

Stress Level Vulnerability Test of University Student's using Machine Learning Approaches

Joyassree Sen

Department of Computer Science and Engineering
Islamic University, Bangladesh

Bipul Roy

Department of Geography and Environment
Islamic University, Bangladesh

Paresh Chandra Barman

Department of Information and Communication Technology
Islamic University, Bangladesh

ABSTRACT

The stress level of mental health problems of the university students have become a major concern Bangladesh. Mental health services for the university students in Bangladesh are critically inadequate, posing a significant challenge to the well-being and academic success of students. The societal stigma and limited access to psychological services worsen the situation. Mental health issues among the students in Bangladesh are a pressing concern, influenced by academic pressure, financial constraints, and societal expectations. Students who do not receive the necessary support may face ongoing mental health challenges that affect their future careers and personal lives. Machine learning (ML) system of mental health vulnerability testing may be helpful to the inadequacy of mental health services. To develop a ML based mental health vulnerability testing support system, the risk factor analysis is a significant and challenging issues. In these work we proposed Mutual Information (MI) and Non-negative Matrix Factorization (NMF) algorithm to identify the significant factors and then stress level vulnerability has been tested using artificial neural network.

In these work at first we collected a database using the well-known DASS-42 (Depression, Anxiety and Stress Scale) [1] questioner along with some campus life factors of the university students. The simple correlation measure shows that the factors are almost equally significant to the stress level, but our proposed MI and NMF algorithms estimate significant ranks of the factors. These ranks are used to reduce the number of factors without affecting the vulnerability detection rate. In the work we have also observed the stress level vulnerability test performance by adding some campus life significant factors of the students and the experimental results show that the stress level vulnerability detection rate improves from 87% to 92%.

Keywords: Mental Health, Machine Learning, Feature selection, Mutual Information, Non-Negative Matrix Factorization (NMF), ANN.

1. INTRODUCTION

Cultural beliefs and a lack of awareness contribute to the stigmatization of mental health, preventing many individuals from seeking help. Furthermore, mental health services in Bangladesh are severely underfunded, with only a small portion of the national health budget allocated to these services. This results in a critical shortage of mental health treatment options.

According to the report of a social and voluntary organization of Bangladesh known as "Aachol Fundation" that has been collected the reports of suicide cases of Bangladeshi Students from different newspapers. They reported that 513 students took their own lives in 2023; in 2022 the number was 532; and in 2021 the number was 528. Among those suicide cases, 98 university students took their own lives in 2023, in 2022 the number was 86 and in 2021 the number was 101. Most of those cases the mental health condition was a major issue.

In study [2] it has been shown that the mental health factors can be influenced by biological, psychological, and environmental issues. Some symptoms and causes are similar to each other, which complicates and makes the accurate diagnosis of mental health problems difficult. As a result, doctors may misdiagnose the issues, and patients may receive incorrect treatment. This situation can make the condition of the mental patient more dangerous, threatening their emotional and behavioral functionality. According to the National Health and Morbidity Survey (NHMS) 2017, depression affects one in five Malaysians. Mental stress affects one in ten persons, while anxiety affects one in two. One group that is particularly vulnerable to mental health issues is students in higher education.

2. RELATED WORKS

In study [3] it has been observed that the incidence of suicide thoughts and behaviors among college students varies from 1.3% to 32.7% worldwide. Negative life experiences, mental illnesses, and prior suicide attempts are the main reasons for suicidal ideation. However, when it came to suicidal ideation, residential location and gender were the least significant determinants. The results of study [4] show that the choice of major and gender significantly impact a student's well-being, and Decision Trees perform better than KNN and SVM in predicting mental health issues. The study [5] identified characteristics linked to suicidal thoughts and behaviors (STBs) in college students by administering a comprehensive set of demographic and clinically relevant self-report measures during the first semester of college and at the conclusion of the year. For the early identification of mental illness, the study [6] involved the collection of data from 30 patients in the southwest region of Nigeria using structured questionnaires with four main sections: 12 questions about demographic information, 6 questions about biological factors, 5 questions about psychological factors, and 5 questions about environmental factors. In study [7] used different machine learning techniques to forecast the incidence of stress among university students in Bangladesh. The data has been collected by asking questions about their anthropometric measures, academic performance, lifestyles, and health-related data. Decision trees, random forests, logistic regression, and support vector machines were used to construct prediction models. According to the findings, Students' pulse rate, systolic and diastolic blood pressures, sleep status, smoking status, and academic background were selected as the important features for predicting the prevalence of stress. The paper has a number of difficulties. Since students were asked to answer a binary response question (Yes or No) and some data are physiological.

