Available at https://journalenrichment.com/index.php/jr/

Enrichment: Journal of Multidisciplinary Research and Development

Grouping Education Students at Pusdikjas Institutions of The TNI-AD's Disjasad Using the K-Means Clustering Method

Imam Wibowo Pramudjianto[1*], Ade Kania Ningsih[2], Agus Komarudin[3]

[1*,2,3]Universitas Jenderal Achmad Yani, Indonesia

*Email: imamwibowo@gmail.com

ARTICLE INFO

Keywords:
education center, K-Means, clustering method.

ABSTRACT

The Military Physical Education Center, known by the abbreviation Pusdikjasmil, is an implementation body that is directly under the TNI AD. Pusdikjas has the task of providing education in the military physical field to support the duties of the Indonesian Army. At the Pusdikjas agency, student data is obtained from the addition of students attending education each year. The process of admitting students to Pusdikjas institutions produces abundant value data and is repeated every year. By using data mining techniques, abundant student data can be grouped and analyzed to find information that is useful for Pusdikjas agencies. The algorithm used in this research is K-means clustering. The data used from 2018 - 2022 was 1428 data on coaching education students, with the variables used being NRP, attitude & and behavior, knowledge & and skills, and physical. Processed using the JavaScript programming language, this research produces optimal clusters, namely 2 clusters. cluster 1 with a total data of 751 students in cluster 1 describes a very satisfactory category of students, and cluster 2 with total data of 676 students. Cluster 2 describes the satisfactory student category. As for testing the quality of the clusters contained in the system using silhouette with results reaching 0.6543347543126016.

INTRODUCTION

The Military Physical Education Center, known by the abbreviation Pusdikjasmil, is an implementation body that is directly under the TNI AD. Pusdikjas has the task of providing education in the military physical field to support the duties of the Indonesian Army (Warih Widiyono & Aris Harliadi, 2011).

At the Pusdikjas agency, student data is obtained from the addition of students attending education each year. The process of admitting students to Pusdikjas institutions produces abundant value data and is repeated every year. By using data mining techniques, abundant student data can be grouped and analyzed to find information that is useful for Pusdikjas agencies. The method used to divide data into several groups is by using the K-Means algorithm (Penentuan et al., 2015). K-Means only takes an average of 1 second while K-medoids take an average of 1 minute 30 seconds (Handayani, 2022). Apart from that, this method is flexible because it can determine the number of clusters to be created. The K-Means algorithm is a non-hierarchical algorithm derived from the data mining clustering method (Nurhayati, 2022). The K-Means algorithm starts by forming a cluster partition at the beginning, then this cluster partition is iteratively corrected until there are no significant changes to the cluster partition (Sulistiyawati & Supriyanto, n.d.)

Several studies have been carried out regarding the use of the k-means algorithm and validation tests using the silhouette coefficient as a reference for further research. This research uses K-Means to analyze groupings of student score data to identify outstanding students. The K-Means method has proven to be effective in grouping student score data into various group sizes. Determining the center point (centroid) at the beginning of the K-Means algorithm has a significant impact on the grouping results, as shown in tests with 25 data that have different centroids (Putu et al., n.d.). The results of this research produced 3 clusters, where Cluster 1 represents the low category with 12 members, Cluster 2 represents the sufficient category with 5 members, and Cluster 3 represents

the high category with 8 members (Saputra & Nataliani, 2021). In this research, k-means can be used. However, k means has a drawback, namely that if the centroid center point is different then the resulting grouping results will also be different (Tri & Rahmayani, 2018).

In further research, the K-Means algorithm was chosen to classify student learning styles. K-Means is used because it can divide data into clusters based on the level of similarity, where data with similar characteristics is grouped. The selection of K-Means is also based on time efficiency, where this algorithm has an average speed of only 1 second, in contrast to the K-Medoids algorithm which requires an average time of 1 minute 30 seconds. In this research, three clusters were produced, with Cluster 1 containing 3 data, Cluster 2 containing 2 data, and Cluster 3 containing 4 data (Handayani, 2022). In research that has been conducted, k-means has advantages in grouping time compared to k-medoid (Sari & Chotijah, 2022).

The next research is to use the k-means method to group several drugs that are almost the same into certain data groups. Drugs that have the same characteristics are then grouped into one cluster, while drugs that have different characteristics are grouped into another cluster. This research also uses the silhouette method which is useful for representing the accuracy of a data object placed in a cluster. The research results show that the k-means clustering method used with several k = 2 clusters, is a cluster with an ideal number that has a silhouette value of 0.4854 (Paembonan et al., 2021a). In this study, experiments were carried out with the number K 2 to K 10, and the most optimal K was K = 2, which from the silhouette coefficient test in this study had a result of 0.4 which could be categorized as the cluster position still not being suitable.

Based on the description above and supported by previous research, this research will apply the clusterization method using the k-means clustering algorithm to group and analyze student data obtained from the TNI-AD Disjasad Pusdikjas agency. The system being built also aims to determine the optimal number of clusters produced by the k-means clustering algorithm and use the Silhouette Coefficient test to determine the quality of the resulting clusters.

METHOD

Data Collection

This research uses a dataset of students who have attended training education, data comes from the Pusdikjas Disjasad TNI-AD agency. The dataset has 13 attributes. The attributes to be selected are NRP, Attitude and behavior Values, Knowledge and skills, and Physical.

Data Preprocessing

Data Cleaning

Data cleaning is the process of analyzing data quality through modifying, changing, or deleting data that is unnecessary, inaccurate, incomplete, or data that has the wrong format, as well as files that are not contained in the dataset so that the resulting data has high quality (Darwis et al., 2021).

