Jurnal Varian
Vol. 6, No. 1, Oktober 2022, Hal. 81–88
e-ISSN: 2581-2017

Machine Learning Prediction Anxiety Levels in the Society Academicians During Covid-19 Pandemic

Angelina Pramana Thenata[1], Martinus Suryadi[2]
[1,2]Department of Informatics, Faculty of Technology and Design, Universitas Bunda Mulia, Tangerang Selatan, Indonesia

Article Info

Article history:

Received : 07-18-2022
Revised : 10-21-2022
Accepted : 11-13-2022

Keywords:

Covid-19;
Anxiety;
SAS;
K-Means;
Confusion matrix.

ABSTRACT

Various sectors in Indonesia have been impacted by the COVID-19 incident, such as the trade, health, entertainment, and social sectors. Although several steps have been taken to minimize the coronavirus's impact, problems still occur, especially in the education sector, which must carry out one of the challenges faced in the learning process during the pandemic. However, the environment and learning process that turned into distance learning caused the interaction with friends to decrease, and academics could only move in a limited space, making them overwhelmed by feelings of anxiety. Anxiety must be detected early and managed properly not to cause mental deterioration. Therefore, the researcher aims to predict academic anxiety based on the self-rating anxiety scale (SAS), demography, family, lifestyle, and employment using k-means. Furthermore, tested the prediction results obtained with a confusion matrix in accuracy, precision, and recall. The test results found the accuracy rate is 99%, precision is 98% (moderate level), 100% (normal level), and recall is 97% (normal level), 100% (moderate level). These results indicate that the k-means on demographic, family, lifestyle, employment, and SAS aspects provide optimal results for predicting the anxiety level of the BM University academic community.

Accredited by Kemenristekdikti, Decree No: 200/M/KPT/2020
DOI: https://doi.org/10.30812/varian.v6i1.2149

Corresponding Author:

Angelina Pramana Thenata,
Department of Informatics, Faculty of Technology and Design, Universitas Bunda Mulia, Tangerang Selatan, Indonesia
Email: angelina.pramana@outlook.com

A. INTRODUCTION

Currently, the coronavirus disease 2019 (COVID-19) is causing a pandemic in almost all parts of the world. This pandemic has had devastating consequences on the global socio-economic and health systems(Marzo et al., 2021). Although the government has taken several steps to limit the impact of COVID-19, overcoming physical health and mental health is a challenge in itself (Vibriyanti, 2020). Some psychiatrists and psychologists note that almost all types of mild to severe mental disorders can occur during this pandemic (Banerjee, 2020). The most common impact of mental disorders is anxiety. This anxiety causes a person to experience anxiety, confusion, frustration, fear, and even causes to feel helpless (Brooks et al., 2020). If the anxiety is not detected as soon as possible, it will interfere with a person's daily life and surroundings.

Meanwhile, using machine learning technology can predict a person's mental health. One of them was carried out by (Wang et al., 2021), who found that using a model that they proposed and optimize with neural network algorithms to process Self-Reporting Inventory data and demographics, family, occupation, lifestyle factors, etc., the mental health prediction accuracy for global medical personnel during the covid -19 pandemic is equivalent to 92.55% (Wang et al., 2021). In addition, Haewon Byeon's research in 2021 found that by using the XGBoost, Support Vector Machine (SVM), LightGBM, Random Forest, and AdaBoost algorithms to process data from Self-rating Anxiety Scale (SAS) and the variables of gender, age, and area of residence obtained, predictions of the accuracy of mental health in the elderly who live alone in South Korea is 87.4% (Byeon, 2021). Furthermore, Anu Priya research in 2020 conducted a comparative study of the KNN algorithm, Random Forest Tree (RFT), Nave Bayes, Decision Tree, and SVM to

predict the mental health of the Indian community. They found the best RFT with an accuracy rate of 71.40% for anxiety, 79.80% for depression, and 72.30% for stress (Priya et al., 2020).