The mathematical models or machine learning approach try to give a coherent view of complex socio-economical, environmental affected person. In this way, the theory and experimental techniques improve in a cooperative manner. A machine learning approach to building early and automated mental health condition classification rules can work efficiently. Given a relatively small set of mutually classified training self-reported survey data, the problem of learning mental health condition rules can be cast as a supervised learning problem. In learning problem feature/factors selection or/and extraction is an important issue. In this work we have been focused on identifying the significant factors that affects the stress of mental health condition of self-reported behavioral or feelings survey data. The self-reported data are non-negative. In the study [8] and [9] it has been shown that for non-negative date Mutual Information (MI) and Non-negative Matrix Factorization (NMF) performs better for significant feature selection. We proposed MI and NMF to select the significant factors for stress level of mental health condition and finally test the performance by single layer perceptron.

3. PROPOSED METHODOLOGY

For significant feature selection of stress level mental health condition we proposed Mutual Information (MI) and Non-negative Matrix factorization algorithm and finally test the vulnerability by simple single layer perceptron. One important characteristics of our survey data is non-negativity. In our previous work [8] and [9] it has been shown that these methods works well to select significant factors for non-negative data.

3.1 Mutual Information

The Mutual Information [9] $I(C, q_i)$ between category C and survey question or risk factor q_i , that is used to rank the factors for classification problem. It is one of the most popular methods for test feature selection. By definition the rank measures the factor or question goodness globally with respect to all categories on average, which is defined as:

$$I(C, q_f) = \sum_{k=1}^{|C|} P(c_k, q_f) \log_2 \frac{P(c_k, q_f)}{P(c_k)P(q_f)} \quad 3.1$$

For continuous random variable [9] the simplified form of the mutual information estimation of a certain factor q_f according to classes vs. bin frequency table could be represented as follows:

$$I(q_f, C) = \sum_{k=1}^K \sum_{j=1}^J \frac{n_{kj}}{N} \log_2 \frac{\frac{n_{kj}}{N}}{\frac{n_k}{N} \cdot \frac{n_j}{N}} = \frac{1}{N} \sum_{k=1}^K \sum_{j=1}^J n_{kj} \log_2 \frac{N \cdot n_{kj}}{n_k \cdot n_j} \quad 3.2$$

where $p(q_f) \approx \frac{n_{\bullet j}}{N} = \frac{n_j}{N}$, where n_j is the number of participants with feature q_f in the j^{th} bin. N is the total number of participants. Similarly, if we have n_k participants for k^{th} category, the class prior probabilities can be calculated from the data sample as: $p(c_k) = \frac{n_{k\bullet}}{N} = \frac{n_k}{N}$; $N = \sum_{k=1}^K n_k$, the joint probability $P(q_f, c_k) \approx \frac{n_{kj}}{N}$; n_{kj} is the number of participants with factor q_f , of k^{th} class in j^{th} bin.

3.2 Non-negative Matrix Factorization (NMF)

Non-negative Matrix Factorization [10] [11] is a useful decomposition method for multivariate data. The NMF algorithm is able to learn parts of faces and semantic features of dataset. This is in contrast to other methods, such as principal components analysis and vector quantization, that learn holistic, not parts-based, representations. Non-negative matrix factorization is distinguished from the other methods by its use of non-negativity constraints. These constraints lead to a parts-based representation because they allow only additive, not subtractive, combinations. When non-negative matrix factorization is implemented as a neural network, parts-based representations emerge by virtue of two properties: the firing rates of neurons are never negative and synaptic strengths do not change sign. The NMF algorithm is as follows:

For a given non-negative $n \times m$ matrix V ; find non-negative factors, W , of $n \times r$ matrix, and H , $r \times m$, such that: $V \approx WH$ or

$$V_{ij} = (WH)_{ij} = \sum_{k=1}^r w_{ik} h_{kj} \quad 3.3$$

Where r is chosen as $r < rm/(n+m)$. For stress factor extraction V be the response factors matrix, $i=1, 2, \dots, n$ the dimension of factors and $j=1, 2, \dots, m$ numbers of total response participants. W is basis matrix where each column represents one basis vector; and H is the encoded matrix which is one-to-one correspondence with a participants of V . Based on KL-divergence minimization the update rules of W and H are defined as

$$h_{aj} = h_{aj} Q_D \left(\left[\mathbf{W}^T \right]_{ia}, \frac{x_{ij}}{\left[\mathbf{W} \mathbf{H} \right]_{ij}} \right) ; \quad w_{ia} = w_{ia} Q_D \left(\frac{x_{ij}}{\left[\mathbf{W} \mathbf{H} \right]_{ij}}, \left[\mathbf{H}^T \right]_{aj} \right) ; \quad 3.4$$

$$w_{ia} = \frac{w_{ia}}{\sum_i w_{ia}}$$

where $Q_D(A, B)_{ij} = \sum_k A_{ik} B_{kj} = AB$, and all the $(.)_{ij}$ indicates that the noted division and multiplications are computed element by element. The significant basis factors have been selected from matrix H .

