Decision Making and Artificial Intelligence Trends

Volume 2, Issue 1, 1-22.

Analyzing Factors Affecting Student Anxiety Using K-means Clustering and Fuzzy Best-Worst Method

https://doi.org/10.22034/dmait.2025.542474.1009

Taha Ahmadi Pargo[1*]

1. Department of School of Industrial Engineering and Management Systems, Amirkabir University of Technology, Tehran, Iran
*Corresponding Author: ta.ahmadi@aut.ac.ir;

ARTICLE INFO

Article history:
Received: 2025/08/20
Revised: 2025/09/10
Accepted: 2025/09/18
Published: 2025/09/23

Keywords:
Student Anxiety, Education, Fuzzy BWM, K-means Clustering.

ABSTRACT

One of the most difficult challenges that students experience is the anxiety that they struggle with during their studies. Various factors affect or are affected by students' anxiety. Things like depression and self-esteem affect students' anxiety, and the level of anxiety also affects them. Also, students' health and their level of success during their studies are affected by their anxiety. In this article, the effects of anxiety on students' lives are first examined using visualization techniques. The effects of self-confidence and depression factors on students' anxiety are then analyzed using statistical approaches. Subsequently, a clustering analysis is performed using an unsupervised K-means machine learning approach. Based on the elbow method and silhouette score methods, an optimal number of k is determined to be five. These five clusters are students with different levels of self-esteem, depression, and anxiety. Then, the importance of external factors and subfactors affecting students' anxiety was investigated. This process was done using the fuzzy best-worst method. The most important category of factors is the environmental category environmental category with a weight of 41.5%. Also, the most significant factor is basic needs with global weight of 13%.

1. Introduction

Student anxiety is one of the major concerns of universities and families in today's societies. Today, anxiety levels in society are higher than in the past (Piao et al, 2024). University students are no exception, and increased levels of anxiety among students can lead to issues like a decline in academic performance (Owens et al., 2012). As students' academic performance declines, the overall standing of universities correspondingly deteriorates. This will also cause losses for countries: future generations who are going to enter the job market or make an impact on society will face problems that will affect their performance in the future.

In addition to its negative effects on academic performance, high anxiety can also affect people's health. High anxiety can cause or exacerbate various diseases, such as digestive diseases (Mayer et al., 2021) and neurological diseases (Prisnie et al., 2018). Such cases cause various societies to suffer losses because the health of individuals in society, especially young people, is one of the important issues that any country seeking sustainability must pursue.

These cases highlight the importance of examining the factors influencing students' anxiety. To explore these factors effectively, it is essential to assess the physiological effects of anxiety on students. For instance, researchers should investigate whether heightened anxiety contributes to physical health issues among students. Additionally, the impact of anxiety on students' academic performance must be evaluated. Therefore, the primary questions to address in research on factors affecting students' anxiety are its effects on their physical well-being and academic outcomes.

To address this question, the effects of two internal factors - self-esteem and depression - are first examined. Two approaches are employed for this purpose: regression analysis and clustering. These steps are carried out based on data visualization and analysis. For the clustering analysis, the K-means method is applied; initially, clustering is performed, and subsequently, the optimal value of K for describing the clusters is determined using the silhouette score and the elbow method. Finally, the identified clusters are analyzed in detail.

In the next step, the Fuzzy Best–Worst Method is employed to assign weights to the importance of external factors, which are categorized into three groups: environmental, academic, and social. First, weights are assigned to each of these three categories, followed by weighting the factors within each category. By multiplying the category weights by the weights of the factors within them, the global weights of the factors are determined. It should be noted that the use of triangular fuzzy numbers allows for handling the ambiguity in experts' opinions, which have been collected based on linguistic terms.

In this paper, the literature review is first presented, followed by the introduction of the research methodologies. Next, the data are analyzed, and the weighting process is conducted. Finally, in the conclusion section, the results are discussed and summarized.

2. Literature review

With technological advancements, modern technologies such as machine learning have become more accessible to researchers, who in turn use these methods to improve their decision-making and analyses. One area where researchers apply these techniques to enhance their analyses in studies related to student anxiety and stress (Ovi et al., 2025). These approaches are used to detect and monitor anxiety and stress among students. One machine learning method, which falls under the category of unsupervised learning, is the K-means clustering method. Researchers have used this approach to predict stress, anxiety, and depression.

In general, machine learning methods are used to analyze relationships within data, identify hidden patterns, and predict issues such as stress and anxiety. However, another aspect of analyzing the existing factors involves assessing the importance of external factors based on expert opinions. Various studies have examined student anxiety from different perspectives and approaches. One such approach used to investigate the factors influencing anxiety and stress is multi-criteria decision-making. For example, weighting methods are applied to assess the importance of various factors that affect students' anxiety and stress. These approaches have also been used to examine factors influencing stress in other occupations (Rajabi et al., 2020) as well as during the COVID-19 pandemic (Gupta et al., 2024).

Multi-criteria decision-making approaches are applied in various contexts, but one important issue is that these methods require inputs that are usually based on expert opinions. These opinions are often expressed in linguistic terms and then converted into usable numerical values, as the final weights are represented numerically. It is worth noting that expert opinions, when described in words, inherently contain ambiguity. To address this ambiguity, researchers employ fuzzy approaches. Fuzzy numbers—such as triangular fuzzy numbers, trapezoidal fuzzy numbers, type-2 fuzzy numbers, and spherical fuzzy numbers—are used to capture and represent the uncertainty involved.

Table 1 presents studies related to the use of machine learning approaches and multi-criteria decision-making weighting methods, with applications in examining factors affecting stress and anxiety, as well as in predicting them and analyzing related data.

Table 1: Reviewed papers related to analyzing influential factors on anxiety and stress, using Multi-Criteria Decision-Making and Machine-learning methods.

Authors	Article Title	Methodology
(Middeldorp et al., 2005)	Familial Clustering of Major Depression and Anxiety Disorders in Australian and Dutch Twins and Siblings	Bivariate analysis

Authors	Article Title	Methodology
(Sau & Bhakta, 2017)	Predicting anxiety and depression in elderly patients using machine learning technology	Bayesian Network, logistic, Multiple layer perceptron, Naïve Bayes, Random forest, Random tree, J48, Sequential minimal optimisation, Random sub-space, K Star
(Pintelas et al., 2018)	A review of machine learning prediction methods for anxiety disorders	Systematic Review
(Rajabi et al., 2020)	Occupational stressors among firefighters: application of multi-criteria decision making (MCDM)Techniques	Fuzzy Delphi, Fuzzy AHP
(Jarmoszewicz, 2020)	Factors Associated with High Preoperative Anxiety: Results from Cluster Analysis	Cluster analysis
(Priya et al., 2020)	Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms	Decision Tree, Random forrest, Support vector machine, Naïve Bayes, K nearest neighbour
(Arif et al., 2020)	Classification of Anxiety Disorders using Machine Learning Methods: A Literature Review	Systematic Review
(Kumar et al., 2020)	Assessment of Anxiety, Depression and Stress using Machine Learning Models	Bayes classification, K-Nearest Neighbor, Neural network, Tree-based classification,
(Liu et al., 2022)	Use of latent profle analysis and K-means clustering to identify student anxiety profles	K-means clustering
(Ahmed et al., 2022)	Machine learning models to detect anxiety and depression through social media: A scoping review	Systematic Review
(Ancillon, 2022)	Machine Learning for Anxiety Detection Using Biosignals: A Review	Systematic Review
(Taylor et al., 2023)	Cluster-based psychological phenotyping and differences in anxiety treatment outcomes	Cluster analysis
(Garg & Bhardwaj, 2024)	Exploring the influence of factors causing stress among doctoral students by combining fuzzy DEMATEL-ANP with a triangular approach	Fuzzy Dematel-ANP
(Prabkaran, 2024)	Ranking Psychological Risk Factors Among College Students Using Interval-Valued Picture Fuzzy Set with Various Operators and MCDM Techniques	Interval-valued picture fuzzy sets, MCDM

Authors	Article Title	Methodology
(Gupta et al., 2024)	Fuzzy AHP Approach for Multi-Criteria Stress Analysis During COVID-19: A Case Study	Fuzzy AHP
(Olthuis et al., 2024)	What Does 'High Anxiety Sensitivity' Look Like? Using Cluster Analysis to Identify Distinct Profiles of High Anxiety Sensitive Treatment-Seekers	Cluster analysis
(Maral & Özdemir, 2025)	A systematic review on multi-criteria decision-making methods in educational research	Systematic Review
(Ovi et al., 2025)	Protecting Student Mental Health with a Context-Aware Machine Learning Framework for Stress Monitoring	Support vector machine, Random forest, Gradient boosting, XGBoost, AdaBoost, Bagging, Ensemble models

Evidently, clustering approaches such as K-means can be used to uncover hidden relationships within data. In this study, the existing data are first examined, and then clustering is applied to analyze the effects of personality factors on anxiety. In the next step, environmental, academic, and social factors that influence individuals' anxiety are analyzed based on expert opinions.

