Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

Analysis of Student Achievement with K-Means on Socioeconomic, Behavioral, and Psychological Factors

Muhammad Iqbal[1], Sardo Pardingotan Sipayung[2], Alex Rikki Sinaga[3], Paska Marto Hasugian[4]
[1]Universitas Panca Budi, [2,3,4]Universias Katolik Santo Thomas Medan Jl. Setiabudi No. 479 F Tanjungsari Medan

Article Info
Keywords:
k-means,
academic achievement,
socio-economic factors,
psychological student behavior
clustering

ABSTRACT
This study aims to analyze students' academic achievement based on socio-economic, behavioral, and psychological factors using the K-Means clustering method. The data used include various variables such as family income, internet access, learning motivation, stress levels, and student attendance. The results of the analysis show that students can be grouped into three different clusters: Cluster 1 consists of students with good socio-economic backgrounds, high motivation, and better academic achievement; Cluster 2 shows students with higher levels of stress that affect their achievement even though they have adequate access to education; and Cluster 3 reflects students from less supportive socio-economic backgrounds, with limited internet access and low study time, so their achievement is lower than other clusters. The Davies-Bouldin Index (DBI) calculation result of 0.63 shows a fairly good cluster separation. This analysis reveals that socio-economic factors have a significant impact on students' academic achievement, while psychological aspects such as motivation and stress levels also play an important role in determining learning success. Intervention programs focused on stress management and increasing access to education are recommended for students from Cluster 2 and Cluster 3 to improve their academic outcomes. This study provides insight into the importance of socio-economic and psychological factors in shaping students' academic achievement.

Corresponding Author:
Muhammad Iqbal
Universitas Pembangunan Panca Budi
muhammadiqbal@dosen.pancabudi.ac.id

INTRODUCTION

Student academic achievement is one of the important indicators in assessing the quality of education in a country. Many factors influence student achievement, including socioeconomic, behavioral, and psychological factors. Socioeconomic factors such as family income and parental education level are often associated with better access to educational resources. On the other hand, behavioral factors such as study habits and self-discipline, as well as psychological factors such as stress levels and motivation, also play an important role in determining students' academic success.

Previous studies have shown that clustering analysis, particularly using the K-Means algorithm, can help identify groups of students with similar characteristics. For example, a study by Kumar and Minz (2014) showed that K-Means is effective in grouping students

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

based on their academic performance, which can then be used to identify groups of students who need special attention. However, most of these studies tend to focus on a single aspect, such as academic or socioeconomic factors, without considering the combination of factors.

This study aims to analyze the effect of a combination of socioeconomic, behavioral, and psychological factors on students' academic achievement using the K-Means algorithm. By identifying groups of students based on a combination of these factors, it is hoped that more comprehensive insights can be obtained regarding the factors that influence student achievement. In addition, this study also aims to provide recommendations for educational interventions that can be applied to improve student achievement in each identified group.

RESEARCH METHODS

This study uses a quantitative approach with the K-Means clustering method to group students based on socioeconomic, behavioral, and psychological factors and analyze the relationship of each factor to academic achievement. The research process is carried out through several stages, including data collection, data processing, cluster analysis, and interpretation of results.

Research Design

This research is descriptive-explorative, which aims to identify student grouping patterns based on characteristics that affect academic achievement. With the K-Means method, students will be grouped into several clusters based on predetermined factors. Each cluster will be evaluated for its relationship to academic achievement to identify groups of students who are vulnerable or in need of further intervention.

Population and Sample

The population in this study were senior high school (SMA) students from various schools in the research area who were selected randomly. The research sample was taken using a purposive sampling technique, with the criteria of students who had complete data.ap regarding socioeconomic background, learning behavior, and psychological factors. The samples taken are expected to represent various socioeconomic backgrounds and other characteristics. The number of samples taken is 10 students.

Research Variables

This study uses three main groups of variables as the basis for grouping students:

a. Socioeconomic Factors: Family economic background, parental education, and access to learning resources (e.g. internet, books, learning tools).
b. Behavioral Factors: Study habits, level of discipline, study time at home, and school attendance.
c. Psychological Factors: Level of learning motivation, stress level, self-confidence, and emotional support from family.