Most previous studies have proposed various algorithm models to predict mental health in the elderly, medical workers, or the general public. However, research has not looked at the conditions in academics who are not immune from the impact of the policy of limiting the spread of the virus through the distance learning system. This system makes interaction with peers less and less, and being trapped in a limited space affects their mental health. Therefore, this study aims to design a predictive system for the anxiety level of the Bunda Mulia (BM) university academic community based on aspects of the self-rating anxiety scale (SAS), demography, family, lifestyle, and employment which was processed using the k-means algorithm. Furthermore, the results of this study are expected to be a reference for analyzing mental health by utilizing machine learning technology and can contribute to further research or similar research.

B. LITERATURE REVIEW

1. Anxiety

Anxiety gives a feeling of worry, fear, and tension. Anxiety is usually temporary, which is caused when an event occurs, such as a presentation in front of many people or when an exam is in progress (Gruda and Ojo, 2022). On the other hand, conditions during a pandemic with limited space, finances, and work increase the anxiety of individuals worried about contracting COVID-19, another serious disease (Sauer et al., 2022). Individuals who often experience anxiety from time to time can have difficulty carrying out daily activities and can even affect the physical such as pain in the chest, can even experience hypertension. Excessive anxiety can make performance in learning achievement at school decrease. Early detection of the disorder helps pave the way for proper treatment and even prevent it before it worsens.

2. SAS

SAS is one of the tools invented by Zung in 1971 to determine an individual's level of anxiety (Hoque et al., 2021). SAS has affective psychometric properties based on the diagnostic criteria (Dunstan et al., 2017). The SAS focuses on general anxiety, which can detect the level of anxiety divided into four, namely normal, moderate, severe, and extreme. Obtained this level of anxiety was obtained from a questionnaire consisting of 20 questions. Questions one to five indicate emotional symptoms of anxiety, and questions six to twenty indicate physical symptoms of anxiety. This question is answered on a Likert scale of 1-4, with a total score of 20-80. A scale of 1 as a little of the time, a scale of 2 as some of the time, a scale of 3 as good part of the time, a scale of 4 as most of the time (Wang and Zhao, 2020). This study uses SAS to analyze the anxiety of BM University students in a pandemic. Each respondent has a total score of less than or equal to 44, which is included in the Normal anxiety category, a score of 4559 is included in the moderate anxiety category, a score range of 60-74 refers to the severe anxiety category, and a score above 75 refers to the extreme anxiety category (Hoque et al., 2021).

3. K-Means Algorithm

Clinical decision-making is based on clinical care guidelines and experience. However, most psychologists only emphasize salient symptoms and ignore relevant information such as age, lifestyle (van Eeden et al., 2021). On the other hand, machine learning can detect complex patterns, more time-efficient than current data processing relies heavily on human decisions (Richter et al., 2021). K-Means is one of the popular clustering algorithms that can detect complex patterns. This algorithm works with a partition system to group data into several clusters. K-Means has a working method which is as follows (Nabila et al., 2021).

1. Set the number of clusters (k) in the dataset
2. Set the centroid as random, and then when iterating use the formula shown in equation (1)

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj} \tag{1}$$

Meanwhile, i and k are cluster indices, j are index variables, V_{ij} is the average of the ith cluster centroid for the key variable, N_i is the number of members of the i cluster, and X_{kj} is the k data value with variable j for the cluster.

3. The distance between the centroid point and the point of each object is calculated by Euclidean Distance (ED). ED can calculate the distance, usually in a straight line between 2 objects in Euclidean space.
4. Objects are grouped based on the object distance to the nearest centroid.

5. Iterate until the centroid is optimal by repeating steps b to d.

The basic K-Means algorithm is to find the center of each data group to group data into one of these groups based on the distance from the center. This clustering concept has been described in Figure 1.

