



K-Means Clustering of Students' Anxiety Levels Based on DASS-42 Scores

Ni Putu Dea Sillviari¹, Ayu Gde Chrisna Udayanie^{2*}

^{1,2*}Sistem Informasi, Universitas Bali Dwipa, Indonesia

¹deasillviari@gmail.com; ^{2*}udayaniechrisna@gmail.com

ARTICLE INFO

Article history:

Received 20 December 2025

Revised 21 January 2026

Accepted 30 January 2026

Keywords: K-Means;

Anxiety;
DASS-42;
Clustering;
Students

ABSTRACT

Anxiety is a psychological state that may adversely influence students' focus, thinking ability, and academic achievement. This study seeks to group students' anxiety levels using the K-Means clustering method based on DASS-42 questionnaire scores. The dataset consisted of responses from 835 tenth-grade students enrolled at a private vocational high school in Gianyar, Bali. In the preprocessing phase, 14 anxiety-related items from the DASS-42 scale were selected, and an overall anxiety score was computed for each participant. The K-Means algorithm was applied with five clusters ($K = 5$) corresponding to anxiety categories: normal, mild, moderate, severe, and extremely severe. The clustering process generated centroid values of 4.20, 9.15, 13.69, 19.30, and 27.50, respectively. The results showed that most students were grouped into the moderate anxiety cluster, representing 32% of the total sample. Meanwhile, 24% were classified as normal, 11% as mild, 19% as severe, and 14% as extremely severe. When compared with the standard DASS-42 classification, the K-Means approach demonstrated greater flexibility than interval-based methods. The findings are expected to help schools better understand students' psychological conditions through computational analysis and support informed educational decision-making.

© 2026 Authors. Licensed under [CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/)

1. Introduction

Anxiety is a psychological condition that can affect students' concentration, cognitive abilities, and academic performance [1]. High levels of anxiety may lead to decreased focus, reduced learning motivation, and negative impacts on students' academic outcomes. Therefore, identifying and understanding students' anxiety levels is essential in the educational context to support an effective and optimal learning process [2].

One of the widely used instruments for assessing psychological conditions, particularly anxiety, is the Depression Anxiety Stress Scales (DASS-42) [3]. This instrument consists of 42 statements designed to quantitatively measure levels of depression, anxiety, and stress [4]. Although DASS-42 has been proven to be valid and reliable, the measurement results are generally treated on an individual basis, where analysis is typically limited to score

calculation and categorical classification without applying computational data processing to uncover broader patterns [5].

Along with the advancement of information technology, data mining approaches can be utilized to process and analyze psychological data in a more comprehensive manner [6]. Data mining enables data exploration to identify patterns, relationships, and specific group characteristics that cannot be obtained through conventional analytical methods [7]. One data mining technique that can be applied is clustering, which aims to group data based on the similarity of their characteristics.

In this study, the K-Means algorithm is employed as a clustering method to group students' anxiety levels based on questionnaire data obtained from DASS-42. The K-Means algorithm is selected due to its ability to effectively and efficiently cluster numerical data [8]. The objective of this research is to classify students' anxiety levels into several groups based on data patterns formed using the K-Means algorithm, thereby providing a more comprehensive overview of students' anxiety conditions and serving as a foundation for decision-making in the educational field. In this study, DASS-42 is used as a baseline for comparison, while K-Means is applied as a tool for data pattern exploration.

2. Method

Data Collection

The research data were collected from students' responses to the Depression Anxiety Stress Scales (DASS-42) questionnaire administered to tenth-grade students at a private vocational high school in Gianyar, Bali. A total of 835 students participated as respondents in this study.

The DASS-42 instrument consists of 42 statements assessed using a 4-point Likert scale ranging from 0 to 3:

- 0 = Never experienced
- 1 = Occasionally
- 2 = Frequently
- 3 = Very frequently

This study specifically concentrates on the anxiety subscale, which comprises 14 items. Consequently, only statements that reflect anxiety-related aspects were included in the analysis process [9]. The collected responses were then converted into numerical data and utilized for further processing and analysis using clustering techniques.