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

Data Collection Instruments

Data was collected using several instruments as follows:

a. Socioeconomic Questionnaire: To measure family economic background, parental education level, and access
a. towards educational facilities.
b. Behavior Questionnaire: To measure students' study habits, discipline, and attendance at school.
c. Psychological Questionnaire: To measure students' levels of motivation, stress, and self-confidence.
d. Academic Grades: Student achievement is measured using report card average grade data as the main indicator of academic achievement.

Data Collection Procedures

The data collection process is carried out in several steps:

a. Distribution of Questionnaires: Questionnaires were distributed to students who were the research samples with guidance from the teacher or supervisor.
b. Academic Grade Data Collection: Student academic achievement data is collected from the average semester grade results. previously taken from school records.
c. Data Verification: Data from questionnaires and academic scores are verified to ensure accuracy and completeness. before being analyzed.

Data Analysis Techniques

The collected data was analyzed using the K-Means clustering method with the following procedure:

a. Data Preprocessing: Questionnaire and academic achievement data are processed to eliminate incomplete or missing data. invalid. The scores of each variable will be normalized to ensure consistency in clustering analysis.
b. Selecting the Number of Clusters: Using the Elbow method to determine the most optimal number of clusters. Elbow method identify the sweet spot where adding clusters does not provide a significant increase in variance within.
c. Grouping with K-Means: Students will be grouped based on the similarity of values from socioeconomic factors, behavior, and psychology. The K-Means algorithm will group the data into a number of clusters by placing students in the cluster with the closest centroid.
d. Cluster Evaluation: After the clusters are formed, analysis is performed to evaluate the characteristics of each cluster and the relationshipwith students' academic achievement. Interpretation of Results: Each cluster was analyzed to find patterns of relationships between socioeconomic, behavioral, and psychological characteristics with academic achievement. Clusters that show academic achievement results low academic performance will be further analyzed to provide intervention recommendations.

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

Validate Results

Validation of cluster results is done by comparing clustering results with academic achievement variables. To evaluate the accuracy of the clustering, correlation analysis is used between academic scores and factors that influence achievement in each cluster.

Conclusion and Recommendations

Based on the results of the clustering analysis, conclusions will be drawn regarding the profile of students in each cluster and the characteristics that influence their academic achievement. Recommendations will also be made for interventions that can be carried out, especially in clusters that show low achievement. This is expected to provide input to schools in developing programs to improve student achievement based on their socioeconomic, behavioral, and psychological profiles. Here is an example of a simple dataset that can be used in the study "Analysis of Student Achievement with K-Means on Socioeconomic, Behavioral, and Psychological Factors". This dataset consists of several columns that cover socioeconomic, behavioral, psychological, and academic achievement factors of students.

Student ID	Parental Income	Parent Education	Internet Access	Study Hours/ Day	Presence (%)	Motivation (1-10)	Stress (1-10)	Self Confidence (1-10)	Average value
S001	Low	2	1	2	90	8	4	7	85
S002	Intermediate	4	1	3	95	9	3	8	90
S003	Intermediate	4	0	1	85	7	6	6	80
S004	Tall	5	1	4	92	9	2	9	93
S005	Low	1	0	1	80	6	7	5	75
S006	Tall	4	1	4	97	10	1	9	95
S007	Intermediate	3	1	3	93	8	4	8	88
S008	Tall	4	1	3	96	9	3	9	92
S009	Low	2	0	2	87	7	6	7	82
S010	Low	2	1	2	88	8	5	6	84

Student ID, Parental Income, Parental Education, Internet Access, Study Hours/Day, Attendance (%) Motivation (1-10), Stress (1-10) Self Confidence (1-10), Average Score Column Explanation:

a. Student ID: Unique identification for each student.
b. Parental Income: Family income category, for example low, middle, or high.
c. Parental Education: Parents' highest level of education (junior high school, high school, bachelor's degree, master's degree).
d. Internet Access: Does the student have access to the internet at home (Yes/No).

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

e. Study Hours/Day: The average number of hours students spend studying each day outside of school.
f. Attendance (%): The rate of student attendance at school in percentage (%).
g. Motivation (1-10): Student learning motivation score, measured from 1 (low) to 10 (high).
h. Stress (1-10): Student stress level, measured from 1 (low) to 10 (high).
i. Self-Confidence (1-10): Student's self-confidence score, measured from 1 (low) to 10 (high).
j. Grade Point Average: A student's average academic grade based on report card or exam scores.

