Enterprise Scale AI Systems and Solutions
CSE – 510

Suicide Prediction Using NLP-ML

Mrunmayee Vijay Rane
UB Person No: 50417094
MS Computer Science and Engineering
School of Engineering and Applied Science
mrane@buffalo.edu

Abstract:
We are using The Diagnostic and Statistical Manual of Mental Disorders (DSM), fifth edition (DSM-5 for diagnostic classification of Mental health disorders according to American Psychiatric Association (APA), for hierarchical classification of Depression, Anxiety, Mania, Substance Abuse and Anger for more accurate and refined results to identify mental health disorders.

DSM - V

We have used the Diagnostic and Statistical Manual of Mental Disorders (DSM), DSM-5 approach for diagnosing various mental health disorders and suicide ideations such as depression, anger, mania, anxiety, and substance abuse in our Aarogya application.

The Diagnostic and Statistical Manual of Mental Disorders (DSM), fifth edition (DSM-5) are the taxonomic and diagnostic tool published in 2013, by the American Psychiatric Association (APA). In the United States, the DSM serves as the principal authority for psychiatric diagnoses. Treatment recommendations, as well as payment by health care providers, are often determined by DSM classifications.

The Diagnostic and Statistical Manual of Mental Disorders (DSM) is a key reference book for mental health professionals. The DSM contains descriptions, symptoms, and other criteria for diagnosing mental disorders. According to the American Psychiatric Association (APA), these criteria for diagnosis provide a common language among clinicians -- professionals who treat patients with mental disorders. By clearly defining the criteria for a mental disorder, the DSM helps to ensure that a diagnosis is both accurate and consistent internationally.

The APA also states that another important role of the DSM is in the area of research i.e., only by having consistent (reliable) diagnoses can researchers compare different treatments for similar patients, determine the risk factors and causes for specific disorders, and determine their incidence and prevalence rates. DSM disorders are also used as the basis for treatment indications by the U.S. Food and Drug Administration (FDA) for clinical Practice Guidelines.

The DSM is not a static document. It has been periodically reviewed and significantly revised since the publication of DSM-I in 1952. New research in neurology, genetics, behavioral sciences, epidemiology

and other scientific areas have dramatically expanded our understanding of mental illnesses and led to changes in the DSM. We are using emerging measures in section III of DSM -5

1. **Online assessment measure:**

   For further clinical evaluation and research, the APA is offering a number of "emerging measures" in Section III of *DSM–5*. These patient assessment measures were developed to be administered at the initial patient interview and to monitor treatment progress, thus serving to advance the use of initial symptomatic status and patient reported outcome (PRO) information, as well as the use of "anchored" severity assessment instruments.

   The measures can be broadly classified into four types:

   - Cross-cutting symptom measure
   - Severity Measures
   - The World Health Organization disability assessment schedule 2.0 (WHODAS 2.0)
   - Personality Inventories.

   Out of which we are using cross-cutting and severity measures. Cross-cutting and Severity measures for diagnostic of depression and anxiety in order to calculate the accurate findings and severity of disorder in a patient. Cross-cutting measures is used for mania and substance abuse identification and for additional areas of inquiries.

   1.1 Cross-cutting symptom measures: Cross-cutting symptom measures may aid in a comprehensive mental status assessment. They are intended to help identify additional areas of inquiry that may guide treatment and prognosis. The cross-cutting measures have two levels: Level 1 questions are a brief survey of 13 domains for adult patients and 12 domains for child and adolescent patients, and Level 2 questions provide a more in-depth assessment of certain domains.

   1.2 Severity measures: Severity measures are disorder-specific, corresponding closely to criteria that constitute the disorder definition. They may be administered to individuals who have received a diagnosis or who have a clinically significant syndrome that falls short of meeting full criteria. Some of the assessments are self-completed, whereas others require a clinician to complete.

