APPLICATION OF THE K-MEANS CLUSTERING METHOD FOR PERFORMANCE ASSESSMENT BASED ON THE COMPETENCY OFEDUCATORS

*Indra Wirayudha Napitupulu[1], Cristian Vieri Sinaga[2], Adzisyah Rahman[3], Steven[4], Marlince Nababan[*5]*
[1,2,3,4,5]Indonesian Prima University
Jl. Sampul No. 1, Sei Putih Tengah, Kec. Medan Petisah, Medan City, North Sumatra 20118E-mail:marlince@unprimdn.ac.id

ABSTRACT- The ability of lecturers to teach for 1 (one) semester needs to be evaluated. Evaluation is a studentassessment through a student academic information system. Review using the system is separate because many students need to care about the filled things. There the author tries to analyze the values from the questionnaire results distributed to students using the Principal Component Analysis (PCA) technique and the K-Means method, where PCA reduces data. At the same time, K-Means assumed the values closest to the study's results and concluded that of the 3 (three) clusters, the maximum distance was 0.5475 in cluster 1 and cluster 3.

Keywords: Data Mining, K-Means, Performance.

1. INTRODUCTION

The field of cooperation is a quality resource— fulfilling the competence of today's college graduates. The Merdeka Campus distributes opportunities for students to take courses outside theirstudy program, study program, and activities outsidethe program, including internships or work practices.Merdeka Learning - Merdeka Campus as an effort toprovide learning choices. MBKM is a policy of the Minister of Education and Culture, which aims for students to explore knowledge helpful in entering theworld of work.

The lecturer performance evaluation process still uses the manual method, so errors occur, such as when filling in the results of lecturer performance. K-Means clustering has the concept that this method is non-hierarchical for grouping data, in which the function is to minimize and maximize variation in onegroup. K-Means clustering is used to cluster the results of lecturer performance evaluations that refer to the four essential competencies, which begin by determining the K value and then calculating thedistance between each data and the Euclidean concept[1].

This research developed a classification method based on K-Means and LVQ grouping. The steps forcollection are: grouping with K-Means. The group continues until it reaches a threshold (certain limits).If the point has been reached and there are still different classes in one cluster, learning is done usingLVQ. The combined accuracy of K-Means and LVQis better than with pure K-Means. The highest averageaccuracy for K-Means and LVQ is 92%, while for pure K-Means, 82%[2].

Academic Assistant Lecturers (DPA) 'sperformance in implementing the Merdeka Campus Program has been carried out every semester.Processing the assessment data manually requires quite a long time. To get good results. Lecturer data was taken as many as 28 data which were then processed using the K-Means method using the RapidMiner Software with the performance evaluation data variable for Academic Assistant Lecturers (DPA). The results of the Academic Lecturer Performance Assessment (DPA) with this method. The K-Means algorithm can help decision- making to make recommendations for assessing lecturer

performance in the following semester (Siti).Data Mining with the K-Means Clustering method, where data is grouped based on the same characteristics, will be entered into groups.[3].

The K-Means method can be applied to determineclothing sales. Where is the function of the k-means method to classify those who are very in demand, in demand, and less in order [4] [5].

The Text Mining method can group these objects into clusters formed by the clustering algorithm. Thisstudy aims to compare 2 clustering algorithms, namely the K-Means and K-Medoids algorithms, which aim to get the proper evaluation in terms of performance and computation time[6] [7].

K-Means Clustering is the application of datamining techniques for decision-making using RapidMiner software. The grouping method used is K-Means Clustering with 3 clusters: on time, notcorrect, and stop out with the application of K-MeansClustering measurement decision making[8][9]. By applying the K-Means algorithm, which is a centroidmodel for predicting student graduation time based onGPA and attendance parameters, it is known that the centroid model is a method that uses a centroid to create clusters and is also a non-hierarchical clustering method[10][11].

Lecturer research onclassification based on scientific work or research published nationally and internationally. Due to the increase in the results of publications and studies by lecturers to deal with the life factors in society. The author applies the K-Means Clustering method, and the method used is literature study, processing research data, designing systems, and testing systems. Prima Indonesia University is one of the largest campuses in Medan, with forty (40) study programs and approximately 600 resources (lecturers). Every semester an evaluation of the lecturer's performance will be carried out. Evaluation is done at the end of the semester.