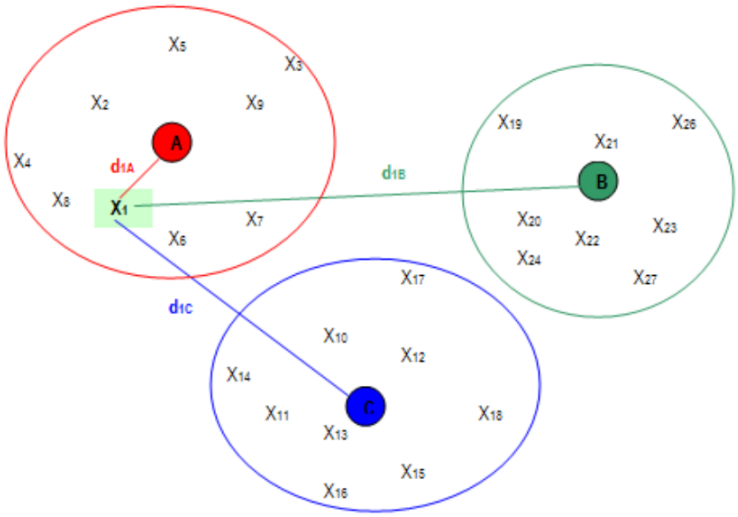

Figure 1. Illustration of How K-Means Works

Figure 1 explains that the closer the data (e.g., X_1) is to one of the existing group centers (e.g., A), the more clarified that data X_1 is a member of a group centered at A, and the more clarified that X_1 is not a member of another group (e.g., B and C) (Rarasati, 2020).

C. RESEARCH METHOD

This research has a research flow, as shown in Figure 2.

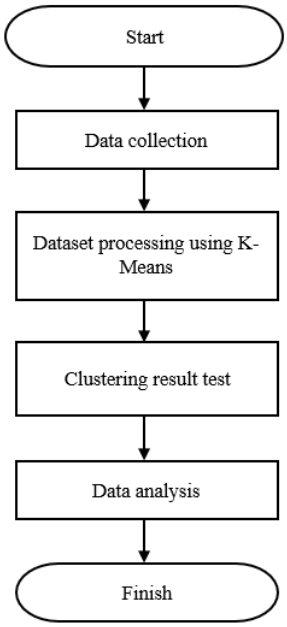

Figure 2. Research Flow

Figure 2 describes the initial stage of this research to collect data, and then it will be processed using a clustering algorithm, namely K-Means. Furthermore, the data processing results will be tested using a confusion matrix, which will then be analyzed based on the data found.

1. Data Collection

Through questionnaires, researchers collect basic academic information and variables that can affect academic and mental status during COVID-19. This study uses a dataset of 246 BM university academic data containing 16 variables (see Table 1) and SAS variables to characterize mental health in the Normal, and Moderate groups. Research has shown that measurable aspects that affect mental status mainly include four things, namely demographics, family, work, and lifestyle.

Table 1. Variable Description

Aspect	Variables
Demography	Gender
	Age
Family	Have minor children
	Have elderly
	Have a family member with a disease that makes them vulnerable to covid-19
	Have a congenital disease or a history of disease that makes them susceptible to covid-19
Lifestyle	Change of routine
	Change of attitude
	Feeling bored at home
	Afraid of being infected with covid-19 when you have a fever / flu / cough
	Afraid to die from covid-19
	Very uncomfortable when I see the news of covid-19
Employment	Can't concentrate well on any work
	Changes in work intensity
	Monthly income
	Monthly outcome

2. Result Testing K-Means

The clustering results obtained at the dataset processing stage using K-Means will be tested using a confusion matrix. This method represents the results of the classification process in True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). The test consists of precision, recall, and accuracy.

1. Precision describes the suitability of true positive data with predictive data. The precision is obtained by using equation (2). This equation compares the true positive classification results with the overall positive classification results (Sitepu et al., 2021).

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

2. Recall describes the success in obtaining information from all available data. This recall value is obtained by using equation (3). This equation compares the classification results according to reality with the sum of the classification results according to reality and negative ones that do not match reality (Chairani et al., 2021).