2. **Measure scoring and interpretations:**

   Assessments are in the form of questionnaire where each question on the measure is rated on a 5-point scale (1=never; 2=rarely; 3=sometimes; 4=often; and 5=always or 0=never, 1=occasionally, 2=half of the time, 3=most of the time, 4= all the time) with a range in score from 0 to 40 or accordingly, with higher scores indicating greater severity of the respective disorder. For example, the assessment may consist of a feeling "I felt worthless" and the patient has to choose how often this symptom has occurred. The clinician is asked to review the score of each question on the measure during the clinical interview and indicate the raw score for each question in the section provided for "Clinician Use." Scores on the individual items should be interpreted independently because each item inquires about the use of a distinct substance. The raw scores on the all the questions should be summed to obtain a total raw score. Next, the T-score table should be used to identify the T-score associated with the total raw score and the information entered in the T-score row on the measure. T score is calculated with respect to the disorders which gives the diagnostic results.

3. **Identification of Depression and Measuring Severity of Depression**:

Depression is one of the most common mental disorders in the United States. The prevalence of at least 1 major depressive episode among US adults aged 18 years or older, was 17.3 million, representing 7.1% of all US adults. Screenings for mental health disorders are as important as high-blood pressure, diabetes and other physical diseases and disorders. Mental health screenings are often the first step in getting help. Not only is clinical depression a serious medical illness, it can lead to suicide and can co-occur and complicate other medical conditions. [1]

In the absence of screening, it is estimated that only 50 percent of patients with major depression are identified [8]. Unless directly asked about their mood, patients omit information about depressive symptoms for a variety of reasons, including fear of stigmatization, belief that depression falls outside the purview of primary care, belief that depression isn't a "real" illness but rather a personal flaw, concerns about medical record confidentiality, and concerns about being prescribed antidepressant medication or being referred to a psychiatrist[9]

(https://genesight.com/blog/healthcare-provider/the-silent-illness-the-importance-of-screening-for-depression-in-primary-care/)1
(https://pubmed.ncbi.nlm.nih.gov/21911763/)9
(https://pubmed.ncbi.nlm.nih.gov/19640579/)8

3.1 PHQ 9:

The Patient Health Questionnaire consists of 9 questions on depression severity identification. Severity of depression using PHQ - 9 is the level 3 diagnostic approach that we are using in addition with cross-cutting measures (level 1 and level 2) to get more fine results. Identification and severity of depression is measured using following algorithm.

Step 1: First the user will fill cross cutting level 1 questionnaire, which consists of 23 questions classified in 13 domains for adults.
Step 2: Calculate individual domain scores
Step 3: Calculate highest domain score.
highest domain score = points scored on Q1+Q2+Q11
Step 4: If the highest domain score is between 0 to 12. Ask user to fill Level 2 Depression questionnaire.
Step 5: Check if the No. of answered questions are greater than 3. If yes proceed to step 6, else proceed with step 8.
Step 6: Check if the No. of answered questions = 8. If yes proceed to step 7, else proceed with step 8.
Step 7: Calculate Raw scores.
Raw score = Sum of points answered on all the questions.
Step 8: Calculate prorated raw score.
Raw score = (Raw Score*(No. of questions answered))/8
Step 9: Ask the user to answer more questions. Calculate prorated raw score.
Raw score = (Raw Score*(No. of questions answered))/8
Step 10: Convert the Raw score into T score using lookup table.

Depression 8b *Short Form Conversion Table*		
Raw Score	**T-score**	**SE***
8	37.1	5.5
9	43.3	3.4
10	46.2	2.8
11	48.2	2.4
12	49.8	2.2
13	51.2	2.0
14	52.3	1.9
15	53.4	1.8
16	54.3	1.8
17	55.3	1.7
18	56.2	1.7
19	57.1	1.7
20	57.9	1.7
21	58.8	1.7
22	59.7	1.8
23	60.7	1.8
24	61.6	1.8
25	62.5	1.8
26	63.5	1.8
27	64.4	1.8
28	65.4	1.8
29	66.4	1.8
30	67.4	1.8
31	68.3	1.8
32	69.3	1.8
33	70.4	1.8
34	71.4	1.8
35	72.5	1.8
36	73.6	1.8
37	74.8	1.9
38	76.2	2.0
39	77.9	2.4
40	81.1	3.4

*SE = Standard Error on T-score metric

Step 11: If Raw score is between 8 to 40. Ask user to fill PHQ 9 questionnaire for calculating severity of depression.
Step 12: Check if the No. of answered questions are greater than 3. If yes proceed to step 13, else proceed with step 15.

