JITEEHA: Journal of Information Technology Application in Education, Economy, Health and Agriculture
Vol. 03, No. 2, June 2026, pp. 49-61, e-ISSN: 3090-3939
https://journal.iteeacademy.org/index.php/jiteeha/

The Student Mental Health Pattern Using Clustering and Classification Approaches

Audrey Suitela[1], Silviana[2], Fahmi Bahaluan[3], Maurecia Tima[4], Indah Dewi Nurhayati[5], Zaenuddin[6]

[1,3,4] Department of Informatics Engineering, University of Widyagama Malang, Jl. Borobudur No. 35 Malang, Indonesia
[2] Department of Industrial Engineering, University of Widyagama Malang, Jl. Borobudur No. 35 Malang, Indonesia
[5,6] Faculty of Economy and Business, University of Widyagama Malang, Jl. Borobudur No. 35 Malang, Indonesia

Article Info

Article history:

Received February 20, 2026
Revised March 24, 2026
Received May 15, 2026

Keywords:

Students Mental Health,
Clustering,
Classification,
Machine Learning,
Social and Demographic Factors

ABSTRACT

Students mental health is a key factor in their academic and social development. However, the patterns and factors that influence mental health in college students are still not fully understood. This study utilizes machine learning-based clustering and classification techniques to identify hidden patterns in college students' mental health data, focusing on social and demographic factors. Using the K-Means algorithm for clustering and Random Forest for classification, we group college students based on their mental health conditions and analyze the associations between variables such as age, marital status, anxiety, and medical history. The process begins with data exploration, followed by data cleaning and feature transformation to ensure optimal input quality. In the clustering stage, we find three main groups of college students with different mental health patterns, which are then used as the basis for a classification model. A Random Forest model is built to predict potential mental disorders, such as depression and anxiety, by identifying the features that have the most influence on the prediction results. The model evaluation shows significant performance with adequate accuracy, where the importance of social factors such as marital status and history of visits to medical professionals is clearly revealed. The results of this study not only offer important insights into students' mental health patterns, but also provide recommendations for university policies in creating an environment that supports students' mental well-being. This combined approach of clustering and classification opens up new opportunities in the application of machine learning for more precise and data-driven mental health analysis.

Corresponding Author:

Audrey Suitela
Department of Informatics Engineering
Faculty of Engineering, University of Widyagama Malang
Jl. Borobudur No. 35 Malang, East Java
Email: audrey@gmail.com

1. INTRODUCTION

Currently, the problem of student mental health is no longer a secret. With the main burden of academic problems, the educational and social demands experienced by students further worsen Student Mental Health. According to research presented by the American Psychological Association (APA), high academic load is one of the main factors causing stress in students. The pressure to achieve certain grades that are considered good for them, and the pressure of their expectations in the future, as well as intense competition among fellow students often lead to significant feelings of overwhelm. The research revealed that 61% of students experienced high anxiety related to the academic load they faced."[1]

In a study conducted by Harvard Medical School, it was shown that the poor mental health of current students can actually have a negative impact and even interfere with student achievement targets and cause stress, anxiety, even in these students, so that it can hinder their ability to concentrate, complete academic assignments on time, and reach their full academic potential. These findings show that mental health is not only a personal issue, but also has a direct impact on students' academic performance[1]. In a survey conducted on college students in the United States during the COVID-19 Pandemic. Showing that as many as 80.57% of respondents reported scores on the Patient Health Questionnaire-9 (PHQ-9 which indicated that students experienced severe depression in addition to the increasingly severe COVID-19 situation at that time added to the level of depression and mental stress of students. So that it takes efforts to overcome this problem that occurs, with regulations that are suitable for appropriate circumstances. [2] [3]

In addition, unhealthy lifestyles and physical activity also affect students' mental health. Like a study conducted on Norwegian students aged 18-35 that measured frequency, intensity, sleep duration, and physical activity. It found that women with low levels of physical activity were almost three times more likely to score high on the HSCL-25, and to self-report depression, compared to women who exercised almost every day. In addition, the study also explained that people with depression also have an influence on suicide attempts and self-harm[4]. Although there have been many studies on mental health, there are still some factors that may not have been explored and some correlations between problems and solution methods that are not suitable for the existing problems, thus raising public questions or not producing appropriate information. To address these challenges, this research aims to identify the mental health patterns of university students by utilizing machine learning algorithms, namely K-Means Clustering, Random Forest Classifier, and Principal Component Analysis (PCA). Each algorithm was selected based on its suitable approach in handling multivariable and complex data, which is a common characteristic of mental health data.

