Healthcare Transformation with Informatics and Artificial Intelligence
J. Mantas et al. (Eds.)

doi:10.3233/SHTI230430

A Machine Learning Study to Predict Anxiety on Campuses in Lebanon

Madhuri MAHALINGAM[a], Manar JAMMAL[b], Reem HOTEIT[c], Dinah AYNA[d], Maya ROMANI[e], Sahar HIJAZI[f], Imad BOU-HAMAD[g], Christo EL MORR[h1]
[a] *SSN College of Engineering, Chennai, Tamilnadu, India*
[b] *School of Information Technology, York University, Toronto, Canada*
[c] *Clinical Research Institute, American University of Beirut, Beirut, Lebanon*
[d] *Department of Psychiatry, Faculty of medicine, American University of Beirut Medical Center, Beirut, Lebanon*
[e] *Department of Family Medicine, Faculty of Medicine, American University of Beirut Medical Center, Beirut, Lebanon*
[f] *Faculty of Social Sciences, Lebanese University, Saida, Lebanon*
[g] *Department of Business Information and Decision Systems, Suliman S. Olayan School of Business, American University of Beirut, Beirut, Lebanon*
[h] *School of Health Policy and Management, York University, Toronto, Canada*
ORCiD ID: Madhuri MAHALINGAM https://orcid.org/0000-0003-4939-0209, Manar JAMMAL https://orcid.org/0000-0002-4833-7644 , Reem HOTEIT https://orcid.org/0000-0001-8139-9321, Dinah AYNA https://orcid.org/0000-0002-9717-5223, Maya ROMANI https://orcid.org/0000-0001-5437-0128, Sahar HIJAZI https://orcid.org/0000-0002-5318-0223, Imad BOU-HAMAD https://orcid.org/0000-0003-0344-673X, Christo EL MORR https://orcid.org/0000-0001-6287-3438

Abstract. University students are experiencing a mental health crisis across the world. COVID-19 has exacerbated this situation. We have conducted a survey among university students in two universities in Lebanon to gauge mental health challenges experienced by students. We constructed a machine learning approach to predict anxiety symptoms among the sample of 329 respondents based on student survey items including demographics and self-rated health. Five algorithms including logistic regression, multi-layer perceptron (MLP) neural network, support vector machine (SVM), random forest (RF) and XGBoost were used to predict anxiety. Multi-Layer Perceptron (MLP) provided the highest performing model AUC score (AUC=80.70%) and self-rated health was found to be the top ranked feature to predict anxiety. Future work will focus on using data augmentation approaches and extending to multi-class anxiety predictions. Multidisciplinary research is crucial in this emerging field.

Keywords. Machine learning, mental health, depression, anxiety, university students

[1] Corresponding Author: Christo El Morr, School of Health Policy and Management York University Toronto, Canada, elmorr@yorku.ca

1. Introduction

University students' mental health is of concern worldwide. Students are facing surges in psychological distress as reported by World Health Organization [1]. Covid-19 pandemic had worsened populations' mental health, including anxiety. The effects were similarly experienced in Lebanon where the situation was aggravated with socioeconomic challenges [2]. An inquiry into students' anxiety is needed to understand, and plan for suitable mental health programs and interventions targeting the student population. If we can predict anxiety, universities can deploy appropriate resources in a tailored manner. We have conducted a survey among university students in Lebanon and applied machine learning approaches [3] to predict a significant level of anxiety.

2. Methods

2.1. Dataset and Feature Selection

The dataset originated from an online survey of 329 students in one public and one private Lebanese university. Data was collected from November 2021 to February 2022. To answer the survey, students had to be either undergraduate and graduate students, with 18 years of age or higher, enrolled between Spring 2021 and Fall 2022. Our target variable is anxiety that was assessed using the Beck Anxiety Inventory (BAI)[4]. A score of 16 is considered the clinical cut-off for anxiety [5]. The anxiety score was divided into "no anxiety" (scores 0 to 15) or "with anxiety" (16 to 63). The data set was split 70% for training and 30% for testing. The variables included in the data set were gender, in a relationship, age, education level, GPA, income, self-rated health, religion importance, believe in conspiracies, adherence to COVID-19 guidelines, COVID-19 infection, having access to private counseling, following healthy diet, cigarette use, shisha use, alcohol use, physical activity habits, and number of sleeping hours.

