Jurnal Teknologi Informasi dan Pendidikan
Volume 18, No. 1, March 2025
https://doi.org/10.24036/jtip.v18i1.860

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

Comparison of Machine Learning for Mental Health Identification (The DASS-21 Questionnaire)

Retno Wardhani[1], Nur Nafi'iyah[1*]✉
[1]Informatics Engineering, Universitas Islam Lamongan, Lamongan, Indonesia
Corresponding Author: mynaff@unisla.ac.id

Article Information

Article history:

No. 860
Rec. June 09, 2024
Rev. February 20, 2025
Acc. February 24, 2025
Pub. February 25, 2025
Page. 612 – 622

Keywords:

- *questionnaire DASS-21*
- *mental health*
- *machine learning*
- *anxiety, depression, stress*
- *identification*

ABSTRACT

Stress appears at almost any age. Stress can disrupt mental and physical balance, and even students can experience it. Early detection of an individual's emotions is crucial. Researchers hope that by taking such actions, an individual can maintain self-control and prevent the stress they are experiencing from worsening. Bodily characteristics such as speech, body language, eye contact, and facial expressions indicate stress, depression, and anxiety. Psychological activities and human life are associated with physiological emotions. The three categories of negative thoughts or sad emotions are stress, anxiety, and depression. This research assesses or finds students who experience anxiety, depression, and stress. This study compares methods for determining mental health through the distribution of DASS-21 scale questionnaires. The researcher classified the research results using Naive Bayes, Decision Tree, k-NN, SVM, and Logistic Regression methods. According to experiments, SVM is effective for identifying mental health anxiety, depression, and stress with accuracy, recall, and precision of respectively *0.86, 0.90, and 0.80. At Universitas Islam Lamongan, 138 engineering faculty students answered the DASS-21 questionnaire.*

How to Cite:

Wardani, R., & Nafi'iyah, N. (2025). Comparison of Machine Learning for Mental Health Identification (The DASS-21 Questionnaire). Jurnal Teknologi Informasi Dan Pendidikan, 18(1), 612-622. https://doi.org/10.24036/jtip.v18i1.860

1. INTRODUCTION

In recent years, mental health has become a subject of concern in every field. Millions of people live with mental illnesses that often go unnoticed [1]. Most of them were children, but most were college students. Anxiety, depression, and stress are mental disorders that

can arise as a result of modern lifestyles [2]. When their environment undergoes physical and emotional changes that require a person to adapt, stress can cause feelings of worry, difficulty concentrating, changes in sleep patterns, and even worsen mental conditions. In addition, stress can cause psychological anxiety [3]. Ultimately, the frequent experience of high levels of anxiety can lead to the development of anxiety disorders. Anxiety disorders, especially when combined with other mental conditions like major depression, can lead to poor physical health, moderate to severe psychological distress, and disability [4]. As a result, high levels of anxiety put pressure on health care resources, mental health, and the overall well-being of society. Mental health resources are inadequate, and funding is currently insufficient in many countries around the world to meet this need [4],[5].

Distress and anxiety are normal reactions to threatening and unexpected situations. Changes in concentration, irritability, anxiety, insomnia, decreased productivity, and interpersonal conflict are some examples of stress reactions in response to problems (Brooks et al., 2020). Many people experience various types of psychological health problems due to modern lifestyles. Psychological problems such as anxiety, depression, and stress have several interrelated characteristics; for example, a person feels sad and lonely. Psychiatrists commonly use questionnaires such as the DASS-42 and DASS-21 to assess anxiety, depression, and stress. This is due to the fact that individuals experiencing this condition are often not open to sharing their feelings with their doctors, friends, or family. Therefore, this research creates a system for assessing mental health anxiety, depression, and stress using the DASS-21 scale. This research was carried out by distributing questionnaires to engineering students at Universitas Islam Lmaongan and analyzing the results. This research also makes a comparison of machine learning methods for identifying and classifying mental health (anxiety, depression, and stress).