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

3. Accuracy describes the level of closeness of the classification results to the actual value so that the accuracy level of the model that performs the classification correctly can be known. The value of this accuracy is obtained by using equation (4). This equation performs a comparison of the correct prediction with the entire data (Fitri et al., 2019).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

D. RESULTS AND DISCUSSION

This study applies the k-means algorithm to classify academics, including normal, moderate, or severe clusters. Prediction results from the dataset can be seen in Table 2.

Table 2. K-Means Clustering Results

	x	y	**Predicted cluster**	**true label**
0	13.501.680	1.269.840	Normal	Normal
1	11.253.672	0.757761	Normal	Normal
2	-0.131533	-0.417015	Moderate	Moderate
3	-0.587468	0.351884	Moderate	Moderate
4	0.894276	13.399.343	Normal	Normal
...	...	...	...	...
95	0.029939	0.880439	Moderate	Normal
96	14.888.931	0.907428	Normal	Normal
97	0.170731	-0.293542	Moderate	Normal
98	0.215456	0.240439	Normal	Normal
99	0.081302	0.423523	Moderate	Normal
...	...	...	...	...
242	12.129.729	0.633838	Normal	Normal
243	11.027.196	0.258259	Normal	Normal
244	0.627447	-0.154538	Normal	Normal
245	-0.408354	-0.478779	Moderate	Moderate

Table 2 presents the results of the dataset collected. The SAS aspect found the academic community with 101 people with normal anxiety levels, and moderate as 145 people. However, the prediction results found that normal anxiety was 98 people, and moderate was 148 people. The results are that the cluster prediction using k-means following the reality is the normal level of as many as 98 people and the moderate level of as many as 145 people. In addition, it was also found that the prediction results that did not match the reality were three people who had the prediction results, including the moderate level. Still, the three people included the normal level. Next, the researcher visualizes the predicted results of the k-means cluster with reality in the form of a scatter plot which can be seen in Figure 3.

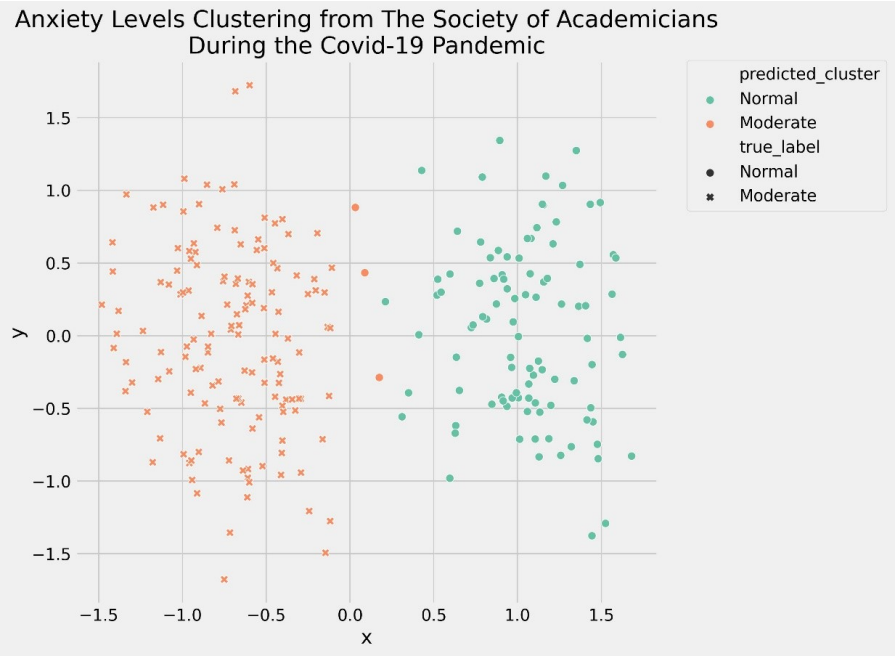

Figure 3. Visualization of Clustering Deployment Using K-Means

Figure 3 shows that green is normal and orange is moderate. Then the actual result with the round symbol is normal, and the cross is moderate. From these results, it can be seen that the green round is a normal prediction and normal reality. The orange cross is a moderate prediction and moderate reality. In addition, the results of this study were tested using a confusion matrix which can be seen in Figure 4. This figure shows the academic community's anxiety prediction system using the k-means algorithm has 99% accuracy, 98% precision at moderate level, 100% precision at normal level, 100% recall at a moderate level, and recall at normal level 97%.