Step 13: Check if the No. of answered questions = 8. If yes proceed to step 14, else proceed with step 15.
Step 14: Calculate Raw scores.
Raw score = Sum of points answered on all the questions.
Step 15: Calculate prorated raw score.
Raw score = (Raw Score*(No. of questions answered))/9
Step 16: Ask the user to answer more questions. Calculate prorated raw score.
Raw score = (Raw Score*(No. of questions answered))/9
Step 17: If Raw score 0<=4: Output: You don't have depression.
If Raw score 5<=9: Output: You have a mild depression.
If Raw score 10<=14: Output: You are diagnosed with moderate depression.
If Raw score 15<=19: Output: You are diagnosed with moderately severe depression.
If Raw score 20<=27: Output: You are diagnosed with severe depression.

4. Identification and Measuring Severity of Anxiety:

Generalized anxiety disorder (GAD) is a chronic, impairing and highly comorbid psychiatric condition afflicting an estimated 2.1% to 3.1% of the U.S. population during any given 12-month period (Grant et al., 2005; Kessler et al., 2005). As public awareness and recognition of GAD continues to grow, clinicians will likely evaluate and treat these patients with increasing frequency. Accurate detection, diagnosis and assessment of disorder severity can inform the clinician's interventions and improve treatment effectiveness. Identification and severity of depression is measured using following algorithm.

(Grant BF, Hasin DS, Stinson FS et al. (2005), Prevalence, correlates, co-morbidity, and comparative disability of DSM-IV generalized anxiety disorder in the USA: results from the National Epidemiologic Survey on Alcohol and Related Conditions. Psychol Med 35(12):1747-1759. Kessler RC, Chiu WT, Demler O et al. (2005), Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. [Published erratum Arch Gen Psychiatry 62(7):709. Merikangas, KR (added).] Arch Gen Psychiatry 62(7):617-627 [see comment].

https://www.psychiatrictimes.com/view/tools-assessing-generalized-anxiety-disorder)

Step 1: First the user will fill cross cutting level 1 questionnaire, which consists of 23 questions classified in 13 domains for adults.
Step 2: Calculate individual domain scores
Step 3: Calculate highest domain score.
highest domain score = points scored on Q6+Q7+Q8
Step 4: If the highest domain score is between 0 to 12. Ask user to fill Level 2 Anxiety questionnaire.
Step 5: Check if the No. of answered questions are greater than 3. If yes proceed to step 6, else proceed with step 8.
Step 6: Check if the No. of answered questions = 7. If yes proceed to step 7, else proceed with step 8.
Step 7: Calculate Raw scores.
Raw score = Sum of points answered on all the questions.
Step 8: Calculate prorated raw score.
Raw score = (Raw Score*(No. of questions answered))/7
Step 9: Ask the user to answer more questions. Calculate prorated raw score.
Raw score = (Raw Score*(No. of questions answered))/7
Step 10: Convert the Raw score into T score using lookup table.

Anxiety 7a
Short Form Conversion Table

Raw Score	T-score	SE*
7	36.3	5.4
8	42.1	3.4
9	44.7	2.9
10	46.7	2.6
11	48.4	2.4
12	49.9	2.3
13	51.3	2.3
14	52.6	2.2
15	53.8	2.2
16	55.1	2.2
17	56.3	2.2
18	57.6	2.2
19	58.8	2.2
20	60.0	2.2
21	61.3	2.2
22	62.6	2.2
23	63.8	2.2
24	65.1	2.2
25	66.4	2.2
26	67.7	2.2
27	68.9	2.2
28	70.2	2.2
29	71.5	2.2
30	72.9	2.2
31	74.3	2.2
32	75.8	2.3
33	77.4	2.4
34	79.5	2.7
35	82.7	3.5

*SE = Standard Error on T-score metric

Step 11: If Raw score is between 7 to 40. Ask user to fill severity of generalized anxiety disorder questionnaire for identification and calculating severity of anxiety.
Step 12: Check if the No. of answered questions are greater than 3. If yes proceed to step 13, else proceed with step 15.
Step 13: Check if the No. of answered questions = 8. If yes proceed to step 7, else proceed with step 8.