By choosing the K-Means clustering approach, it is possible to identify the grouping of students into several groups based on the similarity of their data patterns, so as to obtain the target achievement value for identifying the mental health level of students. PCA is applied to reduce the dimensionality of complex data, which not only improves the efficiency of data processing but also facilitates the visual interpretation of clustering results. In addition, Random Forest was used to classify and predict students' mental state based on identified significant socio-demographic features. Random Forest as a decision tree-based ensemble algorithm has the advantage of handling non-linear data and is able to provide a clear interpretation of the importance of each feature in predicting mental state. Not unlike other countries, Indonesian students also have the potential to experience health problems, which have been reported and analyzed at the provincial level in Indonesia. Where out of 34 provinces, 7 provinces were identified as being in the high category when divided into 3 groups of cluster levels. [5]

In this case study discussion, data was obtained from the Student Mental Health dataset downloaded from the official Kaggle website. The dataset contains social and demographic information of students, with several questions referring to Mental Health questions and corresponding information by respondents. Furthermore, the data will be further processed by handling missing values, categorical data transformation, and data normalization. After the data is ready, it will enter several data mining processes, including using the K-Means algorithm to group students based on the response results and PCA is used to visualize the cluster results in order to obtain further information and can be developed for further research or answer the problems being experienced. Random Forest algorithm was selected to correlate the relationship between sociodemographic variables such as age, gender, major, and grade with students' mental health conditions including depression, anxiety, and panic attacks. The model was evaluated using accuracy, precision area under the curve (AUC) metrics of the ROC curve to ensure optimal performance in classification and prediction.

The author intends for this study to achieve the main objectives. First, to identify the pattern of students' mental health based on social and demographic factors through a clustering approach using K-Means. Second, analyze the relationship between socio-demographic variables and mental health conditions using Random Forest to identify the most influential factors. Third, simplify the dimensions of complex data using PCA to improve the efficiency and interpretation of analysis results. Fourth, provide data-driven recommendations for educational institutions to support more proactive and adaptive mental health policies. The results of this study are expected to address students' mental health problems and patterns and support better decision-making in the field of education and mental health. By using a data-driven approach, this research provides a strong scientific basis for educational institutions to design targeted interventions that can improve students' overall mental well-being. [2], [6], [7], [8].

2. METHOD

2.1 Research Design

This research is expected to yield identifiable and analyzable information regarding students' mental health patterns, particularly in detecting indications of depression based on survey data containing social, demographic, and mental health information. Machine learning algorithms based on the Random Forest Classifier, K-Means Clustering, and Principal Component Analysis (PCA) were used in this study. The steps in this research design include data collection, preprocessing, model implementation, and evaluation and validation of the results.
A general outline of the research design can be seen in Figure 2.1 below:

