International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(9s) 2026 pp. 1158-1170
https://doi.org/10.70917/ijcisim-2026-3544
Article

An Empirical Evidence of Higher Model Efficiency of Machine Learning Model than A Classical Multiple Linear Regression Model

Sabharwal Alka[1], Goyal Babita[2*], Sehgal Shikha [3], Kujal Swati[4], Vidushi[5]

[1]Department of Statistics, Kirori Mal College, University of Delhi, India
Orcid ID: 0000-0002-8252-8284
[2]Department of Statistics, Ramjas College, University of Delhi, India
Orcid ID: 0000-0003-3977-5690
[3]Department of Statistics, Kirori Mal College, University of Delhi, India
Orcid ID: 0000-0002-2333-9264
[4]Department of Statistics, Kirori Mal College, University of Delhi, India
Orcid ID: 0009-0005-3248-7698
[5]Department of Statistics, Kirori Mal College, University of Delhi, India
Orcid ID: 0009-0001-4732-6345
[*]**Corresponding author:** goyalbabita@gmail.com

Abstract: Psychological health of young adults has always been an area of concern for the stakeholders, partially due to the fact that the psychological problems might remain unnoticed unless sever enough. One of the ways of extracting the information about the general psychological health is through questionnaires administered on groups of individuals. Choice of an appropriate statistical technique in analysing this information is as important as the selection of the questionnaire. If more than one technique fit the situation then a natural course of action is to compare the two in terms of some standard parameters. In the present study, data was collected on students enrolled in various colleges affiliated to University of Delhi, India. Information was sought on their psychological wellness and some stresses viz. the academic stress, social stress and financial stress, along with some demographic information such as the year of course and gender. However, no significant effect of gender and the year of course was found. A classical multiple linear regression (MLR) model and machine learning multiple linear regression (ML-MLR) models with varying train data sizes were applied to the data to estimate/predict wellness scores on the basis of the independent predictors. The three stress factors were all significant. The R2 of the MLR model was 0.5207 and the model was significant with p-value of the F-statistic < 0.001. Both the techniques produced models with estimating/predictive efficiency at least 82%. However, the ML-MLR models had the predictive efficiency at least equal to or more than the estimating efficiency of the classical MLR model. The best results were obtained for train data size 70%, a prevalent size selection in machine learning models. The study reinforced the strength of machine learning models over the classical methods.
Keywords: Multiple Linear Regression, Machine Learning, Psychological Wellness Score, Predictive Efficiency, Young adults, Stress

1. INTRODUCTION

Post Covid, there is an increased awareness about health and health issues. In general, health is a state of physical, mental and social wellbeing. Any changes in any of the three health aspects affect the overall health. Whereas physical health issues are more visible and are easier to be monitored, mental and social wellbeing are more intricate in nature. In particular mental health which, in turn, consists of psychological and emotional health, is more difficult to monitor because of its multidimensional nature [9] and any psychological issues might go unnoticed unless they reach a high severity level. It has been observed that post covid, the mental health issues have been on a rise due to various reasons [6]. According to WHO, post covid era has seen up to 25% increase in mental health issues. The young people and women are the worst affected segment of the society [24]. According to a report by NAMS task force on mental stress**,** despite identification of the problem, there have been huge gaps in tackling the issues of mental health [16].

According to WHO, psychological problems are more prevalent among young adults than any other segment of the society [25]. A frequent manifestation of psychological issues, the stress among college students has always been a concern for academicians, doctors, parents, researchers and policymakers. Stress is the physical, mental or emotional inability to cope with the demands or expectations [2]. Long-term exposure to excessive stress may result in poor functioning and negative health outcomes [8]. Researchers have identified several reasons associated with stress among college going students such as psychological, social, financial, and academic pressures [1,7,10,17]. Another reason of psychological health issues among young adults is their transition from adolescence to adulthood when they learn to be independent, free from the controlled environment of their schools and homes [12,14,26]. In a previous study, the authors explored the effect of parental mental health on the wellbeing of adolescents [3].

Psychological health issues are difficult to identify as people do not come forward to report their problems. Further, at initial stages there are no visible symptoms. As such, information is collected through standardized and validated questionnaires. The same questionnaire(s) can be administered on groups in form of surveys, or a number of questionnaires can be administered to a potential subject in order to gauge the nature and extent of problem. A number of questionnaires are available which are used under different circumstances. Some of them, such as Brief Psychiatric Rating Scale (BPRS) can only be administered by an expert [23], several of them are self-reporting, i.e. to be filled by the respondent themselves. These self-reporting questionnaires can be of general nature such as General Health Questionnaire (GHQ) and its different versions, Strength and Difficulty Questionnaire (SDQ) etc. or can be designed to measure the extent of a specific issue such as Perceived Stress Scale (PSS), the Depression Anxiety Stress Scale (DASS), to name a few [15]

Use of an appropriate analytical technique to analyse the data is as important as a good questionnaire. Selection of technique(s) is a function of the objective. In mass studies, the objective is generally identification and/or classification of an attribute on the basis of response of the questionnaire items. In the former case, if the number of response items is not very large or similar items can be clubbed under a common head, then multiple linear regression (MLR) is an efficient tool to determine the relationship between the target variable and independent predictors. Clubbed with machine-learning (ML), MLR is able to extract a model which may be more efficient than an only MLR model. Both these techniques have been used heavily in medical and bio-medical data [3,19-21]. The authors of this study used MLR to estimate and predict the onset of diabetic nephropathy on the basis of independent predictors, fasting blood glucose, systolic and diastolic blood pressures, low density lipoprotein, and with duration of disease and age at which diabetes was diagnosed as covariates [5]. In the latter case of classification problem, non-parametric regression methods are generally applicable [19].