RESULTS AND DISCUSSION

This study uses the K-Means clustering algorithm to group students based on socioeconomic, behavioral, and psychological factors. The results of the analysis provide insight into the influence of these factors on students' academic achievement.To provide a clear understanding, the dataset used contains 10 students with five different attributes: Parental Education, Internet Access, Study Hours/Day, Attendance, Motivation, Stress, Self-Confidence and Average Grades.

Table 1. Dataset of Students' Socioeconomic, Behavioral, and Psychological Factors

Student ID	Parental Income	Parent Education	Internet Access	Study Hours/ Day	Presence (%)	Motivation (1-10)	Stress (1-10)	Self Confidence (1-10)	Average value
S001	Low	2	1	2	90	8	4	7	85
S002	Intermediate	4	1	3	95	9	3	8	90
S003	Intermediate	4	0	1	85	7	6	6	80
S004	Tall	5	1	4	92	9	2	9	93
S005	Low	1	0	1	80	6	7	5	75
S006	Tall	4	1	4	97	10	1	9	95
S007	Intermediate	3	1	3	93	8	4	8	88
S008	Tall	4	1	3	96	9	3	9	92
S009	Low	2	0	2	87	7	6	7	82
S010	Low	2	1	2	88	8	5	6	84

K-Means Clustering

The application of k = 3 for the results of Cluster 1, Cluster 2, and Cluster 3. However, the initial Centroid of the dataset used must be determined beforehand. The initial Centroid determination technique used is a random technique. The initial Centroid selected is c1 from student data 004, c2 from student data 002, and c3 from student data 005.

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

Table 2. Initial centroids

Cluster 1	5	1	4	92	9	2	9	93
Cluster 2	4	1	3	95	9	3	8	90
Cluster 3	1	0	1	80	6	7	5	75

Determine the Euclidean distance from each point to the Centroid. The Euclidean distance is obtained from the root (√) of the subtraction of the value of each dataset point (p) to each initial Centroid value (q) which is given a power according to the formula, then written:

$$d = \sqrt{(x_1 - c_1)^2 + (x_2 - c_2)^2 + \ldots (x_n - c_n)^2}$$

$$\text{Jarak ke Cluster } 1 = \sqrt{(2-5)^2 + (1-1)^2 + (2-4)^2 + (90-92)^2 + (8-9)^2 + (4-2)^2 + (7-9)^2 + (85-93)^2} = 9.5$$

$$\text{Distance to Cluster 2} = 5.3 = \sqrt{(2-4)^2 + (1-1)^2 + (2-3)^2 + (90-95)^2 + (8-9)^2 + (4-3)^2 + (7-8)^2 + (85-90)^2}$$

$$\text{Distance to Cluster 3} = 14.8 = \sqrt{(2-1)^2 + (1-0)^2 + (2-1)^2 + (90-80)^2 + (8-6)^2 + (4-7)^2 + (7-5)^2 + (85-75)^2}$$

Example of calculating the Euclidean distance of Employee1 point to the initial Centroid above, also applies to Student 002 to Student 010 points. Cluster determination from the initial determination is done by comparing the values of the three distances. The result of c1 = 9.5, the result of c2 = 5.3, and the result of c3 = 14.8, from the three values, the result of c2 = 5.3 is the smallest value selected as a Cluster. Determining the cluster by comparing the smallest value or from the nearest Centroid can be seen in the following table:

Table 3. Determination of Euclidean Distance and Cluster

Cluster 1	Cluster C2	Cluster C3
9.5	5.3	14.8
4.7	0.0	22.3
16.1	10.6	7.9
0.0	4.0	23.3
23.3	16.1	0.0
5.7	4.0	27.9
6.1	2.2	19.2
4.5	2.0	24.5
13.6	8.7	10.3
11.4	7.5	12.5

Extract initial Cluster results for determining new Centroids The third group of initial

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

Clusters is separated so that each member is displayed.

