# Detection of possible vulnerability to Depression before its onset

## Project-Based Internship 2022 Report

Submitted to

## CS Department, ABESEC

**Duration: Six Weeks**

**By**

**Vartika: 1900321290067**
**Sanidhiya Sharma: 1900321290051**
**Sanjana Jain:1900321290053**
**Ayushi Gupta:1900321290029**

**Under the guidance of:-**
**1) MR. ASHWIN PERTI**
**2) MS. SAPNA JAIN**

## ABES ENGINEERING COLLEGE, GHAZIABAD



**AFFILIATED TO**
DR. A.P.J. ABDUL KALAM TECHNICAL UNIVERSITY, U.P., LUCKNOW
(Formerly UPTU)

# TABLE OF CONTENTS

# STUDENT'S DECLARATION

We hereby declare that the work being presented in this report entitled **DETECTION TO POSSIBLE VULNERABILITY TO DEPRESSION BEFORE ITS ONSET** is an authentic record of my / our own work carried out under the supervision of Mr Gopal Gupta/ Mr Ashwin Perti/ Ms Sapna Jain.

The matter embodied in this report has not been submitted by me / us for the award of any other degree.

**Date:**                                                                 **Signature of students**

# CERTIFICATE

This is to certify that the Project Report entitled **DETECTION TO POSSIBLE VULNERABILITY TO DEPRESSION BEFORE ITS ONSET** which is submitted by Ayushi Gupta, Sanjana Jain, Sanidhiya Sharma and Vartika in partial fulfillment of the requirement for the summer internship of Data Analysis and Machine Learning using Python in Department of Computer Science of ABES Engineering College, is a record of the candidate own work carried out by him under my/our supervision.

**Signature of HOD**                         **Supervisor 1: Mr Ashwin Perti**
**(Dr. Pankaj Sharma)**
**(Computer Science)**                       **Supervisor 2: Sapna Jain**

**Date.: -**

# ACKNOWLEDGEMENT

*It gives us a great sense of pleasure to present the report of the Project Based Internship 2022 undertaken during B-tech 3rd year. We owe special debt of gratitude to Gopal Sir/Ashwin Perti/Sapna Jain facilitators for HET Data Science Training for their constant support and guidance throughout the course of our work. Their constant motivation have been a source of inspiration for us. It is only their cognizant efforts that our endeavors have seen light of the day.*

*We also do not like to miss the opportunity to acknowledge the motivation of Computer Science Department to provide us the opportunity to undergo this training.*

*Signature:*

*Name     : Sanjana Jain*

*Roll No  : 1900321290053*

*Date     :*

*Signature:*

*Name     :Ayushi Gupta*

*Roll No : 190031290067*

*Date     :*

*Signature:*

*Name      :        Vartika*

*Roll No  : 1900321290067*

*Date     :*




*Signature:*

*Name: Sanidhiya Sharma*

*Roll No  : 1900321290051*

*Date     :*

# ABSTRACT

As per WHO, Depression is a leading cause of disability worldwide and is a major

contributor to the overall global burden of disease, with estimated lifetime prevalence rates

of 20%. Whilst the majority of individuals with depression and anxiety are treated in

primary care settings, over 50% of people are not recognized or adequately treated.

With the rising spread of COVID19 and the peak in the awareness of mental health,

there is a vast increase in the subclinical population. The earlier the detection,

the earlier can intervention stunt the symptoms

Our novel approach is a feature extraction technique that allows us to select meaningful features indicative of depressive symptoms from longitudinal data Our work will have significant implications for the detection of health outcomes using longitudinal behavioral data and limited ground truth. By detecting change and predicting symptoms several weeks before their onset, our work also has implications for preventing depression.

# LIST OF TABLES

# LIST OF FIGURES

Description of fig                                          pgno.