3.3 Single Layer Perceptron (SLP)

The selected significant features using MI and NMF has been considered as the input of the SLP layer, and its synaptic weight \mathbf{U} can be trained based on the gradient descent error minimization rule as

$$u_{ij} \leftarrow u_{ij} + \text{delta} \cdot x' \text{ where } \text{delta} = \eta(o_d - o)(1 - o)^2 \quad 3.5$$

O_d is the desired output and O is actual output of the SLP layer. The basic model of vulnerability test X-SLP network is shown in Fig. 1, where X represents for selected feature by using MI score or NMF algorithm.

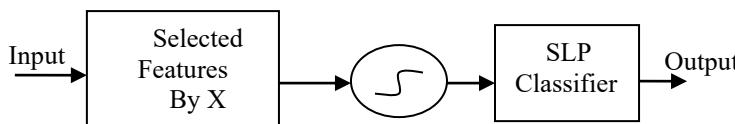


Fig. 1 : A simple model of X-SLP vulnerability test network

3.4 Data set and preprocessing

At the beginning of this work, we have discussed the issue with relevant personals such as a psychologist who provides mental health services in Bangladesh University of Engineering and Technology (BUET), Faculties of the Department of Psychiatry, Faridpur Government Medical Collage, Bangladesh and the Faculties of the Department of Psychology, University of Rajshahi, Bangladesh. From those discussions we realize that, there are lot of inadequacy in mental health support service systems in tertiary level educational organization in Bangladesh, in the early stage of mental health illness students are not willing to share with others, there are some statistical analysis of mental health condition based on some survey data, those survey questioner are not unique and standard i.e., the risk factors of the survey are biased on the researcher view point, finally if the question's number of a questioner is large participants become impatience to response properly. Based on above expertise opinions- we have collected data based on most popular self-reported DASS-42 (Depression, Anxiety and Stress Scale) [1] questionnaire. We use 14 factors

for stress level along with propose 15 campus life factors of the university students. With proper concern and getting the permission of the head of different departments in Islamic University, Bangladesh, we have collected the data. Out of 800 participants 578 participants have been completely filled-up and 222 participants partially filled-up the data. The collected data has been level as vulnerable or not vulnerable based on DASS stress score.

We randomly select 70% responses for training and 30% for testing.

4. RESULT AND DISCUSSION

At first we have observed the correlation of the risk factors with stress level vulnerability of the mental health condition. The result has been shown in figure 4.1, where all factors are positively correlated with stress level and the values are almost similar, i.e., all factors have similar impact on stress vulnerability.

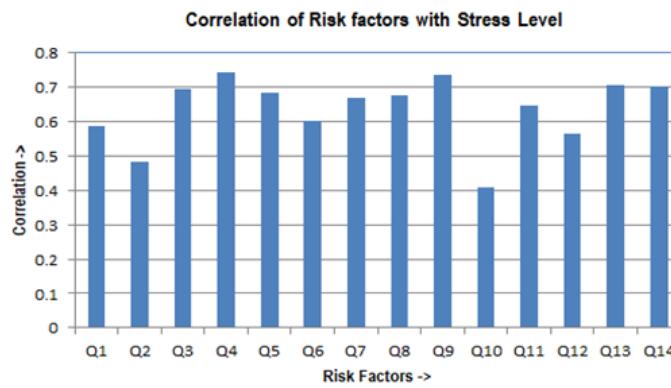


Fig. 4.1: Correlation among Stress level and risk factors of DASS questioner

4.1 Simulation Result significant factor selection by MI and NMF

Based on the proposed formula of mutual information (MI) we estimate the MI score of each DASS stress factors and proposed campus life factors corresponding to the level of stress vulnerability as shown in figure 4.2 and 4.3 respectively.