The present study aims at documenting the general psychological health of young adults in presence of various stress causing factors around them. The data for the study was collected through a survey conducted in online-offline mode on the students of different colleges affiliated to the University of Delhi, India between September-November 2025. The questionnaire used was based on the one used by Yikealo et al [27]. The questionnaire assessed perceived stress in four areas of an individual's life: Academic Stress, Social Integration, Financial Stability and Well-being. In addition to these items, demographic questions about gender, year of course and course name etc. were also asked. A total of 296 responses were collected. The aim of the study was to profile the wellness score (a proxy for presence of any psychological problems) in terms of common stress factors viz, academic, social and financial stresses of young adults. These stress factors were chosen among several as the studies indicate them to be most common stress factors for young adults [10,14]. The second objective of the study was to explore the effectiveness of machine learning MLR model. The generally used bifurcation of data in ML methods is 70% train data and 30% test data. This bifurcation was empirically validated by running the ML-MLR model several times, with varying train and test data sizes [22].

Besides Introduction, the paper has three more sections: Material and methods, Results; and the Discussion.

2. MATERIAL AND METHODS

2.1 Material

The data was collected through a survey conducted in online/offline mode among the students of various colleges affiliated to University of Delhi, India, which consisted of a psychiatric questionnaire adapted from a standardised tool, along with some demographic details such as gender, course, and year of study and six items about the financial status of the respondents. The psychiatric questionnaire was about Academic Stress, Social Integration, and Overall Well-being or psychological well-being, with 10 each items respectively in each. A total of 296 responses was obtained.

Following is the detailed break-up of the questionnaire (Table 1):

Table 1: The scale break-up of the questionnaire

S. No.	Scales	Number of items	Optional Ques. (if any)
1	Demographic Information	11	1
2	Academic Stress	10	0
3	Financial Stability	6	1
4	Social Integration	10	0
5	Overall Well-being	10	0
	Total	54	2

The scoring scheme was a 1-5 Likert scale where '1'-Never; '2'- Rarely; '3'- Sometimes; '4'- Frequently; and '5'- Always.

A pilot study was conducted to ensure the respondents' understanding of the questionnaire. Minor revisions were made after the pilot study.

Inclusion and Exclusion Criteria:

Inclusion criterion: Participants were included in the study if they:

- Provided informed consent
- Were currently enrolled in a college
- Completed at least 80% of the questionnaire

Exclusion criterion: Participants were excluded if they

- reported a history of serious illness or severe physical injury.

No participants reported substance or drug abuse to the best of the researchers' knowledge.

2.2 Methods

Data reliability was tested using Cronbach's alpha and Guttman's Lambda (λ_6) given by

$$\text{Cronbach's alpha,} \alpha = \frac{k}{k-1} \times \frac{1 - \sum {\sigma_i}^2}{\sigma_t^2} \;;$$

where k is the number of test items; ${\sigma_i}^2$ is the variance of the ith item and ${\sigma_t}^2$ is the variance of the total score of the observed data.

$$\text{Guttman's Lambda,} \lambda_6 = 1 - \frac{\left[\sum_{i=1}^{k}(1 - R_i^2)\right]}{\sigma_t^2}$$

R_i^2 is the coefficient of determination of the ith item in its regression with all the other items.

The values of both the statistics lie between 0 and 1. High values of these statistics indicate high internal consistency of the data.

Anova was used for four group comparison for the year of course and Welch t-test was used for two-group comparison of gender (it does not make any assumption about equality of variances).

Welch t-test

Welch t-test is a test for comparing the group means of two groups when the variances of the two groups are not known to be equal (so that a student's t-test is not applicable). For two groups of sizes n_1 and n_2, the welch t-statistic is given by

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t_\upsilon$$

where $\bar{X}_i, s_i^2; i = 1,2$ are the group mean and group variance for two groups; $n_i; i = 1,2$ is the size of the i^{th} group [11].

The degree of freedom υ which is generally a non-integer, is obtained as a solution of Welch Satterthwaite equation given by

$$\upsilon = \left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2 \left(\frac{s_1^4}{n_1^2(n_1 - 1)} + \frac{s_2^2}{n_2^2(n_2 - 1)}\right)^{-1}$$

Multiple Linear Regression (MLR) Model

A general linear regression model for regressing a response variable (Y) on a set of independent predictors $(X_1, X_2, ..., X_k)$ is given by

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k + \varepsilon$$

where β_0 is the intercept term, $\beta_1, \beta_2, ..., \beta_k$ are coefficients of regression to be estimated from the data and ε is the error term. For this study, the model takes the form

$$Wellness\text{-}score = \beta_0 + \beta_1 Academic\text{-}score + \beta_2 Social\text{-}score + \beta_3 Financial\text{-}score + \varepsilon$$

Prior to fitting of the model, the assumptions of the model were tested. The estimating efficiency of the model was obtained as a ratio of correct estimations to the total observations.

