Mini project Final Evaluations

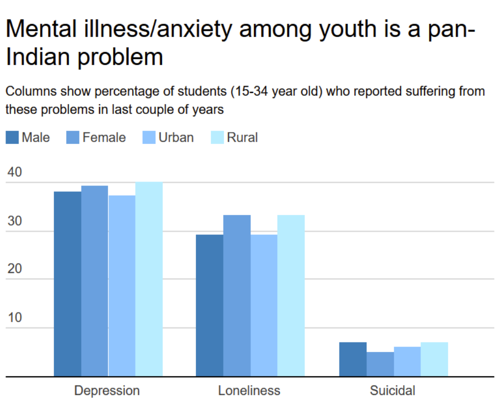
Analysing depression from User's Social Network Footprint

Final report MP10

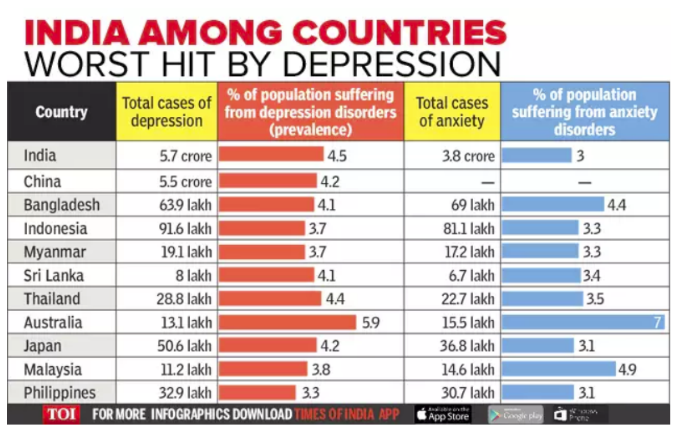
Team Members:

- Prajneya Kumar
- Nikhil Chandak
- Arvindh A
- Shri Vidhatri M M

Introduction and Motivation



[link]



INDIA AMONG COUNTRIES WORST HIT BY DEPRESSION

Country	Total cases of depression	% of population suffering from depression disorders (prevalence)	Total cases of anxiety	% of population suffering from anxiety disorders
India	5.7 crore	4.5	3.8 crore	3
China	5.5 crore	4.2	—	—
Bangladesh	63.9 lakh	4.1	69 lakh	4.4
Indonesia	91.6 lakh	3.7	81.1 lakh	3.3
Myanmar	19.1 lakh	3.7	17.2 lakh	3.3
Sri Lanka	8 lakh	4.1	6.7 lakh	3.4
Thailand	28.8 lakh	4.4	22.7 lakh	3.5
Australia	13.1 lakh	5.9	15.5 lakh	7
Japan	50.6 lakh	4.2	36.8 lakh	3.1
Malaysia	11.2 lakh	3.8	14.6 lakh	4.9
Philippines	32.9 lakh	3.3	30.7 lakh	3.1

TOI FOR MORE INFOGRAPHICS DOWNLOAD TIMES OF INDIA APP

[link]

Depression is a common mental disorder characterized by persistent sadness and a lack of interest or pleasure in previously rewarding or enjoyable activities. Depression is a condition that is marked by sadness, emptiness, hopelessness, and loss of interest for most of the day. Other indicators can include: significant weight loss or gain, insomnia or hypersomnia, fatigue/loss of energy, psychomotor agitation or retardation, feeling worthless, excessive guilt, inability to concentrate, thoughts of death, and suicide ideation

Experts believe there is more stress today than in previous generations. Stress triggers depression and mood disorders. Those predisposed to it by their creative wiring or genes are pretty much guaranteed some symptoms of depression at the confusing and difficult time of adolescence. Lack of community and family support, less exercise, no casual and unstructured technology-free play, less sunshine and more computer have factored into the equation.

Further, in such chaotic times, COVID-19 has raised anxiety levels among many persons. Psychiatrists say they see more people with depression and anxiety while every psychiatric disorder worsens in the midst of the second wave of the pandemic. This surge in depression and anxiety, while worrying, is not surprising given the numerous challenges the pandemic has posed to so many of us. People who reached out to a media outlet, *MNT,* spoke about recurrent anxiety, depression, panic, loneliness, and isolation. Readers mentioned several reasons for their anxieties, including fear for one's health and the health of a loved one, loss of income, being alone, and having too many parental responsibilities, to name only a few.