3. Methodology

In this section of the article, we examine the methods used for the required analyses. First, the K-means clustering method, an unsupervised learning approach, along with the silhouette score and the elbow method, are discussed. Then, the Fuzzy Best–Worst Method (FBWM) using triangular fuzzy numbers is examined.

3.1. K-means Method

K-means is an iterative method used for clustering. This method uncovers hidden patterns that might not be found otherwise. The steps of this method are as follows (Hartigan & Wong, 1979):

1. *Choose the number K.*
2. *The initial cluster centroids are selected randomly from the dataset (Randomly select K points as the initial cluster centers.)*
3. *Assign each data point to the nearest centroid, based on its Euclidean distance.*
4. *Calculate the centroid of each cluster and set it as the new cluster center.*
5. *Repeat steps 3 and 4 until the cluster centers no longer change, i.e., the change in their positions falls below a predefined threshold or the maximum number of iterations is reached.*

The number of points, K, must be provided as input to this method. Determining the optimal K for effective clustering is done through various approaches, with the elbow method and the silhouette score method being among the most commonly used in this field(Kodinariya & Makwana, 2013).

In the elbow method, a graphical approach, inertia is first calculated. Inertia for each cluster is determined by computing the Euclidean distances between each point and the cluster center, then summing these distances for each cluster. The total inertia is calculated by combining the inertia of all clusters. Inertia can be computed using Equation (1). As the number of clusters (K) increases, inertia decreases, but the computational load increases. Therefore, the inertia for different values of K is plotted on a graph, and the point resembling an "elbow" is identified, where inertia decreases rapidly before this point but slows down afterward, while computational load increases. Choosing this point may also depend on the personal judgment of the person performing the calculations. In Equation (1), K represents the total number of clusters, c denotes a single cluster from the set of clusters, xi represents point i (which is a member of cluster c), and μc indicates the centroid of cluster c.

$$Inertia = \sum_{c=0}^{K} \sum_{x_i \in c} \left(\left\| x_i - \mu_c \right\|^2 \right) \tag{1}$$

As mentioned, the elbow method is a graphical approach, and selecting the appropriate number of K depends on the judgment of the person performing the calculations. Therefore, using non-graphical methods alongside this approach is recommended. The silhouette score method is one such approach, which indicates how close each point is to other points in its own cluster and how well it is distinguished from points in other clusters. Higher silhouette scores are better. To calculate this score, the S_i for each data point must be computed, and the silhouette score is derived from the average S_i of all data points. To calculate the S_i, Equation (2) is used.

$$S_i = \begin{cases} \dfrac{b_i - a_i}{max\{a_i, b_i\}}, & i > 1 \\ 0 & , i = 1 \end{cases} . \tag{2}$$

In Equation (2) (previous formula), a_i represents the cohesion of each cluster, indicating how close data point i is to other data points within the same cluster. b_i shows how far data point i is from data points in other clusters, effectively measuring the separation of each data point from points in other clusters. Equations (3-4) are used to calculate a_i and b_i.

$$a_i = \frac{1}{|C| - 1} \sum_{j \in C,\ i \neq j} dist(i, j) \tag{3}$$

As shown in Equation (3), a_i represents the average distance between point i and all other points in cluster C (j points). Also, indicates the number of members present in cluster C.

$$b_i = \min_{W \neq C} \frac{1}{|W|} \sum_{j \in W} dist(i, j) \tag{4}$$

Additionally, Equation (4) is used to calculate . b_i represents the minimum of average distances from point i (Which is in cluster C) to j points in other clusters (W clusters).

3.2. Fuzzy Best-Worst Method

Fuzzy theory, introduced by Zadeh (Zadeh, 1966), is an effective approach for modeling uncertainties in decision-making environments. It can be integrated with various weighting methods proposed in multi-criteria decision-making approaches to model uncertainty in the weighting process, which may arise from experts' verbal scales. One widely used approach is the triangular fuzzy numbers method. Equation (5) represents a triangular fuzzy number (TFN), and Equation (6) represents the membership function of a triangular fuzzy set. Also, it is possible to convert a crisp number C to a fuzzy number, using equation (7).

$$\tilde{f} = \left(f_l, f_m, f_u \right) \tag{5}$$

$$\mu_{\tilde{A}}(x) = \begin{cases} \dfrac{x - f_l}{f_m - f_l}, \; f_l \le x < f_m \\ \dfrac{f_u - x}{f_m - f_u}, \; f_m \le x < f_u \\ 0, x \langle \, f_l \, or \, x \rangle \, f_u \end{cases} \tag{6}$$

$$\tilde{C} = \left(C, C, C \right) \tag{7}$$

The four basic operations on two fuzzy numbers can also be calculated using Equations (Moslem et al., 2020) (8-11).

$$\tilde{f} + \tilde{g} = \left[f_l + g_l, f_m + g_m, f_u + g_u \right] \tag{8}$$

$$\tilde{f} - \tilde{g} = \left[f_l - g_u, f_m - g_m, f_u - g_l \right] \tag{9}$$

$$\tilde{f} \times \tilde{g} = \left[f_l \times g_l, f_m \times g_m, f_u \times g_u \right] \tag{10}$$

$$\tilde{f} \div \tilde{g} = \left[f_l \div g_u, f_m \div g_m, f_u \div g_l \right] \tag{11}$$

Also, it is possible to defuzzify a fuzzy number using Equation (12), in which R(*f*) is the deffuzified value of fuzzy number .

$$\mathrm{R}\left(f \right) = \frac{f_l + f_m + f_u}{3} \tag{12}$$

Based on the provided equations, fuzzy theory can be applied to various weighting methods. One of the most popular weighting methods among researchers is the Best-Worst Method (BWM), introduced by Rezaei (Rezaei, 2015). It is possible to implement a fuzzy approach in this method as well. In this method, the expert first identifies the best criterion, the worst criterion, and performs pairwise comparisons of the best criterion with other criteria, and compares each criterion to the worst criterion. Then, using the mathematical model provided in Equations (13-18), the weighting is carried out (Amoozad Mahdiraji, 2018).

$$min \, \xi^* \tag{13}$$

s.t.

$$\left| \frac{\tilde{w}_{best}}{\tilde{w}_j} - \tilde{A}_{bj} \right| \le \left(\xi^*, \xi^*, \xi^* \right), \; \forall j \in 1, 2, \ldots, n \tag{14}$$

$$\left| \frac{\tilde{w}_j}{\tilde{w}_{worst}} - \tilde{A}_{jw} \right| \le \left(\xi^*, \xi^*, \xi^* \right), \; \forall j \in 1, 2, \ldots, n \tag{15}$$

$$w_{l,j} \le w_{m,j} \le w_{u,j}, \;\; \forall j \in 1, 2, \ldots, n \tag{16}$$

$$\sum\nolimits_{j=1}^{n} \mathrm{R}\left(\tilde{w}_j \right) = 1, \;\; \forall j \in 1, 2, \ldots, n \tag{17}$$

$$w_{l,j} \geq 0, \quad \forall j \in 1,2,\ldots,n. \tag{18}$$

In this model, $\tilde{w} = (w_{l,j},\ w_{m,j},\ w_{u,j})$ is the fuzzy weight of each criterion, $\tilde{A}_{bj}$ is the fuzzy pairwise comparison between the best criteria and the criterion *j*, and $\tilde{A}_{jw}$ is the fuzzy pairwise comparison between the criterion *j* and the worst criterion. Conversion of linguistic terms to fuzzy numbers is as follows in Table 2.