Machine Learning (ML) Model

A regression model through machine learning (using R) was used for varying train data sizes several times. Each time, the frequencies of the three categories of wellness score was predicted for test data and efficiency of the model was computed.

3. RESULTS

A total of 296 college students participated in the study, out of which 273 were undergraduate students and 23 were postgraduates. There were 229 female students (77.4 %), and 67 (22.6%) male respondents. Proportion was those residing in hostels was almost 65%.

Participants responded to each item based on their personal experiences using a five-point Likert scale ranging from 1 (Never) to 5 (Always), except for two questions in the "Financial Stability" section, which were scored 1-3. To ensure scoring consistency, positively worded items were reverse-coded. Higher cumulative scores indicated higher levels of perceived stress. Participants were given the option to skip up to two questions if they felt uncomfortable answering them. The data reliability was tested using Guttman's Lambda (0.933) and Cronbach's alpha (0.911).

For all the four score, descriptive statistics were obtained (Table 2). The central values were close to the 'Normal' and 'Moderate' categories of scores and were almost similar.

Table 2: Descriptive statistics of four scores: Academic, Financial, Social and Wellness scores

Statistic	Academic Score	Financial Score	Social Score	Wellness Score
Mean	29.07	12.43	23.73	26.76
Median	30.00	13.00	24.00	27.00
S.D.	6.58	3.50	6.73	7.49
Range	32.00	18.00	35.00	37.00

The total wellness scores were categorized into three classes: (i) mild or normal (denoted by 0) for score range < 23; (ii) moderate (denoted by 1) for score range between 24-37; and (iii) high or poor (denoted by 2) for score range >38. High scores correspond to high levels of stress.

3.1 Multiple Linear Regression (MLR) Model

One of the objectives of the study was to examine if the specific stresses (viz. Social, Academic and Financial) affect the over all psychological well-being of students and to what extent. For this it was decided to fit a MLR model to the wellness score with Academic score, Social score and Financial score as independent predictors.

To decide about the covariates for the MLR model viz. Impact of the total years of course and Gender was examined graphically and through statistical tests. First, the wellness score was categorized using "year of course" as the grouping variable (Figure 1).

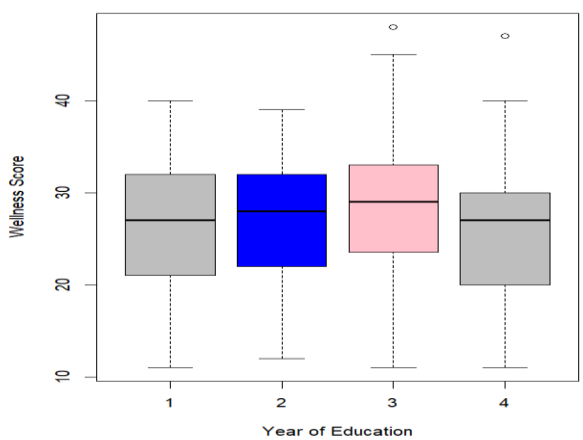

Figure 1: Comparison of wellness scores of students by the year of education

The p-value of the comparison statistic was 0.189 (F(3,292)= 1.602) thus accepting the null hypothesis that the wellness score is not affected by the years of course.

The box-plot of wellness score categorized by gender again revealed no impact of gender. (Figure 2):

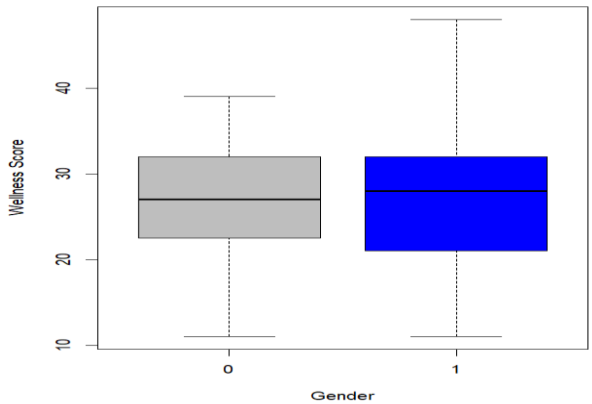

Figure 2 : Comparison of wellness score by gender

Welch t-test for difference of means for the two gender groups yielded a p-value of 0.362 (t113.76 = .91529). Neither year of course nor gender can be taken as a covariate in determining the extent of relationship of wellness scores on the other three scores.