The results of this study are in line with Ray Mondow Sagala, who researched predictions of student graduation (Graduated with good grades, Passed with sufficient grades, Failed) based on assignment scores, unit tests, mid-tests, attendance, residence status, language, and final assignments with using the k-means algorithm, the accuracy value is 93% (Sagala, 2021). In addition, research on grouping motorcycle data based on consumer needs (cheap, standard, and expensive) using the k-means algorithm, which has 300 datasets, gets 81% accuracy, 76% precision, and recall (Dinata et al., 2020). Thus, the k-means algorithm on demographic, family, lifestyle, employment, and SAS aspects provide optimal results for predicting the anxiety level of the BM University academic community. This study has limitations on the prediction system, which does not have a recommendation feature for academics with a moderate level of anxiety. Therefore, future research is expected to add features to these recommendations and also be able to compare several supervised learning algorithms, such as naive Bayes, supporting vector machines, or k-Nearest Neighbor (k-NN), to predict anxiety in the academic community.

```
Confusion Matrix

[[145   0]
 [  3  98]]

Accuracy: 0.99

Micro Precision: 0.99
Micro Recall: 0.99
Micro F1-score: 0.99

Macro Precision: 0.99
Macro Recall: 0.99
Macro F1-score: 0.99

Weighted Precision: 0.99
Weighted Recall: 0.99
Weighted F1-score: 0.99

Classification Report

              precision    recall  f1-score   support

    Moderate       0.98      1.00      0.99       145
      Normal       1.00      0.97      0.98       101

    accuracy                           0.99       246
   macro avg       0.99      0.99      0.99       246
weighted avg       0.99      0.99      0.99       246
```

Figure 4. Test Results Using the Confusion Matrix

E. CONCLUSION AND SUGGESTION

The Covid-19 pandemic has made academic conditions not escape the impact of the policy of limiting the spread of the virus through the distance learning system, which impacts mental health, especially anxiety. Therefore, the researcher developed an anxiety level prediction system using the k-means algorithm. The survey results obtained 101 Normal and 145 Moderate. However, the prediction system with k-means got total predictions and aligned with reality were 98 Normal and 145 Moderate. In addition, this study conducted a test using a confusion matrix which obtained 99% accuracy, 100% normal level precision, 98% moderate level, 97% normal recall level, and 100% moderate recall level. These results indicate that the k-means algorithm's application on demography, family, lifestyle, employment, and SAS has been optimal. Furthermore, future studies will compare several supervised learning algorithms, such as Naive Bayes, SVM, or k-Nearest Neighbor (k-NN), to predict anxiety in the academic community by adding a recommendation feature for academies at moderate anxiety levels.

ACKNOWLEDGEMENT

Acknowledgments are addressed to the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia, which has supported this research in the author's novice lecturer research grant program.

REFERENCES

Banerjee, D. (2020). The covid-19 outbreak: Crucial role the psychiatrists can play. *Asian journal of psychiatry*, 50:102014.

Brooks, S. K., Webster, R. K., Smith, L. E., Woodland, L., Wessely, S., Greenberg, N., and Rubin, G. J. (2020). The psychological impact of quarantine and how to reduce it: rapid review of the evidence. *The lancet*, 395(10227):912–920.

Byeon, H. (2021). Exploring factors for predicting anxiety disorders of the elderly living alone in south korea using interpretable machine learning: A population-based study. *International Journal of Environmental Research and Public Health*, 18(14):7625.

Chairani, R. C., Irawan, B., and Setianingsih, C. (2021). Klasifikasi data politik pada media sosial dengan algoritma naã ve bayes. *eProceedings of Engineering*, 8(2).