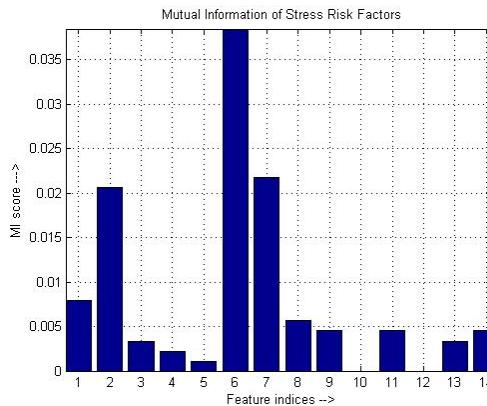


Figure 4.2: Mutual information of DASS risk factors

It has been observed that the most significant risk factors for stress vulnerability are no. 6, 7, 2, 1, 8, 9 and so on. From 4.3 it has been observed that the campus life significant factors for stress vulnerability are 15, 13, 10, 14 and so on. Similarly we estimated the NMF score of DASS factors and proposed campus life factors with three basis vectors. The result showed in figure 4.4 and 4.5 respectively.

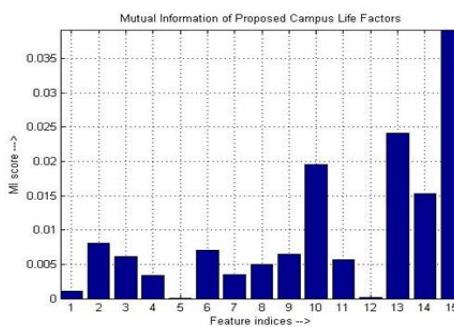


Figure 4.3: Mutual information of campus life risk factors

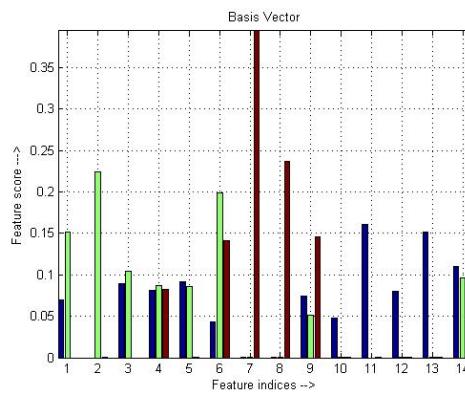


Figure 4.4: MNF score of DASS risk factors

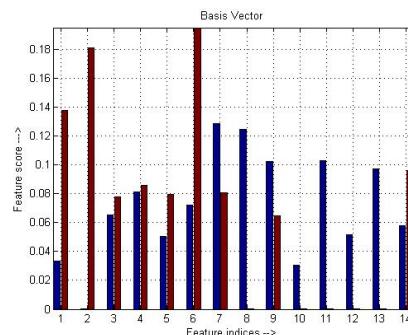


Figure 4.5: MNF score of campus life risk factors

Based on the high rank of MI score and NMF score different numbers of risk factor have been selected as input neurons of the SLP to test the stress level vulnerability. The result has been shown in table 4.1 and in figure 4.6 and 4.7 respectively.

Table 4.1: Stress vulnerability Detection Rate is in %.

No. of selected risk factors	Factor selected by MI		Factors selected by NMF	
	Training	Testing	Training	Testing
5	83.43	82	83.43	82
6	84.57	83	84.57	83
7	86.29	84	86.29	84
8	90.57	86	90.57	86
9	90.29	85	90.29	85
10	90.57	86	90.57	86
11	93.71	87	93.71	87
12	93.71	86	93.71	86
13	96.00	86	96.00	86
14	94.86	87	94.86	87

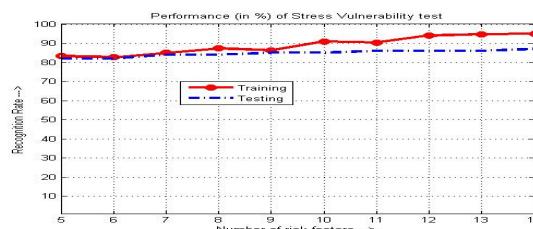


Figure 4.6: Stress Vulnerability DR vs. No. of risk factors ranked by MI

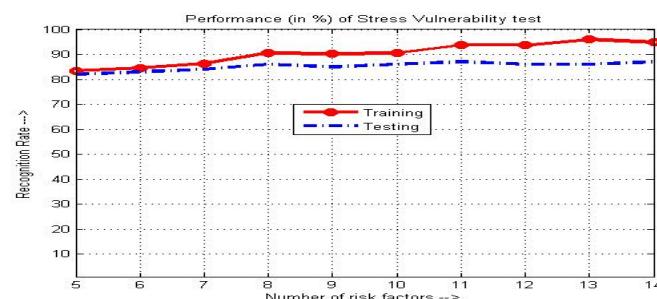


Figure 4.7: Stress Vulnerability DR vs. No. of risk factors ranked by NMF

Finally from dataset few low ranked DASS factors such as factor no. 5, 10 and 12 has been excluded and high ranked factors of proposed factor no. 10, 13 and 15 has been included in the input neuron of SLP model. The vulnerability detection rate has been shown in table 4.2. From this result it has been observed that by including the campus life proposed factors the overall vulnerability detection rate of mental health conditions has been improved.

