ISSN No:-2456-2165

Predicting Stress, Anxiety and Depression Among the University Students of India Post-Covid

Shikha Verma
Department of Computer Science
Ram Lal Anand College,
University of Delhi
New Delhi, India

Aman Goyal
Department of Computer Science
Ram Lal Anand College,
University of Delhi
New Delhi, India

Santushti Gandhi
Department of Computer Science
Ram Lal Anand College,
University of Delhi
New Delhi, India

Abstract:- Objective: The study aimed to assess the reliability and validity of the Psychological factors, namely Depression, Anxiety, and Stress Scale-21 (DASS-21), among University students in Delhi.
Methods: The DASS-21 questionnaire was administered by conducting a survey where around 100 samples were randomly selected. A comparison and training model was formed using the benchmark dataset along with the original data collected in the study. Three supervised machine learning models were trained on the same. The best model was selected and tested on the originally collected data.
Results: A training accuracy of 100%, 98%, and 98% were achieved for the Stress, Anxiety and Depression scale, respectively, using the Random Forest algorithm. The models were cross-validated using 10-fold cross validation. A cross-validation score of 99%, 98% and 99% were achieved for the scales.
Conclusion: The factors selected using machine learning techniques affect an individual's severity. To further verify these factors, practitioners were engaged to identify the specific features that influence these psychological parameters. These results helped to understand the importance and use of machine learning techniques for analyzing the severity of stress, anxiety and depression scales amongst individuals. The testing accuracy achieved was similar to the training accuracy indicating the model did not have any anomaly and could be used for predicting the severity of stress, anxiety and depression among university students in India.

***Keywords**:- Stress, Anxiety, Depression, Correlation, Supervised Machine Learning, Testing Accuracy.*

PUBLIC SIGNIFICANCE STATEMENT

Mental Health is an essential aspect whose importance in the context of higher education cannot be understated. This study on the students of the University of Delhi throws light on a problem that is becoming common each passing day. As per the National Crime Records Bureau's (NCRB) Accidental Deaths & Suicides in India (ADSI) report, over 13,000 students died in 2021 in India at the rate of more than 35 every day, a rise of 4.5 percent from the 12,526 deaths in 2020 with 864 out of 10,732 suicides being due to "failure in examination." The findings of this study have significant implications for both the student body and the administration of the university. By identifying the variables causing stress, anxiety, and depression among students can give us crucial insights into building tailored interventions and support systems. These initiatives could include enhancing counselling services, assisting with efforts to raise mental health awareness, and developing a friendlier and stress-resistant school climate.

This study concludes by issuing an urgent appeal for reform at the University of Delhi and beyond. We can cultivate a more compassionate and empowered generation of people who are better able to negotiate the hurdles of both academics and life by understanding the struggles students have with their mental health and taking proactive measures to offer the necessary support.

I. INTRODUCTION

Depression, stress, and anxiety are common emotions that arise from life's challenges. Depression is characterized by low mood and loss of interest, while stress results from feeling overwhelmed due to pressure. Excessive stress can lead to distress and exhaustion. Anxiety is the feeling that something terrible is about to happen and can be general or specific to a situation, place, or thing. Understanding these distinctions and seeking help when necessary is essential to ensure timely and effective treatment. Early detection is critical, and prevention is better than cure. [1].

The Depression, Anxiety, Stress Scales, commonly known as DASS, is a scale which consists of self-report items. DASS-21 is a questionnaire which consists of 21 questions that the user must answer. The questionnaire is divided into three scales Depression, Anxiety and Stress, and it consists of equal numbers of questions for each scale. The score of each scale is added, which is then used to determine the severity of Depression, Anxiety and Stress of an individual [2].

Machine Learning is a subfield of Artificial Intelligence defined as a machine's capability to imitate human behaviour and predict unforeseen events. It plays a pivotal role in the field of Medicine, Sales and Operations, Customer Service and many more. Machine Learning is of great help in the field of psychology. It allows psychologists to organize data, describe data, and make inferences based

ISSN No:-2456-2165

on the data [3]. All these steps will enable a psychologist to understand the patients better and help in developing better tests and practices to help the patients. In the following sections, we will look at the work done in psychology using machine learning.