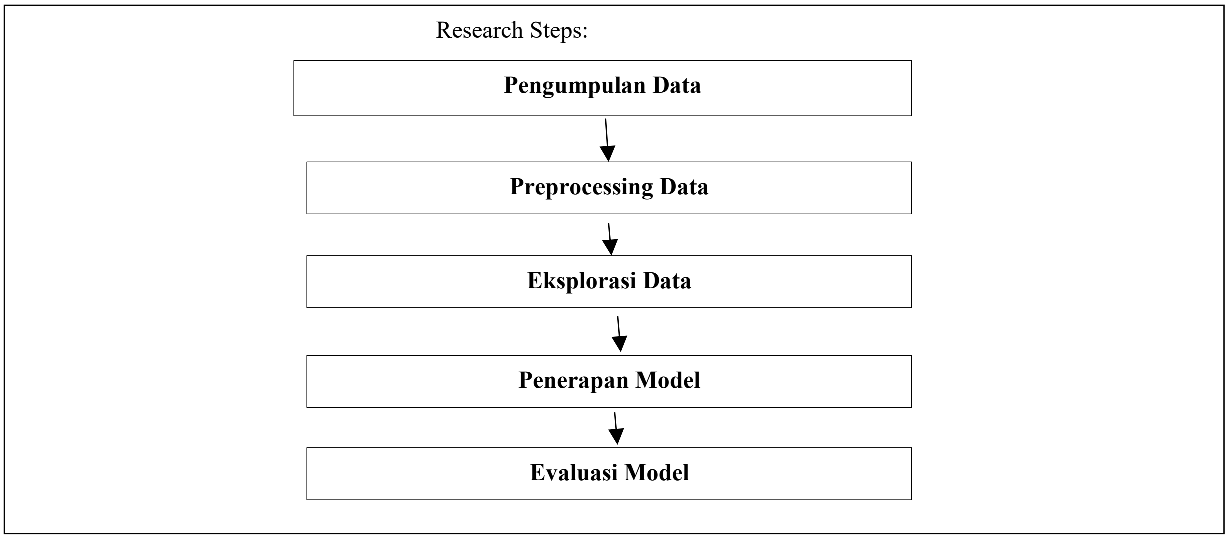

Figure 1. Research Method

1. Information Collection
   This information was obtained from the student mental health section of the official Kaggle website.
2. Data Preprocessing
   a. Handle missing values
   b. Applying the One-Hot Encoding technique to convert categorical data into numerical data
   c. Applying StandardScaler to normalize the data.
3. Data Eksplorasi
   a. Determine how the variables are correlated.
   b. Using histograms and heat maps to visualize the data to understand its distribution.
4. Model Application
   a. Using K-Means Clustering to group the data.
   b. Using PCA to reduce data dimension for pattern analysis and visualization.
   c. Developing a prediction model for depression classification using Random Forest Classifier..
5. Model Assessment
   a. Analyzed model performance using metrics such as F1-score, recall, accuracy, and precision.
   b. Using ROC curves, feature importance analysis, and confusion matrix to validate the findings.

2.2 Information gathering

Student Mental Health.csv, the dataset used in this study, includes 1017 records of students with various social, psychological, and demographic characteristics. Among the important characteristics in this dataset are:

Age: Age of the respondent

Gender: Gender of the respondent Material Status: Marital status: Marital status of the respondent.

- Age : Age of Respondent
- Gender : Gender of Respondents
- Marital Status: Marital status of respondents.
- Academic Achievement: Academic achievement of respondents.
- Are you depressed? Depression status depends on the answer given by the respondent.
- Do you experience anxiety? respondent's anxiety status.
- Respondent status on panic attacks: Do you experience them?

The Kaggle platform, one of the largest open data repositories worldwide, is where the dataset was obtained.

2.3 Preparing Data

Ensuring that the data is clean, free of empty values, and has a uniform data type is the goal of data preparation. The following are the steps involved in data preprocessing:

1. Verifying the Data Type

To ensure compliance with the required format, the data type of each column in the data set is verified. If mixed or incompatible data is found, a data type conversion is performed..

2. Ignoring to Handle Value

The following methods are used to handle missing values in the data set:

- The average value (mean) of each column is entered into the numeric column.
- The mode value of the column is used to fill the categorical column.

```
# Mengisi missing value untuk data numerik
numerical_columns = data.select_dtypes(include=['number']).columns
data[numerical_columns] =
    data[numerical_columns].fillna(data[numerical_columns].mean())

# Mengisi missing value untuk data kategorikal
categorical_columns = data.select_dtypes(include=['object']).columns
for col in categorical_columns:
    data[col].fillna(data[col].mode()[0], inplace=True)
```

Categorical Data Coding is to guarantee interoperability with machine learning algorithms, the one-hot coding method converts categorical inputs into numerical data.

```
# One-hot encoding untuk data kategorikal
data = pd.get_dummies(data, drop_first=True)
```

2.4 Suggested Model/Technique

Three main techniques were combined in this study: PCA, K-Means Clustering, and Random Forest Classifier.

Random Forest Classifier

Based on socioeconomic and demographic characteristics, student sadness was categorized using Random Forest. To improve prediction accuracy, this technique builds multiple decision trees and combines them. The following is the main formula of Random Forest:

$$H(x) = \frac{1}{N}\sum_{i=1}^{N} h_i(x)$$

Where:

H(x) is the final prediction result

N is the number of decision trees
h_i (x) is the prediction of the i-th tree

K-Means Clustering
K-Means is used to cluster students based on their mental health characteristics. This algorithm works by minimizing the distance between data points and cluster centers.