Table 4.2 Vulnerability Detection for mixed datasets

Risk factors		
	Training	Testing
DASS 14 factors only	93.70	87
Mixed 14 factors	93.70	92

From table 4.2 it has been observed that for mixed dataset the vulnerability detection rate of Stress level improves from 87% to 92%.

5. CONCLUSION AND FUTURE WORKS

For risk factor analysis of survey data which are non-negative and sparse, Mutual Information (MI) and Independent Non-negative Matrix Factorization (INMF) algorithm can be used. These algorithms are useful to reduce the less significant factors without affecting the detection performance. To provide the mental health services to university students along with DASS factors we can used some campus life factors such as [(1) how satisfied are you with the teaching methods of your department? (2) How satisfied are you with your curriculum? (3) How satisfied are you with your academic progress? (4) How satisfied are you with virtual games as a means of relieving fatigue or improving the mind? (5) How concerned about the future (career) after education? (6) How concerned are you about the outcome of your relationship? (7) How many hours do you exercise/exercise? With a reliable mental health database, the machine learning system of mental health condition vulnerability detection can be a supportive system for university authority. This automation system will be helpful to reduce the inadequacy of mental health services in Bangladesh as well as in the world.

One major limitation is to collect reliable mental health dataset. The motivation and awareness should be developed among the students during the survey. In future, with a reliable and large sample size dataset machine learning algorithm for mental health vulnerability detection system should be fine-tuned.

REFERENCES

1. Lovibond, S.H. & Lovibond, P.f. (1995). Manual for the Depression anxiety Stress Scales. (2nd Ed) Sydney: Psychology Foundation.
2. N. S. Mohd Shafiee and S. Mutalib, "Prediction of Mental Health Problems among Higher Education Student Using Machine Learning," *Int. J. Educ. Manag. Eng.*, vol. 10, no. 6, pp. 1–9, Dec. 2020, doi: 10.5815/ijeme.2020.06.01.
3. S. Y. Chan and C. K. Chng, "Analysis of Risk Factors Affecting Suicidal Ideation among Public University Students in Malaysia using Analytic Hierarchy Process," *ASM Sci. J.*, vol. 17, 2022, doi: 10.32802/asmcj.2022.1289.
4. F. Sahlan et al., "Prediction of Mental Health Among University Students Article in International Journal on Perceptive and Cognitive Computing," 2021. [Online]. Available: <https://www.researchgate.net/publication/353480574>
5. N. Kirlic et al., "A machine learning analysis of risk and protective factors of suicidal thoughts and behaviors in college students," *J. Am. Coll. Heal.*, vol. 71, no. 6, pp. 1863–1872, 2023, doi: 10.1080/07448481.2021.1947841.
6. M. Priscilla Dooshima, "A Predictive Model for the Risk of Mental Illness in Nigeria Using Data Mining," *Int. J. Immunol.*, vol. 6, no. 1, p. 5, 2018, doi: 10.11648/j.iji.20180601.12.
7. R. Rois, M. Ray, A. Rahman, and S. K. Roy, "Prevalence and predicting factors of perceived stress among Bangladeshi university students using machine learning algorithms," *J. Heal. Popul. Nutr.*, vol. 40, no. 1, Dec. 2021, doi: 10.1186/s41043-021-00276-5.
8. Joyassree Sen, Bipul Roy and Paresch Chandra Barman, Feature Selection approaches for Non-negative data by modified MI, NMF algorithm, *Journal of Science, Islamic University*, Vol-1, No. 2, 2020; pp. 45- 54
9. Joyassree Sen, Paresch Chandra Barman, Identifying the Significant Genes of Leukemia Data Using NMF, *Journal of Science Islamic University*, Vol-1, No.1, ISSN 1994-0368, 2019, pp. 69-80.
10. D. D. Lee and H. S. Seung: Algorithms for non-negative matrix factorization. In *Advances in Neural Information Processing 13 (Proc. NIPS*2000)*, MIT Press, 2001.
11. D. D. Lee and H. S. Seung: Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(1999):788–791.