II. RELATED WORK

Shayan et al. [4] validated the DASS-42 questionnaire in the Dari region of Afghanistan, while J.A. Ademuyiwa and A. A. Adetunji [5] surveyed 200 staff members at the Federal Polytechnic in Ile-Oluji, Nigeria. Priya et al. [6] used various machine learning algorithms to determine the best one for predicting psychological problems, and Al-Wesabi et al. [7] used the DASS-21 questionnaire to evaluate stress, anxiety, and depression severity during COVID-19. Mary and Jabasheela [8] used the DASS-21 questionnaire to evaluate different Machine Learning techniques for classification in a study involving 600 students from Puducherry, India. Marouf et al. [9] examined the relationship between the Big Five Personality traits and perceived stress in Bangladeshi computer science undergraduate students using machine learning techniques. Singh [10] analyzed a dataset of 42,000 instances acquired from an online source containing 170 features with 42 questions from the DASS-42 questionnaire, and Srinath et al. [11] used an online questionnaire to collect data from 30,776 participants between 2017 and 2019. Lastly, Chiong et al. [12] aimed to detect the severity of depression using pre-processed social media text data from Twitter, and Sun et al. [13] evaluated depression levels using DASS-42.

III. PROPOSED WORK

This study focused on determining the severity level of stress, anxiety and depression scale using the Depression, Anxiety and Stress Scale questionnaire (DASS-21). The training data was collected from online sources, and the testing data was collected from 100 participants via Google Forms. Three machine learning algorithms – namely Gaussian Naïve Bayes, Logistic Regression and Random Forest were used for the classification.

A. DASS-21 Survey

The data for this study was collected through the DASS-21 questionnaire. The questionnaire comprises 21 questions. Each scale of Stress, Anxiety and Depression is allocated seven questions. The options available for each question in the DASS questionnaire [14] are shown in Figure 1 below:

The rating scale is as follows:

0 Did not apply to me at all
1 Applied to me to some degree, or some of the time
2 Applied to me to a considerable degree or a good part of time
3 Applied to me very much or most of the time

Fig 1. Values corresponding to the DASS-21 questionnaire.

The questions asked in the questionnaire [14] are described in Figure 2.

Depression	Anxiety
3. I could not seem to experience any positive feeling at all.	2. I was aware of dryness of my mouth.
5. I found it difficult to work up the initiative to do things.	4. I experienced breathing difficulty.
10. I felt that I had nothing to look forward to.	7. I experienced trembling (e.g., in the hands).
13. I felt down-hearted and blue.	9. I was worried about situations in which I might panic and make a fool of myself.
16. I was unable to become enthusiastic about anything.	15. I felt I was close to panic.
17. I felt I was not worth much as a person.	19. I was aware of the action of my heart in the absence of physicalexertion.
21. I felt that life was meaningless.	20. I felt scared without any good reason.

Stress
1. I found it hard to wind down.
6. I tended to over-react to situations.
8. I felt that I was using a lot of nervous energy.
11. I found myself getting agitated.
12. I found it difficult to relax.
14. I was intolerant of anything that kept me from getting on with what I was doing.
18. I felt that I was rather touchy.

Fig 2. The questions for the DASS-21.

After the data collection, the participant's responses to each Stress, Anxiety and Depression scale were grouped separately. They were then calculated by simply adding up the values associated with each question using the below formula [15]:

$$score = Sum\ of\ rating\ points\ of\ each\ class * 2 \quad (1)$$

B. Score Parameters

The condensed form of the original questionnaire is called the DASS-21. The dataset used was a benchmark dataset. The original dataset consists of DASS-21 questions and has data about an individual's personal information such as age, gender, ethnicity, etc. The severity of stress, anxiety and depression is categorized into five categories and is presented in the figure below [14].

Meaning	Depression	Anxiety	Stress
Normal	0-9	0-7	0-14
Mild	10-13	8-9	15-18
Moderate	14-20	10-14	19-25
Severe	21-27	15-19	26-33
Extremely severe	28+	20+	34+

Fig 3. The score range for the DASS-21 subscales

C. Feature Selection

The original dataset consisted of various features pivotal to the study, such as gender, age etc. However, some attributes, such as the screen size of the user's device and which hand is dominant, were deemed non-essential. So, to find this, a correlation between the attributes and the target column was found. The questions about stress, anxiety and depression were stored in different dictionaries. The severity of Stress, Anxiety and Depression amongst the individuals was calculated using formula 1. We created three columns to hold the stress, anxiety and depression scores achieved by an individual based on the Taylor Manifest Anxiety Scale. We categorized them based on the ranges mentioned in Figure 3. Except for the target column, the data was standardized using standardscaler().