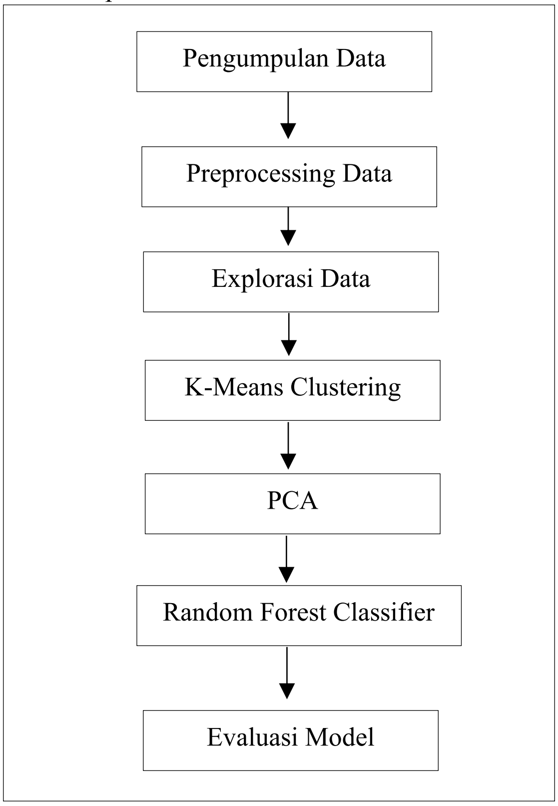

Figure 2. Clustering

The objective function of K-Means is:

$$J = \sum_{i=1}^{k} \sum_{j=1}^{n} ||x_j^{(i)} - c_i||^2$$

Where :

- the objective function is denoted by J.
- The number of clusters is k.
- The jth data in the i-th cluster is x_j^((i)).
- The center of the i-th cluster is c_i.

2.5 Principal Component Analysis
PCA reduces the dimensionality of the data and describes the clustering results in two dimensions. For PCA to work, the data is transformed into a new orthogonal feature space.
2.3 Testing Procedures and Experiments
Flowchart of the research process in Figure 2.2 below:
Eksperimen dilakukan dengan membagi data menjadi data training dan data testing. Data training digunakan untuk melatih model, sedangkan data testing digunakan untuk evaluate the performance of the model. The testing techniques performed are as follows:

- Using confusion matrix to evaluate Random Forest's performance in classification.
- Using accuracy, precision, recall, and F1-score metrics.
- Using silhouette score to evaluate the quality of K-Means clustering results.

2.6 Evaluation and Validation of Results

Evaluation of the results is done by comparing the model's predicted results with the actual data. In addition, ROC curves were used to evaluate the model's ability to distinguish between positive and negative classes. evaluating the model's ability to

```
#Prediksi probabilitas menggunakan data uji
y_prob = model.predict_proba(X_test)[:, 1]

#Hitung ROC curve
fpr, tpr, thresholds = roc_curve(y_test, y_prob)
roc_auc = auc(fpr, tpr)

#Visualisasi ROC Curve
plt.figure(figsize=(8, 6))
plt.plot(fpr, tpr, label=f'ROC curve (area = {roc_auc:.2f})', color='blue')
plt.plot([0, 1], [0, 1], 'k--', color='red')  # Garis diagonal sebagai baseline
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('Receiver Operating Characteristic (ROC) Curve')
plt.legend(loc="lower right")
plt.grid(True)
plt.show()
```

By comparing the experimental findings with previous research and analyzing the findings in relation to student mental health applications, validation was completed

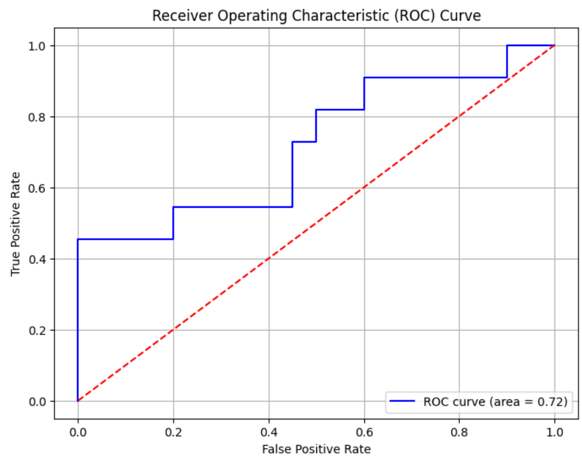

Figure 3 Receiver Operating Characteristic (ROC) Curve

3. RESULTS AND DISCUSSION

3.1 Data Description

The dataset used in this study was collected from 101 university students and contains information about their mental health problems as well as other demographics. Eleven columns make up the data, and each column describes the following features.:

a. Time stamp: The time when the questionnaire was completed.
b. Select gender: The gender of the respondent.
c. Age: Age of the respondent.
d. What study program are you currently enrolled in? The study program currently being followed by the respondent.
e. The academic year you are currently enrolled: the academic year currently enrolled by the respondent.