Normalization

A technique for organizing data into several tables so that user needs can be met. Normalization is carried out on data by dividing the data values about the range of data values (maximum value minus (-) minimum value) (Yunita, 2018). Normalization aims to:

1. Delete duplicate data
2. Minimize complex data
3. Makes it easy to modify data

$$x_n = \frac{x_{0-x_{min}}}{x_{max}-x_{min}} \quad \text{.....(1)}$$

Information:
Xn = normal data value
X0 = actual data value
Xmin = minimum value of overall actual data
Xmax = maximum value of the overall actual data

Data Analysis Using the K-Means Clustering Method

At this stage, we will analyze the results obtained from clustering. The data will be analyzed using the K-Means clustering approach to provide results from the grouping of outstanding students in coaching education.

Validation Test

This validation test was carried out to evaluate the results obtained from the clustering process, to determine the optimal K value. At this stage, the Silhouette Coefficient method is used. This method is a method for evaluating clusters through a combination of the cohesion method and the separation method. Cohesion measurement is done by calculating the number of objects in a cluster, while separation measurement is done by calculating the average distance between each object in one cluster and the closest other cluster. The distance between the data is calculated using the Euclidean distance formula. To provide information about the quality of grouping results in the clustering process (Paembonan et al., 2021b).

Analysis of Cluster Results

Cluster analysis is a statistical analysis that is technically used to allocate a collection of objects into two or more groups based on the same characteristics of each object. Cluster analysis aims to select objects in various groups that have different properties from one group to another. In cluster analysis, each group has homogeneous characteristics between members within the group, or in other words, the objects formed in the group have very little chance of varying. The overall cluster solution depends on the variables used which are the basis for assessing similarity. Reductions and additions to related variables affect the results of cluster analysis. Cluster analysis in general can also be considered as one of the activities in order to analyze the process of forming clusters (Nurissaidah Ulinnuh & Rafika Veriani, n.d.).

RESULTS AND DISCUSSION

In this chapter, implementation, and testing will be discussed on systems or software that have been previously built based on the designs made in the previous chapter.

Implementation

Implementation is an application of a plan that has been made previously, applying the results of the design in Chapter 3.

Preparation for Implementation

Implementation Preparation In this research, web-based software was built based on needs using the JavaScript programming language, Visual Studio tools and using the MongoDB database, and Brave browser as the Web browser medium.

Database implementation

The database implementation was made based on the previous database design and implemented using Monggo DB. The results of the database design are as follows:

Student Table

The results of designing the student table database in Chapter 3 were implemented in Monggo db and the table structure can be seen in Figure 1.

siswas

	_id ObjectId	n1 Double	n2 Double	n3 Double	n4 Double
1	ObjectId('64a519c238775643840…	0.6846846846846865	0.5569655557105008	0.7162499999999999	0.7162499999999999
2	ObjectId('64a519c238775643840…	0.662162162162162	0.5336075861107237	0.734375	0.734375
3	ObjectId('64a519c238775643840…	0.5945945945945951	0.47378329382233964	0.7604166666666666	0.7604166666666666
4	ObjectId('64a519c238775643840…	0.5698198198198203	0.4468693348208064	0.7943749999999999	0.7943749999999999
5	ObjectId('64a519c238775643840…	0.587837837837838	0.4395481801701296	0.7864583333333334	0.7864583333333334
6	ObjectId('64a519c238775643840…	0.5405405405405421	0.48507878956909745	0.71875	0.71875
7	ObjectId('64a519c238775643840…	0.493243243243243	0.5055780225909916	0.6458333333333334	0.6458333333333334
8	ObjectId('64a519c238775643840…	0.6418918918918941	0.4882164272765305	0.6589583333333332	0.6589583333333332
9	ObjectId('64a519c238775643840…	0.6238738738738733	0.4746199972109884	0.6614583333333334	0.6614583333333334
10	ObjectId('64a519c238775643840…	0.5202702702702711	0.43857202621670616	0.7083333333333334	0.7083333333333334
11	ObjectId('64a519c238775643840…	0.6103603603603625	0.4959559336215313	0.609375	0.609375
12	ObjectId('64a519c238775643840…	0.5788288288288308	0.49581648305675663	0.6068749999999999	0.6068749999999999
13	ObjectId('64a519c238775643840…	0.5090090090090105	0.4375958722632827	0.6875	0.6875

Figure 1 Student table

Results Table

The results of designing the database result table in Chapter 3 are implemented in Monggo DB and the table structure can be seen in Figure 2

results

	_id ObjectId	cluster Int32	clustersilhouettescores Double	silhouettescore Double	k Int32
1	ObjectId('64a6866ec3feb926b03…	2	0.24004794701437765	0.3191469271761764	3
2	ObjectId('64a6866ec3feb926b03…	2	0.24004794701437765	0.3191469271761764	3
3	ObjectId('64a6866ec3feb926b03…	1	0.4129838282340275	0.3191469271761764	3
4	ObjectId('64a6866ec3feb926b03…	1	0.4129838282340275	0.3191469271761764	3
5	ObjectId('64a6866ec3feb926b03…	1	0.4129838282340275	0.3191469271761764	3
6	ObjectId('64a6866ec3feb926b03…	1	0.4129838282340275	0.3191469271761764	3
7	ObjectId('64a6866ec3feb926b03…	2	0.24004794701437765	0.3191469271761764	3
8	ObjectId('64a6866ec3feb926b03…	2	0.24004794701437765	0.3191469271761764	3
9	ObjectId('64a6866ec3feb926b03…	2	0.24004794701437765	0.3191469271761764	3
10	ObjectId('64a6866ec3feb926b03…	1	0.4129838282340275	0.3191469271761764	3
11	ObjectId('64a6866ec3feb926b03…	2	0.24004794701437765	0.3191469271761764	3
12	ObjectId('64a6866ec3feb926b03…	2	0.24004794701437765	0.3191469271761764	3
13	ObjectId('64a6866ec3feb926b03…	1	0.4129838282340275	0.3191469271761764	3

Figure 2 Results table

Interface implementation

Implementation of the Dashboard Page

The following is the display for the dashboard page which will be accessed first by the user when the user enters the system for grouping high achieving students which uses the K-Means Clustering algorithm (Ibnu Aji Pamungkas & Wasis Djoko Dwijoyo, 2020), for the implementation of the dashboard page can be seen in Figure 4.3

Figure 3. dashboards

Implementation of Data Upload Page Interface

The following displays the data upload page which will be accessed by the user when uploading data into the system. For the implementation of the data upload page, see Figure 4.