Preprocessing Data

The data pre-processing stage was conducted to ensure data quality prior to the clustering process. Questionnaire response data were compiled into a Microsoft Excel file. Subsequently, the anxiety level of each respondent was calculated by summing the scores of all items included in the DASS-42 anxiety subscale [10].

The total anxiety score for each respondent was calculated using the following formula [11]:

$$Anxiety\ Score = \sum_{k=1}^{14} X_k \quad (1)$$

Where X_k represents the score of the k-th anxiety item [12]. The aggregated anxiety item scores were then recapitulated and saved in .csv format. The resulting numerical data were subsequently used as input for the clustering process [13]. In cases where missing or incomplete data were identified, such records were excluded to prevent interference with the analysis and to ensure more reliable clustering results.

K-Means Clustering

The method applied in this study is the K-Means algorithm, which is a clustering technique in data mining used to group data based on similarity levels. This algorithm aims to minimize the distance between data points within the same cluster while maximizing the distance between different clusters [14].

In this study, the dataset consists of a single numerical variable, namely the total anxiety score derived from the Anxiety subscale of DASS-42. Although only one anxiety score variable is used, the K-Means algorithm remains appropriate for identifying natural cluster boundaries formed from the data distribution, which do not necessarily follow the fixed DASS-42 interval classifications. This data-driven approach enables pattern-based grouping and provides an additional perspective compared to rule-based categorization.

The number of clusters used in this study is set to five clusters ($K = 5$), representing the levels of anxiety: normal, mild, moderate, severe, and extremely severe. The stages of the K-Means algorithm implemented in this research are described as follows:

1. Determining the number of clusters as $K = 5$, representing anxiety levels: Normal, Mild, Moderate, Severe, and Extremely Severe.
2. Calculating the distance between each data point and the centroid. Since the data contain only one variable (anxiety score), the Euclidean distance calculation is simplified as follows [15]:

$$d(x_i, c_j) = |x_i - c_j| \quad (2)$$

x_i = anxiety score of the i -th respondent

c_j = centroid of the j -th cluster

3. Computing the distance between each data point and the centroid using the Euclidean Distance method [16].
4. Updating the centroid based on the average anxiety score within each cluster, calculated as:

$$c_j = \frac{1}{n_j} \sum_{i=1}^{n_j} X_{ij} \quad (3)$$

X_{ij} = nilai data ke- i pada cluster ke- j

5. Assigning each data point to the cluster with the nearest centroid.
6. Updating the centroid values based on the mean of the data points in each cluster.
7. Repeating the distance calculation and centroid update process until no changes occur in cluster assignments or the results reach a convergent state.

DASS-42 Categorization

In addition to clustering, anxiety levels were classified using the DASS-42 score intervals [17]. This classification serves as a comparative reference for interpreting the clustering outcomes. The DASS-42 categories were not used as training labels; rather, they were applied solely for post-clustering interpretation.

Table 1. Assessment Indicators

Category	Score Range
Normal	0 – 7
Mild	8 – 9
Moderate	10 – 14
Severe	15 – 19
Extremely Severe	≥ 20

3. Result and Discussions

K-Means Clustering Results

The application of the K-Means algorithm with $K = 5$ to student anxiety score data produced five distinct clusters representing anxiety levels ranging from low to very high. The centroid value of each cluster reflects the average anxiety score of students within that group.

Table 2. Anxiety Centroid Cluster

Cluster	Centroid Skor Kecemasan
1	4,20
2	9,15
3	13,69
4	19,30
5	27,50

As shown in the table, the centroid values increase progressively across clusters, indicating that the K-Means algorithm successfully differentiated student anxiety levels based on score similarity.