The assumptions of the MLR model were examined before fitting the model:

Normality of the response variable- The skewness of the wellness score was -0.17784 (Close to 0) and kurtosis 2.6635 (close to 3). The Kolmogorov- Smirnov test with D-statistic 0.067 and p-value 0.092 (>0.05) confirmed the normality of the score. A Q-Q plot of the 'Wellness score' corroborated the result (Figure 3):

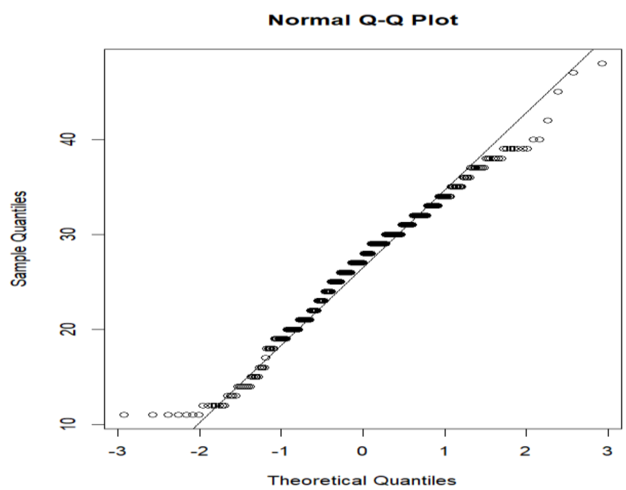

Figure 3: Q-Q plot to test normality of wellness score

(i) Linear association of response variable with independent predictors- The correlations of Wellness scores (along with their significance) with predictors established the linear relationships between the response variable and the predictors (Table 2):

Table 2: Linear association between the response variable and the predictors and VIF of predictors

	Academic Score	Financial Score	Social Score	Wellness Score	VIF
Academic Score	1.0000				1.3938

Financial Score	0.4331	1.0000			1.4004
Social Score	0.4789	0.4827	1.0000		1.4762
Wellness Score	0.5841 (<0.001)	0.4849 (<0.001)	0.6354 (<0.001)	1.0000	

(ii) Multicollinearity- The impact of multi-collinearity is not significant as the correlations among the predictors are not very high (<0.5). Also, VIFs corresponding to the three predictors were all less than 2 (Table 2).

(iii) Heteroscedasticity- Direct tests for heteroscedasticity are all based on residuals obtained after fitting the regression model. However, absence of outliers in response variables is an indication of variance within control limits and hance absence of heteroscedasticity. The box plot of standardized wellness scores found all standardised scores within limits (Figure 4):

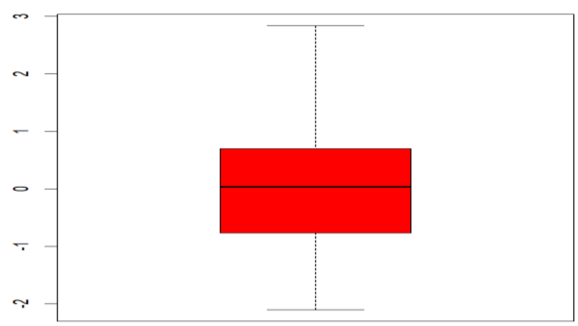

Figure 4: Box plot of standardised wellness scores showing no outliers

(iv) Autocorrelation- Autocorrelation was not an issue in this data as it was a cross-sectional data collected at one point of time.

With all assumptions verified, an MLR model was fitted to the data. The Table 3 below presents the results of the MLR model (with the (psychological) wellness score as response variable and the Academic, Financial, and Social scores as independent predictors).

Table 3: Multiple Linear Regression model with wellness score as the response variable

Coefficients	**Estimates**	**Standard error**	**t-value**	**Pr(>|t|)**
Intercept	1.2921	1.5147	0.853	0.3943
Academic score	0.36965	0.05448	6.786	<0.001
Financial Score	0.3142	0.1025	3.064	0.002
Social score	0.4558	0.0549	8.310	<0.001
Residual standard error (292 degrees of freedom)				5.215
Multiple R^2	0.5207		**Adjusted R^2**	0.5158
$F_{(3,292)}$	105.7		**p-value**	<0.001

Multiple R2 of 0.5207 indicated 52% of variability of the wellness score due to the three significant predictors (p-value < 0.001) which suggests that there might be some other factors such as environmental factors, physical health etc. contributing to the variability in the wellness score.

The plots of the model corroborated the good fit of the model. (i) The residuals were randomly scattered around 0 (Top left) and (ii) were normally distributed (Top right). (iii) The scale location graph (Bottom left) ruled out presence of heteroscedasticity and (iv) the residual-leverage graph indicated at no significant outliers (Bottom right) (Figure 5):