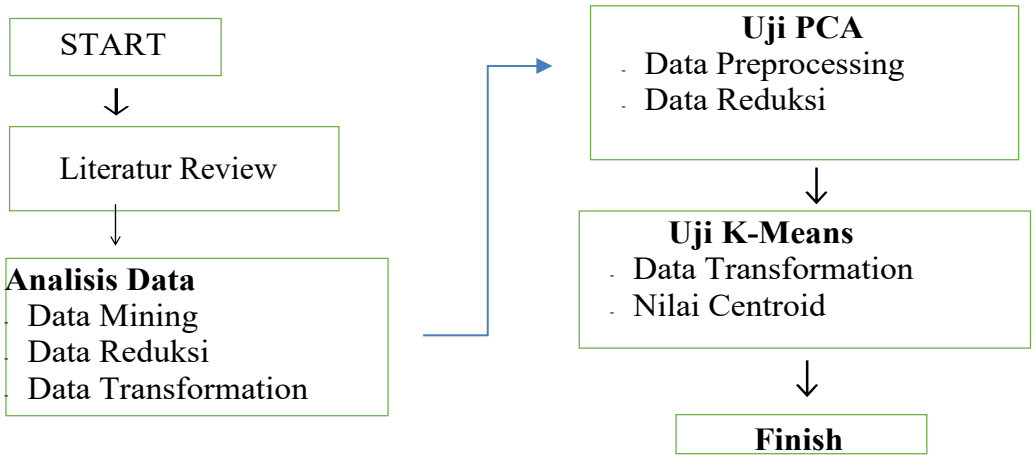
The ability of lecturers to teach for 1 semester is assessed through a system that students evaluate. Assessment using the system is not joint because many students don't care about the filled things. Some authors try to analyze the values from the results of the questionnaires distributed to students using the K-Means method so that the authors determine the topic "Application of PCA Techniques and the K-Means Method in Evaluation Analysis of Lecturer Performance Assessment. "

2. RESEARCH METHODS

2.1 Literature review

A literature study or literature review is a research method that collects, evaluates, and summarizes literature or reference sources relevant to the topic or research question being studied.

Reference sources used in literature studies can be journal articles, books, theses, dissertations, or other information sources relevant to the research topic in Figure 2.1.



2.2. Data Mining

Data mining is part of the Knowledge Discovery in Database (KDD) process. With data mining, we can classify, predict, estimate, and get other information.Many can be used in data mining.

2.3. PCA technique

PCA in a set of N test samples where each sample testis a set of n, i.e., training xi is denoted by:

$$xi = (xli,x2i, \ldots, xni) ; i \in 1,N$$

[1] The mean or average of a set of training samples is mathematically expressed:

$$x = 1/N \Sigma\_(i=1)\text{^}n \; xi \ldots\ldots. [2]$$

Then the deviation can be calculated from the mean of each sample tested:

$$xi = xi - x \ldots\ldots. [3]$$

[3]This deviation shows how far each sample differs from the mean at a point. The 4n×4n covariancematrix of S shows the extent to which all the differences come from the standard as:

$$s= 1/N \sum\_(i=1)\text{^}n\delta xi \; \delta x\_i\text{^}T \ldots\ldots [4]$$

This covariance matrix component shows the numberof samples that exist. Sample form as:

$$xi = x + Pw[5]$$

P = (p1, p2,...pt) is the matrix of the first eigenvector,and w= (w1, w2, ..., wt) T is the weight vector.

2.4. K-Means method

The algorithm used. Namely, K-Means Clustering is the shortest distance value system [2]. The stages in clustering the K-Means method are:
Determine the value of K, where K is the number ofclusters to be formed.
Determine the initial center point of each group. Calculate the distance for each input data for eachcentroid using the Euclidean Distance formula:
$D(x,y)=\sqrt{((X\_(1-) Y\_1)+(X\_2-Y\_2))2}$

(1) Information;
D = Distance
x = Data
y = Centroids

Clustering data based on their proximity to thecentroid.
Recalculate the cluster center with the current clustermembers.
Compute each object again using the new clustercenter.