Table 2: Converting linguistic terms to fuzzy numbers.

Linguistic Term	Triangular fuzzy number
Equally Important (EI)	(1, 1, 1)
Weakly Important (WI)	(2/3,1,1.5)
Fairly Important (FI)	(1.5,2,2.5)
Very Important (VI)	(2.5, 3,3.5)
Absolutely Important (AI)	(3.5,4,4.5)

Also, it is possible to calculate consistency ration as follows:

$$Consistency\ Ratio = \frac{\xi^*}{Consistency\ Index} \tag{19}$$

Consistency index should be obtained from Table 3.

Table 3: Consistency indices in fuzzy BWM.

$\tilde{A}_{Bw}$	Equally Important	Weakly Important	Fairly Important	Very Important	Absolutely Important
Consistency index	3.00	3.80	5.29	6.69	8.04

4. Case Study

This paper utilizes a dataset extracted from a questionnaire. The dataset is published on Kaggle under the Apache 2.0 license (Ovi, 2025). It contains information on anxiety levels, depression and self-esteem, physical conditions such as the occurrence of headaches, as well as sleep quality and academic status. Additionally, it includes the effects of various factors such as living conditions, safety, peer pressure, and other related factors. In this study, this dataset is employed for data analysis purposes.

At the beginning, Figures 1, 2, and 3 illustrate the relation between anxiety and experiencing headaches, sleep quality, and academic performance. In these figures, anxiety is represented by a numerical value, where higher numbers indicate higher levels of anxiety.

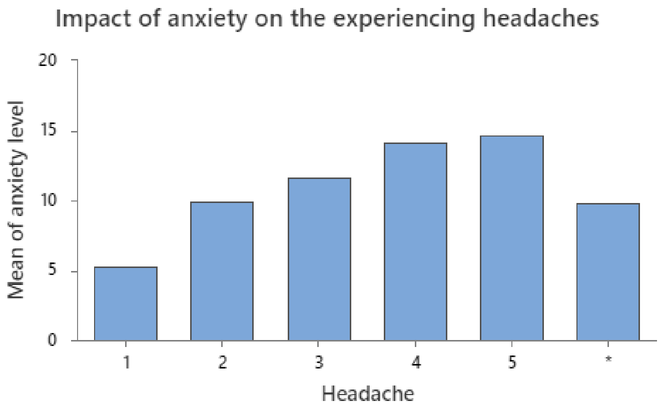

Figure 1: Impact of anxiety on the experiencing headaches (1 to 5: slight to extreme, *: missing value)

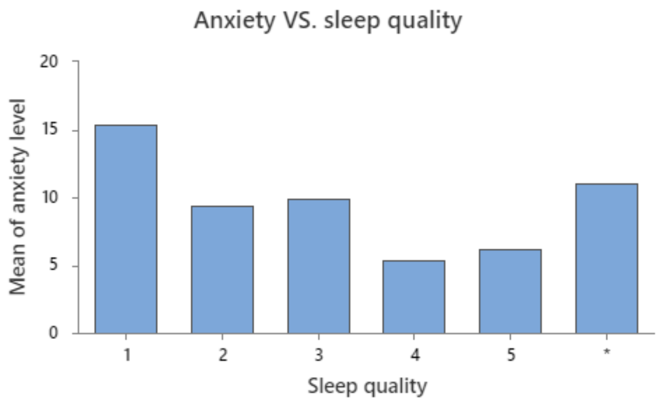

Figure 2: Anxiety VS. sleep quality (1 to 5: Poor to Well, *: missing value)

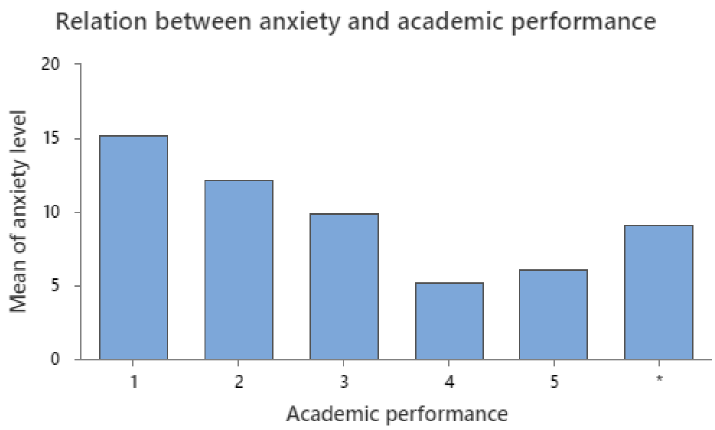

Figure 3: Relation between anxiety and academic performance (1 to 5: Poor to Weel, *: missing value)

As shown in Figure 1, higher levels of anxiety are associated with a greater likelihood of experiencing headaches, indicating a relationship between anxiety and headache occurrence. Figure 2 demonstrates an inverse relationship between sleep quality and anxiety levels. Similarly, Figure 3 illustrates an inverse relationship between anxiety and academic performance, meaning that as anxiety increases, academic performance decreases. These findings highlight the impact of increased stress levels on human life, particularly in terms of academic performance and health.

4.1. The Relationship Between Self Esteem, Depression, and Anxiety

In this section, data analysis is conducted on three factors: anxiety, self-esteem, and depression. First, a regression analysis is performed, followed by a clustering analysis. According to the regression analysis, the results of which are presented in Table 4, self-esteem and depression have significant effects on the response variable, anxiety. Since the p-value is nearly equal to zero, the null hypothesis of no effect for these two factors is rejected. In general, as self-esteem increases, anxiety decreases, and as depression increases, anxiety increases. However, despite the significance of these two factors, the fitted regression equation shows a low R-squared value. This indicates that the model is not able to adequately explain the variability in the data. Moreover, applying linear regression to any data requires the presence of assumptions in the data that do not necessarily exist in data extracted from questionnaires. Therefore, the use of other techniques is also recommended for data analysis. Table 5 shows the summary of the fitted regression model.

Table 4: Coefficients in fitted regression model

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	8.764	0.539	16.26	0.000	
self-esteem	-0.1405	0.0206	-6.81	0.000	1.23
depression	0.3038	0.0238	12.78	0.000	1.23

Table 5: Regression model summary

S	R-sq	R-sq(adj)	R-sq(pred)
5.44021	25.85%	25.70%	25.37%

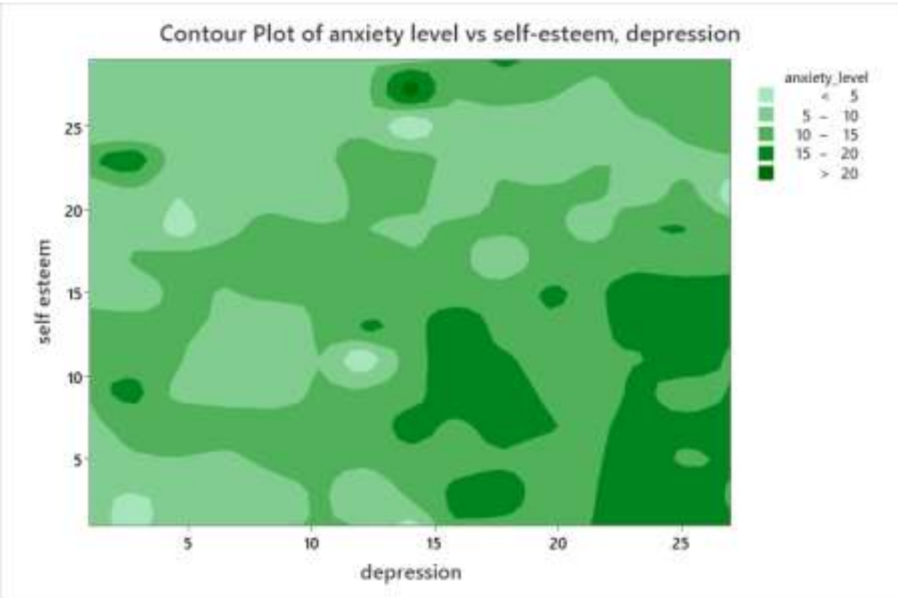

Figure 4: Contour plot of anxiety level (From low to high, with the green spectrum ranging from light to dark) VS. self-esteem (From low to high along the Y-axis) and depression (From low to high along the X-axis)

Figure 4 also shows the contour plot of anxiety levels versus self-esteem and depression, which, like the regression equation, states that in general, with increasing depression, the level of anxiety increases, and with increasing the level of self-esteem, the level of anxiety decreases.