# CHAPTER 1

## INTRODUCTION

### 1.1    Problem Definition

Humans are, by nature, becoming ambitious nowadays and seek every possible opportunity to grow now believe them to be part and parcel of professional life. The World Health Organization (WHO) has observed that depression is the most prevalent mental disorder affecting more than 300 million people worldwide, and the severity of the issue has led many health researchers to focus their studies on this area. Differentiating anxiety, depression, and stress from one another are problematic for machines; hence, an appropriate learning algorithm is required for an accurate diagnosis. According to WHO, a healthy person possesses a healthy brain along with physical wellness. Depression and anxiety are leading causes of disability worldwide, with estimated lifetime prevalence rates of 20%. Whilst the majority of individuals with depression and anxiety are treated in primary care settings, over 50% of people are not recognized or adequately treated. Given the adverse health outcomes and costs associated with untreated conditions and the recent increase in the prevalence of common mental disorders, adequate diagnosis and timely treatment of depression and anxiety has become an urgent priority

### 1.2    Motivation

The major problem that everyone is facing is the unawareness of being mentally unfit or treating it as taboo. It is important to identify it and then find its cause and treat it on an immediate basis. This is the key motivation for our project to make people aware of their mental well-being without any interference and make them understand the key

difference between depression, anxiety, and stress. The main symptoms of depression from a clinical point of view are loss of memory; lack of concentration; an inability to make decisions; loss of interest in recreational activities and hobbies including sex; overeating and weight gain; low appetite and weight loss; feelings of guilt, worthlessness, helplessness, restlessness and irritation; as well as suicidal thoughts. These symptoms were found to have a significant effect on important areas of an individual's life – such as in education, employment and social activities, and this provides a vital clue for forming a clinical diagnosis. The symptoms of GAD (Generalised Anxiety Disorder) are irritability, nervousness, fatigue, insomnia, gastro intestinal problems, panic, and a sense of impending danger, increased heart rate, sweating, rapid breathing and difficulty concentrating. The symptoms of stress are feeling upset or agitated, an inability to relax, low energy levels, chronic headaches, frequent overreaction and persistent colds or infections. Thus, stress, anxiety and depression have many common symptoms including insomnia, chest pain, fatigue, increased heart rate and inability to concentrate, all of which makes classification challenging for machines.

.

### 1.3 Objective of the Project:

- To predict the **anxiety level** of the user accurately.
- A machine learning model to calculate anxiety and depression score by performing four internationally acclaimed anxiety predicting test.
- Perform **regular analysis** to draw the conclusion .
- Different **engaging activities** to help the users to **interact** .
- **Consultation** available based on the score calculated.
- ChatBot : **online companion** to share your feelings

**1.4**	**Scope of the Project:**
In this model, machine learning algorithms can be applied to determine five different severity levels of anxiety, depression and stress. Five different classification techniques can be applied – Decision Tree (DT), Random Forest Tree (RFT), Naïve Bayes, Support Vector Machine (SVM) and K- Nearest Neighbour (KNN). Different researchers have applied different machine leaning algorithms for the prediction of psychological disorders, and the performances of different algorithms have been found to vary, depending on the scenario; no fixed algorithm has been determined as most suitable in all cases. Thus, all the machine learning algorithms can be applied to identify the symptoms of anxiety, depression and stress.

# CHAPTER 2

# RELATED WORK

## 1. INTRODUCTION

### 1.1 Overview

## 2. METHODOLOGY

### 2.1 Dataset Description

**Method**

This data was collected with an online version of the Depression Anxiety Stress Scales (DASS), see http://www2.psy.unsw.edu.au/dass/

The survey was open to anyone and people were motivated to take it to get personalized results. At the end of the test, they also were given the option to complete a short research survey. This dataset comes from those who agreed to complete the research survey and answered yes to the question "Have you given accurate answers and may they be used for research?" at the end.

This data was collected from 2017 - to 2019.

The following items were included in the survey:

Q1 I found myself getting upset by quite trivial things.

Q2 I was aware of dryness of my mouth.

Q3 I couldn't seem to experience any positive feeling at all.

Q4 I experienced breathing difficulty (eg, excessively rapid breathing, breathlessness in the absence of physical exertion).

Q5 I just couldn't seem to get going.

Q6 I tended to over-react to situations.

Q7 I had a feeling of shakiness (eg, legs going to give way).

Q8 I found it difficult to relax.

Q9 I found myself in situations that made me so anxious I was most relieved when they ended.

Q10 I felt that I had nothing to look forward to.

Q11 I found myself getting upset rather easily.

Q12 I felt that I was using a lot of nervous energy.

Q13 I felt sad and depressed.

Q14 I found myself getting impatient when I was delayed in any way (eg, elevators, traffic lights, being kept waiting).

Q15 I had a feeling of faintness.

Q16 I felt that I had lost interest in just about everything.

Q17 I felt I wasn't worth much as a person.

Q18 I felt that I was rather touchy.