2.5. Minitab Software

This research uses Minitab software where thissoftware is open-source.

3. RESULTS AND DISCUSSION

3.1. Literature review

This study uses questionnaire data from students in Medan. The divided questionnaire was 20 questions about the learning system in the network in Table 3.1.

Table 3.1. Questionnaire Variables

Number	Information
1	Mastering the Classroom Atmosphere.
2	Communicative Or Easy To Understand.
3	In accordance with the conceptstaught.
4	Target Reached
5	Answer Student Questions.
6	Have Authority
7	Wisdom in Making Decisions.
8	Have Ethics
9	Can Control Yourself
10	Be an example of attitude andbehavior.
11	Conformity of Material with CourseObjectives.
12	Knowledge Field Ability
13	Ability to Manage ClassroomLearning
14	It can Give Student Motivation.
15	Capacity in Curriculum Development.
16	Able to Receive Criticism, Suggestions, And Opinions FromStudents.
17	Friendliness and Politeness inCommunicating.
18	Consistent Words In Action.
19	Not Racist
20	Lecturer Appearance

The sample questionnaire distributed to students is anexample in Table 3.2.

Table 3.2. Sample Student Questionnaire

Menguasai Suasana Kelas	7	5	5	5	5	7	7	9	3	7	3	5	3
Komunikatif Atau Mudah Dipahami.	7	5	5	7	5	7	7	9	5	7	3	9	5
Sesuai Dengan Konsep Yang Diajarkan	9	7	5	3	5	3	7	9	7	7	3	9	5
Tercapai Sasaran	9	7	5	3	5	3	7	9	7	7	3	9	5
Menjawab Pertanyaan Mahasiswa/ Siswi.	3	5	5	3	5	7	7	7	7	7	3	7	7
Memiliki Kewibawan	9	9	5	5	5	7	9	9	7	7	5	9	7
Kearifan Dalam Mengamb Il Keputusan	9	7	5	5	5	7	5	9	5	7	5	9	5
Memiliki Etika	9	7	5	7	7	7	7	9	7	7	3	9	9
Dapat Mengendalikan Diri	7	7	5	5	5	7	7	9	5	7	5	9	5
Menjadi Contoh Bersikap Dan Berperilaku.	7	5	5	5	5	7	7	9	5	7	5	9	5
Kesesuaian Materi Dengan Tujuan Mata Kuliah.	7	5	5	5	7	5	7	9	5	7	5	9	5
Kemampuan Bidang Ilmu	9	5	5	5	5	5	7	9	5	7	5	9	5
Kemampuan Mengelola Pembelajarandi Kelas	9	7	5	5	5	7	7	9	7	7	5	9	5
Dapat Memberi Motivasi Mahasiswa.	9	7	5	5	5	5	7	9	5	7	5	9	5
Kemampuan Dalam Pengembangan Kurikulum	9	7	5	5	7	5	7	9	5	7	3	7	7
Mampu Menerima Kritik, Saran, Dan Pendapat	3	7	5	5	5	3	5	7	7	7	5	7	5
Keramahan Dan Kesopanan Dalam Berkomunikasi.	9	7	5	7	7	7	7	9	7	7	7	9	5
Konsisten Kata- Kata Dalam Tindakan	9	5	5	5	5	7	5	9	5	7	3	9	5
Tidak Rasis	9	9	5	5	7	7	9	9	7	7	5	9	7
Penampilan Sesuai Dosen	9	7	7	5	5	7	9	9	7	7	7	7	7

3.2. Data Mining

Data Mining is a technology containing a large amount of data that produces information or as decision support. Meanwhile, data mining has 7 (seven) functions: description, classification, clustering, association, sequencing, forecasting, and prediction. Some researchers define 3 (three) goals: strengthening the analyzing hypothesis and finding new, unexpected relationships.