Data Distribution in Each Cluster

The number of students assigned to each cluster is presented in the following table:

Table 3. Number of Data per Cluster

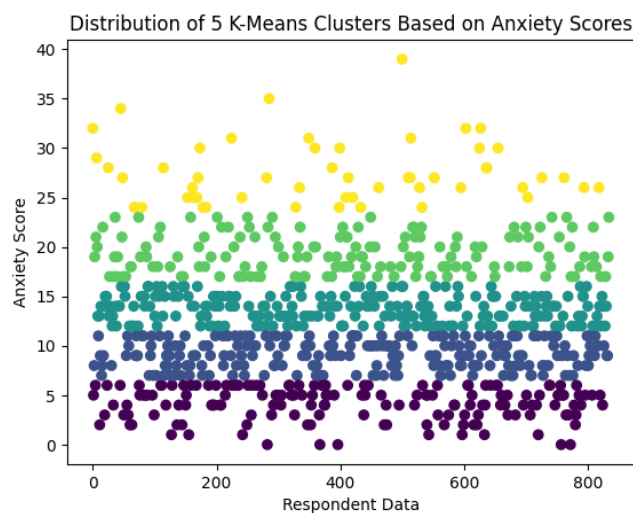
Cluster	Number of Students	Percentage
1	161	19.3%
2	233	27.9%
3	235	28.1%
4	153	18.3%
5	53	6.3%
Total	835	100%

Discussions

These results indicate that most students are concentrated in clusters 2 and 3, which correspond to mild to moderate anxiety levels. Cluster 5 contains the smallest number of students, suggesting that only a limited proportion of students experience very high anxiety. The clustering outcome also reveals an uneven distribution of students across clusters. Clusters with moderate anxiety centroids contain more members than those representing very low or very high anxiety. This pattern suggests that the majority of students experience anxiety at a moderate level.

Cluster Distribution Visualization

The cluster distribution plot (Figure 1) illustrates the grouping pattern of student data based on anxiety scores. Different colors represent the five anxiety clusters generated by the K-Means algorithm.

**Figure 1.** Distribution of 5 K-Means Clusters

The figure displays the results of K-Means clustering based on student anxiety scores. Each point represents a respondent, where the horizontal axis indicates the respondent index and the vertical axis represents the anxiety score. Different colors distinguish the five resulting anxiety clusters. The visualization shows a clear and gradual grouping pattern from low to high anxiety scores. Clusters with low anxiety scores fall within the range of 0–

10, moderate anxiety clusters are within 10–15, and high anxiety clusters exceed 15. This distribution demonstrates that the K-Means algorithm consistently separates student anxiety levels based on score similarity. Furthermore, the relatively even spread of data points along the respondent index suggests that anxiety levels are not influenced by data collection order but are determined solely by anxiety score values. This confirms that the clustering results reflect the natural structure of the data and can be used to comprehensively describe student anxiety conditions. Overall, the findings reinforce that K-Means effectively maps anxiety levels in a gradual and logical manner.

Comparison of K-Means Clusters with DASS-42 Categories

For comparison purposes, anxiety levels were also categorized using the DASS-42 standard. The distribution of these categories is shown in the following table:

Table 4. Comparison of K-Means Clusters with DASS-42 Categories

DASS Category	Number of Students	Percentage
Normal	200	24.0%
Mild	92	11.0%
Moderate	268	32.1%
Severe	159	19.0%
Extremely Severe	116	13.9%
Total	835	100%

Based on these results, the Moderate category is the most dominant, followed by the Normal category. In addition to clustering, anxiety data were categorized using DASS-42 rules based on the total questionnaire score. The categorization results indicate that most respondents fall into the Normal and Moderate categories. When K-Means clustering results are compared with DASS-42 categories, a one-to-one correspondence between clusters and categories is not observed. Several clusters span the same DASS category range. This discrepancy arises from differences in methodological approaches between clustering and rule-based classification.