2.2. Algorithm Selection

We have created a dichotomous target variable (Anxiety: Yes or No) and we've investigated the following algorithms logistic regression, multi-layer perceptron (MLP) neural network, support vector machine (SVM), random forest (RF) and XGBoost based on good performance of those algorithms in predicting anxiety [6]. All models were developed using Python Scikit-Learn (version 1.1.1). For each algorithm, parameters were fine-tuned using a grid search approach that comprehensively searches for the best hyperparameters. Model performance was measured by the area under the receiver operating characteristic curve (AUC). Imputation was performed replacing the missing values with using the mode, and one-hot-encoding was applied on categorical data.

3. Results

The mean age for the sample was 24.99 years and the standard deviation 7.39. Mild to moderate depression, anxiety, and stress were reported by the majority of participants (52.3%, 42.9%, and 61.7%, respectively), while severe depression, severe anxiety, and

high stress were reported by 24.6%, 29.3%, and 27.6%, respectively. For each of the developed model, we calculated the AUC and accuracy.

AUC value for MLP was 80.70%, followed by Logistic regression (Lasso) 77.25 %, Support Vector Machine (Lasso) 76.01%, Random Forest 74.75%, and XGBoost 72.58%.

SVM had the highest accuracy (69.70%), followed by random forest (67.68), logistic regression (67.67%), MLP (67.50%), and XGBoost (63.64%).

In addition, we performed a feature selection using the Random Forest feature importance ranking method. Self-rated health was found to be the top ranked feature with 100% importance followed by Age at 64%. The remaining features were below 30%.

4. Discussion

An AUC value between 70% and 80% is considered as acceptable, while an excellent test would have an AUC value between 80% and 90% [7]. The best AUC was achieved by MLP (80.1%); also, MLP accuracy (67.50%) was comparable to the highest accuracy for SVM (69.70%). In our study, we are aiming at having a quasi-diagnostic model; in such cases, AUC is a best measurement for performance. Hence, MLP neural network model is the model of choice as it has a highest AUC. In the only published study related to anxiety among 1172 university students in China using the Self-Rating Anxiety Scale [8], a multiclassification using XGBoost produced an accuracy close to 80%, which is comparable to our MLP model despite the difference in the measurement scale and the classification approach (multi-class in the study vs. binary in ours); however, the AUC was not reported.

One of the limitations of our study is the relatively limited number of respondents. However, one of the key advantages of MLP neural network is its ability to be trained well on small datasets and generate good results [9]. The developed model led to good performance for the prediction of the risk of anxiety among university students and can help university councillors to plan for tailored interventions (e.g., e-mental health).

Self-rated health was by far the most important feature (100%) in relation to anxiety, followed by age (64%). All other features had importance measurement below 30%. It will be interesting to study the predictive ability of these two features alone or in conjunction with sleeping hours on predicting symptoms of anxiety; this could give a powerful insight, using very few data items, that might be used to enhance universities' engagement with students on the topic of mental health and programming counselling services. There is an great opportunity for machine learning to play in addressing mental health challenges on campuses [10].

5. Conclusion

We have explained and discussed the first steps towards building a model to predict anxiety for university students. MLP provided the highest performances in terms of AUC and an acceptable accuracy. Mental health interventions that are based on virtual care might benefit much from machine learning models. In our future work, we focus on using data augmentation to enhance the results and extend to multi-class anxiety predictions.

Acknowledgment

The work was supported partly by a Mitacs Globalink Research Internship award.

References

[1] Auerbach RP, Mortier P, Bruffaerts R, Alonso J, Benjet C, Cuijpers P, et al. WHO World Mental Health Surveys International College Student Project: Prevalence and distribution of mental disorders. J Abnorm Psychol. 2018;127(7):623-38. Epub 2018/09/14. doi: 10.1037/abn0000362. PubMed PMID: 30211576; PubMed Central PMCID: PMC6193834

[2] Abed AE, Razzak RA, Hashim HT. Mental Health Effects of COVID-19 Within the Socioeconomic Crisis and After the Beirut Blast Among Health Care Workers and Medical Students in Lebanon. Prim Care Companion CNS Disord. 2021;23(4). Epub 20210715. doi: 10.4088/PCC.21m02977. PubMed PMID: 34265874; Younes S, Safwan J, Rahal M, Hammoudi D, Akiki Z, Akel M. Effect of COVID-19 on mental health among the young population in Lebanon. Encephale. 2021. Epub 20210909. doi: 10.1016/j.encep.2021.06.007. PubMed PMID: 34583829; PubMed Central PMCID: PMC8426189.