f. What is your GPA? Grade Point Average (GPA) of the respondent.
g. Marital status: Marital status of the respondent.
h. Do you have depression? This question asks if the respondent has depression.
i. Do you experience anxiety? This question asks if the respondent experiences anxiety.
j. Do you suffer from panic attacks? Does the person answering this question have panic attacks?
k. Have you ever sought treatment from a specialist? Has the respondent ever sought
l. Do you feel anxious? Does the respondent feel anxious?
m. Do you suffer from panic attacks? Does the person answering this question have panic attacks?
n. Have you ever sought treatment from a specialist? Respondent's history of seeking professional help for mental health problems.

Initially, there were 101 categorical values in the Other column and 100 numerical values in the Age column. To ensure data consistency before additional analysis, the median value was inserted into the Age column, which had one blank value (NaN).

3.2 Characteristic Evaluation

To understand the distribution and nature of the data used, a descriptive analysis was conducted. The descriptive analysis of the Age column yielded the following findings:
- Total Information: 100

The average age is 20.53 years.
2.50 years is the standard deviation.
Age Range: 18-24 years old
The majority of responses were between the ages of 18 and 23, which is a common age range for undergraduate and postgraduate students, indicating a fairly even distribution of student ages.
There is a concentration of students in the 19 and 20 years age range, as seen in the peak of the age distribution graph in Figure 3.1.

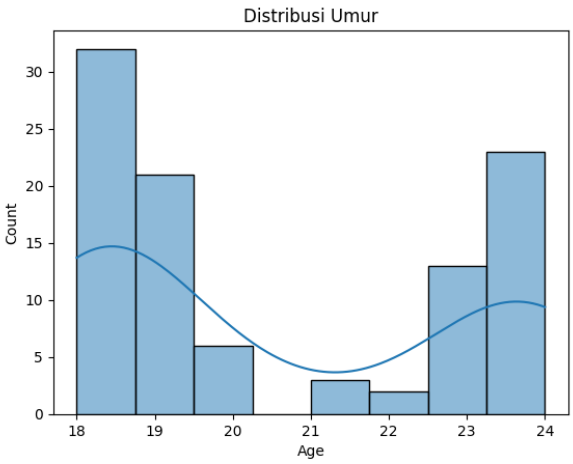

Figure 4. Age Distribution

3.3 Variables and Correlations

Correlation analysis was performed to determine the relationship between the variables in the dataset, and the results are shown in Figure 3.2. Based on the correlation map, there is a significant positive relationship between anxiety and depression, which means that students who feel anxious also often experience depression.
Depression and age have a negative relationship, which indicates that older age can reduce the likelihood of developing depression.
Anxiety and depression have a negative relationship with GPA, which suggests that better mental health may be associated with improved academic performance.

To understand the relationship between variables in the dataset, a correlation analysis was conducted, which is shown in Figure 3.2. The correlation heatmap shows that:

- Depression and Anxiety have a significant positive correlation, indicating that students who experience depression also tend to experience anxiety.
- Age has a negative correlation with Depression, suggesting that higher age may be associated with a lower risk of depression.
- CGPA showed negative correlations with Depression and Anxiety, indicating that better academic performance may be associated with better mental health..

Gambar 3.2(Heatmap Korelasi Antar Variabel)