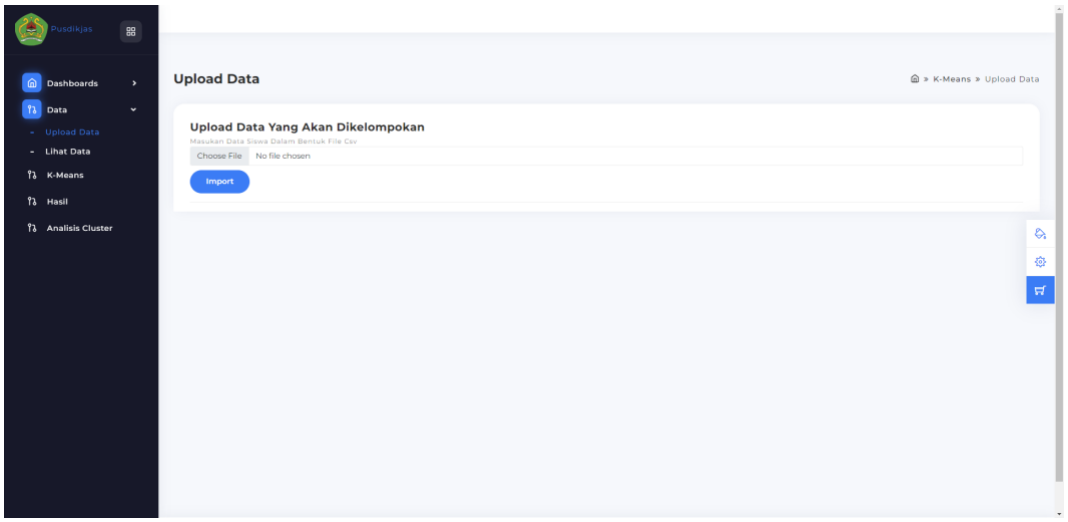

Figure 4. data uploads

Implementation of the Data View Page Interface

The following is the display for the data viewing page which will be accessed by the user when viewing data that has previously been uploaded to the system. To implement the data viewing page, see Figure 5.

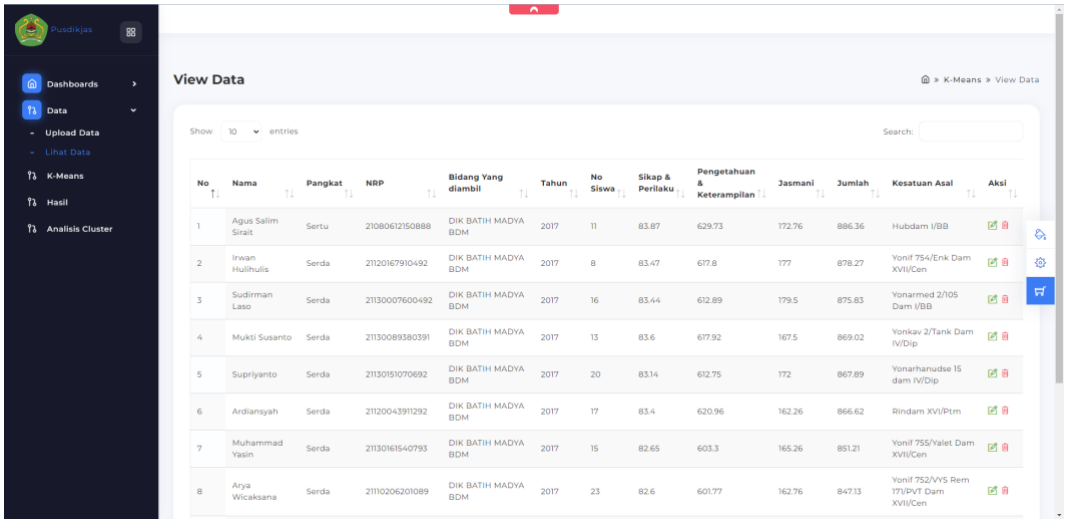

Figure 5. see data

Implementation of Data Cleaning Page Interface

The following displays the data cleaning page which will be accessed by the user when cleaning unused data and attributes. The implementation of the data cleaning page can be seen in Figure 6.

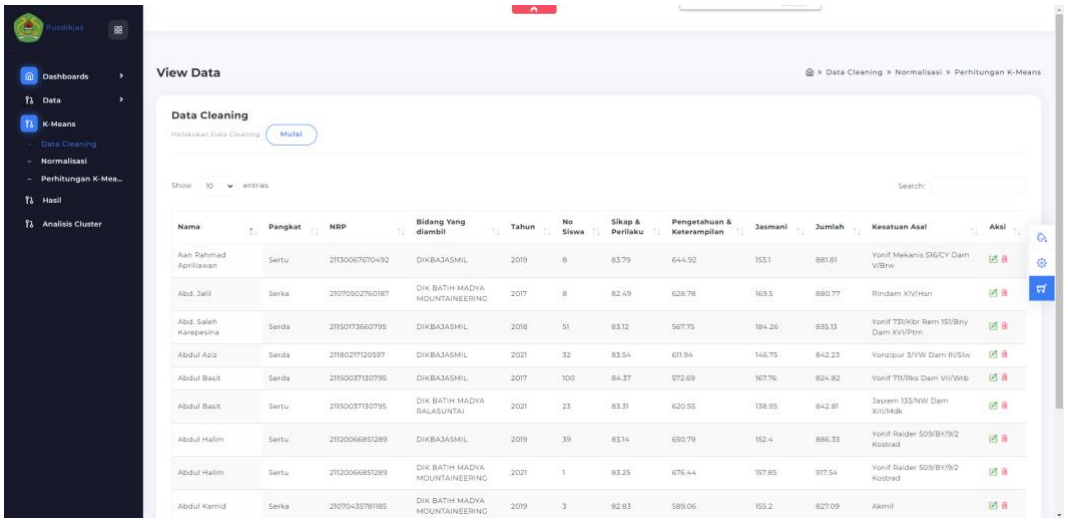

Figure 6. Data cleaning

Implementation of Normalization Page Interface

The following displays the normalization page which will be accessed by the user when normalizing a dataset by changing the dataset value to no more than 1. To implement the normalization page, see Figure 7.