Cluster 1: Student 004

Cluster 2: Students 001, 002, 006, 007, 008, 009, 010

Cluster 3: Students 003, 005

From the three groups above, each value in the Cluster at each Student point is added up and divided by the average of the sum to create three new Centroids. Calculate the average value of each variable for each cluster to obtain a new centroid:

a. New centroid of Cluster 1: Fixed (because there is only 1 data, namely Student 004)
b. New Centroid of Cluster 2 : (Students 001, 002, 006, 007, 008, 009, 010)
(2+4+3+3+4+2+2)/7, (1+1+1+1+1+0+1)/7, (2+3+4+3+3+2+2)/7, (90+95+97+93+96+87+88)/7, (8+9+10+8+9+7+8)/7, (4+3+1+4+3 +6+5)/7, (7+8+9+8+9+7+6)/7, (85+90+95+88+92+82+84)/7 = (3.0, 0.9, 2.7, 92.3, 8.4, 3.7, 7.7 , 8.8)
c. New Centroid of Cluster 3 : (Students 003 and 005)
(4+1)/2, (0+0)/2, (1+1)/2, (85+80)/2,(7+6)/2, (6+7)/2, (6 +5)/2, (80+75)/2 = (2.5, 0.0, 1.0, 82.5, 6.5, 6.5, 5.5, 77.5)

The results of all the initial Clusters are used for the new Centroid which continues the algorithm calculation.

Table 4. New Centroids of Iteration 1

Cluster 1	5.0	1.0	4.0	92.0	9.0	2.0	9.0	93.0
Cluster 2	3.0	0.9	2.7	92.3	8.4	3.7	7.7	88.0
Cluster 3	2.5	0.0	1.0	82.5	6.5	6.5	5.5	77.5

Iteration and Convergence

Iteration is a reworking of the K-Means algorithm as from the beginning of the calculation but using a new Centroid. While convergence states that the calculation iteration stops. Convergence marking occurs if the last two iterations have the same number of Cluster members. If there is no similarity, the iteration is repeated up to a maximum of 30 iterations. In addition, convergence is stopped if the number of iterations is sufficient but the number of Cluster members is not the same. In this algorithm calculation, convergence occurs when the number of Clusters in iteration 4 is the same as the number of Clusters in iteration 3.

Table 5. Convergence from iteration 1 to iteration 4

Cluster A	C2	C2	C3	C1	C3	C2	C2	C2	C2	C2
Iteration 2	C2	C2	C3	C1	C3	C1	C2	C1	C2	C2
Iteration 3	C2	C1	C3	C1	C3	C1	C2	C1	C2	C2
Iteration 4	C2	C1	C3	C1	C3	C1	C2	C1	C2	C2

K-means Cluster Results

The process of calculating distance, assigning clusters and calculating new centroids is repeated until there is no change in cluster assignment and the iteration stops at Iteration 5, the results are as follows:

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

Table 4. K-Means Cluster Results

Student ID	Parental Income	Parent Education	Internet Access	Study Hours/Day	Presence	Motivation	Stress	Self-confident	Average	Cluster	Information Cluster
001	Low	2	1	2	90	8	4	7	85	C2	Intermediate
002	Intermediate	4	1	3	95	9	3	8	90	C1	Tall
003	Intermediate	4	0	1	85	7	6	6	80	C3	Low
004	Tall	5	1	4	92	9	2	9	93	C1	Tall
005	Low	1	0	1	80	6	7	5	75	C3	Low
006	Tall	4	1	4	97	10	1	9	95	C1	Tall
007	Intermediate	3	1	3	93	8	4	8	88	C2	Intermediate
008	Tall	4	1	3	96	9	3	9	92	C1	Tall
009	Low	2	0	2	87	7	6	7	82	C2	Intermediate
010	Low	2	1	2	88	8	5	6	84	C2	Intermediate

Silhouette Score Evaluation

This evaluation is done on the Cluster results from the K-Means algorithm. Before starting the measurement. Use the Euclidean distance between each data point that has been obtained previously.

