

Predictive Analysis on Depression among University Students in Bangladesh

Syed Aref Ahmed, Khondokar Jamal E Mustafa , Ibtesum Arif & MD. Nafis Tahmid

Supervisor: Md. Shahriar Rahman Rana , Co-Supervisor: Rafeed Rahman

Department of Computer Science and Engineering, Brac University



Abstract

The identification of depression is done by medical practitioners based on mental status questionnaires and the patient’s self-reporting. Apart from the methods being highly dependent on the patient’s current mood, people who go through mental dis- orders seek mental help reluctantly. Universities always promise scholars a promising career in their domains. However, the academic competition, peer pressure, isolation and many other factors could put a stu- dent in a state of depression. In this research, we propose a big data analytics template to detect depression among university stu- dents. Asserting again, since isolation and separation are believed to have the most dramatic effect on the pupils, the framework also models the correlation between these factors and depression. To conclude, the journal evaluates the performance of the proposed framework on a massive real dataset collected from different university students of Bangladesh and proves that the accuracy of the machine learning models outperforms traditional techniques for detecting depression in univer- sities.

1 Problem Statement

We are investigating the factors that may be contributing to the widespread problem of depression among college students in an effort to find solutions to this critical issue. Using a dataset, our primary goal is to determine whether or not individuals are suffering from depression. We will use a variety of machine learning models in order to find which one is the most accurate. In the end, our goal is to use this model to perform predictive analysis in order to identify university students in Bangladesh who are suffering from depression.

Research Objective

- Provide insights to improve student welfare sector in universities in Bangladesh
- Analyze patterns of various factors that has a high chance to lead university students to de- pression.
- Enhance focus on mental health of university students
- Develop a model which can detect depression

2 Workflow Methology



Figure 1: Number of images in training, testing & validation.

Dataset

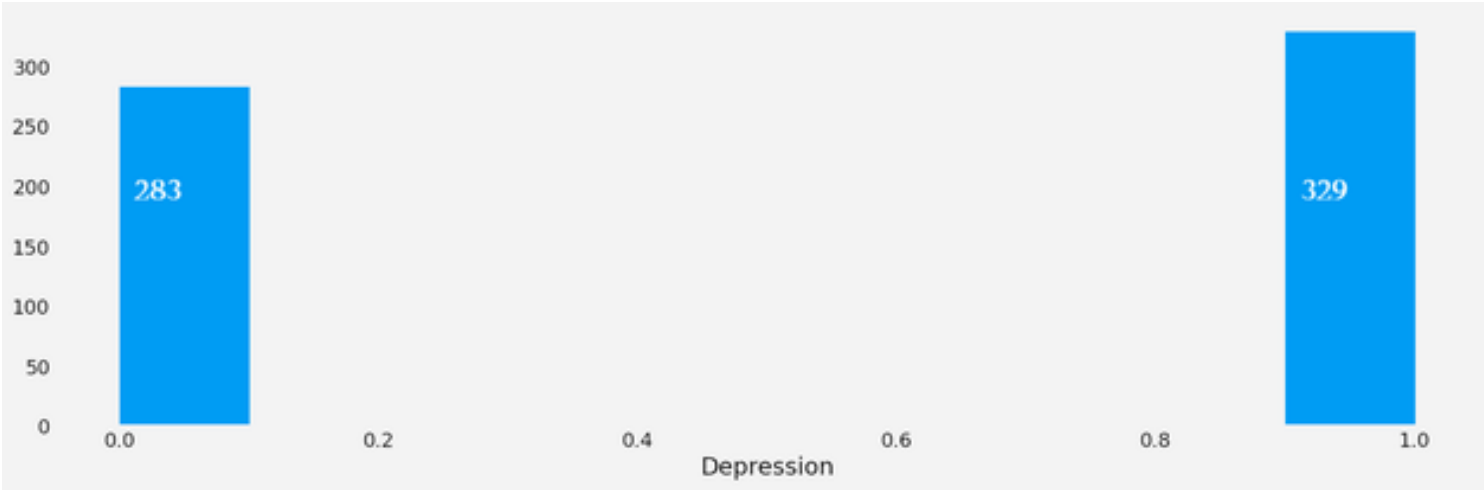


Figure 2: Distribution of test subjects with and with no depression.

- A dataset for mental health analysis of university students was created via an anonymous Google Form survey with 612 responses.

- The dataset includes 26 queries, with one for depression detection and others as features, en- compassing binary, multiclass, and numeric data.

- Data preprocessing involved renaming features, handling missing values (mode imputation for one column), and encoding binary features as 0s and 1s

- Discrete numeric features like age and sleeping hours were grouped, and onehot encoding applied to some integer-valued and multiclass features. Seven multiclass categorical features were label encoded.

- After preprocessing, the dataset contains binary and discrete numeric values, ready for feature scaling and machine learning model implementation.

Experiment Result



Figure 3: Workflow accuracy of each model.