Much health-related research has been carried out, for example, assessing the mental health of medical personnel in Lamongan district based on questionnaires and the DASS-21 scale by distributing questionnaires to all medical personnel in Lamongan with questions based on the DASS-21 scale [6]. The results of the medical personnel questionnaire of 2149 respondents were assessed and identified using machine learning methods (Naive Bayes, k-NN, Decision Tree, SVM, Logistic Regression, and Backpropagation). Furthermore, research related to the mental health of Lamongan Islamic University Engineering students was done by distributing questionnaires and academic tests. With the results of the questionnaire from Universitas Islam Lamongan Engineering students, there were 857 respondents. The results of the questionnaire analyzed the factors that cause students stress, namely the amount of studying in one week and clustering mental health using the k-means method [7].

By looking at behavior, body language, and facial expressions, researcher can find out students' mental health conditions from an early age. As university educators, researcher have the capacity to understand students' mental health. Educators can convey the results of the processed data to the parents or students involved because they can help

with analyzing the results of the questionnaire questions that have been processed. Treating students' mental health from the start can reduce stress and other negative factors. This research compares machine learning methods for identifying mental health (anxiety, depression, and stress).

Machine learning methods (Naive Bayes, Neural Networks, and k-NN) were used to identify anxiety, depression, and stress in data from 39,776 respondents collected via online questionnaires from 2017 to 2019 [2], [8]. Machine learning methods (Naive Bayes, Decision Tree, Random Forest, and SVM) were used to identify anxiety from respondent data from questionnaire engineering students [9], [10], [11]. The methods that are often compared for classification cases are Naive Bayes, Random Forest, SVM [12]. The SVM method has good results in prediction and classification [13], [14]. One clustering method that has good results is K-Means [15].

2. RESEARCH METHOD

The researchers collected the dataset by distributing a Google Form questionnaire to informatics engineering students at the Faculty of Engineering, Universitas Islam Lamongan. Table 1 contains questions that adapt the DASS-21 assessment scale.

Table 1. DASS-21 Questions

No	Questions
1	S I found it hard to wind down
2	A I was aware of dryness of my mouth
3	D I couldn't seem to experience any positive feeling at all
4	A I experienced breathing difficulty (e.g. excessively rapid breathing, breathlessness in the absence of physical exertion)
5	D I found it difficult to work up the initiative to do things
6	S I tended to over-react to situations
7	A I experienced trembling (e.g. in the hands)
8	S I felt that I was using a lot of nervous energy
9	A I was worried about situations in which I might panic and make a fool of myself
10	D I felt that I had nothing to look forward to
11	S I found myself getting agitated
12	S I found it difficult to relax
13	D I felt down-hearted and blue
14	S I was intolerant of anything that kept me from getting on with what I was doing
15	A I felt I was close to panic
16	D I was unable to become enthusiastic about anything

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

17	D I felt I wasn't worth much as a person
18	S I felt that I was rather touchy
19	A I was aware of the action of my heart in the absence of physical exertion (e.g. sense of heart rate increase, heart missing a beat)
20	A I felt scared without any good reason
21	D I felt that life was meaningless

Table 1 groups the questions into A (anxiety question numbers 2, 4, 7, 9, 15, 19, 20), D (depression question numbers 3, 5, 10, 13, 16, 17, 21), and S (stress question numbers 1, 6, 8, 11, 12, 14, 18). After processing the data, the researcher used only 138 respondents for the mental health classification process, categorizing them into anxiety, depression, and stress. The questionnaire added questions about age, working status, and marital status. Figure 1 shows the stages of data processing. Figure 1, researcher converted the original data into numerical values derived from the respondents' answers. Each DASS-21 scale question has an answer (0: never; 1: occasionally; 2: sometimes; 3: often). Next, researcher add up the respondents' numerical answers based on the anxiety, depression, and stress question groups. Then, the researchers create a mental health class label based on the highest number in each anxiety, depression, and stress class. Figure 1 provides the stages of class label creation, and Table 2 displays the data on the number of class labels. Table 2 shows that the lowest total data for each label is anxiety, with 15 rows per respondent. Then, in Figure 2, the stages of this research compare machine learning methods (Naive Bayes, Decision Tree, k-NN, SVM, and Logistic Regression) for mental health classification (anxiety, depression, and stress). The data in Table 2 is divided into training and testing data, with a training division of 80% from 138 rows totaling 110 rows and a testing division of 20% from 138 rows totaling 28 rows.