Figure 5 presents a 3D scatter plot, where each point is positioned based on anxiety level, self-esteem level, and level of depression. As can be seen, there appear to be hidden patterns in these plots, which may be due to the presence of clusters within the data.

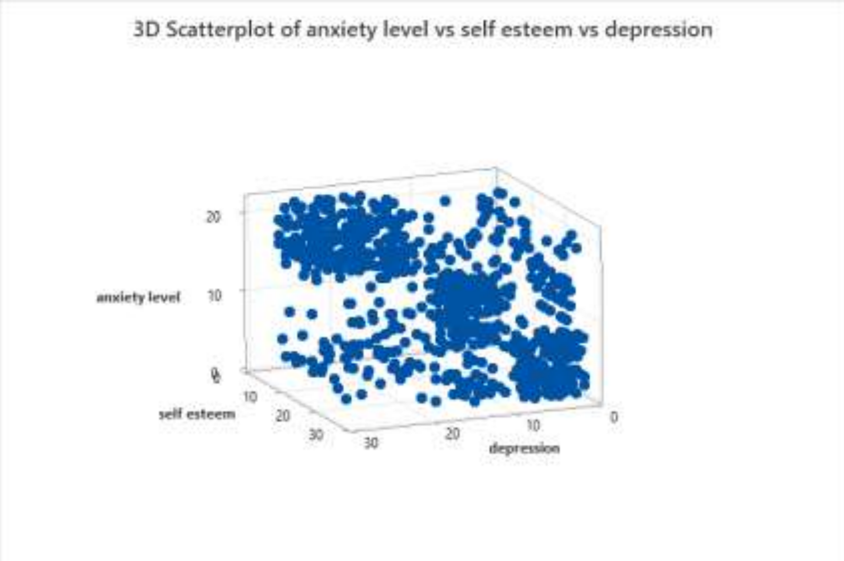

Figure 5: 3D scatter plots, each point is positioned based on anxiety level (Low to high), self-esteem level (Low to high), and level of depression (Low to high)

Based on these observations, K-means clustering was applied to the data. For the clustering process, rows containing missing values were first removed. Different clusterings with varying values of k were tested. Figure 6 shows the elbow plot, indicating that K values of 4 and 5 appear to be suitable for the number of clusters. Figure 7 presents the silhouette scores, which suggest that the highest score occurs at 3 clusters, although this number may not adequately capture all the existing clusters. Therefore, K = 5 is selected, as it has the second-highest silhouette score and performs well according to the elbow plot, being close to the elbow point.

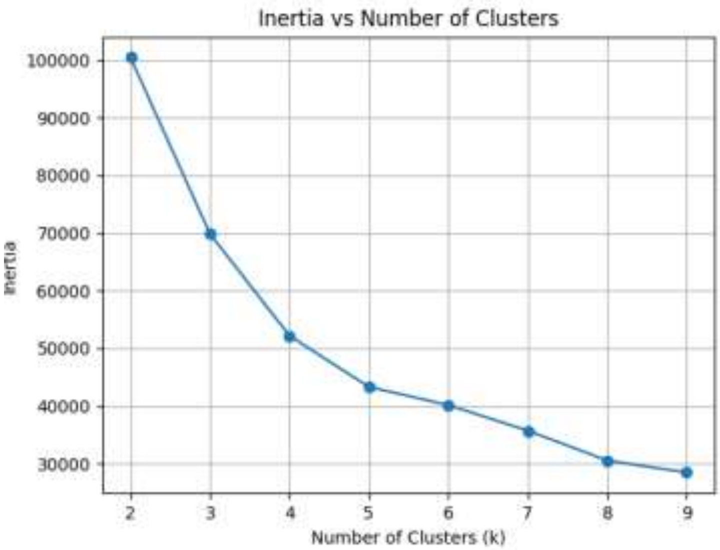

Figure 6: Finding optimal k using elbow method. K = 4 and K = 5 are visually determined as "good" number of K

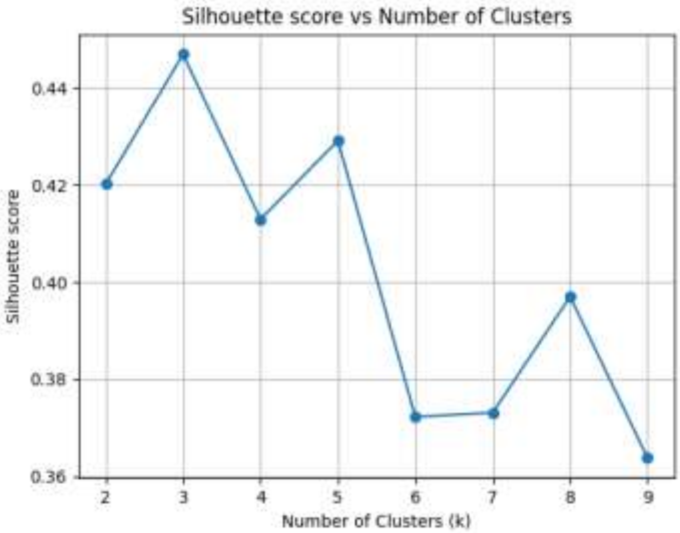

Figure 7: Silhouette scores for different numbers of K

Based on these considerations, the K parameter is set to 5. The number of data points in each cluster is shown in Table 6. Also, the centroids of each cluster are specified in Table 7.

Table 6: Number of datapoints within each cluster

Cluster code	n
1	248
2	260
3	290
4	76
5	133

Table 7: Centroid of each cluster

Cluster	1	2	3	4	5
Self esteem	19.9677	7.823077	25.4931	7.763158	4.729323
Depression	11.75	21.46923	5.048276	19.25	4.639098
Anxiety level	12.16532	17.62308	4.62069	2.776316	8.578947

According to these results, Cluster 1 consists of individuals with relatively high self-esteem, relatively low depression, and relatively low anxiety, which is expected since low depression and high self-esteem are associated with lower anxiety levels.

Cluster 3 comprises individuals with the highest self-esteem, low depression, and very low anxiety. Compared to Cluster 1, these individuals have higher self-esteem and lower depression, resulting in even lower anxiety levels. Cluster 2 includes individuals with low self-esteem, relatively high depression, and relatively high anxiety, essentially representing characteristics opposite to those in Cluster 3.

Cluster 4 includes individuals with low self-esteem, relatively high depression, but low anxiety. This is the rarest cluster, containing the fewest individuals. External factors, such as the use of anti-anxiety medication, should also be examined for this cluster, as they may explain why anxiety is very low despite low self-esteem and high depression. Considering issues such as the use of various medications that may even be prescribed for non-mental illnesses but have mental effects is needed to improve this paper in future works.

Finally, Cluster 5 represents individuals with relatively low self-esteem but low depression. Despite their lower self-esteem, their anxiety remains low (Maybe due to their low depression levels).