Step 14: Calculate Raw scores.
Raw score = Sum of points answered on all the questions.
Step 15: Calculate prorated raw score.
Raw score = (Raw Score*(No. of questions answered))/10
Step 16: Ask the user to answer more questions. Calculate prorated raw score.
Raw score = (Raw Score*(No. of questions answered))/10
Step 17: Calculate average score.
Average score = Raw score/10
Step 18: If Average score =0: Output: You don't have an anxiety.
If Average score =1: Output: You have a mild anxiety.
If Average score =2: Output: You are diagnosed with moderate anxiety.
If Average score =3: Output: You are diagnosed with severe anxiety.
If Average score =4: Output: You are diagnosed with extreme anxiety.

5. **Identification of Anger**:

Anger is a completely normal, usually healthy, human emotion. But when it gets out of control and turns destructive, it can lead to problems—problems at work, in your personal relationships, and in the overall quality of your life. And it can make you feel as though you're at the mercy of an unpredictable and powerful emotion. Persistent Anger can lead to increased risks of hypertension and stroke, heart disease, gastric ulcers, and bowel diseases, as well as slower wound healing and a possible increased risk of some types of cancers.
Hence, we used following Algorithm based on DSM-V for identification of Anger.

(*https://www.apa.org/topics/anger/control*)

Step 1: First the user will fill cross cutting level 1 questionnaire, which consists of 23 questions classified in 13 domains for adults.
Step 2: Calculate individual domain scores
Step 3: Calculate highest domain score.
highest domain score = points scored on Q3.
Step 4: If the highest domain score is between 0 to 4 both including. Ask user to fill Level 2 Anger questionnaire.
Step 5: Check if the No. of answered questions are greater than 1. If yes proceed to step 6, else proceed with step 8.
Step 6: Check if the No. of answered questions = 5. If yes proceed to step 7, else proceed with step 8.
Step 7: Calculate Raw scores.
Raw score = Sum of points answered on all the questions.
Step 8: Calculate prorated raw score.
Raw score = (Raw Score*(No. of questions answered))/5
Step 9: Ask the user to answer more questions. Calculate prorated raw score.
Raw score = (Raw Score*(No. of questions answered))/5.
Step 10: Convert the Raw score in to T score using lookup table.

Raw Score	T-Score	SE
5	32.9	5.3
6	38.1	4
7	41.3	3.7
8	44	3.5
9	46.3	3.4
10	48.4	3.3
11	50.5	3.3
12	52.6	3.2
13	54.7	3.2
14	56.7	3.2
15	58.8	3.2
16	60.8	3.2
17	62.9	3.2
18	65	3.2
19	67.2	3.2
20	69.4	3.3
21	71.7	3.3
22	74.1	3.3
23	76.8	3.4
24	79.7	3.5
25	83.3	3.9

Step 11: If T score <=55.0: Output: You don't have Anger.
If T score 55.0<=59.9: Output: You are diagnosed with mild anger.
If T score 60.0<=69.9: Output: You are diagnosed with moderate anger.
If T score > 70.0 Output: You are diagnosed with severe anger.

6. Identification of Mania:

Mania is a period of extreme high energy or mood associated with bipolar disorder. Everyone's moods and energy levels change throughout the day and over time. But mania is a serious change from the way a person normally thinks or behaves, and it can last for weeks or even months. It makes sense that this could cause serious problems in a person's relationships, work, and school.
The complexity and severity of symptoms in mania often necessitate hospitalization and require rapid and effective symptom control. Agitation, irritability and severe aggression are common elements of mania and are associated with self-harm. We used following Algorithm based on DSM-V for identification of Mania.

Step 1: First the user will fill cross cutting level 1 questionnaire, which consists of 23 questions classified in 13 domains for adults.
Step 2: Calculate individual domain scores
Step 3: Calculate highest domain score.
highest domain score = points scored on Q4+Q5.
Step 4: If the highest domain score is between 0 to 8. Ask user to fill Level 2 Mania questionnaire.
Step 5: Check if the No. of answered questions are greater than 2. If yes proceed to step 6, else proceed with step 8.
Step 6: Check if the No. of answered questions = 25. If yes proceed to step 7, else proceed with step 8.
Step 7: Calculate Raw scores.
Raw score = Sum of points answered on all the questions.
Step 8: Calculate prorated raw score.
Raw score = (Raw Score*(No. of questions answered))/25
Step 9: Ask the user to answer more questions. Calculate prorated raw score.
Raw score = (Raw Score*(No. of questions answered))/25

Step 10: If Raw score <=5: Output: You are less likely to be associated with significant symptoms of mania.
If Raw score>6: Output: Report: score of 6 or higher indicates a high probability of a manic or hypomanic condition and may indicate a need for treatment and/or further diagnostic workup.