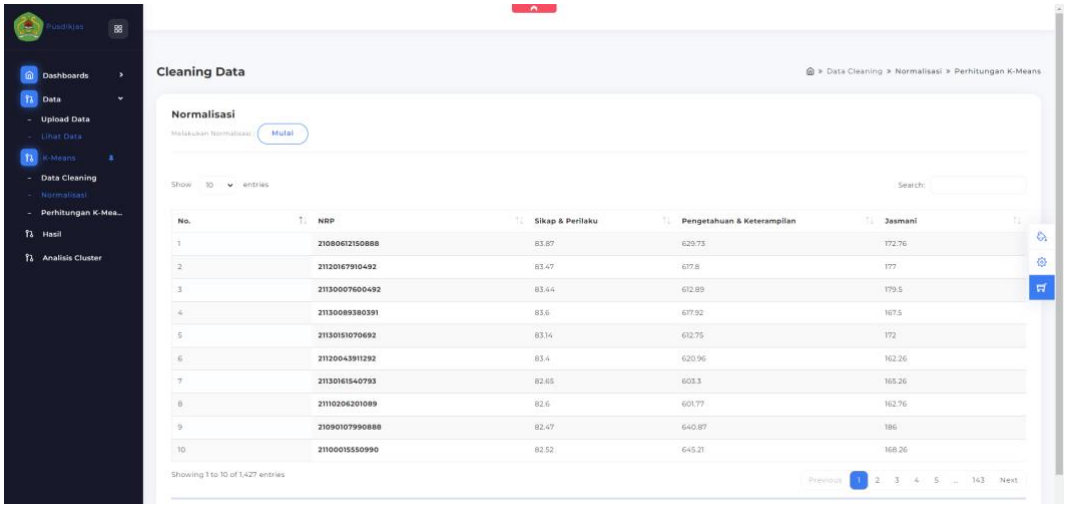

Figure 7. Normalization

Implementation of the K-Means Calculation Page Interface

The following is the display for the k-means calculation page which will be accessed by the user when carrying out calculations using the k-means clustering algorithm, for The implementation of the k-means calculation page can be seen in Figure 8.

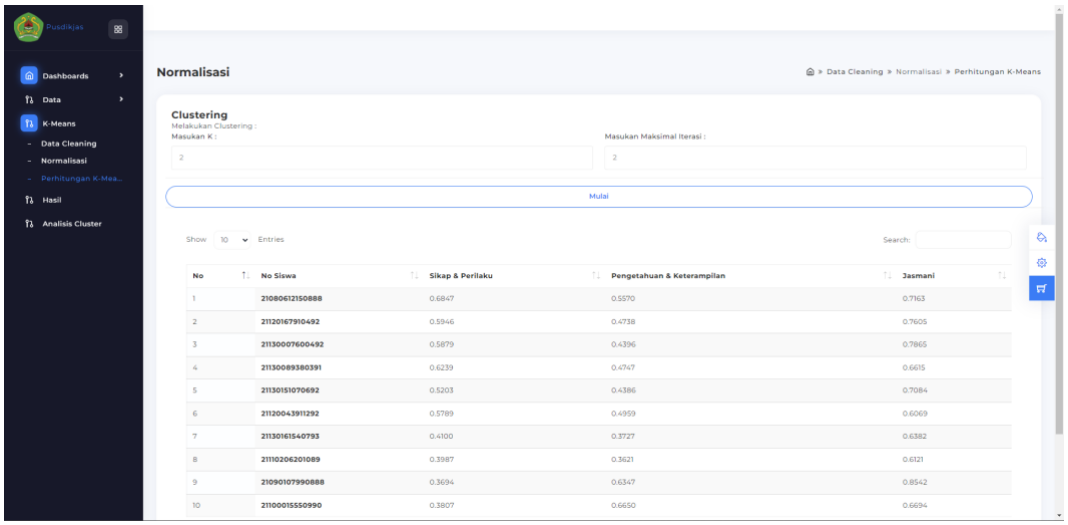

Figure 8. k-means calculations

Implementation of Results Page Interface

The following displays the results page which the user will access when they want to see the results of grouping using the k-means clustering algorithm, for The implementation of the results page can be seen in Figure 9