Q19 I perspired noticeably (eg, hands sweaty) in the absence of high temperatures or physical exertion.

Q20 I felt scared without any good reason.

Q21 I felt that life wasn't worthwhile.

Q22 I found it hard to wind down.

Q23 I had difficulty in swallowing.

Q24 I couldn't seem to get any enjoyment out of the things I did.

Q25 I was aware of the action of my heart in the absence of physical exertion (eg, sense of heart rate increase, heart missing a beat).

Q26 I felt down-hearted and blue.

Q27 I found that I was very irritable.

Q28 I felt I was close to panic.

Q29 I found it hard to calm down after something upset me.

Q30 I feared that I would be "thrown" by some trivial but unfamiliar task.

Q31 I was unable to become enthusiastic about anything.

Q32 I found it difficult to tolerate interruptions to what I was doing.

Q33 I was in a state of nervous tension.

Q34 I felt I was pretty worthless.

Q35 I was intolerant of anything that kept me from getting on with what I was doing.

Q36 I felt terrified.

Q37 I could see nothing in the future to be hopeful about.

Q38 I felt that life was meaningless.

Q39 I found myself getting agitated.

Q40 I was worried about situations in which I might panic and make a fool of myself.

Q41 I experienced trembling (eg, in the hands).

Q42 I found it difficult to work up the initiative to do things.

Each item was presented one at a time in a random order for each new participant along with a 4 point rating scale asking the user to indicate how often that had been true of them in the past week where

1 = Did not apply to me at all

2 = Applied to me to some degree, or some of the time

3 = Applied to me to a considerable degree, or a good part of the time

4 = Applied to me very much, or most of the time

This response is stored in variable A (e.g. Q1A). Also recorded was the time taken in milliseconds to answer that question (E) and that question's position in the survey (I).

These other durations were also recorded (measured on the server's side):

introelapse The time spent on the introduction/landing page (in seconds) testelapse The time spent on all the DASS questions (should be equivalent to the time elapsed on all the indiviudal questions combined) surveyelapse The time spent answering the rest of the demographic and survey questions

On the next page was a generic demographics survey with many different questions.

The Ten Item Personality Inventory was administered (see Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A Very Brief Measure of the Big Five Personality Domains. Journal of Research in Personality, 37, 504-528.):

TIPI1 Extraverted, enthusiastic.

TIPI2 Critical, quarrelsome.

TIPI3 Dependable, self-disciplined.

TIPI4 Anxious, easily upset.

TIPI5 Open to new experiences, complex.

TIPI6 Reserved, quiet.

TIPI7 Sympathetic, warm.

TIPI8 Disorganized, careless.

TIPI9 Calm, emotionally stable.

TIPI10 Conventional, uncreative.

The TIPI items were rated "I see myself as:" _ such that

1 = Disagree strongly

2 = Disagree moderately

3 = Disagree a little

4 = Neither agree nor disagree

5 = Agree a little

6 = Agree moderately

7 = Agree strongly

The following items were presented as a check-list and subjects were instructed "In the grid below, check all the words whose definitions you are sure you know":

VCL1 boat

VCL2 incoherent

VCL3 pallid

VCL4 robot

VCL5 audible

VCL6 cuivocal

VCL7 paucity

VCL8 epistemology

VCL9 florted

VCL10 decide

VCL11 pastiche

VCL12 verdid

VCL13 abysmal

VCL14 lucid

VCL15 betray

VCL16 funny

A value of 1 is checked, 0 means unchecked. The words at VCL6, VCL9, and VCL12 are not real words and can be used as a validity check.

A bunch more questions were then asked:

education "How much education have you completed?", 1=Less than high school, 2=High school, 3=University degree, 4=Graduate degree urban "What type of area did you live in when you were a child?", 1=Rural (country side), 2=Suburban, 3=Urban (town, city) gender "What is your gender?", 1=Male, 2=Female, 3=Other engnat "Is English your native language?", 1=Yes, 2=No age "How many years old are you?" hand "What hand do you use to write with?", 1=Right, 2=Left, 3=Both religion "What is your religion?", 1=Agnostic, 2=Atheist, 3=Buddhist, 4=Christian (Catholic), 5=Christian (Mormon), 6=Christian (Protestant), 7=Christian (Other), 8=Hindu, 9=Jewish, 10=Muslim, 11=Sikh, 12=Other orientation "What is your sexual orientation?", 1=Heterosexual, 2=Bisexual, 3=Homosexual, 4=Asexual, 5=Other race "What is your race?", 10=Asian, 20=Arab, 30=Black, 40=Indigenous Australian, 50=Native American, 60=White, 70=Other voted "Have you voted in a national election in the past year?", 1=Yes, 2=No married "What is your marital status?", 1=Never married, 2=Currently married, 3=Previously married family size "Including you, how many children did your mother have?" major "If you attended a university, what was your major (e.g. "psychology", "English", "civil engineering")?"