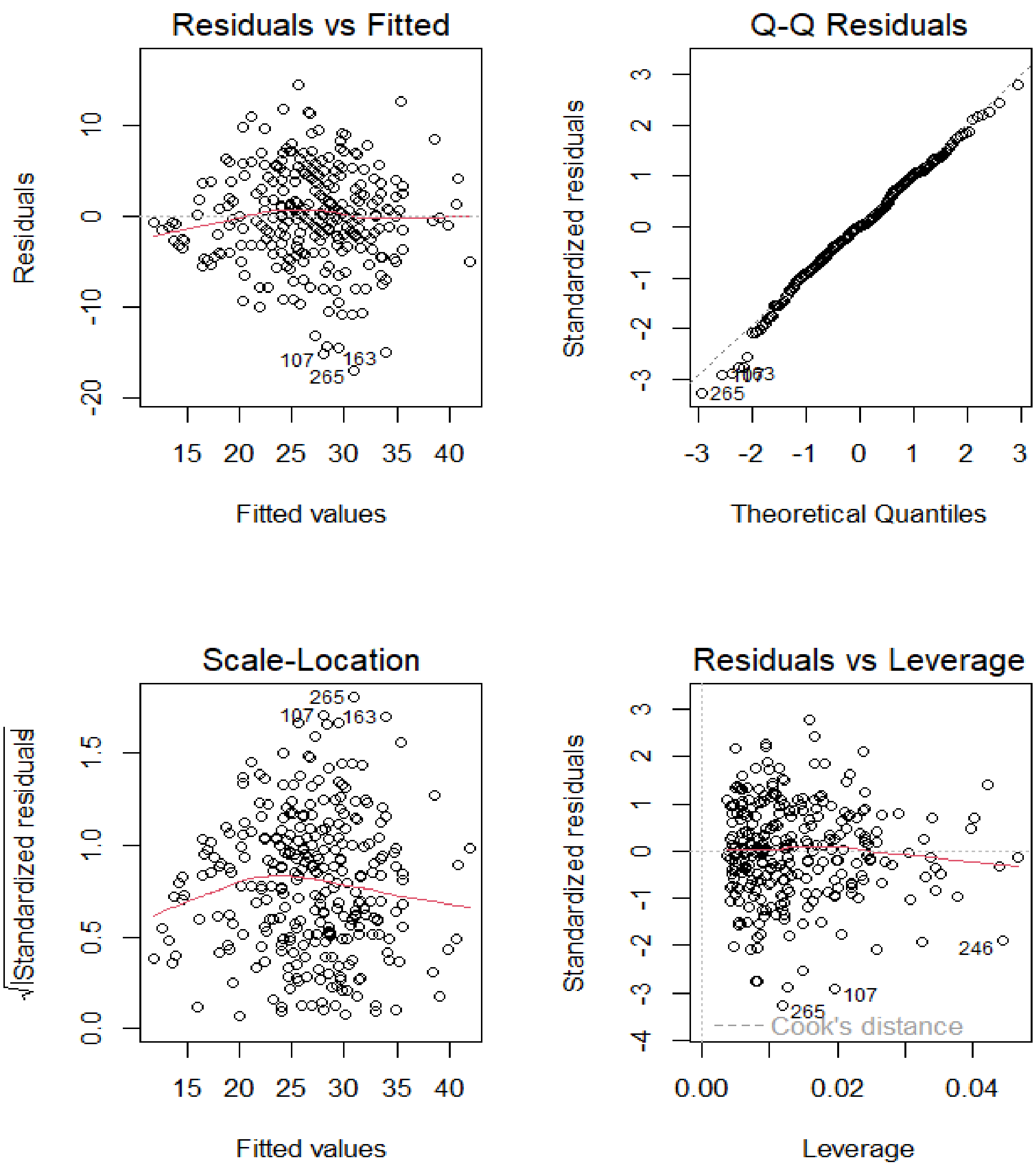

Figure 5: Model plots of the classical regression model with Wellness score as the response variable

Finally, the model was used to estimate the frequencies of the three classes of stress: 0 (Normal),1(Moderate), 2 (High). Efficiency of the model was 82.77%. The results are tabulated below (Table 4):

Table 4: Estimated frequencies of the three categories 0,1,2, through MLR

	Estimated categories		
	0	1	2

Observed categories	0	45	9	0
	1	17	194	1
	2	0	24	6

- Categories: 0-Normal/Mild, 1- Moderate, 2- Higher/Poor

3.2 Machine Learning Model

Next, the regression models through machine learning with varying train data sizes were obtained. For each sample, predicted frequency for each class (of the test data) and the model prediction efficiency (after combining the results of train and the test data) was calculated. The results are presented in Table 5 below:

Table 5: Predicted frequencies and prediction efficiencies for various train data sizes for machine learning model

Train data size	Observed categories	Estimated categories			Predictive Efficiency (%)
		0	1	2	
50%	0	45	9	0	84.80
	1	11	200	1	
	2	0	24	6	
50%	0	43	11	0	80.41
	1	21	189	2	
	2	0	24	6	
50%	0	40	14	0	82.43
	1	13	198	1	
	2	0	24	6	
60%	0	41	13	0	80.74
	1	18	192	2	
	2	0	24	6	
65%	0	48	6	0	83.44
	1	18	193	1	
	2	0	24	6	
70%	0	45	9	0	84.12
	1	13	198	1	
	2	0	24	6	
75%	0	45	9	0	83.44
	1	15	196	1	
	2	0	24	6	
80%	0	48	6	0	83.44
	1	18	193	1	
	2	0	24	6	
85%	0	49	5	0	83.78
	1	18	193	1	
	2	0	24	6	
87%	0	45	9	0	83.19
	1	16	195	1	
	2	0	24	6	
90%	0	45	9	0	82.77
	1	17	194	1	
	2	0	24	6	
93%	0	45	9	0	83.11
	1	16	195	1	
	2	0	24	6	
95%	0	45	9	0	82.77

	1	17	194	1	
	2	0	24	6	

Except for the case when train data size was 60%, the machine learning model had better prediction than the estimation by the full data model. Maximum efficiency was when the train data size was 70%, which also is the data size generally recommended in machine learning models. The models' predictive efficiencies have been plotted for various train data sizes (Figure 6).