1. a(i): The average distance between data point i and all other data points in its own Cluster.

$$a(1) = \frac{\sum \text{Euclidean distance to other members in C1}}{\text{Number of other members in C1}}$$

Ais the average distance of student 002 from all other students in cluster C1. Students in cluster C1 other than Student 002 are:

a. Student 004
b. Student 006
c. Student 008

Distance between Student 002 and other students in cluster C1:
Distance between Student 002 and Student 004: 4.69
Distance between Student 002 and Student 006: 6.00
Distance between Student 002 and Student 008: 2.45

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

Average distance (a):

$$a = (4.69 + 6.00 + 2.45) /3 = 4.38$$

b is the average distance between student 002 and all students in the nearest cluster. The nearest cluster for Student 002 is C2. Students in cluster C2 are:

a. Student 001
b. Student 007
c. Student 009
d. Student 010

Distance between Student 002 and other students in cluster C2:
Distance between Student 002 and Student 001: 7.62
Distance between Student 002 and Student 007: 3.32
Distance between Student 002 and Student 009: 12.17
Distance between Student 002 and Student 010: 9.95

Average distance (b):

$$b = (7.62 + 3.32 + 12.17 + 9.95) /4 = 8.26$$

s(i): Silhouette Score for data point i, calculation starts from Student 002 s(i) The formula for calculating Silhouette Score:

$$S = \frac{b - a}{\max(a,b)}$$

With value:
a = 3.80
b = 8.26
S = (8.26 – 3.80) / max(3.80, 8.26) = 0.47
So, the Silhouette Score for Student 002 is 0.47

Next, by using the values a(i) and b(i), all s(i) are obtained for all data points resulting from the K-Means Cluster.

Table 5. Silhoutte Score Cluster K-Means

Class	Student	a(i)	b(i)	S(i)
Cluster 1	002	4.38	8.26	0.47
	004	4.94	10.12	0.51
	006	5.22	13.86	0.62
	008	4.09	11.33	0.64
Cluster 2	001	4.04	10.02	0.60
	007	6.84	5.59	-0.13
	009	5.63	14.45	0.61
	010	6.42	12.16	0.47
Cluster 3	003	7.87	7.33	-0.07
	005	7.87	14.22	0.45

2. Average Silhouette Score s(i)
   a. Cluster 1
      = (0.47 +0.51 + 0.62 + 0.64) /4 = 0.56
      The average Silhouette Score for Cluster 1 is 0.56, which indicates that

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

Studentsin this Cluster have higher quality. These higher scores indicate that they are closer to the cluster center.

b. Cluster 2

The average Silhouette Score for Cluster 2 is 0.39, indicating that students in this cluster have lower Clusterization quality than Cluster 1. Student 001 and Student 009 have higher Silhouette Scores, indicating that they are better with Clusterization in Cluster 2.

c. Cluster 3

The average Silhouette Score for Cluster 3 is 0.19, indicating that students in this Cluster have a low level of Clusterization quality. Student 005 has a higher Silhouette Score and Student 003 has a lower Silhouette Score.

Characteristics and Discussion of Each Cluster

Cluster 1: High Achievement

Characteristics: This cluster is dominated by students with high socioeconomic backgrounds, where parents' income is relatively high and parents' education level is at least S1. Students in this cluster have full access to the internet and other learning resources. They also show good learning behavior with an average of 3-4 hours of study time per day and almost perfect school attendance (95%). Psychological factors show high motivation and self-confidence, while their stress levels are relatively low.

Academic Achievement: The average academic scores in this cluster range from 88-92, indicating that students from better socioeconomic backgrounds tend to have higher academic performance. Discussion: Students in this cluster have adequate resources, both economically and in terms of access to educational facilities, which support their academic achievement. Good family support and access to technology such as the internet also help to increase motivation and learning effectiveness.

Cluster 2: Intermediate Achievement

Characteristics: This cluster consists of students with middle socioeconomic backgrounds. Parents' education varies from high school to undergraduate, and internet access is not always available. Students' study hours in this cluster range from 2-3 hours per day with a fairly good school attendance rate (90%). Learning motivation is still quite high, but their stress levels are higher than students in the first cluster.

Academic Achievement: The average score of students in this cluster is 80-85, which indicates fairly good academic achievement but not as high as students in the first cluster. Discussion: Students in this cluster have relatively high motivation to learn, although access to educational resources such as the internet may not always be optimal. These factors make their achievement at a medium level. Higher levels of stress are also a challenge, which may affect their ability to maximize academic achievement.

Cluster 3: Low Achievement

Characteristics: This cluster consists of students from families with low socioeconomic

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

backgrounds. Most parents' education is only up to junior high or high school level, and students in this cluster generally do not have access to the internet at home. Their average study hours are only 1-2 hours per day, with a lower attendance rate than the other two clusters (75-85%). High stress levels and low self-confidence are also typical of students in this cluster.