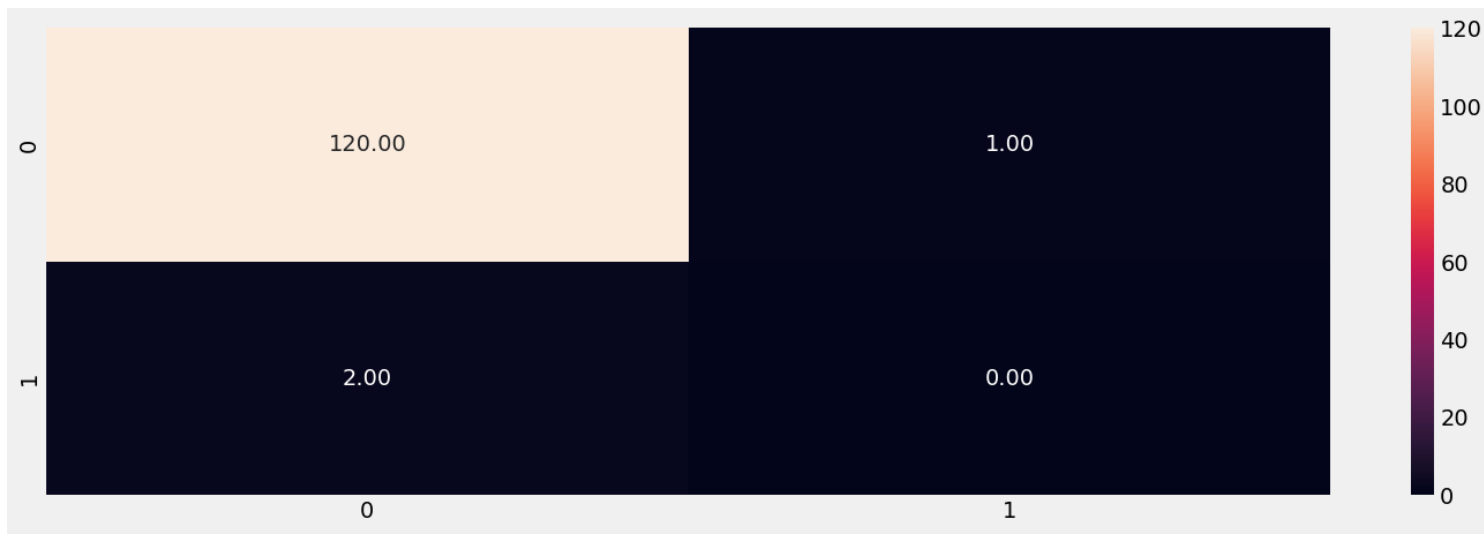


Figure 4: Confusion Matrix of Gradient Boost.

	LR	DT	RF	Ada	GB	SVM	KNN	NB
0	1	0.99	1	1	0.99	0.99	1	0.99
1	0.67	0.4	1	1	0	0.67	1	0
Avg	0.83	0.69	1	1	0.5	0.83	1	0.5
Weighted-Avg	0.99	0.98	1	1	0.98	0.99	1	0.98

Figure 5: F1 score of different models.

3 Result Analysis

All of the models had very high validation accuracy, with the lowest being 97.6% for Decision Tree and highest 100% for Random Forest, AdaBoost and KNN. In testing dataset, majority of the samples belonged to the negative class, but our model detected true negatives with very high precision. As a result weighted F1_score was very high for all the algorithms, lowest being 98% for Decision Tree and highest being 100% for Random Forest, AdaBoost and KNN.

4 Related Works

- Jasiya Fairiz Raisa, M. Shamim Kaiser & Mufti Mahmud, "A Machine Learning Approach for Early Detection of Postpartum Depression in Bangladesh," 2022
- Ahnaf Atef Choudhury, Md. Rezwan Hassan Khan, Nabuat Zaman Nahim, Sadid Rafsun Tulon, Samiul Islam, Amitabha Chakrabarty, "Predicting Depression in Bangladeshi Under- graduates using Machine Learning," IEEE, 2019.

5 Conclusion

With the use of machine learning algorithms and statistical models, this research effectively cre- ated a predictive model which can detect depression among university students in Bangladesh. The predictive model has good accuracy scores on its training dataset. The dataset was created via surveys, the survey was distributed to many university students in Bangladesh.

6 Reference

- Arkaprabha Sau and Ishita Bhakta, "Predicting anxiety and depression in elderly patients using machine learning technology," Creative Commons, 2017.
- Umme Marzia Haque ,Enamul Kabir, Rasheda Khanam, "Detection of child depression using machine learning methods," Plos One, 2021.
- Rumana Rois, Manik Ray, Atikur Rahman Swapan K. Roy , "Prevalence and predicting fac- tors of perceived stress among Bangladeshi university students using machine learning algo- rithms," Journal of Health Population and Nutrition, 2021