3.2 Principal Component Analysis (PCA)

PCA is a mathematical method that can be applied to reduce the number of dimensions in a data set. PCA tasks can clarify quickly to visualize and analyze. As for the formula.

$$X = \frac{\sum_{i=1}^{m} X_1}{m} \quad (1)$$

Then determine the covariance matrix;

$$C_y = \frac{1}{m} YY^T \quad (2)$$

$$Y = (X - X) \quad (3)$$

The associated eigenvector values are entered into the transformation matrix W by projecting the original data X.

$$Y = WX \quad (4)$$

3.4. Data analysis using PCA

There are 20 (twenty) variables used so that when PCA is reduced, it produces values , as shown in Table 3.3. Where the highest eigenvalue is 5.8 while the lowest eigenvalue is 0.7126. In the second set, thehighest eigenvalue is 0.6, while the minimum value is 0.3. While the cumulative value for the first component is 0.738. In calculating the proportion of the maximum value in set 1 is 0.736 as below:

- Proporsi kumulatif komponen utama pertama = 0,291
- Proporsi kumulatif komponen utama kedua = 0,291 + 0,064 = 0,355
- Proporsi kumulatif komponen utama ketiga = 0,355 + 0,060 = 0,415
- Proporsi kumulatif komponen utama keempat = 0,415 + 0,056 = 0,471
- Proporsi kumulatif komponen utama kelima = 0,471 + 0,051 = 0,522
- Proporsi kumulatif komponen utama keenam = 0,522 + 0,050 = 0,572
- Proporsi kumulatif komponen utama ketujuh = 0,572 + 0,048 = 0,620
- Proporsi kumulatif komponen utama kedelapan = 0,620 + 0,044 = 0,663
- Proporsi kumulatif komponen utama kesembilan = 0,663 + 0,038 = 0,701
- Proporsi kumulatif komponen utama kesepuluh = 0,701 + 0,036 = 0,736

Table 3.3. Eigenanalysis of the Correlation Matrix

Eigenvalue	5,8206	1,2802	1,1949	1,1201	1,0274	0,9941	0,9527	0,8707	0,7501	0,7126
Proportion	0,291	0,064	0,060	0,056	0,051	0,050	0,048	0,044	0,038	0,036
Cumulative	0,291	0,355	0,415	0,471	0,522	0,572	0,620	0,663	0,701	0,736
Eigenvalue	0,6932	0,6779	0,6479	0,5748	0,5332	0,4956	0,4614	0,4544	0,3979	0,3401
Proportion	0,035	0,034	0,032	0,029	0,027	0,025	0,023	0,023	0,020	0,017
Cumulative	0,771	0,805	0,837	0,866	0,893	0,917	0,940	0,963	0,983	1,000

The eigenvectors of the 20 variables used in PCA form 20 PCAs, namely PC1, PC2, and so on. Of the20 PCs, there is 1 PC that influences it, namely PC1 of 0.253 against PC4 of 0.301, as in table 3.4, and the variable that does not have the most influence isPC20, as shown in Figure 3.1.

Table 3.4. Eigenvectors Value

Variables	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Master the class atmosphere.	0.253	-0.156	0.100	0.301	-0.200	-0.226	-0.047
Communicative	0.226	-0.214	0.317	0.298	0.002	-0.136	0.276
Directed/clear	0.207	-0.284	0.078	0.396	0.112	0.388	-0.139
Target achieved	0.263	-0.102	0.169	0.053	0.062	-0.170	0.121
Ability to answer questions	0.192	-0.145	-0.055	-0.266	0.351	-0.369	0.300
Have authority	0.248	-0.050	-0.109	0.056	-0.148	0.141	0.331
Wisdom in the making decisions	0.214	-0.004	-0.435	0.149	0.324	-0.102	-0.169
Have Ethics	0.219	0.106	-0.219	0.025	0.315	-0.095	0.298
Can control yourself	0.229	-0.001	-0.117	0.196	0.129	0.294	-0.159
Be an example of attitude and be	0.216	-0.093	-0.292	0.030	-0.267	-0.138	0.058
Suitability of the material with the purpose	0.229	0.077	-0.423	-0.188	-0.298	0.038	-0.154
Material suitability	0.243	-0.269	-0.143	-0.027	-0.101	0.096	-0.183
Knowledgeability	0.232	0.137	0.218	-0.134	0.124	0.393	0.006
Ability to manage classes	0.250	-0.211	0.140	-0.404	-0.027	0.023	-0.101
Motivating students.	0.222	-0.231	0.202	-0.523	-0.136	0.111	0.002
Ability in my development	0.167	0.146	0.366	-0.004	0.347	-0.211	-0.454
Able to accept criticism, suggestions, d	0.221	0.349	-0.061	-0.023	0.025	-0.306	-0.269
Friendliness and courtesy in b	0.245	0.326	0.056	-0.051	-0.109	0.126	-0.188
Not Racist	0.176	0.489	0.031	-0.040	0.139	0.302	0.393
Good Appearance	0.192	0.334	0.234	0.173	-0.469	-0.217	0.085