Table 2. Original Data Label Class

Class	Total
Anxiety (0)	15
Depression (1)	43
Stress (2)	80
Total	138

	Prediction	
Actual	Positive	Negative
Positive	TP	FN
Negative	FP	TN

This research process compares machine learning methods (Naive Bayes, Decision Tree, k-NN, SVM, Logistic Regression) with evaluation using accuracy, recall, and precision as in Equation 1, Equation 2, and Equation 3 [2].

Jurnal Teknologi Informasi dan Pendidikan
Volume 18, No. 1, March 2025
https://doi.org/10.24036/jtip.v18i1.860

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$recall = \frac{TP}{TP + FN} \quad (2)$$

$$precision = \frac{TP}{TP + FP} \quad (3)$$

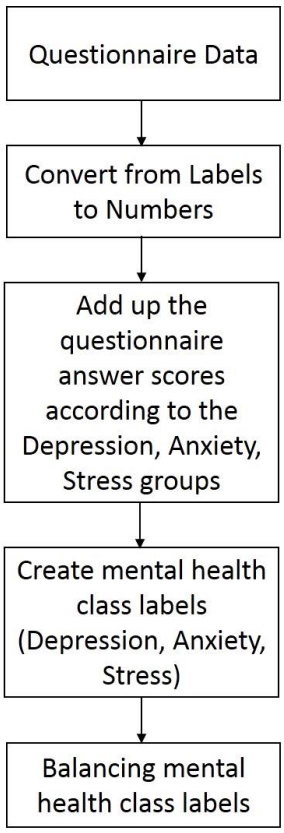

Figure 1. Stages of Data Processing

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

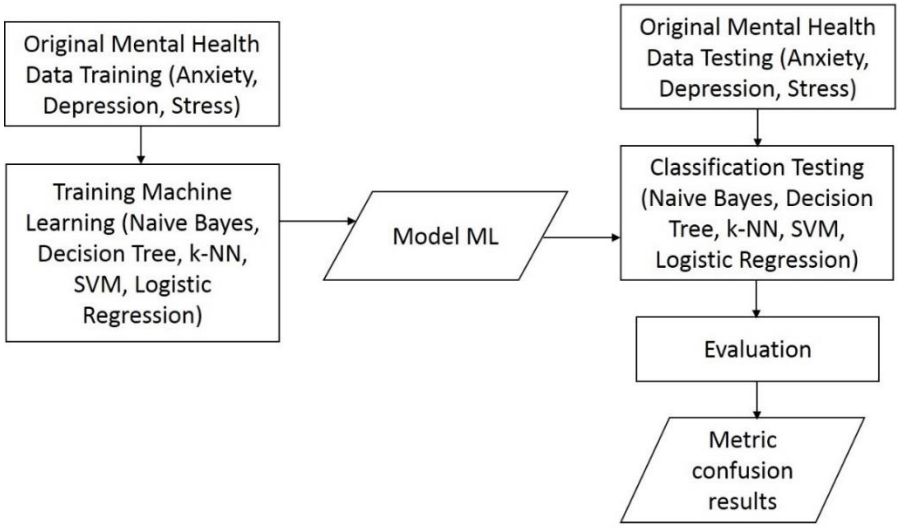

Figure 2. Research Stages

3. RESULTS AND DISCUSSION

Questionnaire data was obtained by distributing DASS-21 questions to respondents from Informatics Engineering students at Universitas Islam lamongan. The questionnaire yielded 140 responses, subsequently processing them into 138 rows. Table 3 displays the analysis of respondents between the ages of 17 and 23 years, along with their probabilities. The majority of these respondents are unmarried and have not yet found employment. Questionnaire answer data was converted from labels to numbers, as in Table 4. Table 4: Each DASS-21 question was converted to numbers: 0: never; 1: occasionally; 2: sometimes; 3: often. Table 4 from the conversion process is summed up according to the anxiety, depression, and stress question groups; the highest total value for each is the output label for mental health (anxiety, depression, and stress), as in Table 5. For example, respondent number 11 from the number of questions in DASS-21 totals anxiety 11, depression 18, and stress 16, so the mental health class output is depression (Table 5).