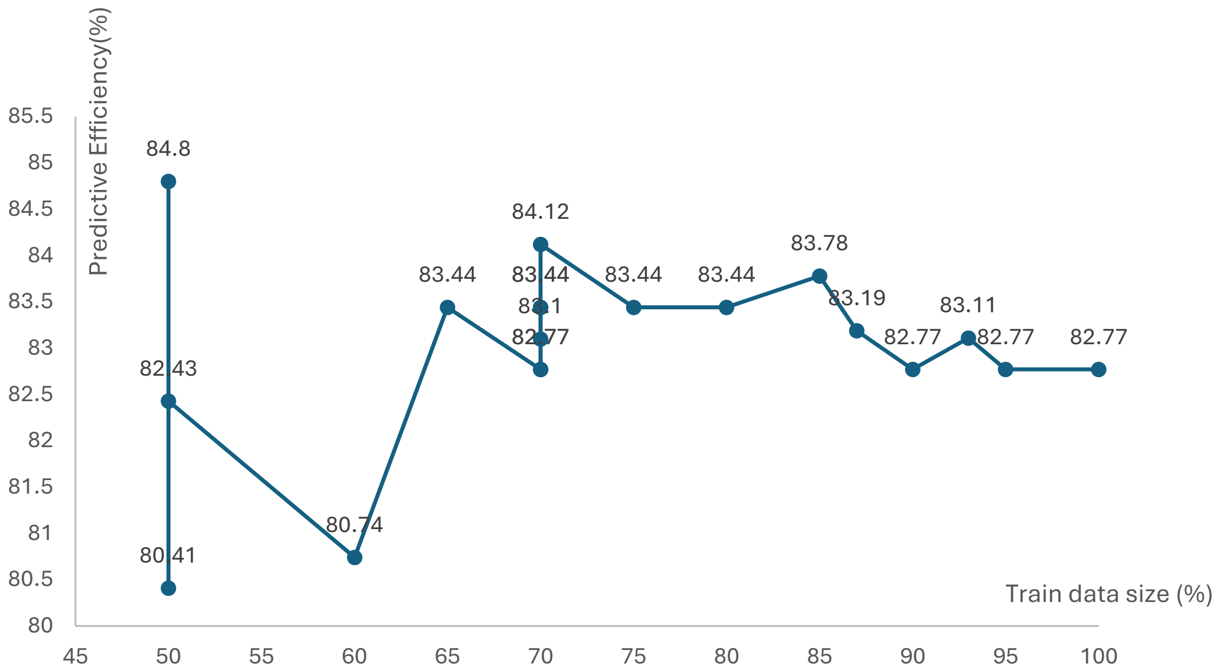

Figure 6: The predictive efficiencies of the machine learning model with different train data sizes

A generally accepted train data size is 70% of the total data. For this data size, the model was run five more times and every time, the ML-MLR model efficiency was al least as good as the MLR model efficiency (Table 6).

Table 6: Predictive efficiencies of the machine learning model with train data size 70%

Train data size70%	**Observed categories**	**Estimated categories**			**Predictive Efficiency (%)**
		0	1	2	
1	0	45	9	0	84.12
	1	13	198	1	
	2	0	24	6	
2	0	45	9	0	82.77
	1	17	194	1	
	2	0	24	6	
3	0	45	9	0	83.44
	1	15	196.	1	
	2	0	24	6	
4	0	45	9	0	83.44
	1	15	196.	1	
	2	0	24	6	
5	0	48	6	0	83.10
	1	19	192	1	
	2	0	24	6	
6	0	44	10	0	82.77

	1	16	195	1	
	2	0	24	6	

4. DISCUSSION

Psychological health issues in society and particularly among young population are increasing. One of the possible reasons is stress from various quarters of life which confound and manifest in psychological health issues [7,10]. In particular, the college going students have to cope with increasing academic expectations, social adaptations, financial challenges, future uncertainties, and changing personal identities throughout this crucial transitional phase [1]. The problem has worsened post Covid-19 period. The psychological health issues among younger population is a never-ending vicious loop. While academic, social and financial stresses affect the total wellness score, a poor psychological health results in poor academic growth, reduced financial stability and higher social stress [1,12,17].

The importance of focusing on student well-being is a concern for researchers, academicians, parents and policymakers. The concern for a sound psychological health has been highlighted in India's National Education Policy (NEP) 2020 which clearly states that education should not only focus on marks but on a more holistic approach encompassing mental health, emotional strength, and life skills.

The data for this study was collected through an online/offline survey conducted among the students of colleges in Delhi, India. One of the objectives of the study was to estimate the overall psychological health (through wellness score) in terms of three major stress factors viz. Academic, Financial and Social stress among students. A modified questionnaire with components of the standard psychological health tool (Yikealo et al [27]) along with some demographic and financial questions was used for data collection.