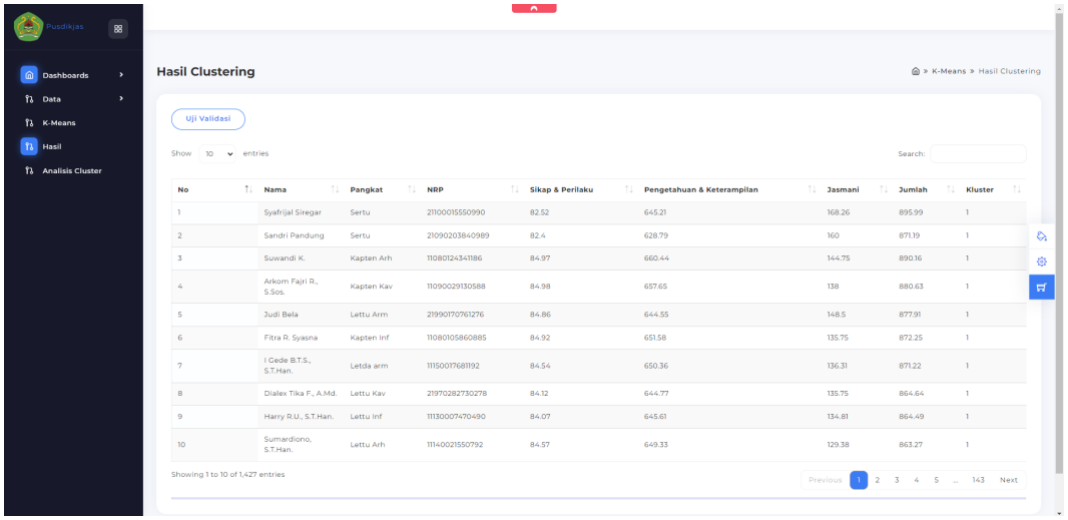

Figure 9. Results

Implementation of Validation Test Page Interface

The following is the display for the validation test page which will be accessed by the user when looking at the results of the cluster quality tester using the silhouette coefficient, for implementation the results page can be seen in Figure 10.

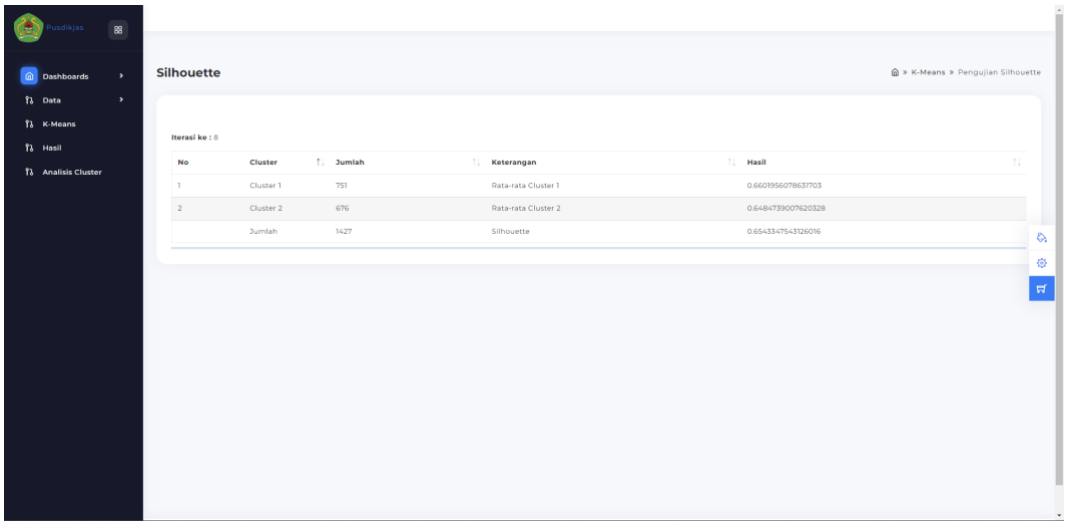

Figure 10. validation tests

Implementation of the Cluster Analysis Page Interface

The following is the display for the cluster analysis page which will be accessed by the user when looking at the results of the resulting cluster categories. For the implementation of the cluster analysis page, it can be seen in Figure 11

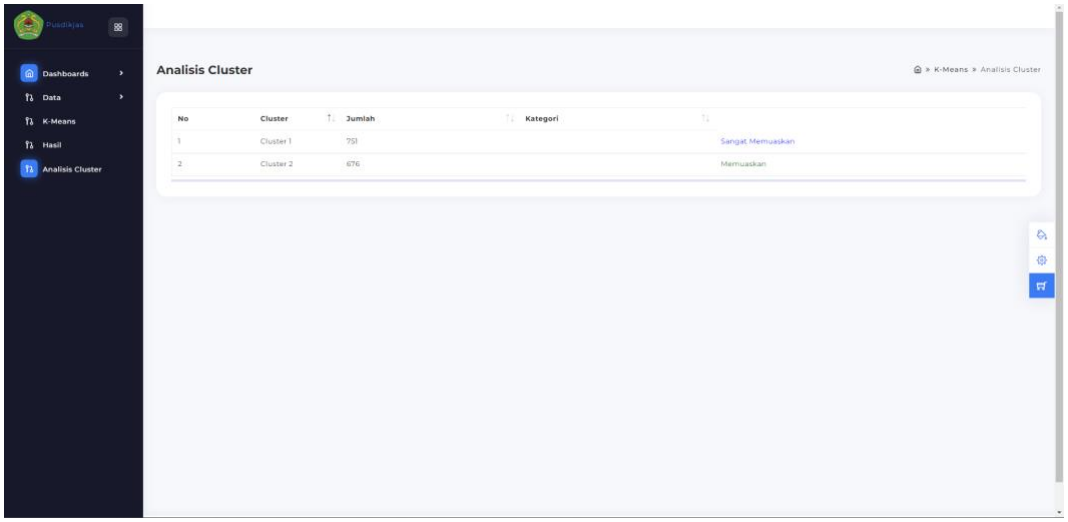

Figure 11 Cluster analysis

Cluster Analysis

In the analysis of the results obtained from the interpretation that has been carried out. The following analysis results from student data can be seen in the table in table 1

Table 1 Cluster analysis

No	NRP	Attitude behavior	Knowledge & skills	Physical	Cluster
1	21010213641281	0.8355	0.8979	0.5052	1
2	21140087470693	0.6216	0.9094	0.6031	1
3	21190163800899	0.5495	0.6756	0.3593	2
4	31020158500282	0.5450	0.5369	0.4625	2

Number of Members of Each Cluster

Table 2 Number of members of each cluster

No	Cluster	Amount
1	Cluster 1	751
2	Cluster 2	676

In table 2 you can see the results of grouping student data produced by the system, in cluster 1, there are 751 students, and in cluster 2, there are 676 students.