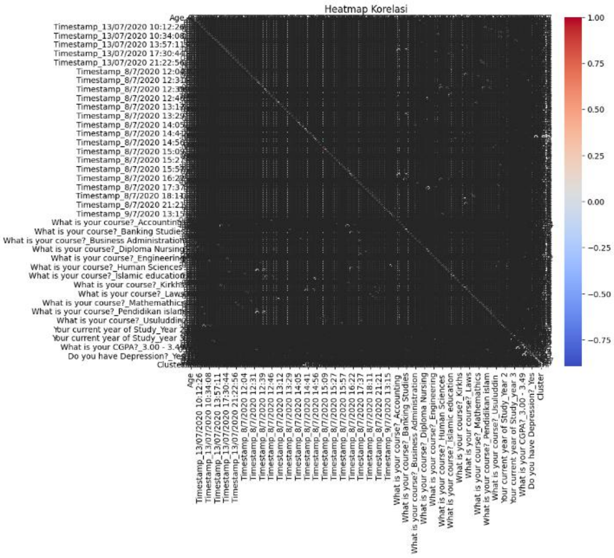

Figure 5. Heatmap Correlation Between Variable

3.4 Clustering with K-Means

Three clusters from the K-Means Clustering technique were used to find patterns or groups in the student mental health data. The following are the steps taken:

1. Data Normalization: To guarantee that each feature has the same scale, the data is normalized using StandardScaler.
2. Clustering: Using the elbow approach, which features a bending point at K=3, the K-Means algorithm is used to determine n_clusters=3..
3. Clustering Visualization Findings: As illustrated in Figure 3.3, the clustered data is represented by reducing the dimensionality of the data into two principal components using Principal Component Analysis (PCA).

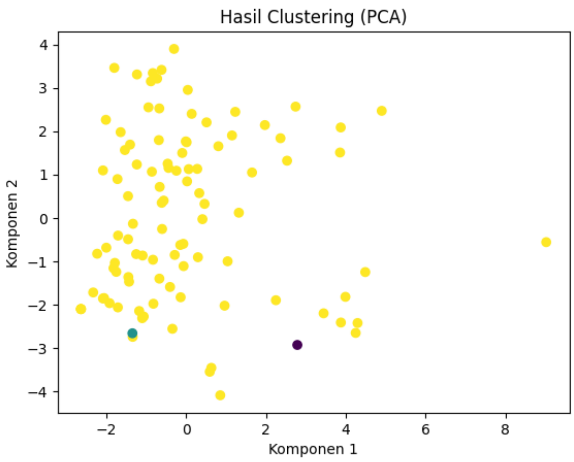

Figure 6. Clustering Resultd with PCA

Cluster Interpretation:

- tudents in Cluster 0 are unlikely to experience mental health problems. This group's anxiety and sadness ratings are low.
- Cluster 1: Students with moderate anxiety and depression scores who are at moderate risk of mental illness.
- Klaster 2: Students with high anxiety and depression scores who are at high risk of mental illness.

To create more customized interventions, these clusters offer valuable insight into how the student body is divided based on their mental health issues.

3.3 The Random Forest classifier was used to predict college students' mental health problems, mainly anxiety, panic attacks, and sadness. A number of performance criteria, such as Confusion Matrix, Classification Report, and ROC, were used to assess the model.

3.5.1 Depression Prediction

Confusion Matrix:

$$\begin{bmatrix} 21 & 2 \\ 5 & 2 \end{bmatrix}$$

Classification Report:

	precision	recall	f1-score	support
False	0.81	0.91	0.86	23
True	0.50	0.29	0.36	7
accuracy			0.77	30
macro av	0.65	0.60	0.61	30
weighted avg	0.74	0.77	0.74	30

Interpretation: With 77% accuracy, the model was able to accurately classify children who were not sad. The accuracy and recall of the model were 50% and 29% respectively, making it less successful in identifying depressed children. This shows how often truly depressed children are misclassified by the program..

3.5.2 Matrix for Predicting Marital Status:

$$\begin{bmatrix} 28 & 0 \\ 1 & 1 \end{bmatrix}$$

Classification Report :

	precision	recall	f1-score	support
False	0.97	1.00	0.98	28
True	1.00	0.50	0.67	2
accuracy			0.97	30
macro av	0.98	0.75	0.82	30
weighted avg	0.97	0.97	0.96	30

Interpretation:

The model's classification accuracy of 97% for students' marital status is impressive. For the affirmative (Yes) class, the precision and recall are 100% and 50%, respectively, indicating that the model is very good at identifying married students, but less good at classifying unmarried students..

3.5.2 Anxiety Prediction

Confusion Matrix :

$$\begin{bmatrix} 21 & 0 \\ 0 & 9 \end{bmatrix}$$

Classification Report :

	precision	recall	f1-score	support
False	1.00	1.00	1.00	21
True	1.00	1.00	1.00	9
accuracy			1.00	30
macro av	1.00	1.00	1.00	30
weighted avg	1.00	1.00	1.00	30

Interpretation: The model classified students' anxiety states with 100% accuracy without any flaws. The model demonstrated outstanding performance in identifying students with and without anxiety, as evidenced by precision, recall, and F1 scores of 1.00 for both classes (False and True).

