Modes method will be used to cluster categorical variables, with observation objects placed in the appropriate groups. This study will explore the number of clusters at 2, 3, 4, and 5 to determine the optimal number of clusters, using accuracy values to compare the effectiveness of each grouping.

Table 7. K-Modes Accuracy Value

Number of Clusters	Accuracy Value
2	0.8462
3	0.5962
4	0.2885
5	0.3462

From Table 7, the highest accuracy value for the K-Modes method is achieved with 2 clusters, which divides the regencies/cities in Central Sulawesi into two groups: one group consisting of 7 regions and the other consisting of 6 regions.

Table 8. Group Members 2 Cluster K-Modes

Cluster	Member
1	Kab. Banggai Kepulauan, Kab. Poso, Kab. Donggala, Kab. Buol, Kab. Tojo Una-Una, Kab. Sigi, Kab. Banggai Laut
2	Kab. Banggai, Kab. Morowali, Kab. Toli-Toli, Kab. Parigi Moutong, Kab. Morowali Utara, Kota Palu

4.4. Ensemble K-Modes

The optimal clustering results for numerical variables show that the best number of clusters is 5, while for categorical variables, the optimal number is 2. After clustering each variable type separately, the next step is to perform a combined clustering to obtain the final optimal grouping for the mixed data. The combined clustering is done by merging the clustering results of the numerical variable (X_1) and the categorical variable (X_2) to form a new variable. The K-Modes method is applied to obtain the final clustering results based on categorical data clustering principles.

Table 9. Ensemble K-Modes Accuracy Value

Number of Clusters	Accuracy Value
2	0.7949
3	0.8462
4	0.2308
5	0.6410

According to Table 9, the highest accuracy value for the Ensemble K-Modes method is achieved with 3 clusters, yielding an accuracy of 0.8462 (84.62%). This study selects 3 as the optimal number of clusters, dividing the regencies/cities in Central Sulawesi into two groups: one with 7 regions and the other with 3 regions.

Table 10. Group Members 2 Cluster Ensemble K-Modes

Cluster	Member
1	Kab. Banggai Kepulauan, Kab. Poso, Kab. Donggala, Kab. Buol, Kab. Tojo Una-Una, Kab. Sigi, Kab. Banggai Laut
2	Kab. Banggai, Kab. Morowali, Kab. Morowali Utara
3	Kab. Toli-Toli, Kab. Parigi Moutong, Kota Palu

4.5. Similarity Weight and Filter Method (SWFM)

In the SWFM method, mixed data clustering is conducted in two stages. The first stage involves clustering each scale using the Fuzzy C-Means method for numerical variables and the K-Modes method for categorical variables. The results of this clustering are treated as a new categorical variable, which is then analyzed using the SWFM method.

Table 11. SWFM Accuracy Value

Number of Clusters	Accuracy Value
2	0.5385
3	0.5769
4	0.7308
5	0.8077

According to Table 11, the accuracy values for the SWFM method across different cluster counts indicate that the selected number of clusters is 5, with an accuracy of 0.8077 (80.77%). This clustering divides the regencies/cities in Central Sulawesi into 5 groups, consisting of 8 regions, 1 region, 2 regions, 1 region, and 1 region, respectively

Table 12. Group Members 5 Cluster SWFM

Cluster	Member
1	Kab. Banggai Kepulauan, Kab. Poso, Kab. Donggala, Kab. Toli-Toli, Kab. Buol, Kab. Parigi Moutong, Kab. Tojo Una-Una, Kab. Banggai Laut
2	Kab. Banggai
3	Kab. Morowali, Kab. Morowali Utara
4	Kab. Sigi
5	Kab. Palu

4.6. Best Mixed-Data Clustering

The best clustering result is obtained by comparing the optimal values from mixed data clustering using the Ensemble K-Modes and SWFM methods. This comparison is based on the accuracy values of the clustering results; the higher the accuracy, the better the clustering quality.

Table 13. Comparison of Mixed Data Clustering Accuracy Values

Method	Number of Clusters	Accuracy Values
<i>Ensemble K-Modes</i>	3	0.8462
SWFM	5	0.8077

From Table 13, the comparison of mixed data clustering accuracy values shows that the highest accuracy is achieved with the Ensemble K-Modes method using 3 clusters, reaching 0.8462. Therefore, it can be concluded that the best mixed data clustering method for classifying the regencies/cities in Central Sulawesi based on underdeveloped area indicators is the Ensemble K-Modes method

5. Conclusion

The best mixed data clustering method shows that the Ensemble K-Modes method provides better results than SWFM, with the highest accuracy at 0.8462. Therefore, the clustering of regencies/cities in Central Sulawesi based on underdeveloped area indicators results in 3 groups. Statistics indicate that Cluster 1, covering 53.84% of the regencies/cities, is in the lowest condition and is categorized as underdeveloped.