Table 3. Statistical Analysis of Questionnaire Data

	Value	Probability
Age	17	0.0072
	18	0.0072
	19	0.1667
	20	0.3188
	21	0.2826
	22	0.1667
	23	0.0507
Work	Yes	0.3043
	No	0.6957
Marital status	Yes	0
	No	1

Table 4. Process of Converting Characters to Numbers

I feel that life is meaningless	Result
sometimes	2
never	0
never	0
sometimes	2
occasionally	3
often	3
sometimes	2
never	0
never	0
never	0

Table 5. Target Class Creation Process

No	A	D	S	Output	Class Label Results
1	17	18	17	depression	1
2	6	9	13	stress	2
3	5	9	16	stress	2
4	17	16	19	stress	2
5	17	15	14	anxiety	0
6	17	19	20	stress	2
7	16	19	19	stress	2
8	12	5	18	stress	2
9	6	6	13	stress	2
10	13	9	17	stress	2
11	11	18	16	depression	1

This research compares machine learning methods (Naive Bayes, Decision Tree, k-NN, SVM, Logistic Regression) in classifying mental health into Anxiety, Depression, and Stress classes. The model evaluation results use accuracy, recall, precision as in Table 6. The best model based on Table 6 is SVM, with accuracy, recall, and precision values respectively 0.86, 0.90, and 0.80.

Table 6. Model Evaluation Results

Method	Accuracy	Recall	Precision
Naive Bayes	0.79	0.72	0.71
Decision Tree	0.61	0.43	0.44
k-NN	0.79	0.70	0.66
SVM	0.86	0.90	0.80
Logistic Regression	0.79	0.56	0.52

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

This research carries out data processing stages by creating balanced class labels. The process of balancing output classes uses the SMOTE (Synthetic Minority Over-sampling Technique) method [16]. The results of balancing the data resulted in each Anxiety, Depression, and Stress class having 80 rows, the total data being 240 rows. Table 7 shows the results of the balanced data evaluation during the testing process. Based on Table 7 with data that is balanced by the output class, the evaluation values (accuracy, recall and precision) increase. The methods that have the highest evaluation are SVM and Logistic Regression, respectively, the accuracy, recall and precision values are 0.96, 0.96, and 0.97. Table 8 is the statistical value of the probability of answering the DASS-21 questionnaire from respondents. There are 21 questions, each question has answers never, occasionally, sometimes and often, and the probability of each answer to the question will be calculated (Table 8). In the xx study comparing the highest accuracy of the Naive Bayes and k-NN methods was the Naive Bayes method (97.44%) [17]. Figure 3 displays an evaluation of several methods for classifying mental health. Figure 3 shows the questionnaire data that has been balanced by target class, then mental health classification testing was carried out, showing that the best methods were SVM and Logistic Regression with accuracy, recall, and precision values of 0.96, 0.96, and 0.97, respectively.