Overall, the first three clusters contain the majority of individuals (These three clusters are also recognizable in Figure 5), while Clusters 4 and 5 together have fewer members than any of the first three. Among the first three clusters, there is a clear ordering: Cluster 3 has the lowest depression, highest self-esteem, and minimal anxiety; Cluster 1 has slightly lower self-esteem, somewhat higher depression, and moderately low anxiety; Cluster 2 has the lowest self-esteem, highest depression, and the highest anxiety among these three clusters.

4.2. Weighting the External Factors

First, three categories of factors are identified, each consisting of a number of factors. These categories include environmental, social, and academic factors. Table 8 shows the pairwise comparisons based on expert judgment, while Table 9 presents the optimal weights for these three categories. According to these results, environmental factors are the most important, followed by social factors. The consistency ratio of these comparisons is low, indicating good consistency in the evaluations.

Table 8: Pairwise comparisons of main three categories

Best to others		
Environmental to:	Environmental	EI
	Academic	FI
	Social	WI
Others to worst		
Environmental to:	Academic	FI
Academic to:		EI
Social to:		WI

Table 9: Weights of main three categories

Factor	Fuzzy Weight			Defuzzified Weight (Out of 1):
Environmental	0.415	0.415	0.415	0.415
Academic	0.24	0.24	0.24	0.24
Social	0.345	0.345	0.345	0.345
Consistency Ratio:		0.061666667		

Next, the weighting of factors within each category is performed, starting with the environmental factors. In the dataset, the considered environmental factors include noise level, living conditions, safety, and basic needs. Figures 8, 9, 10, and 11 illustrate the effects of these environmental factors.

As shown in Figure 8, higher noise levels are associated with increased anxiety. Interestingly, Figure 9 shows that improving living conditions initially reduces anxiety, but at the highest level of living conditions, anxiety unexpectedly rises. According to Figure 10, higher safety levels correspond to lower anxiety, and Figure 11 shows that improving basic needs also leads to reduced anxiety.

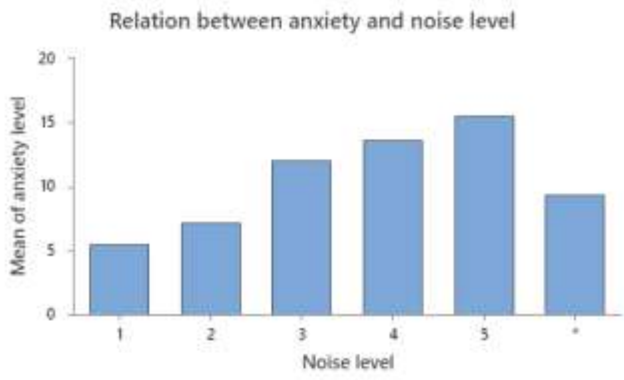

Figure 8: Relation between anxiety and noise level (1 to 5: Slight to Extreme, *: missing value)

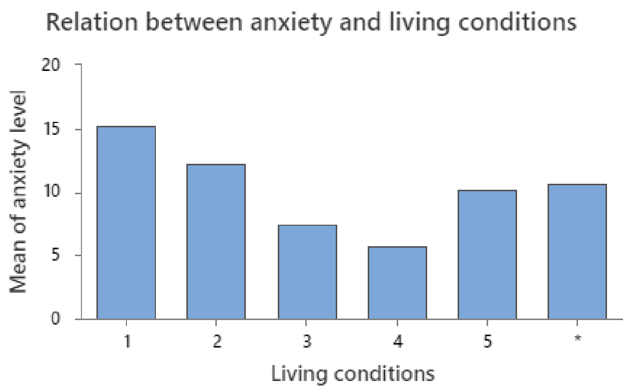

Figure 9: Relation between anxiety and living conditions (1 to 5: Poor to Well, *: missing value)

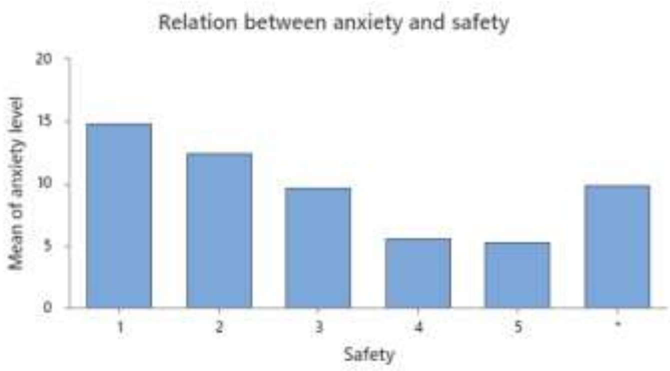

Figure 10: Relation between anxiety and safety (1 to 5: Low to High, *: missing value)

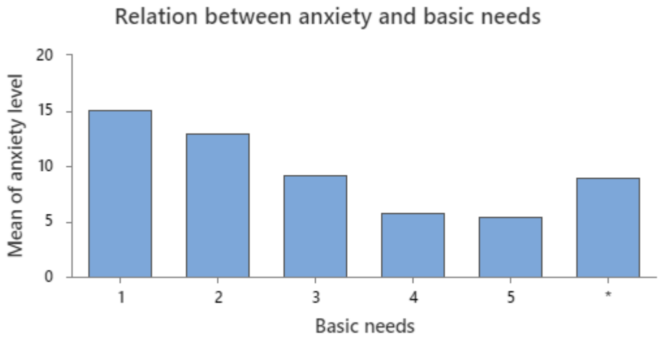

Figure 11: Relation between anxiety and basic needs (1 to 5: Poor to Well, *: missing value)

The expert comparisons for the environmental sub-criteria are presented in Table 10, and the resulting weights are shown in Table 11. The consistency ratio of these comparisons is low, indicating good consistency. Among these factors, basic needs are identified as the most important factor in this category.

Table 10: Pairwise comparisons of environmental factors

Best to others		
Basic Needs to:	Noise Level	AI
	Living Condition	WI
	Safety	WI
	Basic Needs	EI
Others to worst		
Noise Level to:	Noise Level	EI
Living Condition to:		VI
Safety to:		VI
Basic Needs to:		AI

Table 11: Weights of environmental factors

Factors	Fuzzy Weights			Defuzzified Weigths (Out of 1):
Noise Level	0.097561	0.097561	0.097561	0.097560976
Living Condition	0.292683	0.292683	0.292683	0.292682927
Safety	0.292683	0.292683	0.292683	0.292682927
Basic Needs	0.317073	0.317073	0.317073	0.317073171
Consistency Ratio:		0.015168062		

The focus now shifts to academic factors. Figures 12, 13, and 14 illustrate the relationship between anxiety levels and academic factors. These factors include study load, the student–teacher relationship, and future career concerns. As shown in Figure 12, higher study load is associated with increased anxiety. Figure 13 indicates that better student–teacher relationships correspond to lower anxiety levels. Additionally, Figure 14 shows that individuals with greater concerns about their future careers experience higher levels of anxiety.