Academic Achievement: The average score of students in this cluster is 65-75, which indicates lower academic achievement. Discussion: Students in this cluster face socioeconomic challenges and lack of access to educational facilities. These limitations affect their study hours and lead to lower academic performance. Psychological factors such as low self-esteem and high stress levels are also major barriers that need to be addressed.

Davies-Bouldin Index (DBI)

Davies-Bouldin Index (DBI) is a metric used to evaluate the quality of clustering results in data analysis. This metric measures how well the separation between clusters is and how compact or homogeneous a cluster is based on the distance between data within and between clusters. DBI is calculated as the average of the ratio of intra-cluster distance and inter-cluster distance for each cluster. Mathematically, the formula is

$$DBI = \frac{1}{n}\sum_{i=1}^{n} \frac{max}{j \neq 1}\left(\quad \frac{si + sj}{dij} \right)$$

Where:

n is the number of clusters.
Si is the intra-cluster variance for cluster i.
Sj is the intra-cluster variance for cluster j.
d(ij) is the distance between the centroids of clusters i and j.
max means taking the largest ratio value between cluster i and other cluster j.

Intra-Cluster Variance

a. Euclidean distance for each student in cluster 1
   Distance between Student 002 and *Centroid* Cluster 1
   Jarak ke Centroid Cluster 1

Siswa 02 =
$\sqrt{(4-4.3)^2 + (1-1)^2 + (3-3.5)^2 + (95-95)^2 + (9-9.3)^2 + (3-2.3)^2 + (8-8.8)^2 + (90-92.5)^2}$
= 2.78

The results of the distance calculation for all students in Cluster 1, Cluster 2 and Cluster 3 can be seen in the table below:

Table 6. Average distance values in Cluster 1

Student	Cluster 1	Average variance
002	2.78	2.76
004	3.20	
006	3.57	
008	1.50	

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

b. Euclidean distance for each student in Centroid cluster 2

Table 7. Average distance values in Cluster 2

Student	Cluster 2	Average variance
001	1.06	3.06
007	5.06	
009	4.08	
010	2.03	

c. Euclidean distance for each student in cluster 3

Table 8. Average distance values in Cluster 3

Student	Cluster 3	Average variance
003	3.94	3.94
005	3.94	

Distance between Centroids

Distance of Cluster 1 Centroid to Cluster 2 Centroid

C1 dan C2

$$= \sqrt{(4.3-2.3)^2+(1-0.8)^2+(3.5-2.3)^2+(95-89.5)^2+(9.3-7.8)^2+(2.3-4.8)^2+(8.8-7.0)^2+(92.5-84.8)^2}$$

= 10,368

Distance of Cluster 1 Centroid to Cluster 3 Centroid

C1 dan C3

$$= \sqrt{(4.3-2.5)^2+(1-0.0)^2+(3.5-1.0)^2+(95-82.5)^2+(9.3-6.5)^2+(2.3-6.5)^2+(8.8-5.5)^2+(92.5-77.5)^2}$$

= 20,628

Distance of Cluster 2 Centroid to Cluster 3 Centroid

C2 dan C3

$$= \sqrt{(2.3-2.5)^2+(0.8-0.0)^2+(2.3-1.0)^2+(89.5-82.5)^2+7.8-6.5)^2+4.8-6.5)^2+(7.0-5.5)^2+(84.8-77.5)^2}$$

= 10,518

Calculate Ratio and DBI

R12 = (2.76 + 3.06)/10.368
= 0.56

R13 = (2.76 + 3.94)/20.628
= 0.32

R23 = (3.06 + 3.94)/10.518
= 0.67

$$DBI = \frac{1}{3}(R_{12} + R_{13} + R_{23})$$

Table 9. Ratio values of each cluster

R	Data to -i			R-Max
	1	2	3	
1	0	0.56	0.32	0.56
2	0.56	0	0.67	0.67
3	0.32	0.67	0	0.67

$$DBI = \frac{1}{3}(0.56 + 0.67 + 0.67) = 0.63$$

Jurnal Info Sains : Informatika dan Sains
Volume 14, Number 04, 2024, DOI 10.54209/infosains.v14i04
ESSN 2797-7889 (Online)
https://ejournal.seaninstitute.or.id/index.php/InfoSains

Evaluation Results

1. Silhouette Score from K-Means Cluster results In the K-Means results, Cluster 1 shows the best Cluster quality with an average Silhouette Score of 0.56, followed by Cluster 2 (0.39) and Cluster 3 (0.19).
2. The DBI value is0.63 indicates that the Clustering performed is quite good at separating the data into well-defined Clusters. This number is relatively low, meaning that the resulting Clusters have high similarities between members within the Cluster and clear differences between Clusters.