D. The Framework of the Proposed Work

The primary aim was to determine the severity of an individual's stress, anxiety and depression. The initial step was creating different models for stress, anxiety and depression. Firstly, the null records were dropped, and the count of each severity for the scales was plotted. This helped us understand the balance among the various severity levels, as shown below in Figure 4.

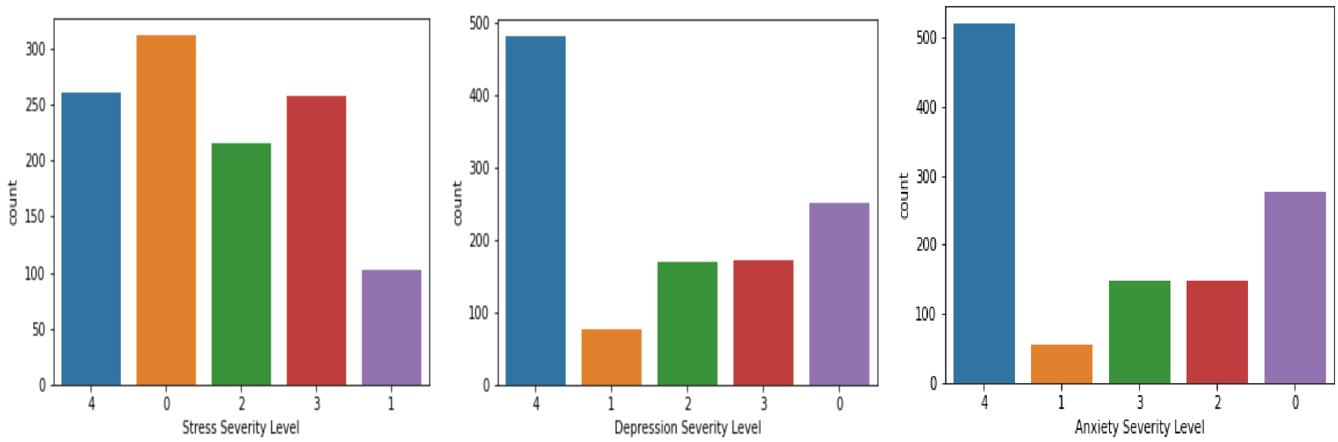

Fig 4. Count plot of the severity of the Stress, Depression and Anxiety scale

From the above graphs, we can infer that the depression and anxiety scale data was imbalanced. The data included cases where most samples had highly severe depression and anxiety. The data was split into training and testing data in a ratio of 75:25. In the next stage, Gaussian Naïve Bayes, Logistic Regression and Random Forest were applied to the training data to help us determine the severity. The performance of the algorithms was tested based on different metrics explained in the following section. The framework of the model is shown below in Figure 5.