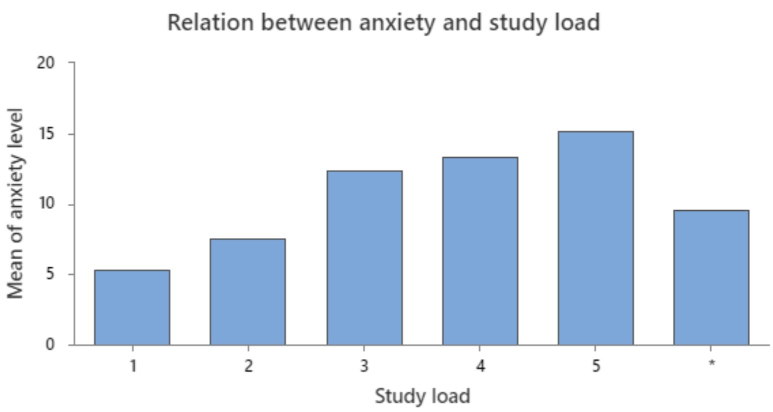

Figure 1۲: Relation between anxiety and study load (1 to 5: Low to High, *: missing value)

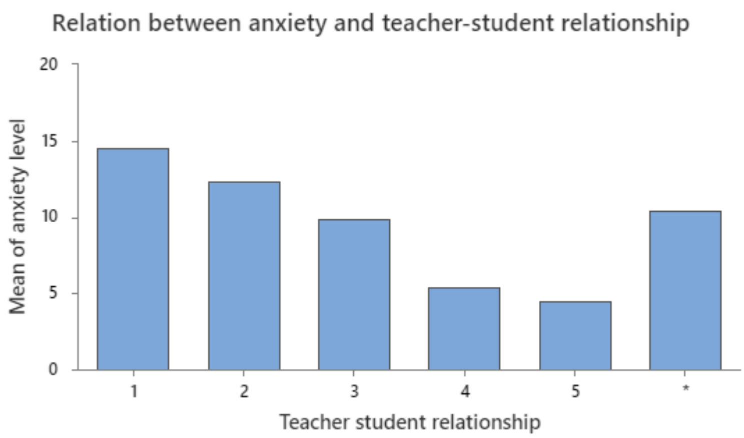

Figure 13: Relation between anxiety and teacher-student relationship (1 to 5: Very Bad to Very Good, *: missing value)

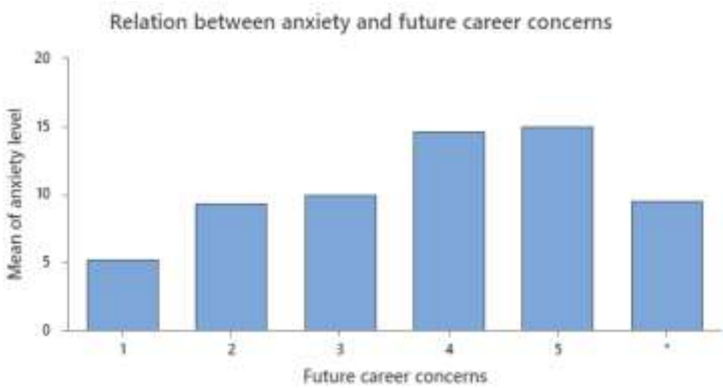

Figure 14: Relation between anxiety and future career concerns (1 to 5: Slight to Extreme, *: null)

Based on these observations, pairwise comparisons of academic factors were conducted by experts, as shown in Table 12. The final weights are presented in Table 13, with the student–teacher relationship identified as the most important factor. The consistency ratio of these comparisons indicates good consistency.

Table ۱2: Pairwise comparisons of academic factors

<table>
<tr><th colspan="3">Best to others</th></tr>
<tr><td rowspan="3">Teacher and Student Relationship to:</td><td>Study Load</td><td>WI</td></tr>
<tr><td>Teacher and Student Relationship</td><td>EI</td></tr>
<tr><td>Future Career Concerns</td><td>VI</td></tr>
<tr><th colspan="3">Others to worst</th></tr>
<tr><td>Study Load to:</td><td rowspan="3">Future Career Concerns</td><td>FI</td></tr>
<tr><td>Teacher and Student Relationship to:</td><td>VI</td></tr>
<tr><td>Future Career Concerns to:</td><td>EI</td></tr>
</table>

Table 13: Weights of academic factors

<table>
<tr><th>Factors:</th><th colspan="3">Fuzzy Weights:</th><th>Defuzzified Weights (Out of 1):</th></tr>
<tr><td>Study Load</td><td>0.398104</td><td>0.398104</td><td>0.398104</td><td>0.398104265</td></tr>
<tr><td>Teacher and Student Relationship</td><td>0.43128</td><td>0.43128</td><td>0.43128</td><td>0.431279621</td></tr>
<tr><td>Future Career Concerns</td><td>0.170616</td><td>0.170616</td><td>0.170616</td><td>0.170616114</td></tr>
<tr><th colspan="2">Consistency Ratio:</th><th colspan="3">0.055292259</th></tr>
</table>

Finally, in this section, the external factors within the social category are examined. This category includes social support, peer pressure, extracurricular activities, and bullying. As shown in Figure 15, higher social support is associated with lower anxiety levels. Figure 16 indicates that increased peer pressure leads to higher anxiety. Figure 17 shows that greater participation in extracurricular activities is linked to increased anxiety, and finally, Figure 18 illustrates that higher levels of bullying correspond to increased anxiety.

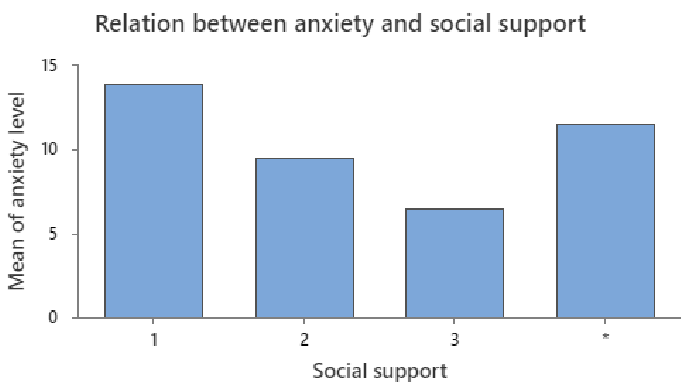

Figure 15: Relation between anxiety and social support (1 to 3: Low to High, *: missing value)

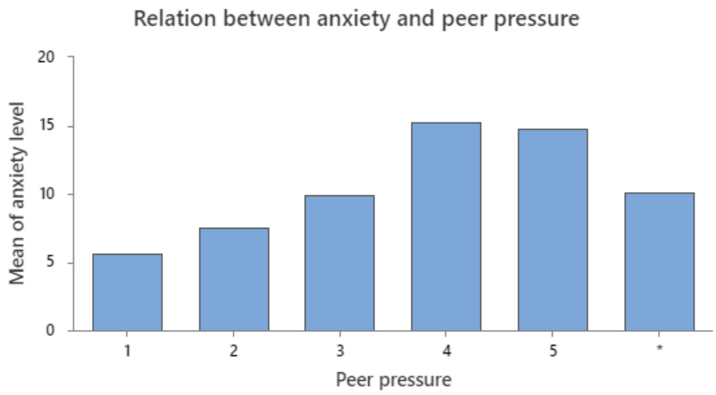

Figure 16: Relation between anxiety and peer pressure (1 to 5: Slight to Extreme, *: missing value)

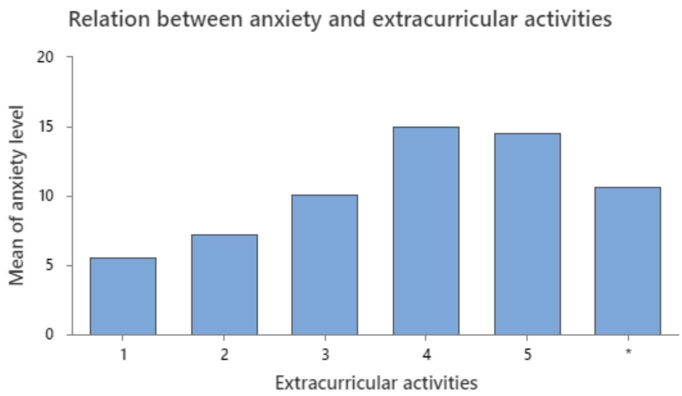

Figure 17: Relation between anxiety and extracurricular activities (1 to 5: Slight to Extreme, *: missing value)

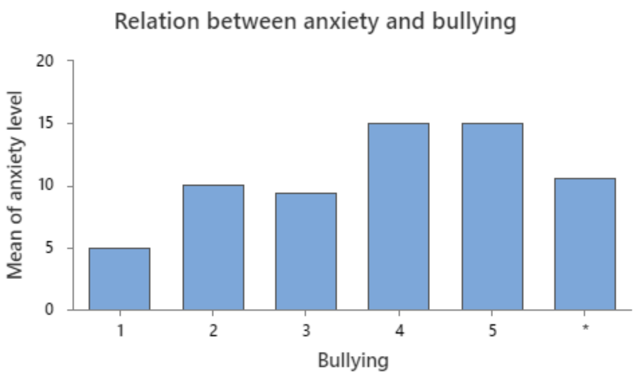

Figure 18: Relation between anxiety and bullying (1 to 5: Slight to Extreme, *: missing value)

Based on these observations, the social factors were weighted. Table 14 presents the pairwise comparisons conducted by experts, while Table 15 shows the weights assigned by the fuzzy best–worst method, with social support receiving the highest weight.