Dinata, R. K., Safwandi, S., Hasdyna, N., and Azizah, N. (2020). Analisis k-means clustering pada data sepeda motor. *INFORMAL: Informatics Journal*, 5(1):10–17.

Dunstan, D. A., Scott, N., and Todd, A. K. (2017). Screening for anxiety and depression: reassessing the utility of the zung scales. *BMC psychiatry*, 17(1):1–8.

Fitri, V. A., Andreswari, R., and Hasibuan, M. A. (2019). Sentiment analysis of social media twitter with case of anti-lgbt campaign in indonesia using naïve bayes, decision tree, and random forest algorithm. *Procedia Computer Science*, 161:765–772.

Gruda, D. and Ojo, A. (2022). All about that trait: Examining extraversion and state anxiety during the sars-cov-2 pandemic using a machine learning approach. *Personality and individual differences*, 188:111461.

Hoque, M. N., Hannan, A., Imran, S., Alam, M. A., Matubber, B., and Saha, S. M. (2021). Anxiety and its determinants among undergraduate students during e-learning in bangladesh amid covid-19. *Journal of Affective Disorders Reports*, 6:100241.

Marzo, R. R., Vinay, V., Bahari, R., Chauhan, S., Ming, D. A. F., Fernandez, S., Fernandez, P. N., Johnson, C. C. P., Rahman, M. M., Goel, S., et al. (2021). Depression and anxiety in malaysian population during third wave of the covid-19 pandemic. *Clinical epidemiology and global health*, 12:100868.

Nabila, Z., Isnain, A. R., Permata, P., and Abidin, Z. (2021). Analisis data mining untuk clustering kasus covid-19 di provinsi lampung dengan algoritma k-means. *Jurnal Teknologi Dan Sistem Informasi*, 2(2):100–108.

Priya, A., Garg, S., and Tigga, N. P. (2020). Predicting anxiety, depression and stress in modern life using machine learning algorithms. *Procedia Computer Science*, 167:1258–1267.

Rarasati, D. B. (2020). A grouping of song-lyric themes using k-means clustering. *JISA (Jurnal Informatika dan Sains)*, 3(2):38–41.

Richter, T., Fishbain, B., Fruchter, E., Richter-Levin, G., and Okon-Singer, H. (2021). Machine learning-based diagnosis support system for differentiating between clinical anxiety and depression disorders. *Journal of Psychiatric Research*, 141:199–205.

Sagala, R. M. (2021). Prediksi kelulusan mahasiswa menggunakan data mining algoritma k-means. *TeIKa*, 11(2):131–142.

Sauer, K. S., Schmidt, A., Jungmann, S. M., Bailer, J., and Witthöft, M. (2022). Do patients with pathological health anxiety fear covid-19? a time-course analysis of 12 single cases during the first wave of the covid-19 pandemic in germany. *Journal of psychosomatic research*, 152:110687.

Sitepu, A. C., Wanayumini, W., and Situmorang, Z. (2021). Determining bullying text classification using naive bayes classification on social media. *Jurnal Varian*, 4(2):133–140.

van Eeden, W. A., Luo, C., van Hemert, A. M., Carlier, I. V., Penninx, B. W., Wardenaar, K. J., Hoos, H., and Giltay, E. J. (2021). Predicting the 9-year course of mood and anxiety disorders with automated machine learning: A comparison between auto-sklearn, naïve bayes classifier, and traditional logistic regression. *Psychiatry Research*, 299:113823.

Vibriyanti, D. (2020). Kesehatan mental masyarakat: mengelola kecemasan di tengah pandemi covid-19. *Jurnal Kependudukan Indonesia*, pages 69–74.

Wang, C. and Zhao, H. (2020). The impact of covid-19 on anxiety in chinese university students. *Frontiers in psychology*, 11:1168.

Wang, X., Li, H., Sun, C., Zhang, X., Wang, T., Dong, C., and Guo, D. (2021). Prediction of mental health in medical workers during covid-19 based on machine learning. *Frontiers in public health*, 9:697850.