Why are ML-based things important for automatic analysis?

- Can find hidden patterns
- Can compute and discover more relations than a human
- Machine learning techniques identify high-quality solutions to mental health problems among Facebook users [link]
- The high volume of data, impossible to manually iterate [pastebin] (Notice the sheer amount of metadata for a single tweet)

- ML makes textual analysis much faster and more efficient than manual processing of texts.

Why is social network footprints more informative for this analysis?

- People post to express their feelings more on social media than in real life nowadays, as it's not face-to-face.
- In recent years, several studies have linked heavy social media use to an increased risk of depression.
- Overall, depression risk rose in tandem with time spent on social media. Compared with the lightest users (2 hours or less per day), the heaviest users (at least 5 hours per day) had three times higher depression risk. Meanwhile, that risk was two times higher among young adults who were active on social media, around 3.5 to 5 hours per day.
- We find that social media contains useful signals for characterizing the onset of depression in individuals, as measured through a decrease in social activity, raised negative effect, highly clustered ego networks, heightened relational and medicinal concerns, and greater expression of religious involvement [link]
- Problematic SNS usage is significantly and positively related to depression and Neuroticism while negatively associated with Agreeableness [link]
- There is a significant positive relationship between problematic SNS usage and depressive symptomatology.
- O'Dea et al. [link] examined that Twitter is progressively researched as a method for recognizing psychological well-being status, including depression and suicidality in the population
- The researchers reported good performance in detecting depression by analyzing a variety of features such as lexical features, including symptom lexicons (De Choudhury et al., 2013a; De Choudhury et al., 2013b; Coppersmith et al., 2014), syntactic features (Nambisan et al., 2015; Gkotsis et al., 2016), sentiment analysis (Wang et al., 2013; Preoţiuc-Pietro et al., 2015), or topic modelling (Resnik et al., 2015; Preoţiuc-Pietro et al., 2015)

- Language in social media activities is known to represent the current state of writers, including their mental health. By analyzing the language used in social media, many researchers have discovered a way to identify depressed individuals.
- Many researchers have demonstrated that utilizing user-created content (UGC) accurately may help decide individuals' psychological wellness levels. For instance, Aldarwish and Ahmad [link] examined that the utilization of Social Network Sites (SNS) is expanding these days, particularly by the more youthful eras, because the accessibility of SNS enables clients to express their interests, sentiments and offer day-by-day schedules [link, link].

Behavioural Analysis of Depression

1. Depression often varies according to age and gender, with symptoms differing between men and women or young people and older adults. Thus it is important to know the target audience before mass analysis of depressive symptoms.
2. Depression often varies according to age and gender, with symptoms differing between men and women or young people and older adults. Thus it is important to know the target audience before mass analysis of depressive symptoms.
3. Dysthymia is a type of chronic "low-grade" depression. It is also called a persistent depressive disorder. More days than not, you feel mildly or moderately depressed, although you may have brief periods of normal mood. This is commonly seen in students and teenagers across the country due to stress, peer pressure and familial issues.

▼ DSM-IV criteria for Major Depressive Episode detection from mild to severe

   1. Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g., feels sad or empty) or observation made by others (e.g. appears tearful) Note: in children and adolescents, can be irritable mood.

2. Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation made by others)
3. Significant weight loss when not dieting or weight gain (e.g., a change of more than 5% of body weight in a month), or decrease or increase in appetite nearly every day
Note: In children, consider failure to make expected weight gains
4. Insomnia or hypersomnia nearly every day
5. Psychomotor agitation or retardation nearly every day (observable by others, not the merely subjective feeling of restlessness or being slowed down)
6. Fatigue or loss of energy nearly every day
7. Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick)
8. Diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others)
9. Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide

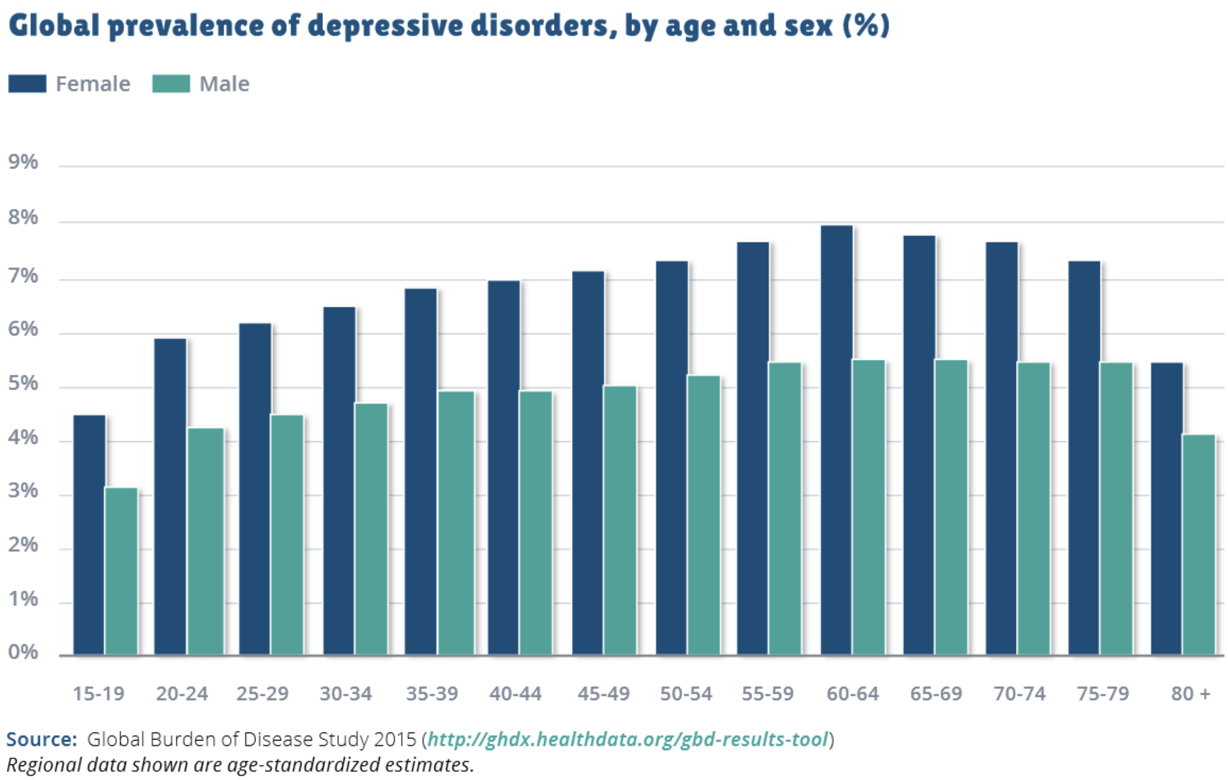
Relevant Work and Statistical Analysis

Mental disorders have a vast literature that is too broad to summarize here. However, most of the work has used clinical trials, cognitive approaches, and medical treatments to address depression directly. Only recently has there been a rise in the use of computational models (especially Machine Learning techniques) to tackle depression and provide more effective intervention techniques. We below provide a brief literature review.

Grover et al. [link] conducted a review in 2010 focusing on research done on various depressive disorders in India. A thorough internet search was done using keywords like depression, life events, prevalence, classification, cultural issues, outcome, prevention, disability and burden, etc., in various combinations. The various search engines like Pubmed, Google Scholar, ScienceDirect, Search Medica, Scopus, And Medknow etc., were used. In addition, a thorough search of all the issues of the Indian Journal of Psychiatry available online was done. The key results they found were:

- Studies have shown that compared to healthy controls and subjects with schizophrenia, and depressed patients have a significantly greater number of life events prior (6-12 months) to the onset of their illness.
- Compared to patients with mild depression, patients with moderate and severe depression tend to use avoidance as a coping strategy more frequently for stressful life events, suggesting that it may be a maladaptive way to cope with the situation, which is responsible for the development of depression.
- Economic and interpersonal relationship difficulties, partner violence, sexual coercion by the partner are common causal factors related to the development of depression in general and depression during the antenatal and postnatal periods.
- Studies report that somatic symptoms are the most common manifestation of depression in India.