The following values were derived from technical information:

country ISO country code of where the user connected from screen size 1=device with a small screen (phone, etc), 2=device with a big screen (laptop, desktop, etc) unique-network location 1=only one survey from user's specific network in the dataset, 2=multiple surveys submitted from the network of this user (2 does not

necessarily imply duplicate records for an individual, as it could be different students at a single school or different members of the same household; and even if 1 there still could be duplicate records from a single individual e.g. if they took it once on their wifi and once on their phone) source how the user found the test, 1=from the front page of the site hosting the survey, 2=from google, 0=other or unknown

## 2.2 Data Visualization

The below graph represents the correlation between various survey questions.
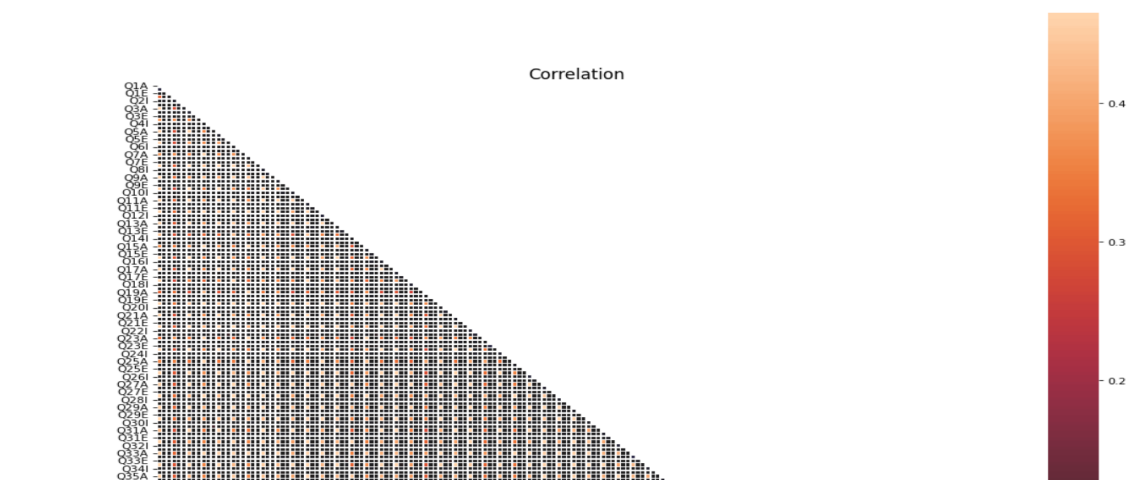


Live dataset on how many participants have taken the survey.

👫 Participants
Took the test

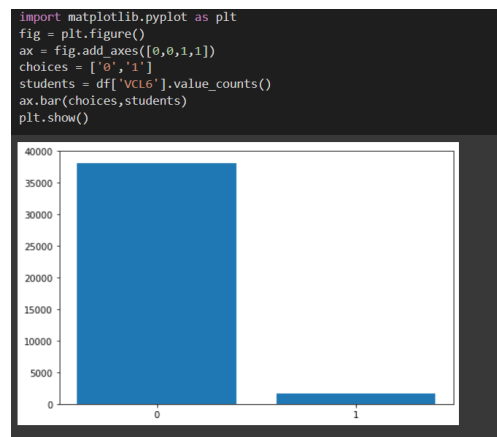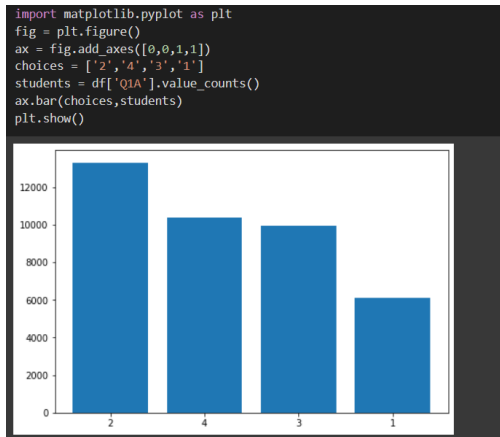# 39.8k