Description of Cluster Analysis Results

Based on table 4.1, in cluster 1, the value of the 1st student on the attitude and behavior variable has a value of 0.8355, which is included in the high value, and for the 2nd student's data, the value of the attitude and behavior variable has a value of 0.6216, which is included in the average value. , then for student 1 the Knowledge & Skills variable has a value of 0.8979 which is a high value. For student 2's data on knowledge & and skills, they got a value of 0.9094, which is considered high, while student 1 in the physical variable had a value of 0.5052, which is considered a medium value. For the second student, the physical variable has a value of 0.6031, which is included in the average value. The results of this analysis describe the student category as very satisfactory.

In cluster 2 the 1st student's score on the attitude and behavior variable has a value of 0.5495 which is included in the medium value and for the 2nd student's data on the attitude and behavior variable the value has a value of 0.5450 which is included in the medium value, then for the 1st student the variable Knowledge & skills have a value of 0.6756 which is included in the moderate value. For the 2nd student's data, the knowledge and skills variable got a value of 0.5369, which is considered medium, for the 1st student, the physical variable had a value of 0.3593, which is a low value. And for the second student, the physical variable has a value of 0.4625, which is a low value. The results of cluster 2 analysis describe the satisfactory student category

Silhouette Coefficient Validation Test Results

Calculations that have been carried out using the K-Means Clustering method are then carried out with a validation test which is used to assess whether the cluster results are good or not. The following is a table of the results of cluster testing using the K-Means algorithm. The cluster testing table can be seen in tables 3 to 11.

Table 3 Silhouette Results K = 2

No	Cluster	Amount	Description	Results
1	Cluster 1	751	Average Cluster 1	0.6601956078631703
2	Cluster 2	676	Average Cluster 2	0.6484739007620328
		1427	Silhouettes	0.6543347543126016

Based on the results of silhouette coefficient testing with the number K=2, the Silhouette coefficient value is 0.6543347543126016.

Table 4. Silhouette K=3 Results

No	Cluster	Amount	Description	Results
1	Cluster 1	425	Average Cluster 1	0.5409536953702614
2	Cluster 2	475	Average Cluster 2	0.729121214121146
3	Cluster 3	527	Average Cluster 3	0.5583870720674422
		1427	Silhouettes	0.6094873271862832

Based on the results of the silhouette coefficient test with the number K = 3, the Silhouette coefficient value is 0.6094873271862832, where at K = 3 the Si value decreases.

Table 5. Silhouette K=4 Results

No	Cluster	Amount	Description	Results
1	Cluster 1	351	Average Cluster 1	0.6202506526874594
2	Cluster 2	398	Average Cluster 2	0.5141267705888787
3	Cluster 3	258	Average Cluster 3	0.7068699161665092

No	Cluster	Amount	Description	Results
4	Cluster 4	420	Average Cluster 4	0.5685496163799284
		1427	Silhouettes	0.6024492389556939

Based on the results of the silhouette coefficient test with the number K = 4, the Silhouette coefficient value is 0.6024492389556939, where at K = 4 there is also a decrease in the Si value.

Table 6 Silhouette K=5 Results

No	Cluster	Amount	Description	Results
1	Cluster 1	245	Average Cluster 1	0.4862338725631343
2	Cluster 2	346	Average Cluster 2	0.49173778992590406
3	Cluster 3	252	Average Cluster 3	0.7386375376266698
4	Cluster 4	405	Average Cluster 4	0.5740578999296464
5	Cluster 5	179	Average Cluster 5	0.622088648008456
		1427	Silhouettes	0.582551149610762

Based on the results of the silhouette coefficient test with the number K = 5, the Silhouette coefficient value is 0.582551149610762, where at K = 5 there is also a decrease in the Si value.

Table 7 Silhouette Results K=6

No	Cluster	Amount	Description	Results
1	Cluster 1	361	Average Cluster 1	0.6131731398242216
2	Cluster 2	273	Average Cluster 2	0.5352900030537002
3	Cluster 3	292	Average Cluster 3	0.6490474533831758
4	Cluster 4	141	Average Cluster 4	0.48500982324067243
5	Cluster 5	158	Average Cluster 5	0.6436263718564331
6	Cluster 6	202	Average Cluster 6	0.48333737579567204
		1427	Silhouettes	0.5682473611923126

Based on the results of the silhouette coefficient test with the number K = 6, the Silhouette coefficient value is 0.5682473611923126, where at K = 6 there is also a decrease in the Si value.

Table 8 Silhouette Results K = 7

No	Cluster	Amount	Description	Results
1	Cluster 1	217	Average Cluster 1	0.5138945417467744
2	Cluster 2	319	Average Cluster 2	0.5769857451457471
3	Cluster 3	199	Average Cluster 3	0.5415744166026828
4	Cluster 4	237	Average Cluster 4	0.581536081326556
5	Cluster 5	151	Average Cluster 5	0.6766815580926132
6	Cluster 6	162	Average Cluster 6	0.6188903628247233
7	Cluster 7	142	Average Cluster 7	0.5978157737102269
		1427	Silhouettes	0.5867683542070462

Based on the results of the silhouette coefficient test with the number K = 7, the Silhouette coefficient value is 0.5867683542070462, where at K = 7 the Si value produced increases but the results remain below 0.6 at the Si value

Table 9 Silhouette results K = 8

No	Cluster	Amount	Description	Results
1	Cluster 1	112	Average Cluster 1	0.540073438978835
2	Cluster 2	234	Average Cluster 2	0.501187234901667
3	Cluster 3	146	Average Cluster 3	0.5508183761931541
4	Cluster 4	261	Average Cluster 4	0.6112575893596949

No	Cluster	Amount	Description	Results
5	Cluster 5	101	Average Cluster 5	0.7129443969351825
6	Cluster 6	153	Average Cluster 6	0.66110652967814
7	Cluster 7	164	Average Cluster 7	0.512465426115482
8	Cluster 8	256	Average Cluster 8	0.6098255667693875
		1427	Silhouettes	0.5874598198664429

Based on the results of the silhouette coefficient test with the number K = 8, the Silhouette coefficient value is 0.5874598198664429, where at K = 8 the result Si remains the same as K = 7.