Graph of distance between centroids

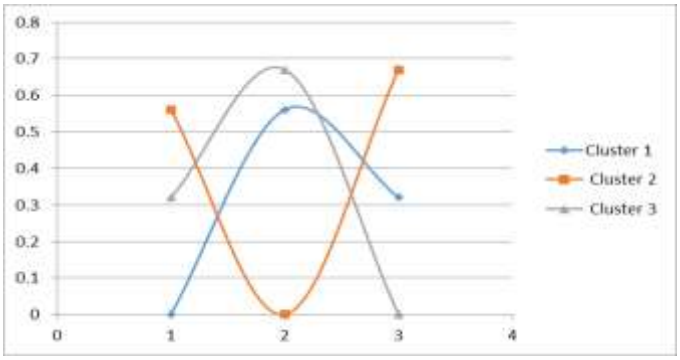

Figure 1. Distance between centroids

CONCLUSION

The conclusion of the research conducted is that the author can draw conclusions based on the results obtained as follows: The Influence of Socio-Economic Factors on Academic Achievement: Students from better socio-economic backgrounds, as seen in Cluster 1, have better access to learning facilities such as the internet and parental support. They also tend to have more optimal study time and higher motivation, which contribute positively to their academic performance. In contrast, students in Cluster 3, who come from lower socio-economic backgrounds, face greater challenges such as lack of internet access and educational support from parents, leading to lower academic achievement. The Role of Behavioral and Psychological Factors: Motivation and self-confidence play a significant role in determining student achievement. Students with high motivation and good attendance, such as those in Cluster 1, show better academic results than students with low motivation. Stress and emotional balance are also important factors. Students in Cluster 2 experienced higher levels of stress, which impacted their academic performance even though they had relatively good access to educational resources. This suggests that stress management and mental health need to be addressed. Cluster Distance Analysis and Davies-Bouldin Index: The results of the calculation of the distance between centroids show that there are quite clear differences between clusters, especially between Cluster 1 (students with good backgrounds) and Cluster 3 (students with less supportive backgrounds). The Davies-Bouldin Index (DBI) value of 0.63 indicates that the clustering performed is quite good, with clear separation between clusters and a fairly large distance between clusters.

Recommendation: Cluster 3 requires greater interventions, such as increased internet access, financial assistance, and motivational and learning support programs to improve academic achievement. Cluster 2 needs to focus on stress management and mental balance, so that students can be more focused and improve their academic performance. Cluster 1, although showing good results, also still requires support so that students can maintain their motivation and academic excellence. This clustering analysis shows that socio-economic, behavioral, and psychological factors have a significant impact on students' academic achievement. Students from better-off backgrounds have greater access to learning resources and higher motivation, while students from less-supportive backgrounds face greater challenges. Appropriate interventions for groups of students in need, especially in Clusters 2 and 3, will go a long way in improving the quality of their education and learning outcomes.

REFERENCES

[1] Kumar, V., & Minz, S. (2014). Cluster analysis using K-Means algorithm. International Journal of Computer Applications, 97(5), 18-24.
[2] Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques (3rd ed.). Morgan Kaufmann Publishers.
[3] Zhang, Y., & Xia, R. (2019). Clustering students based on socioeconomic and behavioral data: A K-Means analysis approach. Journal of Educational Data Science, 5(3), 212-229
[4] Al-Jabri, I. M., & Roztocki, N. (2015). Adoption of ERP systems: Does information transparency matter?. Telematics and Informatics, 32(2), 300-310.
[5] Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1(2), 224-227.
[6] Suparman, L. (2020). Pengaruh motivasi dan stress belajar terhadap prestasi siswa. Jurnal Pendidikan, 10(1), 45-56.
[7] Rokach, L., & Maimon, O. (2005). Clustering methods. In O. Maimon & L. Rokach (Eds.), Data Mining and Knowledge Discovery Handbook (pp. 321-352).