In their previous studies on the psychological health of young adults during Covid-19 period, the authors of the study found out the impact of the present year of education (a proxy for total years of education before they complete their present degree) on the psychological health but no impact of gender. [4]. However, in the present study there was no impact of either the years of education (as measured through the "Year of course") or gender. This may an indication of high stress levels across total years of education and gender.

An MLR approach has widely been used in bio-medical data when the aim is to model a target variable in terms of independent predictors [5]. In order to meet the objective of the study, a multiple linear regression model was fitted on the given data after duly testing all the assumptions and it yielded

The highest coefficient is of Social score, indicating that while Academic and Financial stresses affect Wellness score significantly, it is the Social score that affects them the most. This may be on account of parental pressure, peer pressure, future uncertainties, society around them etc. It has been observed that stress level of students in unstable societies is higher than that in stable societies [8,12]. The multiple correlation coefficient of value 0.5207 (Adj R2=0.5158) is an indicator of a good relationship between the predictors and the response variable as the data is cross-sectional. Further there may be other stress factors which have not been included in the study such as physical health, family structure and support, parental health [3, 22] etc.

A machine learning model is one which picks up a certain proportion of the data randomly (called the train data), fits the model on this data and validates the model on the remaining data (called the test data). Machine Learning (ML) models learn from the structure of the data and improves itself if needed. Larger is the train data size, better is the fitted model. However, there is always a trade-off between the train data size and test data size. A model with larger (train) data may be an overfitted model with limited prediction power and hence may not work well with the test data. On the other hand, a smaller train data may result in underfitting of model and hence may not be able to pick the effect of some of the independent predictors. Further there is no thumb rule guiding the size of train or the test data. As such, the experimenter may have to run the ML model several times before reaching at an optimal train data size. Figure 6 is pictorial empirical evidence of the ML-MLR algorithm. The model was run several times with varying train data sizes. At 50%, it was run thrice and yielded the highest (84.8%) and the lowest efficiencies (80.41%) of the test runs. The variation in the efficiencies was quite high. As the train data size increased, the efficiencies improved up to size 70%. Beyond this size, either the efficiency did not improve or declined. The reason for decline is that the overall model efficiency is derived from efficiency of the model on train data as well as from test data. As such an optimal train data size is needed which should leave enough observations in the test data also. Table 6 demonstrates that the ML model with 70% train data has efficiency at least as that of the full model. Further the efficiencies at 70% depict lower variations (from 82.77% to 84.12% (Figure 7)) [22].

Figure 7: The predictive efficiencies of the machine learning model with train data size 70%

A good quality data is the backbone of any statistical analysis. Equally essential is the choice of a good statistical model for extracting the story hidden in data. A MLR model is good in extracting the effect of independent predictors on the overall psychological health of students. However, fragmenting the data appropriately achieves the deeper insight into the data. A machine learning model is able to improve the correctly predicted categories substantially, without compromising the essential characteristics of the data.

The results of the study can be used to identify rising stress levels among the young adults in educational institutions. However, the results may not be applicable to youth of the same age group who are not a part of any educational institution as their stress markers may be different. This may be considered as a limitation of the study.