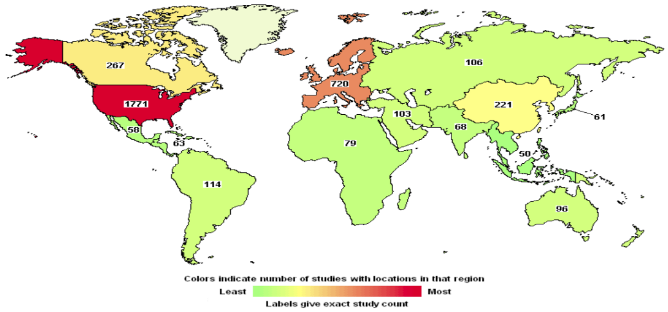
Another interesting fact discovered during the literature review was that while the rate at which people are getting affected by depression is highest among teenagers, the prevalence of depression, however is highest among the elderly, as shown in the histogram below:



Given the high prevalence of depressive disorders among the older age groups, Pilania et al. [link] conducted a systematic review and meta-analysis to estimate the prevalence of depression among the elderly population in India, lack of information on the magnitude of depression among them.

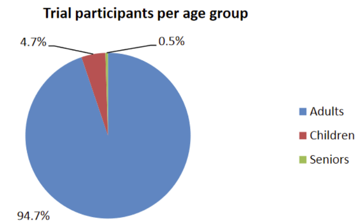
In association with WHO, Cesar et al. [link] had extended a background paper on depression [2004 Priority Medicines for Europe and the World from 2004] describing demographic trends and the burden of disease of major depression disorder around the world. As we can witness, India has a particularly less number of research studies done on depression, thereby motivating the cause for our report.

Figure 6.15.16: Geographic map number of studies on depression, per region worldwide.



Source: clinicaltrials.gov[216]

Figure 6.15.17: Percentages of participants in depression trials per age group.



Source: clinicaltrials.gov[216]

Approach and Extrapolation of Behavioural Analysis

Proceeding further from the behavioural analysis and relevant work discussed earlier, we narrowed out four major problems and missing gaps that persisted in existing studies. These were:

1. Accuracy
   There is no true way to find out about "true negatives" and "true positives" with given datasets, particularly in samples of social media data. This is because the data found or collected via websites or apps such as Twitter, Facebook etc., is usually unidentified.

2. Data Typology
   In most of the studies, the data that is typically studied consists of textual and image data. We believe that analysing the social network of a user might give further insight than just data about posts.

3. Traceable Depth
   Doctors most of the time need to know about the factors or parameters that lead to the final prediction of depression to analyse the situation better. Providing simple boolean data about a user(whether they are depressed or not) might not suffice and beats the original cause of taking up the project.

4. Lack of Real-Time Interventions
   Even after predicting or analysing users regarding their mental health issues, all effort goes in vain if immediate actions are not taken against the same. A

mechanism to inform doctors/close family members and friends based on the analysis might also improve the interactivity and indulgence of the project.