Variables	PC8	PC9	PC10	PC11	PC12	PC13	PC14
Master the class atmosphere.	0.082	-0.039	-0.191	0.309	-0.127	-0.268	-0.093
Communicative	0.220	-0.061	-0.046	-0.384	-0.234	0.020	0.154
Directed/clear	0.045	-0.104	-0.197	-0.275	0.071	0.260	-0.112
Target achieved	-0.048	0.175	0.241	0.016	0.547	0.041	-0.355
Ability to answer questions	-0.253	-0.250	-0.053	-0.148	-0.149	-0.251	-0.378
Have authority	-0.116	-0.270	0.169	0.408	-0.234	0.373	0.024
Wisdom in the making decisions	0.256	-0.132	-0.344	0.189	0.194	0.135	0.016
Have Ethics	0.394	0.369	0.187	0.118	-0.076	-0.133	0.390
Can control yourself	-0.298	0.539	0.217	0.006	-0.255	-0.133	-0.272
Be an example of attitude and be	-0.508	0.153	-0.123	-0.329	0.027	-0.054	0.460
Suitability of the material with the purpose	0.087	-0.061	-0.302	-0.088	-0.124	-0.114	-0.241

Material suitability	-0.093	-0.318	0.517	0.205	0.152	-0.109	0.153
Knowledgeability	-0.030	-0.284	-0.104	-0.056	0.287	-0.497	0.277
Ability to manage classes	0.075	0.350	-0.213	0.072	0.277	0.160	0.043
Motivating students.	0.180	0.061	-0.020	-0.015	-0.268	0.230	0.039
Ability in my development	-0.303	-0.026	-0.140	0.268	-0.267	0.084	0.206
Able to accept criticism, suggestions, d	0.042	-0.152	0.303	-0.374	0.096	0.371	0.049
Friendliness and courtesy in b	0.313	-0.057	0.210	-0.173	-0.245	-0.226	-0.187
Not Racist	-0.220	-0.035	-0.179	0.027	0.047	0.227	-0.076
Good Appearance	0.006	0.100	-0.126	0.187	0.164	-0.036	-0.032

Variables	**PC15**	**PC16**	**PC17**	**PC18**	**PC19**	**PC20**
Master the class atmosphere.	-0.105	0.494	0.379	0.123	-0.255	-0.024
Communicative	-0.054	0.039	-0.103	-0.223	0.450	-0.270
Directed/clear	0.255	-0.136	-0.124	0.202	-0.415	0.087
Target achieved	-0.411	0.104	-0.342	0.029	0.004	0.163
Ability to answer questions	0.260	-0.213	0.101	0.056	-0.111	0.025
Have authority	-0.314	-0.320	0.074	-0.233	-0.173	-0.057
Wisdom in the making decisions	-0.066	-0.125	0.164	0.014	0.442	0.253
Have Ethics	0.156	0.020	-0.210	0.046	-0.329	0.005
Can control yourself	0.101	-0.005	0.190	-0.308	0.189	0.083
Be an example of attitude and be	-0.213	-0.073	0.027	0.267	-0.000	0.149
Suitability of the material with the purpose	-0.020	0.160	-0.499	-0.278	-0.093	-0.235
Material suitability	0.374	0.140	-0.123	0.199	0.257	-0.196
Knowledgeability	-0.114	-0.031	0.110	-0.369	-0.092	0.139
Ability to manage classes	0.033	-0.203	0.298	0.030	-0.029	-0.532
Motivating students.	0.076	0.228	-0.018	0.033	0.098	0.566
Ability in my development	-0.086	-0.039	-0.354	0.042	-0.020	-0.070
Able to accept criticism, suggestions, d	0.074	0.187	0.294	-0.284	-0.211	-0.005
Friendliness and courtesy in b	-0.300	-0.316	0.086	0.490	0.063	-0.051
Not Racist	0.115	0.405	0.014	0.294	0.173	-0.166
Good Appearance	0.474	-0.341	-0.075	-0.068	0.082	0.210