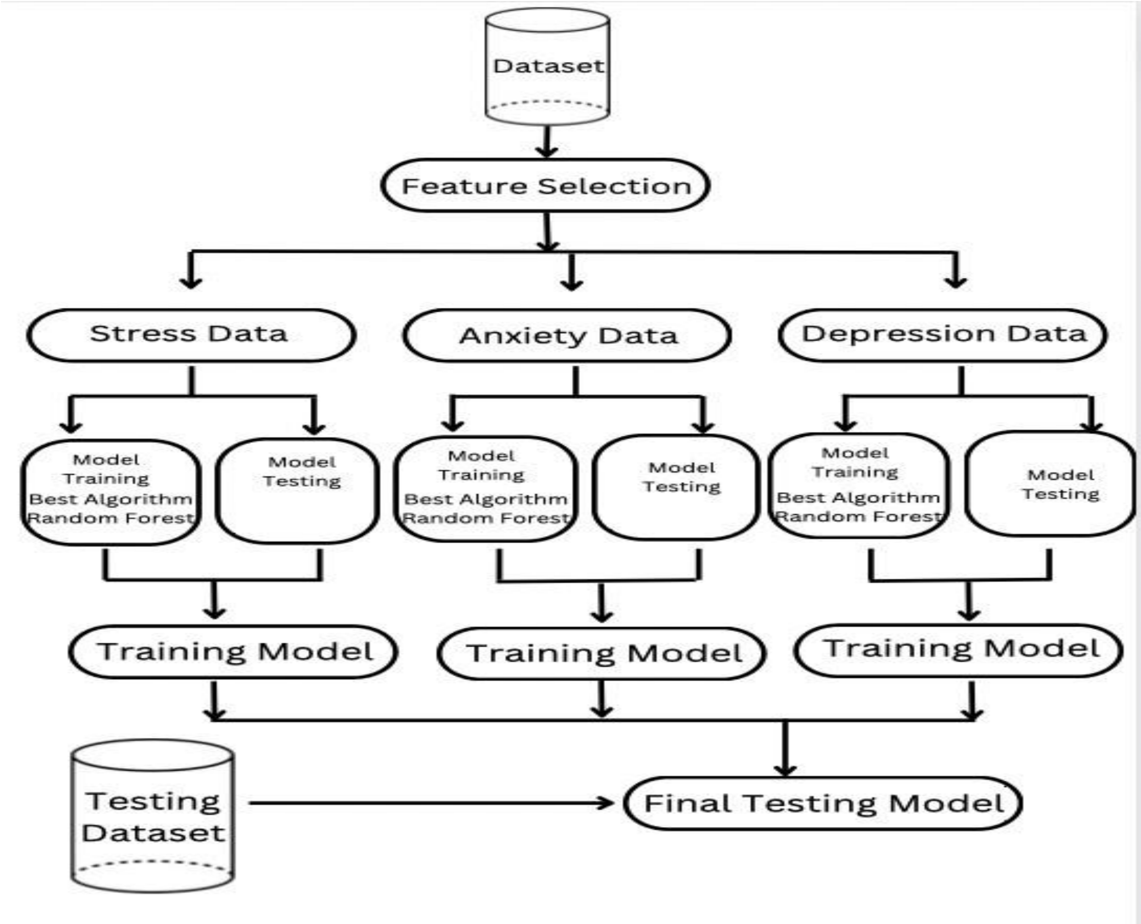

Fig 5. The framework of the model

E. *Hyperparameter Tuning for the Model*

Hyperparameter Tuning is a concept where the existing machine learning models are improved using a set of optimal hyperparameter values. Combining hyperparameters helps maximize the model's performance and minimize the loss to produce better results with fewer errors [16]. The technique used in this study was GridSearchCV which helped to find the optimal hyperparameters for all the algorithms shown in Table 1 below.

Table 1. The value of hyperparameters for the different models

Hyperparameter	Models	Value
var_smoothening	Gaussian Naïve Bayes	1e-06
multi_class	Logistic Regression	multinomial
n_jobs	Logistic Regression	-1
criterion	Random Forest	entropy
max_depth	Random Forest	10
n_estimators	Random Forest	100
random_state	Random Forest	100

IV. RESULTS AND DISCUSSION

The quality of a machine learning model and its ability to generalize to new data is assessed with the help of performance metrics. The study evaluated the three supervised machine learning algorithms using four metrics, Accuracy, Cross-Validation, Precision, and Matthews Correlation Coefficient (MCC). The cross-validation score for each algorithm was determined for each scale separately, and the best model was selected for each scale based on these metrics. The results for different metrics for all the subscales are shown below in Table 2.

Table 2. Result for different metrics for DASS-21 subscales

Algorithm	Scale	Accuracy	Cross-Validation	Precision	MCC
Gaussian Naïve Bayes	Stress	0.9	0.89	0.98	0.87
	Anxiety	0.98	0.97	0.97	0.97
	Depression	0.82	0.81	0.96	0.77
Logistic Regression	Stress	0.93	0.91	0.98	0.91
	Anxiety	0.91	0.92	0.9	0.87
	Depression	0.93	0.91	0.96	0.9
Random Forest	Stress	1.0	0.99	1.0	0.97
	Anxiety	0.98	0.98	0.96	0.98
	Depression	0.98	0.99	0.96	0.97

As inferred from the above observations, Random Forest was the best-performing algorithm for all the scales based on all the metrics.