❓ Questions
On the test

# 172

The below bar chart represents the responses of various users according to the survey for the following question:
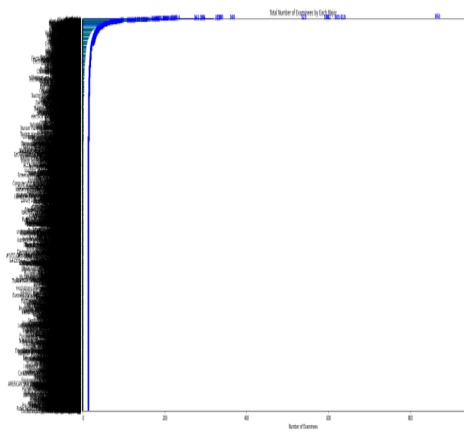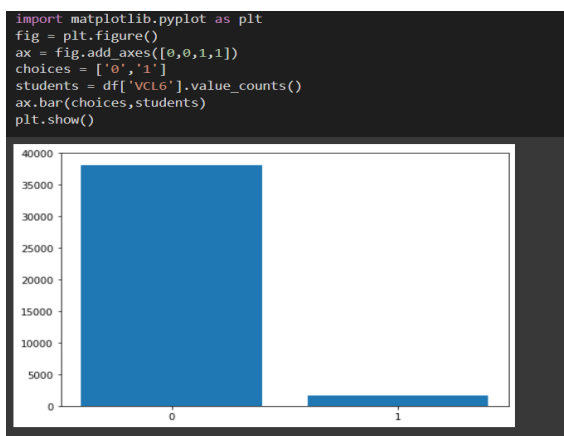
Q1:I found myself getting upset by quite trivial things.

The responses can be submitted as:

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

```python
import matplotlib.pyplot as plt
fig = plt.figure()
ax = fig.add_axes([0,0,1,1])
choices = ['2','4','3','1']
students = df['Q1A'].value_counts()
ax.bar(choices,students)
plt.show()
```



```python
import matplotlib.pyplot as plt
fig = plt.figure()
ax = fig.add_axes([0,0,1,1])
choices = ['0','1']
students = df['VCL6'].value_counts()
ax.bar(choices,students)
plt.show()
```



The graph is to describe a word that does not exist and people were asked whether they know the meaning of that word or not. It tells us how people are not true to themselves as few of the users have chosen 1, i.e., they know the meaning of the word which does not exist.
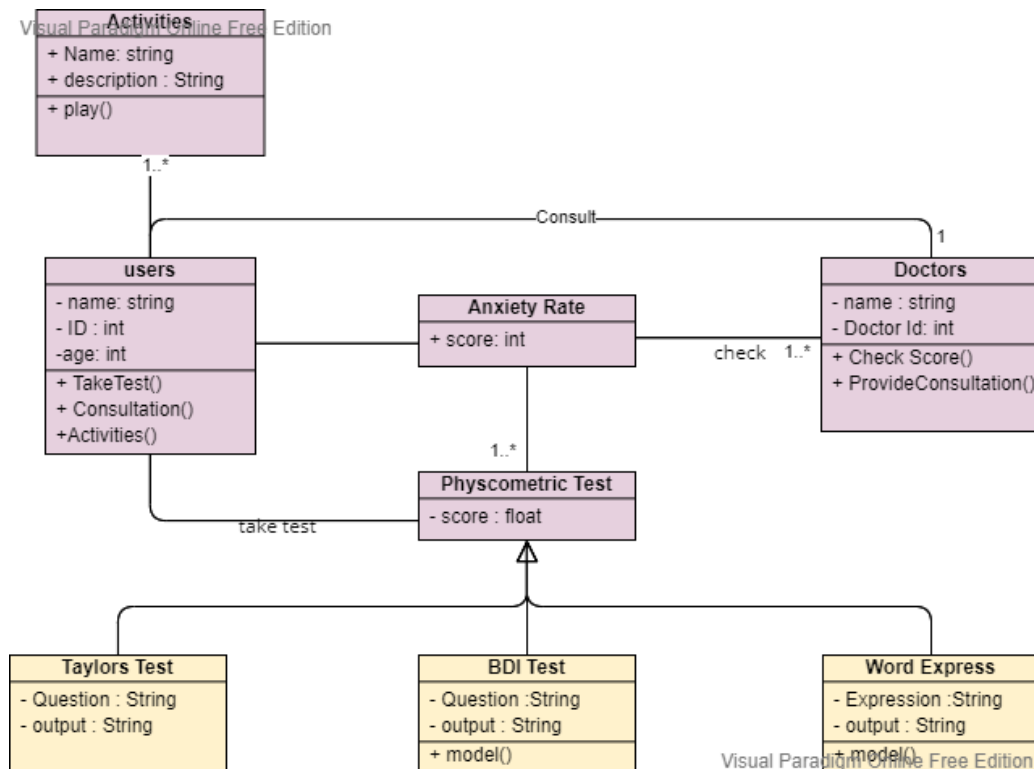
```python
import matplotlib.pyplot as plt
fig = plt.figure()
ax = fig.add_axes([0,0,1,1])
choices = ['0','1']
students = df['VCL6'].value_counts()
ax.bar(choices,students)
plt.show()
```