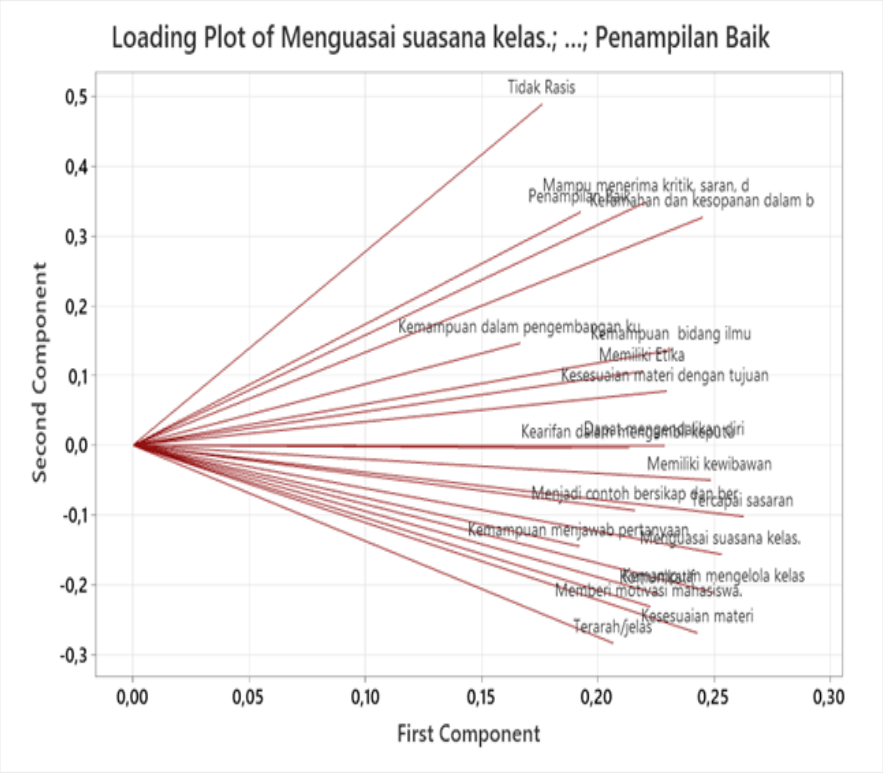


Figure 3.1. Loading Plot PC

3.5. Data Analysis Using C-Means

In this study, researchers used variables from thePC1, PC2, PC4, PC5, and PC6 results. The K-meansmethod describes the number of observations in eachcluster, namely the total sum of squares, the distancefrom the centroid, and the maximum value of the centroid. This study uses the value of K = 3. Cluster 1 has 7 (seven) observations with a total sum ofsquares value of 0.794, centroid distance of 0.33, andmaximum centroid value of 0.406.

Cluster 2 has 9 (nine)words with a total sum of squares value of 1.286, centroid distance of 0.364, and maximum centroid value of 0.571. Cluster 3 has 4 (four) observations with a total sum of squares value of 0.491, centroid distance of 0.323, and maximum centroid value of 0.472, as shown in table 3.5.

Table 3.5 Final Partition Table

	Number of Observations	Within cluster sumof squares	The average distance from the centroid	Maximu m distance from the centroid
Cluster1	7	0.794	0.030	0.406
Cluster2	9	1,286	0.364	0.571
Cluster3	4	0.491	0.323	0.472

Table 3.6 explains the average in each cluster (variable dimension) where for cluster 1 (one), the centroid value is 0.22 in C21, in C24, the centroid value is 0.426, the centroid value for C22 is 0.1124, the variable C25 is -0.2093 and unstable C26 is - 0.0831.