Table 3. Results for DASS-21 subscales for validation data.

Algorithm	Anxiety	Stress	Depression
Gaussian Naïve Bayes	0.989	0.967	0.989
Logistic Regression	0.733	0.8	0.733
Random Forest	0.989	1.0	0.967

Random Forest had the best testing accuracy for the scales, and Gaussian Naive Bayes had the same accuracy for Anxiety. Similar testing and training accuracy indicate no over- or under-fitting, making models perform similarly on new and trained data. Random Forest outperformed others as it is an ensemble of Decision Trees, using a subset of the data to make different predictions. Majority voting is used to make final predictions based on the output of each tree.

V. CASE STUDY

We analyzed our machine learning model's performance by engaging practitioners in studying the patterns observed. The practitioners were given access to all the features present in the dataset and were asked to identify which features were linked to an individual's levels of stress, anxiety, and depression. The factors selected by the practitioners for the depression scale are shown below in Table 4.

Table 4. Stress, Anxiety and Depression features selected by the practitioner

Stress	Anxiety	Depression
Finding hard to wind down	Being aware of dryness of mouth	Using a lot of negative energy
Worried about situations where I might panic and make a fool of myself	Experiencing breathing difficulty	Not experience any positive feelings at all
Getting agitated	Over-reacting to situations	Feeling that there is nothing to look forward to
Feeling close to panic	Feel like trembling	Feeling downhearted and blue
Intolerant of things that kept me from getting on with what I was doing	Aware of the action of my heart in the absence of physical exertion	Difficult to work up the initiative to do things
Scared without any good reason	Difficulty in relaxing	No enthusiasm about anything
Feeling that I was rather touchy		Feeling that I am not worth much as a person
		Feeling that life wasn't worthwhile.

The above table shows a pattern in the factors selected by the practitioners for DASS-21 subscales. The practitioners' selection of factors for each scale showed the following patterns :

- Physical characteristics are more related to anxiety due to the release of cortisol during the body's "fight or flight" response [17].
- Self-esteem and self-confidence were more related to depression, while behaviour-based parameters were more related to stress.
- These factors can also be interconnected, such as fear causing restlessness and hindering activity and focus.
- These findings suggest a complex relationship exists between physical, psychological, and behavioural factors in the development of anxiety, depression, and stress.

Further research can explore the interplay between these factors and their impact on mental health outcomes.

VI. CONCLUSION & FUTURE WORK

The study used machine learning algorithms to identify severity levels of anxiety, stress, and depression using the DASS-21 questionnaire. Three supervised ML techniques were employed, with Random Forest showing the best performance for DASS-21 subscales' training data. Testing accuracy of 0.75, 1.0, and 0.995 was achieved for stress, anxiety, and depression, respectively. Random Forest and Gaussian Naïve Bayes performed best for the testing data, while Logistic Regression had a significant drop in accuracy. Metrics such as precision, Matthew correlation coefficient, and cross-validation were used to evaluate model performance due to the data's imbalanced nature.

The primary objective of our paper was to investigate the levels of stress, anxiety, and depression among students. Upon analyzing the survey results, we noted that the participants fell within the age range of 18-22. Expanding our study to include various age groups would allow for a comparative analysis of the prevalence and severity of stress, anxiety, and depression across different populations. This would provide valuable insights into the role of age concerning stress, anxiety, and depression, as well as the variations in the frequency of these conditions among different age groups based on their responsibilities.

To verify these factors affecting the psychological parameters, we sought the input of practitioners who identified the features that impacted these parameters. The DASS-21 questionnaire reported in this study is a valid and reliable tool for assessing the severity of stress, anxiety and depression amongst university students in Delhi. This questionnaire can serve as a baseline for a more comprehensive validation analysis of the magnitude of stress, anxiety and depression amongst university students in India and help analyze the reason behind the development of stress, anxiety and depression.

REFERENCES

[1]. "Anxiety, Depression, Stress: Why the Differences Matter | Premier Health." https://www.premierhealth.com/your-health/articles/women-wisdom-wellness-/anxietydepression-stress-why-the-differences-matter (accessed Sep. 20, 2022).