Table 10 Shilhouette Results K = 9

No	Cluster	Amount	Description	Results
1	Cluster 1	249	Average Cluster 1	0.49736783578699084
2	Cluster 2	131	Average Cluster 2	0.5787025673762227
3	Cluster 3	133	Average Cluster 3	0.6052501904339977
4	Cluster 4	179	Average Cluster 4	0.5796066613738844
5	Cluster 5	90	Average Cluster 5	0.6908474812928835
6	Cluster 6	115	Average Cluster 6	0.7076690009953249
7	Cluster 7	214	Average Cluster 7	0.5163300103048898
8	Cluster 8	156	Average Cluster 8	0.6095784429794553
9	Cluster 9	160	Average Cluster 9	0.5386794294696198
		1427	Silhouettes	0.5915590688903632

Based on the results of the silhouette coefficient test with the number K = 8, the Silhouette coefficient value is 0.5915590688903632, where at K = 9 the result is that Si increases.

Table 11 Silhouette K = 10 results

No	Cluster	Amount	Description	Results
1	Cluster 1	151	Average Cluster 1	0.5319192906693954
2	Cluster 2	127	Average Cluster 2	0.5868631690387353
3	Cluster 3	129	Average Cluster 3	0.5283037881561756
4	Cluster 4	74	Average Cluster 4	0.5697797286472401
5	Cluster 5	215	Average Cluster 5	0.5814127542457758
6	Cluster 6	252	Average Cluster 6	0.6868887390051833
7	Cluster 7	107	Average Cluster 7	0.7146909970369411
8	Cluster 8	150	Average Cluster 8	0.5043195489402023
9	Cluster 9	88	Average Cluster 9	0.6059696323947181
10	Cluster 10	134	Average Cluster 10	0.5334677320461945
		1427	Silhouette	0.5843615380180561

Based on the results of the silhouette coefficient test with the number K = 10, the Silhouette coefficient value is 0.5843615380180561, where at K = 10 the result is that Si decreases. The conclusion from 10 Cluster testing experiments that have been carried out using the silhouette coefficient results in the largest Si value, namely in cluster 2 which produces Si 0.6543347543126016.

Software Testing

Software testing is a software quality testing technique in research on the achievement student grouping system on data from students attending education at the TNI-AD Disjasad Education Center using k-means with black box testing.

Test Method

The testing method used is black box testing. The black box testing method is to test software functionality and look for errors by testing the software interface.

Testing Stages

The steps taken when testing the software in this research include:

1. Determining quality testing objectives
2. Determine the category of quality test results
3. Design quality tests based on use case grouping
4. Implementation of quality testing
5. Conclusions and results of quality testing

Testing objectives

The purpose of quality testing of the system being built can be seen in Table 4.7

Table 12 Testing Objectives

No	Use Case	Purpose
1	Upload data	Test the software's ability to upload student data and then save it to the database
2	View data	Test the software's ability to view student data that has been stored in the database
3	Data cleaning	Test the software's ability to delete data that cannot be used in the dataset
4	Normalization	Test the software's ability to normalize the value of the dataset to be no greater than 1
5	K-Means Calculation	Testing the software's capabilities in the process of grouping student data
6	Results	Test the software's ability to view cluster results
7	Validation test	test Testing software capabilities when conducting cluster quality testing
8	Cluster analysis	Testing the software's ability to view cluster category results

Quality Test Result Category

In determining the categories used in quality testing, they are divided into 2 categories, namely:

a. In accordance
   Software that has been tested is included in the "Successful" category if its function and quality are in accordance with the planning objectives and its use.
b. It is not in accordance with
   Software that has been tested is included in the "Failed" category if its function and quality do not match the planning objectives and its use.

Quality Testing Scenarios

Software testing scenarios are designed to look for errors or faults in the system, thereby creating quality testing scenarios with incorrect and correct steps or stages of system use. The quality testing scenario can be seen in Table 4.8

Table 13 Quality Testing Scenarios

Use Case	Function Name	Test Code	Test Case
Upload data	Upload data	KU.101	Test software capabilities when the user uploads data
View data	View data	KU.102	Testing software capabilities when the user wants to view data.

Use Case	Function Name	Test Code	Test Case
Data cleaning	Cleaning	KU.103	Testing software capabilities when the user wants to carry out data cleaning
Normalization	Normalization	KU.104	Testing software capabilities when the user wants to normalize
K-means calculation	Algorithm calculation	KU.105	Testing software capabilities when grouping student data
Results	Results	KU.106	Test the software capabilities when the user sees the grouping results.
Validation Test	Shillhoute coefficient	KU.107	Tests software capabilities when the user sees the cluster quality results
Cluster analysis	cluster analysis	KU.108	Testing software capabilities when the user will see the results from the cluster categories

Software Testing

Testing of software is carried out in order to test the comparison of the suitability of the system being built with the system being designed so as to produce test results that are suitable and those are not suitable, Software testing can be seen in Table 4.9