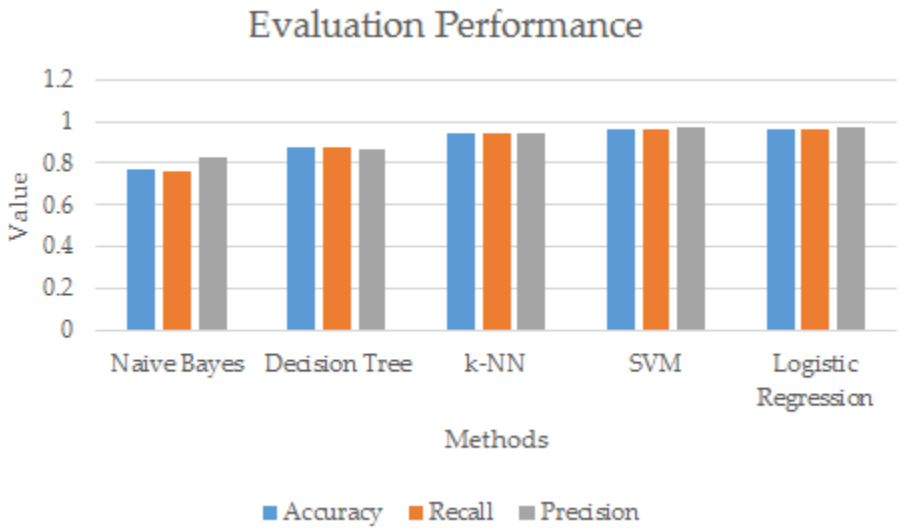

Figure 3. Evaluation Method

Table 7. Balanced Data Model Evaluation Results

Method	Accuracy	Recall	Precision
Naive Bayes	0.77	0.76	0.83
Decision Tree	0.88	0.88	0.87
k-NN	0.94	0.94	0.94
SVM	0.96	0.96	0.97
Logistic Regression	0.96	0.96	0.97

Table 8. Questionnaire Statistical Data

Question	never	occasionally	sometimes	often
1	0.107	0.59	0.200	0.100
2	0.186	0.56	0.171	0.079
3	0.293	0.53	0.143	0.036
4	0.571	0.37	0.050	0.007
5	0.243	0.57	0.157	0.029
6	0.264	0.51	0.171	0.057
7	0.357	0.44	0.150	0.050
8	0.243	0.63	0.100	0.029
9	0.236	0.49	0.179	0.093
10	0.400	0.44	0.143	0.014
11	0.164	0.59	0.164	0.079
12	0.214	0.66	0.093	0.029
13	0.257	0.59	0.129	0.021
14	0.264	0.54	0.093	0.100
15	0.643	0.33	0.029	0.000
16	0.400	0.48	0.107	0.014
17	0.457	0.41	0.064	0.064
18	0.493	0.34	0.093	0.079
19	0.300	0.54	0.107	0.050
20	0.357	0.50	0.121	0.021
21	0.614	0.27	0.057	0.057

4. CONCLUSION

This research compares machine learning methods (Naive Bayes, Decision Tree, k-NN, SVM, Logistic Regression) to classify mental health into three labels, namely Anxiety, Depression, and Stress. The results of trials of machine learning methods for classifying mental health show that the best method is SVM with accuracy, recall, and precision values of 0.86, 0.90, and 0.80. When the questionnaire data was carried out to balance the label classes (Anxiety, Depression, and Stress), the results of the evaluation of the best machine learning methods in classifying mental health, namely the SVM and Logistic Regression methods, had accuracy, recall, precision values of 0.96, 0.96, and 0.97.

ACKNOWLEDGEMENTS

Thank you to the research and development of Universitas Islam Lamongan.

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

REFERENCES

[1] R. Rosdiana, A. Hasibuan, A. Chairani, M. Daud, and M. Sayuti, "Stress Diagnosis System Using Fuzzy Logic Method," *Motiv. J. Mech. Electr. Ind. Eng.*, vol. 4, no. 3, pp. 355–366, 2022, doi: 10.46574/motivection.v4i3.162.

[2] P. Kumar, S. Garg, and A. Garg, "Assessment of Anxiety, Depression and Stress using Machine Learning Models," in *Procedia Computer Science*, 2020, vol. 171, doi: 10.1016/j.procs.2020.04.213.

[3] R. V. Shah *et al.*, "Personalized machine learning of depressed mood using wearables," *Transl. Psychiatry*, vol. 11, no. 1, 2021, doi: 10.1038/s41398-021-01445-0.

[4] Y. H. Sun, H. Luo, and K. Lee, "A Novel Approach for Developing Efficient and Convenient Short Assessments to Approximate a Long Assessment," *Behav. Res. Methods*, pp. 2802–2828, 2022, doi: 10.3758/s13428-021-01771-7.

[5] A. Y. Mughal *et al.*, "A systematic review of validated screening tools for anxiety disorders and PTSD in low to middle income countries," *BMC Psychiatry*, vol. 20, no. 1. 2020, doi: 10.1186/s12888-020-02753-3.

[6] R. Wardhani and N. Nafi'iyah, "Identification Of Mental Health Workers In Lamongan With Machine Learning," *J. Teknol. Inf. Univ. Lambung Mangkurat*, vol. 8, no. 2, 2023, doi: 10.20527/jtiulm.v8i2.160.