3.5.3 Predicting Panic Attacks

Confusion Matrix :

$$\begin{bmatrix} 18 & 4 \\ 8 & 0 \end{bmatrix}$$

Classification Report :

	precision	recall	f1-score	support
False	0.69	0.82	0.75	22
True	0.00	0.00	0.00	8
accuracy			0.60	30
macro av	0.35	0.41	0.38	30
weighted avg	0.51	0.60	0.55	30

Interpretasi:

The model showed an accuracy of 60%, with poor performance in identifying students experiencing panic attacks. The precision and recall for the positive (True) class was 0.00, indicating that the model failed to classify students who actually experienced a panic attack.

3.5.3 Professional Treatment Search Prediction

Matriks Kebingungan:

$$\begin{bmatrix} 27 & 0 \\ 3 & 0 \end{bmatrix}$$

Classification Report :

	precision	recall	f1-score	support
False	0.90	1.00	0.95	27
True	0.00	0.00	0.00	3
accuracy			0.90	30
macro av	0.45	0.50	0.47	30
weighted avg	0.81	0.90	0.85	30

Interpretasi:

The model's accuracy rate of 90% indicates how well it can identify students who choose not to seek professional help. However, with precision and recall rates of 0.00 each, the model was unable to identify students who sought professional help.

3.5.3 Depression-Based Anxiety PredictionDepression-Based Anxiety Prediction

Confusion Matrix :

$$\begin{bmatrix} 21 & 0 \\ 0 & 9 \end{bmatrix}$$

Classification Report :

	precision	recall	f1-score	support
False	1.00	1.00	1.00	21
True	1.00	1.00	1.00	9
accuracy			1.00	30
macro av	1.00	1.00	1.00	30
weighted avg	1.00	1.00	1.00	30

Interpretation: Using data on depression, the model classified anxiety states with 100% accuracy without any flaws. With a precision, recall and F1-score of 1.00 for both classes, the model performed perfectly in this area.

3.3 Analisis Feature Importance

Finding the most important features for prediction is one of the benefits of the Random Forest algorithm. An examination of the importance of features reveals just thattersebut:

- The most significant predictor for sadness, anxiety, and panic attacks was student age.
- Do you suffer from anxiety? Yes: Depression is significantly affected by anxiety.
- What grade point average (GPA) do you have? 3.00 - 3.49: Students' mental health is significantly affected by their academic success as well.
- Select your gender: The risk of anxiety and depression varies by gender.
- Marital status_Yes: Marital status has an impact on mental health conditions, although the effect varies depending on the target variable.

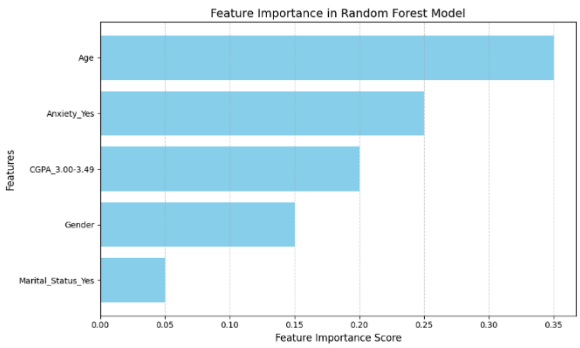

Figure 7. Feature Importance pada Model Random Forest

3.2 Discussion

The results of the analysis show some important findings related to students' mental health patterns. The following is a discussion of the results obtained:

A. Age Distribution

The majority of respondents were between 18 and 23 years old, which is a crucial time for social and academic growth. Age is one of the most important characteristics in the prediction model, which is in line with the idea that younger people are more likely to experience anxiety and depression.

B. Relationship between Variables

Anxiety and depression showed a strong positive relationship, indicating that students who suffer from one mental illness are more likely to suffer from the other.

This is consistent with research showing that anxiety and depression are often related and can exacerbate each other's symptoms [1], [3].

C. K-Means Clustering Results

Students' mental health status varied significantly when they were divided into three clusters. Cluster 0 is low-risk, which means little mental health care is needed. More care should be given to Cluster 1, which is medium-risk, to stop the emergence of more severe mental illnesses. The main focus of mental health interventions and support initiatives should be on the high-risk Cluster 2.