Table 14: Pairwise comparisons of social factors

Best to others		
Social Support to:	Social Support	EI
	Peer Pressure	WI
	Extracurricular Activities	AI
	Bullying	FI
Others to worst		
Social Support to:	Extracurricular Activities	AI
Peer Pressure to:		VI
Extracurricular Activities to:		EI
Bullying to:		FI

Table 15: Weights of social factors

Factors:	Fuzzy Weights:			Deffuzified Weights (Out of 1):
Social Support	0.359116	0.359116	0.359116	0.359116022
Peer Pressure	0.331492	0.331492	0.331492	0.331491713
Extracurricular Activities	0.110497	0.110497	0.110497	0.110497238
Bullying	0.198895	0.198895	0.198895	0.198895028
Consistency Ratio:		0.017179297		

Based on the weights of the main categories and the factors within each, the final weight of each factor is calculated, as presented in Table 16 and Figure 19. Moreover, Table 17 ranks the factors based on their global weights.

Table 16: Global weights of the factors

Category	Category weight	Factor	Factor weight (In category)	Factor weight (Global)
Environmental	0.415	Noise Level	0.097560976	0.040487805
		Living Condition	0.292682927	0.121463415
		Safety	0.292682927	0.121463415
		Basic Needs	0.317073171	0.131585366
Academic	0.24	Study Load	0.398104265	0.095545024
		Teacher and Student Relationship	0.431279621	0.103507109
		Future Career Concerns	0.170616114	0.040947867
Social	0.345	Social Support	0.359116022	0.123895028
		Peer Pressure	0.331491713	0.114364641
		Extracurricular Activities	0.110497238	0.038121547
		Bullying	0.198895028	0.068618785

Table 17: Factors ranked by their global weights

Factor	Factor weight (Global)	Rank
Basic Needs	0.131585366	1
Social Support	0.123895028	2
Living Condition	0.121463415	3

Factor	Factor weight (Global)	Rank
Safety	0.121463415	3
Peer Pressure	0.114364641	5
Teacher and Student Relationship	0.103507109	6
Study Load	0.095545024	7
Bullying	0.068618785	8
Future Career Concerns	0.040947867	9
Noise Level	0.040487805	10
Extracurricular Activities	0.038121547	11

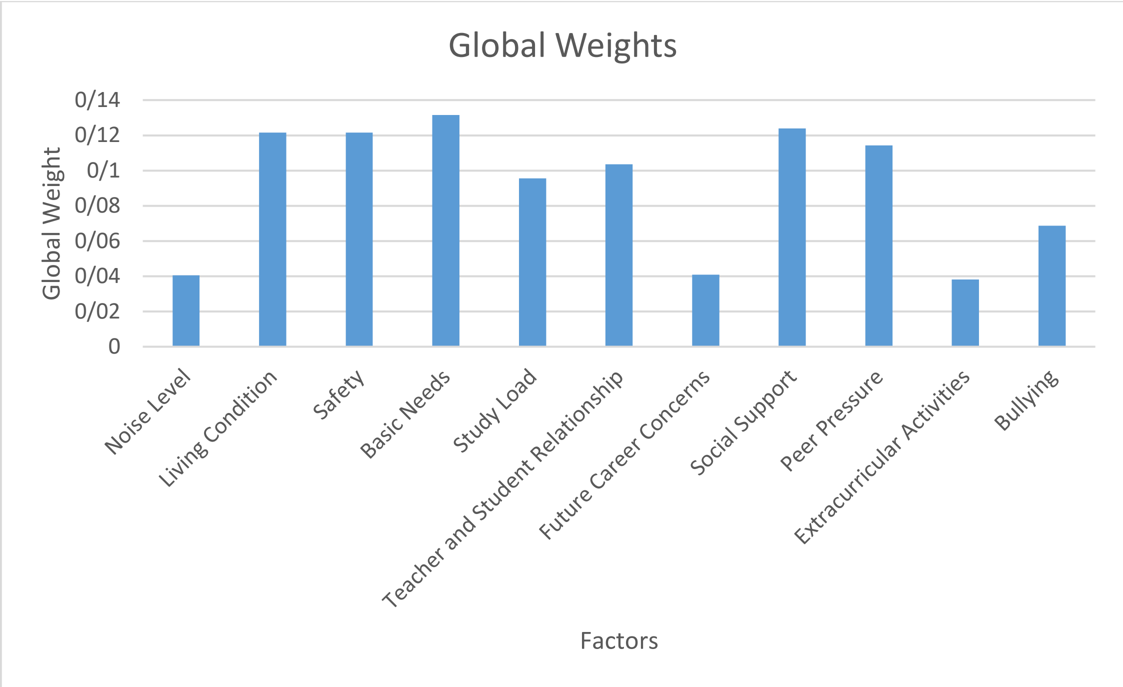

Table 19: Global weights of the factors

Based on these results, basic needs are identified as the most important factor overall, while extracurricular activities are considered the least important factor.

5. Conclusion

In this study, data available on Kaggle was used to evaluate and examine internal and external factors affecting anxiety levels. Initially, the effects of anxiety on academic performance, sleep quality, and headache experience were investigated. According to the review, there is generally a positive correlation between anxiety levels and the experience of headaches, meaning that as anxiety increases, the occurrence of headaches also rises. Additionally, a negative relationship exists between anxiety levels and both academic performance and sleep quality, indicating that increases in anxiety are associated with decreases in these two variables. Based on these observations and considering the impact of anxiety on individuals' daily lives, the factors influencing stress were analyzed.

First, internal factors - self-esteem and depression - were examined. A regression analysis indicated that self-esteem has a positive correlation with anxiety level (which means higher self-esteem leads to lower anxiety level), while depression has a negative correlation with anxiety levels (higher depression leads to higher anxiety). However, since the regression model could not adequately capture the variability in the data, a clustering analysis was conducted using the K-means method. Based on the elbow method and silhouette scores, five clusters were selected for the K parameter. These clusters were then analyzed in terms of their locations and the number of data points contained within each cluster. The unsupervised K-means technique revealed hidden patterns among self-esteem, depression, and anxiety.

Next, the effects of external factors were examined. Given that fuzzy theory allows for handling ambiguity, the fuzzy best–worst weighting method was applied. First, three categories of factors - environmental, social, and academic - were weighted, with environmental factors found to be the most important. The weight of the environmental factor was assigned as 0.415, the academic factor as 0.24, and the social factor as 0.345. These weights are numerical values between 0 and 1, and the sum of all weights must equal 1. Subsequently, the factors within each category were weighted, and the global weight of each factor was calculated by multiplying the category weight by the weight of each factor within that category. Finally, basic needs received the highest global weight among all factors considered. The second rank is assigned to social support, followed by a shared third rank between safety and living conditions. The lowest weight is assigned to extracurricular activities.

The work examines a real dataset and employs suitable statistical methods. The use of clustering to identify groups of students, along with the weighting of external factors related to anxiety, constitutes the main contributions of this study. One of the limitations of this study is related to the dataset; despite positive aspects such as good usability, it does not cover certain factors, such as medication usage. Also, the following are proposed as future directions:

- Utilize different datasets to perform clustering analysis.
- Apply alternative machine learning approaches, such as neural networks, decision trees, and random forests (supervised learning approaches).
- Explore other weighting methods, including objective approaches.
- Future studies should investigate additional factors that may influence the datasets, such as the potential effects of medications.
- Conduct further analysis of individuals within different clusters to examine other potential contributing factors in more detail.