[3] El Morr C, Ali-Hassan H. Healthcare, Data Analytics, and Business Intelligence. Analytics in Healthcare: A Practical Introduction. SpringerBriefs in Health Care Management and Economics. Cham, Switzerland: Springer International Publishing; 2019. p. 1-13; El Morr C, Jammal M, Ali-Hassan H, El-Hallak W. Machine Learning for Practical Decision Making: A Multidisciplinary Perspective with Applications from Healthcare, Engineering and Business Analytics. Price CC, editor. Cham, Switzerland: Springer International Publishing; 2023. 250 p

[4] Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. Journal of consulting and clinical psychology. 1988;56(6):893. doi: https://psycnet.apa.org/doi/10.1037/0022-006X.56.6.893; Beck AT, Steer RA. Relationship between the Beck anxiety inventory and the Hamilton anxiety rating scale with anxious outpatients. Journal of Anxiety Disorders. 1991;5(3):213-23. doi: https://doi.org/10.1016/0887-6185(91)90002-B.

[5] Beck AT, Epstein N, Brown G, Steer R. Beck anxiety inventory. Journal of consulting and clinical psychology. 1993.

[6] Hornstein S, Forman-Hoffman V, Nazander A, Ranta K, Hilbert K. Predicting therapy outcome in a digital mental health intervention for depression and anxiety: A machine learning approach. Digit Health. 2021;7:20552076211060659. Epub 20211129. doi: 10.1177/20552076211060659. PubMed PMID: 34868624; PubMed Central PMCID: PMC8637697; Nemesure MD, Heinz MV, Huang R, Jacobson NC. Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence. Sci Rep. 2021;11(1):1980. Epub 20210121. doi: 10.1038/s41598-021-81368-4. PubMed PMID: 33479383; PubMed Central PMCID: PMC7820000; Su D, Zhang X, He K, Chen Y. Use of machine learning approach to predict depression in the elderly in China: A longitudinal study. J Affect Disord. 2021;282:289-98. Epub 20201230. doi: 10.1016/j.jad.2020.12.160. PubMed PMID: 33418381; Zhang Y, Wang S, Hermann A, Joly R, Pathak J. Development and validation of a machine learning algorithm for predicting the risk of postpartum depression among pregnant women. J Affect Disord. 2021;279:1-8. Epub 20200930. doi: 10.1016/j.jad.2020.09.113. PubMed PMID: 33035748; PubMed Central PMCID: PMC7738412.

[7] Mandrekar JN. Receiver Operating Characteristic Curve in Diagnostic Test Assessment. Journal of Thoracic Oncology. 2010;5(9):1315-6. doi: https://doi.org/10.1097/JTO.0b013e3181ec173d. PubMed PMID: 20736804

[8] Wang C, Zhao H, Zhang H. Chinese College Students Have Higher Anxiety in New Semester of Online Learning During COVID-19: A Machine Learning Approach. Front Psychol. 2020;11:587413. Epub 20201203. doi: 10.3389/fpsyg.2020.587413. PubMed PMID: 33343461; PubMed Central PMCID: PMC7744590.

[9] Olson M, Wyner AJ, Berk R. Modern Neural Networks Generalize on Small Data Sets. Proceedings of the 32nd International Conference on Neural Information Processing Systems. {Red Hook, NY, USA}: Curran Associates Inc.; 2018. p. 3623–32 , numpages = 10; Pasini A. Artificial neural networks for small dataset analysis. Journal of Thoracic Disease. 2015;7(5):953-60

[10] El Morr C. Virtual Communities, Machine Learning and IoT: Opportunities and Challenges in Mental Health Research. International Journal of Extreme Automation and Connectivity in Healthcare. 2019;1(1):4-11. doi: 10.4018/ijeach.2019010102