The suggestion given by the author based on the results of this study is that the government must prioritize improving the economy and quality of human resources in each region in order to remove these regions from underdeveloped status. For further researchers, further exploration can be carried out on additional variables that have the potential to influence the underdeveloped status of the region.

References

- Anggara Cahya Putra, & Dwi Hartomo, K. (2021). Optimalisasi Penyaluran Bantuan Pemerintah Untuk UMKM Menggunakan Metode Fuzzy C-Means. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 5(3), 474–482. <https://doi.org/10.29207/resti.v5i3.2980>
- Arifah, E. D., Irawan, M. I., & Mukhlash, I. (2017). Aplikasi Metode Fuzzy Mamdani Dalam Penentuan Jumlah Produksi (Application of Fuzzy Mamdani Method in the Determination of Total Production). *Majalah Ilmiah Matematika dan Statistika*, 17(2), 79–90. <https://jurnal.unej.ac.id/index.php/MIMS/index>
- Costales, J., Catulay, J. J. J., Costales, J., & Bermudez, N. (2022). Kaiser-Meyer-Olkin Factor Analysis: A Quantitative Approach on Mobile Gaming Addiction using Random Forest Classifier. *ACM International Conference Proceeding Series*, March, 18–24. <https://doi.org/10.1145/3546157.3546161>
- Fajrianti, F., Bustan, M. N., & Tiro, M. A. (2019). Penggunaan Analisis Cluster K-Means dan Analisis Diskriminan Dalam Pengelompokan Desa Miskin di Kabupaten Pangkep. *VARIANSI: Journal*

- of Statistics and Its application on Teaching and Research, 1(2), 7.
<https://doi.org/10.35580/variasiunm9355>
- He, Z., Xu, X. i, & Shengchun Deng. (2005). Clustering mixed numeric and categorical data : A Cluster Ensemble Approach. Departement of Computer Science and Engineering, Harbin Institute of Technology Harbin 150001, P.R. China, 5(1), 1689–1699.
<https://doi.org/https://doi.org/10.48550/arXiv.cs/0509011>
- Himayati, H., Switrayni, N. W., Komalasari, D., & Fitriyani, N. (2020). Analisis Rotasi Ortogonal pada Teknik Analisis Faktor Menggunakan Metode Procrustes. *Eigen Mathematics Journal*, 1(2), 45.
<https://doi.org/10.29303/emj.v1i2.66>
- Huang, Z., & Ng, M. K. (1999). A Fuzzy k -Modes Algorithm for Clustering Categorical Data. 7(4), 446–452.
- Irabawati, N., Wahyuningsih, S., & Syoer, R. R. (2016). Perbandingan Metode C-Means dan Fuzzy C-Means Dalam Pengelompokan Wilayah Desa/Kelurahan di Kabupaten Kutai Kartanegara. *Jurnal Eksponensial*, 7(1), 67–76.
<http://jurnal.fmipa.unmul.ac.id/index.php/exponensial/article/view/47%0Ahttp://jurnal.fmipa.unmul.ac.id/index.php/exponensial/article/download/47/24>
- Kementerian Desa, P. D. T., & Indonesia, D. T. R. (2020). Ini Daerah Tertinggal Menurut Perpres. 2024 Kemendes.go.id. <https://www.kemendes.go.id/berita/view/detil/3261/ini-daerah-tertinggal-menurut-perpres>
- Kusuma, D. T., & Sangadji, I. B. M. (2016). Segmentasi Penilaian Kompetensi Alumni Stt-Pln Menggunakan Model Klaster Fuzzy Clustering Means (Fcm). *Kilat*. <https://stt-pln.ejournal.id/kilat/article/view/685>
- Mara, N., & Intisari, N. S. (2013). Pengklasifikasian Karakteristik Dengan Metode K-Means Cluster Analysis. In *Buletin Ilmiah Mat. Stat. dan Terapannya (Bimaster) (Vol. 02, Nomor 2)*.
- Muhardi, M. (2019). Penentuan Penerima Beasiswa Dengan Algoritma Fuzzy C-Means. *Jurnal Informasi dan Komputer*, 7(2), 37–46. <https://doi.org/10.35959/jik.v7i2.148>
- Nurfitri Imro'ah, I. A. N. N. D. (2019). Analisis Cluster Non-Hirarki Dengan Metode K-Modes. *Bimaster: Buletin Ilmiah Matematika, Statistika dan Terapannya*, 8(4), 909–916.
<https://doi.org/10.26418/bbimst.v8i4.36633>
- Orpin, A. R., & Kostylev, V. E. (2006). Towards a statistically valid method of textural sea floor characterization of benthic habitats. 225, 209–222.
<https://doi.org/10.1016/j.margeo.2005.09.002>
- Paruntu, J. G., Hatidja, D., & Langi, Y. A. . (2021). Keputusan Mahasiswa dalam Memilih Indekos dengan Analisis Faktor. *Jurnal Ilmiah Sains*, 21(2), 119.
<https://doi.org/10.35799/jis.v21i2.28868>
- Purnamasari, S. M. (2011). Analisis Kelompok (Cluster Analysis). Program, 18209007, 1–6.
<http://informatika.stei.itb.ac.id/~rinaldi.munir/Probststat/2010-2011/Makalah2010/MakalahProbststat2010-030.pdf>
- Purnomo, & Sutadji, E. (2022). Analisis Data Multivariat. Omera Pustaka.
- Purwanto. (2004). Analisis Faktor : Konsep, Prosedur Uji dan Interpretasi. Analisis Faktor: Konsep, Prosedur Uji dan Interpretasi, 12, 153–169.
- Putri Kartikasari, Y. A., Agus Pranoto, Y., & Rudhistiar, D. (2021). Penerapan Metode K-Modes Untuk Proses Penentuan Penerima Bantuan Langsung Tunai (Blt). *JATI (Jurnal Mahasiswa Teknik Informatika)*, 5(1), 389–397. <https://doi.org/10.36040/jati.v5i1.3300>
- Reddy, J., & Kavitha, B. (2012). Clustering the Mixed Numerical and Categorical Datasets Using Similarity Weight and Filter Method. *International Journal of Database Theory and Application*, 5(1), 121–133.
- Rencher, A. C. (1998). *Multivariate Statistical Inference and Application*. Wiley-Interscience Publication.