Table 5. Comparative Analysis

Cluster	Normal	Mild	Moderate	Severe	Extremely Severe	Total
1	150	11	0	0	0	161
2	50	81	102	0	0	233
3	0	0	166	69	0	235
4	0	0	0	90	63	153
5	0	0	0	0	53	53
Total	200	92	268	159	116	835

Comparative Analysis

When comparing the clustering results with DASS-42 categories, the following patterns are observed:

- Clusters 1 and 2 are predominantly associated with the Normal and Mild categories
- Cluster 3 corresponds mainly to the Moderate category

- Clusters 4 and 5 are closely related to the Severe and Extremely Severe categories. However, not all clusters align perfectly with DASS-42 categories. This occurs because:
 - DASS-42 uses fixed score intervals
 - K-Means forms clusters based on the natural structure of the data

Therefore, K-Means offers a more flexible approach to understanding variations in student anxiety levels.

Implications of the Findings

The clustering results can be utilized by educational institutions to:

- Identify groups of students with high anxiety levels
- Provide appropriate psychological interventions or counseling support
- Develop learning strategies that are better aligned with students' mental health conditions

4. Conclusion

Based on the findings of this study, the K-Means algorithm with five clusters ($K = 5$) successfully categorized students' anxiety levels using scores from the DASS-42 anxiety subscale. The centroid values of each cluster exhibited a clear and progressive increase—4.20, 9.15, 13.69, 19.30, and 27.50—reflecting anxiety levels ranging from normal to extremely severe. The clustering results indicate that the largest proportion of students belonged to cluster 3, representing moderate anxiety, with 235 students (28.1%) of the total respondents. This was followed by the mild anxiety cluster (cluster 2), which included 233 students (27.9%), the normal anxiety cluster (cluster 1) with 161 students (19.3%), the severe anxiety cluster (cluster 4) comprising 153 students (18.3%), and the extremely severe anxiety cluster (cluster 5) with 53 students (6.3%).

When compared with the standard DASS-42 classification, the data distribution showed that most students were categorized as having moderate anxiety, totaling 268 students (32.1%), followed by the normal category with 200 students (24.0%), severe anxiety with 159 students (19.0%), extremely severe anxiety with 116 students (13.9%), and mild anxiety with 92 students (11.0%). These results demonstrate that the clustering approach using the K-Means algorithm provides a more flexible and data-driven mapping of students' anxiety levels than the interval-based categorization of DASS-42. The findings may serve as a foundation for schools to identify groups of students with varying anxiety levels and to develop more targeted psychological intervention and support strategies. Future studies are recommended to incorporate additional psychological variables from the DASS-42, such as depression and stress, as well as to evaluate other clustering techniques in order to obtain more comprehensive and accurate grouping outcomes.

References

- [1] D. Muhammad A *et al.*, "Anxiety of Final Semester Students: Mini Review," *Ahmad Dahlan Med. J.*, vol. 2, no. 2, pp. 85–92, 2021, doi: 10.12928/admj.v2i2.5432.
- [2] N. P. D. Sillviari, I. M. Candiasa, and G. Indrawan, "Classification of Anxiety Levels in Vocational Students Through Life Story Analysis Using Multi-class SVM," vol. 15,