D. Random Forest Prediction Model:

Depression: The effectiveness of this model in diagnosing depression is poor, although it has high accuracy in identifying students who are not depressed. Class imbalance or small sample size may be the cause..

- Anxiety: The model's perfect performance in categorizing anxious adolescents suggests that the included criteria are very important for predicting anxiety.
- Panic Attacks and Treatment Seeking: The model did a poor job in categorizing students who experienced panic episodes and those who sought medical help. This suggests that more data is needed or a different algorithm that better considers minority classes is used.

E. Feature Importance

Feature importance analysis shows that Age, Anxiety, and CGPA are the main features that affect the mental health condition of college students. This is in line with previous studies which show that age and academic performance can have a significant impact on mental health [1], [2].

F. Limitations of the Study:

- Sample Size: The results of this study have limited generalizability due to the small sample size of 101 entries. Results from future research with a larger sample size will be more reliable.
- Data Imbalance: Model performance is affected by significant class imbalance of certain target variables. It is recommended to consider strategies such as over-sampling or using algorithms that are more resistant to data imbalance.
- Additional Variables: The variables included in the dataset are the only variables used in this investigation. More information can be obtained from additional variables such as environmental influences, social support, and academic pressure.

3.3 Conclusion

In this conclusion, we have outlined the results of the analysis of student mental health data using the clustering and classification approach. Key findings include:

Age Distribution: The majority of students are at an age vulnerable to mental disorders.
Correlation: There is a significant relationship between depression and anxiety.
Clustering: Identification of three distinct mental health risk clusters among students.
Prediction: The Random Forest model showed varying accuracy depending on the target variable, with the best performance in predicting anxiety.
Feature Importance: Age, anxiety, and academic achievement are the main factors influencing students' mental health conditions.

The results of this study provide a strong basis for the development of more targeted mental health interventions in higher education settings. By understanding students' mental health patterns, educational institutions can design more effective support programs to improve students' academic and personal well-being.

REFERENCES

[1] Kumparan, “Tantangan Kesehatan Mental dalam Perguruan Tinggi Generasi Z,” *Kumparan*, 2024. https://kumparan.com/pengetahuan-umum/tantangan-kesehatan-mental-dalam-perguruan-tinggi-generasi-z-22uz9ylAyda

[2] W. E. Copeland *et al.*, “Impact of COVID-19 pandemic on college student mental health and wellness,” *J. Am. Acad. Child Adolesc. Psychiatry*, vol. 60, no. 1, pp. 134–141, 2021.

[3] X. Wang, S. Hegde, C. Son, B. Keller, A. Smith, and F. Sasangohar, "Investigating mental health of US college students during the COVID-19 pandemic: Cross-sectional survey study," *J. Med. Internet Res.*, vol. 22, no. 9, p. e22817, 2020.

[4] M. Grasdalsmoen, H. R. Eriksen, K. J. Lønning, and B. Sivertsen, "Physical exercise, mental health problems, and suicide attempts in university students," *BMC Psychiatry*, vol. 20, pp. 1–11, 2020.

[5] S. Wulandari, H. S. Husin, and W. R. Sukmaningsih, "DATA MINING K-MEANS: CLUSTERING HEALTH AND COMPLAINTS RESIDENT IN INDONESIA," *Int. J. Heal. Sci.*, vol. 1, no. 3, pp. 57–63, 2021.

[6] D. M. Low, L. Rumker, T. Talkar, J. Torous, G. Cecchi, and S. S. Ghosh, "Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on reddit during covid-19: Observational study," *J. Med. Internet Res.*, vol. 22, no. 10, p. e22635, 2020.

[7] R. Rismayani, S. Alam, A. H. Endang, S. Y. Hasyrif, and N. Erdianza, "Identifikasi Gangguan Kesehatan Mental Pada Remaja Generasi Z Menggunakan Artificial Neural Network," *JUSTIN (Jurnal Sist. dan Teknol. Informasi)*, vol. 12, no. 4, pp. 776–783.

[8] A. Alfitha, S. Seruni, and W. D. Werdani, "Pengaruh Banyaknya Tugas Terhadap Kesehatan Mental Mahasiswa Universitas Pendidikan Indonesia," *Med. Nutr. J. Ilmu Kesehat.*, vol. 1, no. 4, pp. 71–80, 2023.