[7] R. Wardhani, N. Nafiiyah, and M. Kemal Farouq, "K-Means Clustering of the Mental Health of Engineering Students at the Universitas Islam Lamongan," 2023, doi: 10.1109/COMNETSAT59769.2023.10420590.

[8] A. Singh and D. Kumar, "Identification of Anxiety and Depression Using DASS-21 Questionnaire and Machine Learning," *Proc. 1st Int. Conf. Adv. Comput. Futur. Commun. Technol. ICACFCT 2021*, pp. 69–74, 2021, doi: 10.1109/ICACFCT53978.2021.9837365.

[9] J. Agarwal and J. Agarwal, "Detection and classification of anxiety students in university students through the application of machine learning," *Procedia Comput. Sci.*, vol. 218, no. 2022, pp. 1542–1550, 2023, doi: 10.1016/j.procs.2023.01.132.

[10] S. M. Et. al., "Mental Health Prediction Models Using Machine Learning in Higher Education Institution," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 5, 2021, doi: 10.17762/turcomat.v12i5.2181.

[11] A. Priya, S. Garg, and N. P. Tigga, "Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms," in *Procedia Computer Science*, 2020, vol. 167, doi: 10.1016/j.procs.2020.03.442.

[12] S. R. Ahmad, N. Insani, and M. Salim, "Analysis of Cyberbullying on Social Media Using A Comparison of Naïve Bayes, Random Forest, and SVM Algorithms," *J. Teknol. Inf. dan Pendidik.*, vol. 17, no. 1, 2024, doi: 10.24036/jtip.v17i1.807.

[13] A. S. Budi, P. H. Susilo, and N. Nafi'iyah, "SVM Algorithm for Predicting Rice Yields," *J. Teknol. Inf. dan Pendidik.*, vol. 13, no. 341, 2020.

[14] M. Yuhendri, H. Hambali, and M. Muskhir, "Permanent Synchron Magnet Motor Speed Observer Based on Least Squares Support Vector Machine Regression," *J. Teknol. Inf. dan Pendidik.*, vol. 13, no. 2, 2020, doi: 10.24036/tip.v13i2.324.

[15] I. Muluk, "Clustering Analysis of Internal and External Factors Affecting Post-Pandemic Study Duration in XYZ Educational Institution Using the Orange Application," *J. Teknol. Inf. dan Pendidik.*, vol. 16, no. 2, 2023, doi: 10.24036/jtip.v16i2.804.

[16] F. A. Bachtiar, I. K. Syahputra, and S. A. Wicaksono, "Perbandingan Algoritme Machine Learning untuk Memprediksi Pengambil Matakuliah," *J. Teknol. Inf. dan Ilmu Komput.*, 2019, doi: 10.25126/jtiik.2019651755.

[17] T. J. Melmambessy, "Analysis of the Opinion Students about The Online Learning System During the Pandemic Using The K-NN and Naïve Bayes Methods," *J. Teknol. Inf. dan Pendidik.*, vol. 16, no. 1, 2023, doi: 10.24036/jtip.v16i1.702.

[18] S. Ahmad, N. Insani, and M. Salim, "Analysis of Cyberbullying on Social Media Using A Comparison of Naïve Bayes, Random Forest, and SVM Algorithms", JTIP, vol. 17, no. 1, pp. 75-86, Jan. 2024, https://doi.org/10.24036/jtip.v17i1.807.

[19] N. Ardelia, L. Lindawati, and N. Anugraha, "Comparison of Automation Deployment Implementation on Google Cloud Virtual Machine Using Deployment Manager and Terraform", JTIP, vol. 17, no. 2, pp. 529-542, Jan. 2025, https://doi.org/10.24036/jtip.v17i2.869.

[20] F. Kamarudin, F. Budiman, S. Winarno, and D. Kurniawan, "Optimizing Classification Algorithms Using Soft Voting: A Case Study on Soil Fertility Dataset", JTIP, vol. 16, no. 2, pp. 255-268, Dec. 2023, https://doi.org/10.24036/jtip.v16i2.800.

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id