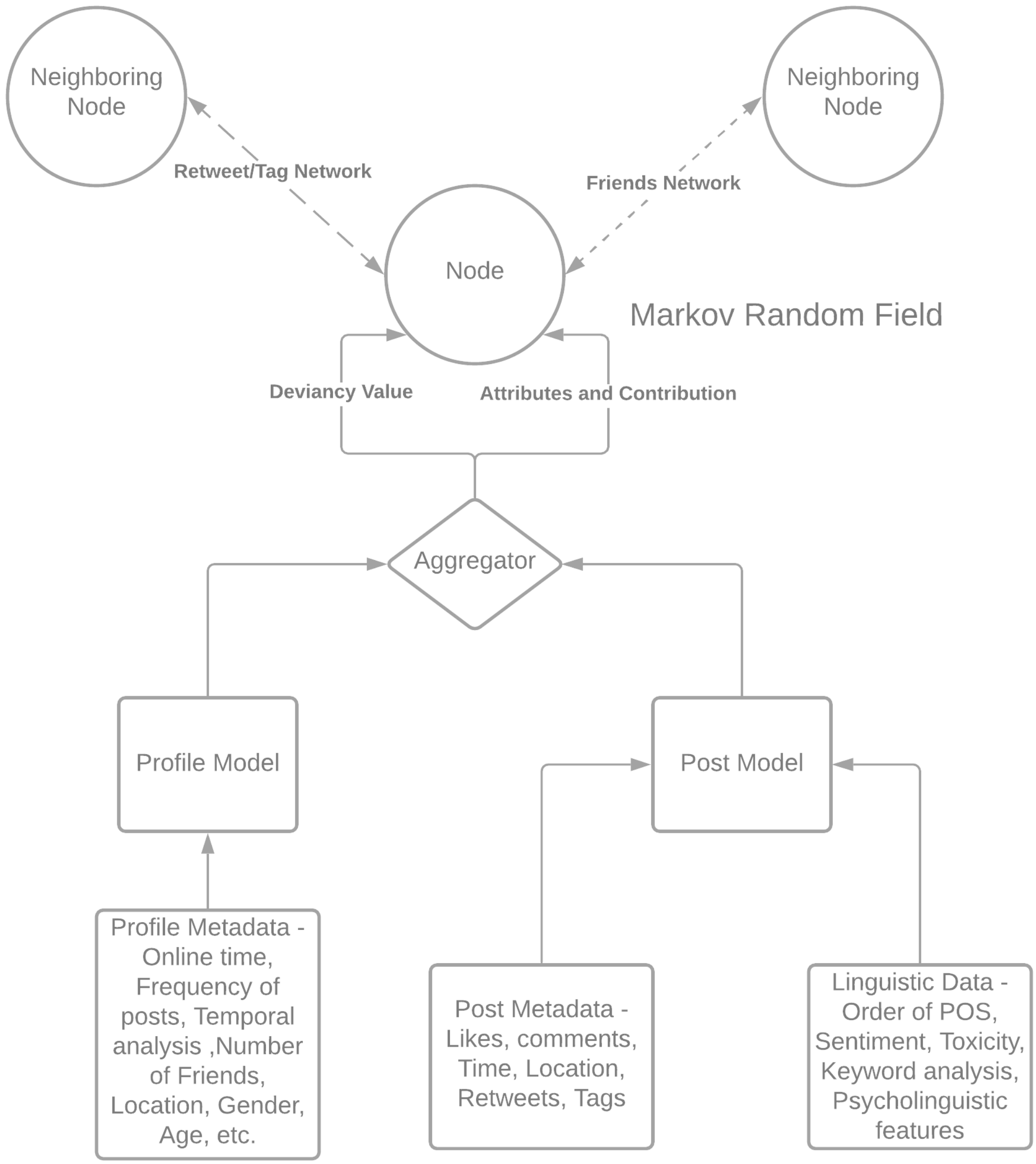
Keeping the above four problems in mind, we draw our approach towards the aforesaid problem in four particular phases. The demarcation of the approach in certain phases is to tackle the problem of Traceable Depth. Categorizing a user into a phase will allow doctors and experts to know about factors leading to the prediction more intricately. The four phases thus are as follows:

1. Phase 1: Detection of Disconnectedness
   Initially, patients showing early signs of depression might show inactivity and disconnectedness from their environment. This can be seen in their social media activity and the active network that they interact with. We maintain two networks of each user: the actual social network and the active social network. The comparison of these two networks over time might help us analyse depression in its growing stages while simultaneously tackling the "Data Typology" problem mentioned earlier.
2. Phase 2: Textual Data and Image Analysis for Global Signs of Depression
   After analysing behaviour as described in the earlier phase, we apply machine learning techniques to actual data posted by the user. We look for usage of words that imply that the user talks about their problems in a global/transitional mode. This includes the usage of generic terms such as: “No one will ever love me;” “I will always fail at everything I try;” What is missing here is the capacity to focus attention on a real situational problem with one individual that occurred at a particular time.
3. Phase 3: Stagnant Behaviour over long periods of time
   Chronically depressed patients are not motivated to change their behaviour. We analyse this behaviour in phase three by looking at user activity over long periods of time. For example, inactivity for a month or two after high activity combined with the other two phases might denote symptoms of depression.
4. Phase 4: Interpersonal Detachment along with signs of Hostility
   There are many signs of interpersonal detachment in patients suffering from chronic depression, which is at times accompanied by hostility towards others. People engaging in social media wars/comment wars after showing signs of

depression from the above three phases have a high probability of leading a sad and depressed life.

It must also be noted that instead of concluding with a yes/no answer to depression, we are concluding with a phase-wise categorization of the user with a certain estimate as to how risky his condition is towards depression. This tackles the problem of accuracy, as discussed earlier.