**Total Number of Examinees by Each Major**

Text(0.5, 1.0, 'Distribution of Spent Time per Question')



Distribution of Spent Time per Question
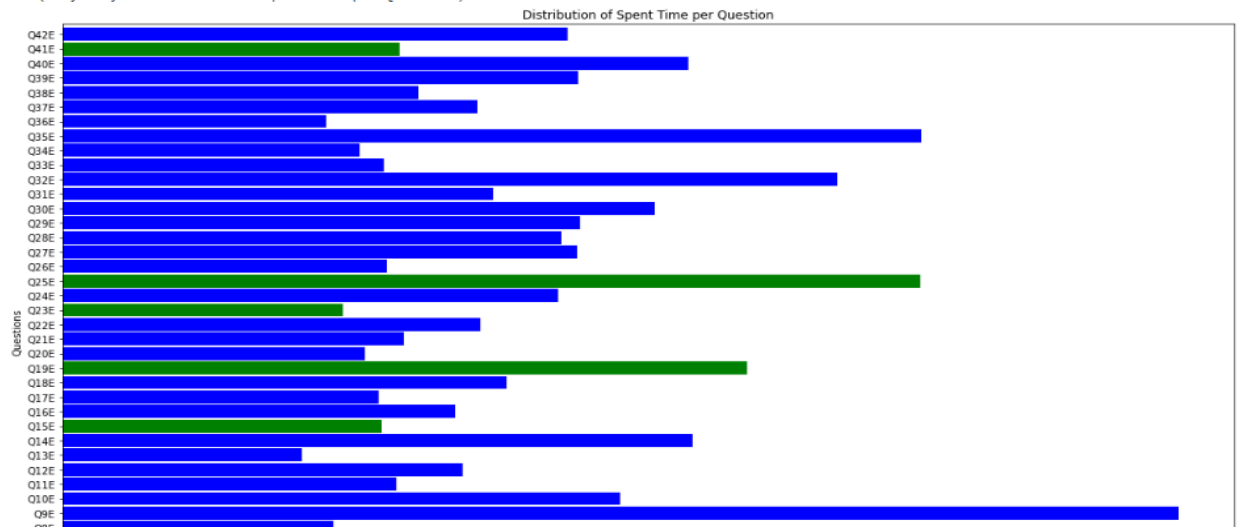
## MODULES:

## CONCLUSION

We tried to understand the factors laying behind Depression, Anxiety and Stress. There are various options; personality type, gender, sexual orientation, and where you live!

## REFERENCES

[1]Sharifa Alghowinem, Roland Goecke, Michael Wagner, Gordon Parkerx, and Michael Breakspear. 2013. Head pose and movement analysis as an indicator of depression. In Proceedings of the Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII '13). IEEE, 283--288

[2]American Psychiatric Association. 2013. Diagnostic and Statistical Manual of Mental Disorders (DSM-5®). American Psychiatric Pub.

[3]Philip I. Chow, Karl Fua, Yu Huang, Wesley Bonelli, Haoyi Xiong, Laura E. Barnes, and Bethany A. Teachman. 2017. Using mobile sensing to test clinical models of depression, social anxiety, state affect, and social isolation among college students. Journal of Medical Internet Research 19, 3 (2017), e62.