References

1. Abouserie, R. (1994). Sources and Levels of Stress in Relation to Locus of Control and Self Esteem in University Students. Educational Psychology, 14(3), 323–330. https://doi.org/10.1080/0144341940140306
2. American Institute of Stress (2012). https://www.stress.org/what-is-stress/
3. Goyal, B., Sabharwal, A. & Dhingra, N.A. (2021). Impact of Parental Psychiatric problems on adjustment behaviour of Adolescents: A Study through adjustment Inventory of school students. International Journal of Scientific Research, 10(1). 1-04. Doi:10.36106/ijsr/4137619.
4. Goyal, B., Sabharwal, A., Chauhan, V. & Joshi, L. M. (2023). Psychological health of Indian youth during COVID-19: a study through three chronological surveys. International Journal of Public Health Science (IJPHS), 12(2). 752-763. 10.11591/ijphs.v12i2.22420.
5. Gurprit, G., Gadpayle. A.K., & Sabharwal, A. (2012). Identifying patients with diabetic nephropathy based on serum creatinine in the presence of covariates in type-2 diabetes: A retrospective study. Biomed Res- India, 23 (4): 615-624. https://www.researchgate.net/publication/286531839_Identifying_patients_with_diabetic_nephropathy_based_on_serum_creatinine_in_the_presence_of_covariates_in_type-2_diabetes_A_retrospective_study.
6. Kathirvel, N. (2022). Post COVID-19 pandemic mental health challenges. Asian J Psychiatr.,53:102430. doi: 10.1016/j.ajp.2020.102430.
7. Kohn, J. P., & Frazer, G. H. (1986). An Academic Stress Scale: Identification and Rated Importance of Academic Stressors. Psychological Reports, 59(2), 415-426. https://doi.org/10.2466/pr0.1986.59.2.415
8. Lazarus, R.S. (1993). From psychological stress to the emotions: a history of changing outlooks. Annual Review of Psychology,44,1–21. https://doi.org/10.1146/annurev.ps.44.020193.000245
9. Levine, G.N., Cohen, B.E., Mensah, Y.C., Fleury, J., Huffman, J. C., Khalid, U., Labarthe, D.R., Lavretsky, H., Michos, E.D., Spatz, E.S., & Kubzansky, L . D. (2021). Psychological Health, Well-Being, and the Mind-Heart-Body Connection: A Scientific Statement From the American Heart Association Circulation, 143(10):e763-e783. https://doi.org/10.1161/CIR.0000000000000947
10. Mofatteh, M. (2025). Risk factors associated with stress, anxiety, and depression among university undergraduate students. AIMS Public Health. 2020 Dec 25;8(1):36-65. doi: 10.3934/publichealth.2021004. PMID: 33575406; PMCID: PMC7870388.
11. Nor Aishah Ahad, Sharipah Soaad Syed Yahaya; Sensitivity analysis of Welch's t-test. AIP Conf. Proc. 10 July 2014; 1605 (1): 888–893. https://doi.org/10.1063/1.4887707
12. Reddy, K. J., Menon, K.R., & Thattil, A. (2018). Academic Stress and its Sources Among University Students. Biomedical & Pharmacology Journal, 11(1), 531–537. DOI: 10.13005/bpj/1404
13. Rodríguez, L.S., & Feig, L. A. (2013). Chronic social instability induces anxiety and defective social interactions across generations. Biol Psychiatry, 73(1), 44-53. DOI: 10.1016/j.biopsych.2012.06.035
14. Ross, S. E. B., Niebling, C., & Heckert, T. M. (1999). Sources of stress among college students. College Students, 33, 312-318. https://www.scribd.com/document/210195964/.
15. Sabharwal, A., Goyal, B. & Joshi, L.M. (2023). An application of ordinal regression to extract social dysfunction levels through behavioral problems. AIMS Public Health, 10(3): 577-592. doi: 10.3934/publichealth.2023041.
16. Sagar, R., Chatterjee, K., Thareja, S., Timothy, A., Yadav, A. S., Yadav, P., et al. (2025). NAMS task force report on mental stress. Ann Natl Acad Med Sci.; 61(1):66-97. doi: 10.25259/ANAMS_TFR_10_2024.
17. Sax, L. J. (1997). Health Trends among College Freshmen. Journal of American College Health, 45(6), 252–264. https://doi.org/10.1080/07448481.1997.9936895
18. Scheepers, D., & Ellemers, N. (2018). Stress and the stability of social systems: A review of neurophysiological research. European Review of Social Psychology, 29(1), 340–376. https://doi.org/10.1080/10463283.2018.1543149).
19. Selk, Leonie & Gertheiss, Jan. (2022). Nonparametric regression and classification with functional, categorical, and mixed covariates. Advances in Data Analysis and Classification. 17. 519-543. 10.1007/s11634-022-00513-7.
20. Semwal, J., Bahuguna, A., Uniyal, A. & Vyas, S.(2023). A Study to Analyse Covid-19 Outbreak Using Multiple Linear Regression: A Supervised Machine Learning Approach. Natl J Community Med.; 14(2):82-89. DOI: 10.55489/njcm.140220232656. https://njcmindia.com/index.php/file/article/view/2656
21. Shakil, M., Ahsanullah, M., Kibria, B.M.G., Singh, J.N., Jabeen, R., Khadim, A. & Sirajo, M. (2024). A Multiple Regression Model for Identifying some Risk Factors affecting the Cardiovascular Health issues in Adults. Jnanabha, 53(2). 113-125. DOI:10.58250/jnanabha.2023.53214

22. Sivakumar, M., Parthasarathy, S., & Padmapriya T. (2024). Trade-off between training and testing ratio in machine learning for medical image processing. PeerJ Comput. Sci. 10:e2245 DOI 10.7717/peerj-cs.2245.
23. Unni, K.E.S., Srivastav, R., Sabharwal, A., & Goyal., B. (2019). Assessing the effect of Parental Psychiatric disorder on Adolescents using Brief Psychiatric Rating Scale-A Case Study, International Journal of Scientific research, 8(7), 41-43. https://indianjournals.com/article/ijrss-9-7-032.
24. World Health Organizatation. (2022). https://www.who.int/news/item/02-03-2022-covid-19-pandemic-triggers-25-increase-in-prevalence-of-anxiety-and-depression-worldwide.
25. World Health Organization. (2025). https://www.who.int/news-room/fact sheets/detail/adolescent-mental-health.
26. Yang, C., Chen, A., & Chen, Y. (2021). College students' stress and health in the COVID-19 pandemic: The role of academic workload, separation from school, and fears of contagion. PLOS ONE, 16(2), e0246676. https://doi.org/10.1371/journal.pone.0246676
27. Yikealo, D., Yemane, B. & Karvinen, I. (2018). The Level of Academic and Environmental Stress among College Students: A Case in the College of Education. Open Journal of Social Sciences, 6(11), 40-57. doi: 10.4236/jss.2018.611004.