- no. 2, 2025, doi: 10.22437/teknopedagogi.v15i2.46957.
- [3] C. Bimantara Putra, "Implementasi Metode Certainty Factor Pada Sistem Penilaian Gangguan Depresi, Kecemasan dan Stress Berdasarkan DASS - 42," no. April, 2023.
- [4] S. Prodi, I. Kedokteran, G. Fakultas, K. Gigi, and U. Gadjah, "The Relationship Between Practice Preparedness and Dentists' Mental Condition in Semarang During Coronavirus Disease 2019 (COVID-19) Pandemic," vol. 9, no. 1, pp. 138–147, 2022.
- [5] L. Oktamarin *et al.*, "Gangguan Kecemasan (Axiety Disorder) Pada Anak Usia Dini," *J. Multidisipliner Bharasumba*, vol. 1, no. 02, pp. 119–134, 2022, doi: 10.62668/bharasumba.v1i02.192.
- [6] C. C. Soen, I. M. Hardjasasmita, and A. E. Ulitua, "Generalized Anxiety Disorder: Diagnosis and Treatment," *J. Muara Med. dan Psikol. Klin.*, vol. 1, no. 2, p. 133, 2022, doi: 10.24912/jmmpk.v1i2.14865.
- [7] H. Kurniawan and S. Defit, "Data Mining Menggunakan Metode K-Means Clustering Untuk Menentukan Besaran Uang Kuliah Tunggal," vol. 1, no. 2, pp. 80–89, 2020.
- [8] P. Alluri, Z. Chen, T. Thesen, N. C. Jacobson, and W. J. Marrero, "Fairness-aware K-means clustering in digital mental health for higher education students: a generalizable framework for equitable clustering," vol. 9, no. 1, pp. 1–11, 2026.
- [9] A. R. Biromo, A. H. Santoso, E. Destra, and L. Anthony, "Skrining DASS-42 dan Edukasi Pencegahan Terjadinya Gangguan Psikologis DASS-42 Screening and Education on Prevention of Psychological Disorders," vol. 4, pp. 80–88, 2025.
- [10] S. Kusumadewi, H. Wahyuningsih, T. Informatika, U. I. Indonesia, U. I. Indonesia, and P. Korespondensi, "MODEL SISTEM PENDUKUNG KEPUTUSAN KELOMPOK UNTUK PENILAIAN GANGGUAN DEPRESI, KECEMASAN DAN STRESS BERDASARKAN DASS-42 GROUP DECISION SUPPORT SYSTEM MODEL FOR ASSESSMENT OF DEPRESSION, ANXIETY AND STRESS DISORDERS BASED ON DASS-42," vol. 7, no. 2, pp. 219–228, 2020, doi: 10.25126/jtiik.202071052.
- [11] S. Juwariyah, A. Hulvi, and N. Riduan, "Mengukur Faktor Demografi Psikologis: Memprediksi Depresi, Kecemasan, dan Stres dengan menggunakan Machine Learning," vol. 13, 2024, doi: 10.34010/komputika.v13i2.11793.
- [12] S. T. Pertiwi, M. F. Moeliono, and L. Kendhawati, "Depresi, Kecemasan, dan Stres Remaja selama Pandemi Covid-19.," vol. 6, no. 2, pp. 72–77, 2021, doi: 10.36722/sh.v.
- [13] M. Sholeh and U. Lestari, "Evaluation of Data Clustering Accuracy using K-Means Algorithm," vol. 2, no. 01, pp. 385–396, 2024.
- [14] N. K. Zuhul *et al.*, "Klasifikasi Emosi Pada Wajah Dengan Menggunakan K-MEANS Clustering dan KDEF," pp. 243–248, 2021.
- [15] D. Pakuan Putra and B. Agus Wardijono, "Analisis Akurasi Penerapan Algoritma Support Vector Machine Menggunakan Kernel Radial Basis Function pada Penentuan Kelayakan Kredit (Studi Kasus German Kredit Data)," *J. Ilmiah KOMPUTASI*, vol. 19, no. 2, pp. 175–180, 2020, [Online]. Available: <http://dx.doi.org/10.32409/jikstik.19.2.2786>
- [16] P. Taha Ahmadi, "Analyzing Factors Affecting Student Anxiety Using K-means Clustering and Fuzzy Best-Worst Method," vol. 2, no. 1, pp. 1–22, 2025, doi: 10.22034/dmait.2025.542474.1009 Taha.

- [17] P. T. Depresi, D. A. N. S. Pada, K. Bank, X. Y. Z. Dengan, and M. Metode, "Penilaian Tingkat Depresi, Kecemasan, dan Stress Pada Karyawan Bank XYZ dengan Menggunakan Metode DASS-42 Test," pp. 29–39, 2023.