- Sari, H. L., & Suranti, D. (2016). Perbandingan Algoritma Fuzzy C-Means (FCM) Dan Algoritma Mixture Dalam Penclusteran Data Curah Hujan Kota Bengkulu. *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, 7–15.
- Sari, R. (2014). Dampak Kebijakan Desentralisasi Fiskal pada Daerah Tertinggal di Indonesia. *Jurnal Ekonomi dan Kebijakan Publik*, 5(1), 79–99.
- Science, D., Baharu, F., Ngurah, G., Wibawa, A., Abapihi, B., & Ihwal, M. (2023). Cluster Analysis On Senior High Schools Of Kendari Using. 1(1), 18–26.
- Setiawan, D., Juwita, R., Lindasari, D., & Rahayu, S. P. (2017). Analisis MANOVA Satu Arah pada Data Komponen Kesehatan Bayi di Pulau Jawa pada Tahun 2013. March.
- Solikin, A. F., Wibowo, F. W., & Kusriani. (2021). Analisis Cluster Data Interkomparasi Anak Timbangan dengan Algoritma Self Organizing Maps. 7, 435–448.
- Suguna, J., & Arul Selvi, M. (2012). Ensemble Fuzzy Clustering for Mixed Numeric and Categorical Data. *International Journal of Computer Applications*, 42(3), 19–23.
<https://doi.org/10.5120/5672-7705>
- Suhaeni, C., Kurnia, A., & Ristiyanti, R. (2018). Perbandingan Hasil Pengelompokan menggunakan Analisis Cluster Berhierarchy, K-Means Cluster, dan Cluster Ensemble (Studi Kasus Data Indikator Pelayanan Kesehatan Ibu Hamil). *Jurnal Media Infotama*, 14(1).
<https://doi.org/10.37676/jmi.v14i1.469>
- Syahputri, N. I., Chiuloto, K., & Harahap, N. N. A. (2022). Analisa Perbandingan Membership Function Fuzzy Tsukamoto dalam Menentukan Dosen Berprestasi. *Blend Sains Jurnal Teknik*, 1(2), 164–170. <https://doi.org/10.56211/blendsains.v1i2.134>
- Tiro, M. A., & Rasyid, A. (2018). *Himpunan dan Logika Kabur*. Andira Publisher Makassar.
- Triyanto, W. A. (2015). Algoritma K-Medoids Untuk Penentuan Strategi Pemasaran Produk. *Simetris : Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, 6(1), 183.
<https://doi.org/10.24176/simet.v6i1.254>
- Verdian, E. (2019). Analisis Faktor yang Merupakan Intensi Perpindahan Merek Transportasi Online di Surabaya. *Agora*, 7(1).
- Yulianto, S., & Hidayatullah, K. H. (2014). Analisis Kluster Untuk Pengelompokan Kabupaten/Kota Di Provinsi Jawa Tengah Berdasarkan Indikator Kesejahteraan Rakyat. *Jurnal Statistika*, 2(1), 1–8.