[2]. "Overview of the DASS and its uses." http://www2.psy.unsw.edu.au/dass/over.htm (accessed Sep. 20, 2022).

[3]. K. Cherry, "Why Are Statistics Necessary in Psychology?," p. 2.

[4]. N. Shayan, A. Niazi, A. WASEQ, and H. Ozcebe, “Depression, Anxiety, and Stress Scales 42 (DASS-42) in Dari-Language: Validity and Reliability Study in Adults, Herat, Afghanistan,” *Bezmialem Science*, vol. 9, pp. 356–362, Jun. 2021, doi: 10.14235/bas.galenos.2020.4250.

[5]. A. A. Adetunji and J. A. Ademuyiwa, "Assessing DASS-42 Models among Polytechnic Staff," *Open Access Library Journal*, vol. 6, no. 10, Art. no. 10, Oct. 2019, doi: 10.4236/oalib.1105334.

[6]. A. Priya, S. Garg, and N. Tigga, "Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms," *Procedia Computer Science*, vol. 167, pp. 1258–1267, Jan. 2020, doi: 10.1016/j.procs.2020.03.442.

[7]. F. N. Al-Wesabi, H. Alsolai, A. M. Hilal, M. A. Hamza, M. Al Duhayyim, and N. Negm, "Machine Learning Based Depression, Anxiety, and Stress Predictive Model During COVID-19 Crisis," *CMC-Comput. Mat. Contin.*, pp. 5803–5820, 2022.

[8]. S. T. A. Mary and L. Jabasheela, "An Evaluation of Classification Techniques for Depression, Anxiety and Stress Assessment," presented at the International Conference for Phoenixes on Emerging Current Trends in Engineering and Management (PECTEAM 2018), Atlantis Press, Feb. 2018, pp. 64–69. doi: 10.2991/pecteam-18.2018.13.

[9]. A. Marouf, A. Ashrafi, T. Ahmed, and T. Emon, "A Machine Learning based Approach for Mapping Personality Traits and Perceived Stress Scale of Undergraduate Students," *International Journal of Modern Education and Computer Science*, vol. 11, pp. 42–47, Aug. 2019, doi: 10.5815/ijmecs.2019.08.05.

[10]. T. Singh, "Prediction of Depression, Stress and Anxiety," *International Journal of Research in Engineering, Science and Management*, vol. 5, no. 4, Art. no. 4, May 2022.

[11]. K. S. Srinath, K. Kiran, S. Pranavi, M. Amrutha, P. D. Shenoy, and K. R. Venugopal, "Prediction of Depression, Anxiety and Stress Levels Using Dass-42," in *2022 IEEE 7th International conference for Convergence in Technology (I2CT)*, Apr. 2022, pp. 1–6. doi: 10.1109/I2CT54291.2022.9824087.

[12]. R. Chiong, G. S. Budhi, S. Dhakal, and F. Chiong, "A textual-based featuring approach for depression detection using machine learning classifiers and social media texts," *Computers in Biology and Medicine*, vol.

ISSN No:-2456-2165

135, p. 104499, Aug. 2021, doi: 10.1016/j.compbiomed.2021.104499.

[13]. Y. H. Sun, Q. Liu, N. Y. Lee, X. Li, and K. Lee, "A novel machine learning approach to shorten depression risk assessment for convenient uses," *Journal of Affective Disorders*, vol. 312, pp. 275–291, Sep. 2022, doi: 10.1016/j.jad.2022.06.035.

[14]. "Depression Anxiety Stress Scales (DASS)." https://www.psytoolkit.org/surveylibrary/depression-anxiety-stress-dass.html (accessed Sep. 20, 2022).

[15]. P. F. Lovibond and S. H. Lovibond, "The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories," *Behav Res Ther*, vol. 33, no. 3, pp. 335–343, Mar. 1995, doi: 10.1016/00057967(94)00075-u.

[16]. "Anyscale - What is hyperparameter tuning?," *Anyscale*. https://www.anyscale.com/blog/whatis-hyperparameter-tuning (accessed Sep. 29, 2022).

[17]. "Physical Symptoms of Anxiety: What Does It Feel Like?" https://www.healthline.com/health/physical-symptoms-of-anxiety#fa-qs (accessed Jan. 06, 2023).