Table 14 Software Testing

No	Test Case	Expected Output	Results Obtained	Results
1	Upload data	Users can upload data into the system	Save the uploaded data	Accordingly
2	View data	Users can view data in the system	Display data that has been stored in the database	Accordingly
3	Data cleaning.	The system displays the results of data cleaning, namely deleting unused values in the dataset	Displays the results of data cleaning by deleting unused values in the dataset	Accordingly
4	Normalization	The system displays the normalization results, namely the changed value to no greater than 1	Displays the normalization results of the dataset by changing the dataset value to no more than 1	Accordingly
5	K-Means Calculation.	The system calculates and displays groups of student data based on the proximity distance between objects and the cluster center	Can calculate and display the results of grouping student data	Accordingly
6	Results.	The system displays the results of grouping student data	Displays the number of students from each cluster	Accordingly
7	Validation test	The system can test cluster quality.	Can display results of cluster quality	Accordingly
8	Cluster analysis	The system can display the results of cluster category analysis.	Can display the results of the cluster category	Accordingly

Next, the conclusion of the black box testing that has been carried out is in a table with a total of 8 functions, namely Upload data, View data, Data cleaning, Normalization, K-Means calculation, Results, and Validation test. So the percentage of function suitability in the system can be calculated as follows:

Number of Test Codes = 8 Test Codes
Test Codes with Corresponding Results = 8 Test Codes
Test Codes with Inappropriate Results = 0 Test Codes
Percentage

$$\text{Percentage} = \frac{(\text{number of test codes} - \text{non} - \text{matching test codes})}{(\text{number of test codes})} \times 100\%$$
$$= \frac{(8-0)}{(8)} \times 100\%$$
$$= 100\%$$

Based on the results of calculations on the system suitability function, it can be concluded that the testing software has been implemented using black box testing, and can be operated with a percentage of 100% meeting the specifications in accordance with the previously determined specifications.

CONCLUSION

This research aims to group education students at the PUSDIKJAS DISJASAD TNI-AD agency using the K-Means Clustering method. The K-Means Clustering method is a data analysis approach used to group data into several groups based on similar characteristics. In this research, the author focuses on applying the K-Means Clustering method to group education students at the PUSDIKJAS DISJASAD TNI-AD agency.

The results of the grouping obtained using data mining techniques with the k-means algorithm in this research produced optimal clusters, namely 2 clusters. cluster 1 with a total of 751 students, cluster 1 describes a very satisfactory category of students, and cluster 2 with a total of 676 students. Cluster 2 describes the satisfactory student category. As for testing the quality of the clusters contained in the system using silhouette with results reaching 0.6543347543126016.

REFERENCES

Darwis, D., Siskawati, N., & Abidin, Z. (2021). *Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional*. *15*(1).

Handayani, F. (2022). Aplikasi Data Mining Menggunakan Algoritma K-Means Clustering untuk Mengelompokkan Mahasiswa Berdasarkan Gaya Belajar. *Jurnal Teknologi Dan Informasi*. https://doi.org/10.34010/jati.v12i1

Ibnu Aji Pamungkas, & Wasis Djoko Dwijoyo. (2020). *Pengembangan Media Pembelajaran Berbasis Mobile Learning Untuk Aktifitas Kesegaran Jasmani Siswa kelas X Sekolah Menengah Kejuruhan*. *2*(5).

Nurhayati. (2022). *Unsupervised_Learning_Technique_Big_Data*. https://www.google.co.id/books/edition/Unsupervised_Learning_Technique_Big_Data/97JmEAAAQBAJ?hl=id&gbpv=0&kptab=overview

Nurissaidah Ulinnuh, & Rafika Veriani. (n.d.). *Analisis Cluster dalam Pengelompokan Provinsi di Indonesia Berdasarkan Variabel Penyakit Menular Menggunakan Metode Complete Linkage, Average Linkage dan Ward*. https://doi.org/10.30743/infotekjar.v5i1.2464

Paembonan, S., Abduh, H., & Kunci, K. (2021a). *Penerapan Metode Silhouette Coeficient Untuk Evaluasi Clutering Obat Clustering; K-means; Silhouette coeficient* (Vol. 6, Issue 2). https://ojs.unanda.ac.id/index.php/jiit/index

Paembonan, S., Abduh, H., & Kunci, K. (2021b). *Penerapan Metode Silhouette Coeficient Untuk Evaluasi Clutering Obat Clustering; K-means; Silhouette coeficient* (Vol. 6, Issue 2). https://ojs.unanda.ac.id/index.php/jiit/index

Penentuan, D., Barang, S., Muningsih, E., & Kiswati, D. S. (2015). Penerapan Metode K-Means Untuk Clustering Produk Online Shop. *Jurnal Bianglala Informatika*, *3*.

Putu, N., Merliana, E., & Santoso, A. J. (n.d.). *Analisa Penentuan Jumlah Cluster Terbaik Pada Metode K-Means Clustering*.

Saputra, E. A., & Nataliani, Y. (2021). Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means. *Journal of Information Systems and Informatics*, *3*(3). http://journal-isi.org/index.php/isi

Sari, M. A. P., & Chotijah, U. (2022). Pengelompokan Anggota Divisi Himpunan Mahasiswa Jurusan Pada Universitas "Xyz" Dengan Metode K-Means Clustering. *Antivirus : Jurnal Ilmiah Teknik Informatika*, *16*(1), 52–62. https://doi.org/10.35457/antivirus.v16i1.2139

Sulistiyawati, A., & Supriyanto, E. (n.d.). *Implementasi Algoritma K-means Clustring dalam Penetuan Siswa Kelas Unggulan*. *15*(2).

Tri, M., & Rahmayani, I. (2018). *Analisis Clustering Tingkat Keparahan Penyakit Pasien Menggunakan Algoritma K-Means (Studi Kasus Di Puskesmas Bandar Seikijang)*.

Warih Widiyono, & Aris Harliadi. (2011). *Sejarah Pusat Pendidikan Jasmani Militer Kodiklat Tni Ad* (Warih Widiyono & Aris Harliadi, Eds.).

Yunita, F. (2018). Penerapan Data Mining Menggunkan Algoritma K-Means Clustring Pada Penerimaan Mahasiswa Baru (Studi Kasus : Universitas Islam Indragiri). In *Jurnal SISTEMASI* (Vol. 7).