7. **Identification of substance abuse:**
Substance use and mental health disorders can create significant family, social, and health problems. However, they are preventable and treatable. Regular screenings can help identify substance use and mental health disorders early and help families access community resources quickly.
Early identification can improve health and relationships as well as help provide safe and nurturing environments. The goal of substance abuse screening is to identify individuals who have or are at risk for developing alcohol- or drug-related problems, and within that group, identify patients who need further assessment to diagnose their substance use disorders and develop plans to treat them.
Assessment is a thorough evaluation designed to definitively establish the presence or absence of a diagnosable drug abuse problem. The results of the assessment also suggest what is likely to be the most appropriate type of treatment when drug problems are evident.

(https://www.ncbi.nlm.nih.gov/books/NBK64820/)

We used following Algorithm based on DSM-V for identification of substance abuse.

Step 1: First the user will fill cross cutting level 1 questionnaire, which consists of 23 questions classified in 13 domains for adults.
Step 2: Calculate individual domain scores
Step 3: Calculate highest domain score.
highest domain score = points scored on Q21+Q22+Q23.
Step 4: If the highest domain score is between 0 to 12. Ask user to fill Level 2 Substance abuse questionnaire.
Step 5: Calculate Raw scores.
Raw score = Sum of points answered on all the questions.
Step 6: If Raw score >1: Output: Report: greater severity and complexity of substance use.

8. **Identification of Anger**:
Anger is a completely normal, usually healthy, human emotion. But when it gets out of control and turns destructive, it can lead to problems—problems at work, in your personal relationships, and in the overall quality of your life. And it can make you feel as though you're at the mercy of an unpredictable and powerful emotion. Persistent Anger can lead to increased risks of hypertension and stroke, heart disease, gastric ulcers, and bowel diseases, as well as slower wound healing and a possible increased risk of some types of cancers.
Hence, we used following Algorithm based on DSM-V for identification of Anger.

(*https://www.apa.org/topics/anger/control*)

Step 1: First the user will fill cross cutting level 1 questionnaire, which consists of 23 questions classified in 13 domains for adults.
Step 2: Calculate individual domain scores

Step 3: Calculate highest domain score.
highest domain score = points scored on Q3.
Step 4: If the highest domain score is between 0 to 4 both including. Ask user to fill Level 2 Anger questionnaire.
Step 5: Check if the No. of answered questions are greater than 1. If yes proceed to step 6, else proceed with step 8.
Step 6: Check if the No. of answered questions = 5. If yes proceed to step 7, else proceed with step 8.
Step 7: Calculate Raw scores.
Raw score = Sum of points answered on all the questions.
Step 8: Calculate prorated raw score.
Raw score = (Raw Score*(No. of questions answered))/5
Step 9: Ask the user to answer more questions. Calculate prorated raw score.
Raw score = (Raw Score*(No. of questions answered))/5.
Step 10: Convert the Raw score in to T score using lookup table.

Raw Score	T-Score	SE
5	32.9	5.3
6	38.1	4
7	41.3	3.7
8	44	3.5
9	46.3	3.4
10	48.4	3.3
11	50.5	3.3
12	52.6	3.2
13	54.7	3.2
14	56.7	3.2
15	58.8	3.2
16	60.8	3.2
17	62.9	3.2
18	65	3.2
19	67.2	3.2
20	69.4	3.3
21	71.7	3.3
22	74.1	3.3
23	76.8	3.4
24	79.7	3.5
25	83.3	3.9

Step 11: If T score <=55.0: Output: You don't have Anger.
If T score 55.0<=59.9: Output: You are diagnosed with mild anger.
If T score 60.0<=69.9: Output: You are diagnosed with moderate anger.
If T score > 70.0 Output: You are diagnosed with severe anger.