References

Ahmed, A., Aziz, S., Toro, C. T., Alzubaidi, M., Irshaidat, S., Serhan, H. A., Abd-alrazaq, A.A., & Househ, M. (2022). Machine learning models to detect anxiety and depression through social media: A scoping review. *Computer Methods and Programs in Biomedicine Update, 2*, 100066. https://doi.org/10.1016/j.cmpbup.2022.100066

Amoozad Mahdiraji, H., Arzaghi, S., Stauskis, G., & Zavadskas, E. K. (2018). A hybrid fuzzy BWM-COPRAS method for analyzing key factors of sustainable architecture. *Sustainability, 10*(5), 1626. https://doi.org/10.3390/su10051626

Ancillon, L., Elgendi, M., & Menon, C. (2022). Machine learning for anxiety detection using biosignals: a review. *Diagnostics, 12*(8), 1794. https://doi.org/10.3390/diagnostics12081794

Arif, M., Basri, A., Melibari, G., Sindi, T., Alghamdi, N., Altalhi, N., & Arif, M. (2020). Classification of anxiety disorders using machine learning methods: a literature review. *Insights Biomed Res, 4*(1), 95-110. DOI:https://doi.org/10.36959/584/455

Garg, S., & Bhardwaj, R. (2024). Exploring the influence of factors causing stress among doctoral students by combining fuzzy DEMATEL-ANP with a triangular approach. *Scientometrics, 129*(8), 4695-4719. https://doi.org/10.1007/s11192-024-05108-x

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-means clustering algorithm. *Journal of the Royal Statistical Society. series c (applied statistics), 28*(1), 100-108. https://doi.org/10.2307/2346830

Hashemkhani Zolfani, S. (2024). Objective weighting in mcdm: A comparative study of critic, itara, merec, and beyond. *Decision Making and Artificial Intelligence Trends, 1*(1), 105-120. https://doi.org/10.22034/dmait.2024.488347.1007

Jarmoszewicz, K., Nowicka-Sauer, K., Zemła, A., & Beta, S. (2020). Factors associated with high preoperative anxiety: results from cluster analysis. *World Journal of Surgery, 44*(7), 2162-2169. https://doi.org/10.1007/s00268-020-05453-x

Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of Cluster in K-means Clustering. International Journal of Advance Research in Computer Science and Management Studies*, 1*(6), 90-95.

Kumar, P., Garg, S., & Garg, A. (2020). Assessment of anxiety, depression and stress using machine learning models. *Procedia computer science, 171*, 1989-1998. https://doi.org/10.1016/j.procs.2020.04.213

Liu, F., Yang, D., Liu, Y., Zhang, Q., Chen, S., Li, W., Ren, J., Tian, X., & Wang, X. (2022). Use of latent profile analysis and K-means clustering to identify student anxiety profiles. *BMC Psychiatry, 22*(1), 12. https://doi.org/10.1186/s12888-021-03648-7

Maral, M., & Özdemir, A. (2025). A systematic review on multi-criteria decision-making methods in educational research. *British Educational Research Journal.* https://doi.org/10.1002/berj.70002

Mayer, E. A., Craske, M., & Naliboff, B. D. (2001). Depression, anxiety, and the gastrointestinal system. *Journal of Clinical Psychiatry, 62*, 28-37.

Middeldorp, C. M., Birley, A. J., Cath, D. C., Gillespie, N. A., Willemsen, G., Statham, D. J., ... & Boomsma, D. I. (2005). Familial clustering of major depression and anxiety disorders in Australian and Dutch twins and siblings. *Twin Research and Human Genetics, 8*(6), 609-615. https://doi.org/10.1375/twin.8.6.609

Moslem, S., Gul, M., Farooq, D., Celik, E., Ghorbanzadeh, O., & Blaschke, T. (2020). An integrated approach of best-worst method (BWM) and triangular fuzzy sets for evaluating driver behavior factors related to road safety. *Mathematics, 8*(3), 414. https://doi.org/10.3390/math8030414

Olthuis, J. V., Connell, E. M., Watt, M. C., & Stewart, S. H. (2024). What does 'high anxiety sensitivity'look like? Using cluster analysis to identify distinct profiles of high anxiety sensitive treatment-seekers. *Cognitive Therapy and Research, 48*(2), 337-349. https://doi.org/10.1007/s10608-024-10476-y

Ovi, M. S. I. (2025, July 24). Student stress monitoring datasets. *Kaggle.* https://www.kaggle.com/datasets/mdsultanulislamovi/student-stress-monitoring-datasets

Ovi, M. S. I., Hossain, J., Rahi, M. R. A., & Akter, F. (2025). Protecting Student Mental Health with a Context-Aware Machine Learning Framework for Stress Monitoring. *ArXiv preprint arXiv:2508.01105.* https://doi.org/10.48550/arXiv.2508.01105

Owens, M., Stevenson, J., Hadwin, J. A., & Norgate, R. (2012). Anxiety and depression in academic performance: An exploration of the mediating factors of worry and working memory. *School Psychology International, 33*(4), 433-449. https://doi.org/10.1177/0143034311427433

Piao, X., Xie, J., & Managi, S. (2024). Continuous worsening of population emotional stress globally: universality and variations. *BMC Public Health, 24*(1), 3576. DOI: https://doi.org/10.1186/s12889-024-20961-4

Pintelas, E. G., Kotsilieris, T., Livieris, I. E., & Pintelas, P. (2018, June). A review of machine learning prediction methods for anxiety disorders. In *Proceedings of the 8th international conference on software development and technologies for enhancing accessibility and fighting info-exclusion* (8-15). https://doi.org/10.1145/3218585.3218587

Prabakaran, R. (2024, October). Ranking Psychological Risk Factors Among College Students Using Interval-Valued Picture Fuzzy Set with Various Operators and MCDM Techniques. In *2024 First international conference on innovations in communications, electrical and computer engineering (ICICEC)* (pp. 1-9). IEEE. https://doi.org/10.1109/ICICEC62498.2024.10808401

Prisnie, J. C., Sajobi, T. T., Wang, M., Patten, S. B., Fiest, K. M., Bulloch, A. G., ... & Jette, N. (2018). Effects of depression and anxiety on quality of life in five common neurological disorders. *General Hospital Psychiatry, 52*, 58-63. DOI: https://doi.org/10.1016/j.genhosppsych.2018.03.009

Priya, A., Garg, S., & Tigga, N. P. (2020). Predicting anxiety, depression and stress in modern life using machine learning algorithms. *Procedia Computer Science, 167*, 1258-1267. https://doi.org/10.1016/j.procs.2020.03.442

Rajabi, F., Molaeifar, H., Jahangiri, M., Taheri, S., Banaee, S., & Farhadi, P. (2020). Occupational stressors among firefighters: application of multi-criteria decision making (MCDM) Techniques. *Heliyon, 6*(4). https://doi.org/10.1016/j.heliyon.2020.e03820

Rezaei, J. (2015). Best-worst multi-criteria decision-making method. *Omega, 53*, 49-57. https://doi.org/10.1016/j.omega.2014.11.009

Sau, A., & Bhakta, I. (2017). Predicting anxiety and depression in elderly patients using machine learning technology. *Healthcare Technology Letters, 4*(6), 238-243. https://doi.org/10.1049/htl.2016.0096

Sharma, H., Sohani, N., & Yadav, A. (2022). Comparative analysis of ranking the lean supply chain enablers: An AHP, BWM and fuzzy SWARA based approach. *International journal of Quality & Reliability Management, 39*(9), 2252-2271. https://doi.org/10.1108/IJQRM-04-2021-0114

Taylor, V. A., Roy, A., & Brewer, J. A. (2023). Cluster-based psychological phenotyping and differences in anxiety treatment outcomes. *Scientific Reports, 13*(1), 3055. https://doi.org/10.1038/s41598-023-28660-7

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control, 8*(3), 338-353. https://doi.org/10.1016/S0019-9958(65)90241-X