Feature extraction: Application of Machine Learning



Structure of the Model Ensemble

- We use Belief Networks and Markov random field to propagate “Deviant” behaviour.
- MRFs are a class of graphical models particularly suited for solving inference problems with uncertainty in observed data. MRFs are widely used in image restoration problems wherein the observed variables are the intensities of each pixel in the image. The inference problem is to identify high-level details

such as objects or shapes. An MRF consists of an undirected graph, each node of which can be in any finite number of states. The state of a node is assumed to statistically depend only upon each of its neighbours and independent of any other node in the graph. The dependency between a node and its neighbours is represented by a Propagation Matrix (ψ), where $\psi(i, j)$ equals the probability of a node being in state j given that it has a neighbour in state i.

- The main motivation behind the usage of MRFs is that some people have a shallow but niche footprint, which may be enough to conclude if a person is deviant but may escape/be misjudged by the model, which can be caught using the propagation.
- Each node also stores the distribution of the features that contributed to the total Deviancy value hence addressing the Traceability concern.
- We can adapt an incremental version of MRF, limiting the propagation but keeps the difference within the error margin, which can drastically increase the computational speed and make it suitable for Real-Time Analysis.
- Here, social media accounts with very little profile and post metadata are considered uncertain, and the "deviancy" values assigned to them are only propagated.
- 2 network layers - Friends layer, Retweet and Tag layer

Profile and Post Metadata

- Metadata like location can also be an important factor in the analysis as certain places could be associated with a higher concentration of depressed individuals.
- People who go through the same experiences in day to day life tend to have a similar outlook of the world, so exploring the friend network of a depressed individual could unlock more potential research.
- Wolfradt and Doll (2001) suggested that gender is a major factor to consider when researching Internet use or use of Social Networking Sites.
- Geographical location was researched as a factor mainly in terms of access to the internet and the “digital divide”, that is the unequal access to computers

and the internet (Dewan & Riggins, 2005); urban areas are more likely to have internet access than rural areas, even though this fact is rapidly changing. A study on Norwegian adolescents (Johansson & Götestam, 2004) showed that the frequency of problematic Internet use was relatively higher in small cities and rural areas than in large cities (more than 150,000 inhabitants).

- Lastly, the number of posts and the temporal analysis of the activeness of an individual could be a key measure as well, as the activity tends to drop a lot or shoot up a lot, seeking attention if one falls into depression.

Linguistic Analysis

- We can use pre-existing vocabulary packages for deep and detailed analysis of the text like the LIWC.

  Psycholinguistic features LIWC is a psycholinguistic vocabulary package made by psychological analysts to perceive the different affective, intellectual, and etymological parts that lie on the user's verbal or written correspondence. It returns more than 70 different factors with a higher level of psycholinguistic features, for example,

  - Psychological process—effective process, social process, cognitive process, perceptual process, biological process, drives, time orientations, relativity, personal concerns
  - Linguistic process—word count, word/sentence, pronoun, personal pronoun, articles, prepositions, auxiliary verbs, adverbs, conjunctions, Negations
  - Other grammar—verbs, adjectives, comparisons, interrogatives, numbers, quantifiers.

  These higher-level categories are also divided into subcategories such as

  - Biological processes—sexual, body, ingestion, and health
  - Affective processes—anxiety, anger, sadness, positive emotion, negative emotion
  - Time orientations—present, past, future
  - Social processes—family, friends, male, female

    - Perceptual processes—see, hear, feel.
- Identify Depressive symptoms using evidence keywords taken from a lexicon of nine groups of depressive symptoms in the Diagnostic and Statistical Manual of Mental Disorders (DSM-V)
- Analyze the sentiment of the posts as depressed people tend to have negative polarity in their posts
- Identify Ruminative thinking patterns as depressed people tend to have repetitive thoughts.
- Looking at the POS (Part of Speech) level in their writings as their writing style tends to contain a different distribution of nouns, verbs, and adverbs and the complexity of sentences (Gkotsis et al., 2016)

Conclusions

Detailed requirements for the analysis of depression in users via their social network footprint has been provided. There has been a mention of how depression is diagnosed clinically and the key takeaways from the clinical diagnosis for the requirement. There has been a mention of missing components in existing research and a systematic approach to tackle the same.

Future work

1. The problem of the lack of real-time interventions is something that is still not addressed by our methodology. This can be improved upon to inculcate an even secure model that handles the depression of people all around the world delicately and smoothly.
2. There is a need to pre-process available data to make a more precise prediction and analysis of depression.

3. Information such as age, gender, location, marital status, the financial situation, along with the data, can be highly beneficial while analysing the data for psychological disorders