Table 3.6. Distances Between Cluster Centroids

	Cluster 1	Cluster 2	Cluster 3
Cluster1	0.000	0.4167	0.5474
Cluster2	0.4167	0.000	0.5111
Cluster3	0.5474	0.5111	0.000

Distances Between Cluster Centroids is a distance matrix between cluster centroids using the K-Means method, for example:

- The distance between cluster 1 and cluster 1 = 0.0000
- The distance between cluster 1 and cluster 2 = 0.4167
- The distance between cluster 1 and cluster 3 = 0.5475.
- So from 3 (three) groups, the maximum distance is 0.5475 in cluster 1 and cluster 3.

CONCLUSIONS ANDRECOMMENDATIONS

Conclusions

There is a conclusion from the author, namely:

1. The number of observations in the cluster affects a lot of data.
2. The sum of squares (SSE) describes the closest value of the data points in the cluster to the centroid.
3. 3. Lower values indicate denser clusters.
4. The maximum distance in the centroid matrix is cluster 1 and cluster 3

Recommendations

There is a suggestion from the author, namely:

1. Increase the number of data sets
2. To reduce the data set, change the PCAtechnique to analytical techniques.
3. Testing data analysis using information-basedsystems

REFERENCE

[1] KK Silalahi, “Simki-Techsain Vol. 01 No. 01 of2017 ISSN : XXXX-XXXX,"*Simki-Techsin*, vol. 01, no. 01, pp. 1–7, 2017.

[2] D. Eka Ratnawati, M., and L. Muflikhah,"Development of Classification MethodsBased on K-Means and LVQ," *J. Technol. inf. and Computing Science.*, vol. 1, no.1,p.1,2014,doi:10.25126/jtiik.20141197.

[3] R. Helilintar and IN Farida, "Application of the K- Means Clustering Algorithm for Prediction ofStudent Achievement Academic Values," *J. Science and Inform.*, vol. 4, no. 2, pp. 80–87, 2018, doi: 10.34128/jsi.v4i2.140.

[4] Normah, B. Rifai, S. Vambudi, and R. Maulana, "Sentiment Analysis of Vtuber Development Using the SMOTE-Based Support Vector Machine Method,"*J.Tek. Computer. AMIK BSI*, vol. 8, no. 2, pp. 174–180, 2022, doi: 10.31294/jtk.v4i2.

[5] E. Elhaik, *Principal Component Analyses (PCA)-based findings in population genetic studies are highly biased and must bereevaluated*, vol. 12, no. 1. Nature PublishingGroup UK, 2022.

[6] Karno Nur Cahyo, Agus Subekti, and Muhammad Haris, "Analysis of Computational Performance and Speed of K-Means and K- Medoids Algorithms in Text Clustering," *Pixel J. Ilm. computer. Graph.*, vol. 15, no. 2, pp. 444–464, 2022, doi: 10.51903/pixel.v15i2.931.

[7] HG Gauch, S. Qian, HP Piepho, L. Zhou, and R. Chen, "Consequences of PCA graphs, SNP codings, and PCA variants for elucidating population structure," PLoS One, vol. 14, no. 6, pp. 1–26, 2019, doi:10.1371/journal.pone.0218306.

[8] A. Rahmini Azhara, R. Anantadira Nugraha Faculty of Economics and Business, J. Information Systems, and A. Piqri Faculty of Economics and Business, "Application of K-Means Clustering Level of Timeliness of Graduation of Informatics Engineering Study Program Students Stmik Nurdin Hamzah Jambi," J. SYMMETRIC, vol. 12, no. 1, pp. 1–15, 2021.

[9] R. Guerra-Urzola, NC de Schipper, A. Tonne, K. Sijtsma, JC Vera, and K. Van Deun, “Sparsifying the least-squares approach to PCA: comparison of lasso and cardinality constraints,” Adv. Anal Data. Classif., vol. 17, no. 1, pp. 269–286, 2023, doi: 10.1007/s11634-022-00499-2.

[10] H. Priyatman, F. Sajid, and D. Haldivany, "Clusterization Using the K-Means Clustering Algorithm to Predict Student Graduation Time,"J. Education and Researcher. inform., vol. 5, no. 1, p. 62, 2019, doi: 10.26418/jp.v5i1.29611.

[11] Ismai, "Website Design as Promotional and Information Media," J. Inform. Pelita Nusant., vol. 3, no. 1, pp. 82–86, 2018.