

RESEARCH

Identifying Factors that Affect Filipino's Mental Health using Topic Modelling and Keyword Extraction in r/mentalhealthph Subreddit.

Learning Team 7*, Hilbert Lazatin†, Claudine Licong† and Matthew Romero†

*Correspondence:

hilbertlazatin.msds2021@aim.edu

Asian Institute of Management,

Paseo de Roxas St., Makati, PH

Full list of author information is

available at the end of the article

†Equal contributor ‡Team number

Abstract

As mental health encompasses psychological, mental, and emotional well-being, it is imperative that an individual has an optimum mental state. In this time of the pandemic, however, worsening mental health has been observed. In this study, we aim to identify several factors that affect Filipinos' mental health during the quarantine. To do this, posts from the r/mentalhealthph forum in Reddit were scraped and afterward cleaned through text normalization and lemmatization. A topic model was then computed for all posts in the r/mentalhealthph forum in Reddit using Latent Dirichlet Allocation. A model containing 5 topics was selected based on the topic coherence metric, which intuitively, relates to the prevalent mental health issues in the Philippines. The posts were then segmented per topic, and keyword extraction using the TextRank algorithm was performed for the segmented posts to identify potential factors that affect the mental health issues identified from topic modeling.

The 5 topics identified from LDA are ADHD, Depression, Anxiety, Seeking Help, and Online Therapy. Among the five topics identified, ADHD was the most common as 52% of the posts were found to belong to such a topic. This is concerning given the 4.1% increase in the number of children with ADHD from 1997 to 2015. We also observed that the number of posts, regardless of the topic, doubled when the nationwide enhanced community quarantine was imposed, which shows the impact of the COVID-19 on the mental health of Filipinos.

Our keyword extraction using TextRank also revealed a few potential factors that affect mental health issues. For example, we observed that "work" is commonly ranked as an important word for depression and anxiety-related posts, which suggests that Filipinos are suffering from burnout in the workplace. This in fact agrees with a study that shows Manila, Philippines as 5th worldwide in terms of burnout rate. We also identified words such as "parents", "family" and "money" as important words for depression related posts. This may imply that family problems and poverty are triggers of depression for Filipinos. Lastly, the results from keyword extraction emphasize the importance of family and friends in the battle against mental health problems.

For future studies, we propose to explore other methods that leverage semantic relationships between words, in extracting factors that affect mental health issues. Additionally, we can translate Filipino posts to English to preserve other useful information. Lastly, other online forums related to mental health can be explored to gather more data.

Keywords: natural language processing; latent dirichlet allocation; keyword extraction; TextRank

Introduction

Mental illness is widespread in the Philippines, making it the country's third most common disability. An estimate of around 6 million Filipinos lives with anxiety and/or depression, which places the Philippines on the third spot in the Western Pacific Region with the highest rate of mental health problems. [1] Barriers to seeking help such as inaccessibility of facilities, hesitation due to prohibitive beliefs, and monetary constraints are contributing to the high rates of mental health problems. This is even worsened by the effects of the COVID-19 lockdowns as Filipinos are forced to abruptly adopt behaviors such as working/studying from home and avoidance of huge social gatherings that blur the lines between home activities and activities which used to be predominantly done outside such as working and schooling.

In this study, we aim to solve the problem of identifying the factors that affect the mental health of Filipinos from various commenters on the Reddit platform, in hopes of discovering insights that will help propose effective interventions in treating and also in preventing mental health problems.

Mental health awareness in the Philippines is relatively new and has been increasing, with the passing of the Republic Act No. 11036 in 2017, known as the Mental Health Act. Given the rate at which the country is progressing in terms of recovering from the pandemic, it is best for the Filipinos to prepare for the mental health problems that may come along with it.

Related Literature

A study shows that there are numerous barriers discouraging Filipinos from seeking help from mental health professionals. These include cultural variables such as stigma, collective beliefs, and shame.[2] Inaccessibility is also a hindrance especially for those Filipinos located in rural areas since most facilities are available in urban areas.[3] Additionally, mental health services may be monetarily prohibitive given that a huge percentage of the population is living below the minimum daily wage, driving people to resort to cheaper alternative medicines, where 50-70% of the population use traditional and complementary medicines.[4] Consequently, many Filipinos resort to traditional and indigenous practices such as folk healing practices instead of having their conditions diagnosed by mental health professionals.[3]

Technological advancements lower these barriers nowadays as people turn to online services to search information and seek help about their mental health problems. A new approach is also being introduced to reduce the prevalence of mental health disorders and improve mental health by way of digital technology- that is the collection, analysis and use of data from digital devices.[5] It asserts for the need to create outcome optimization teams by integrating data from various sources and using advanced data analytics and new designs to develop strategies to increase reach, uptake, use, and outcomes for both preventive and treatment interventions.

With numerous online sites that serve as avenues for people to share about their mental health concerns, a study was made employing topic modelling to extract quantitative objective description of a mental health Internet Support Group (ISG). [6] Results show that users engage in topics with a greater focus on experiential knowledge, disclosure and informational support, a pattern resembling the clinical

symptom-focused approach to recovery. Results suggest that super users serve the role of emotionally supportive companions who are ‘active help providers.’

Various mental health concerns are raised in online sites such as work-related ones, which construe the study that says working for long hours or irregular hours deteriorates the mental health of workers, but its impact is likely to differ significantly across job types such as that of white-collar and blue-collar jobs.[7] The most pressing concern nowadays is the effect of the COVID-19 pandemic. A study on Filipino students revealed that the majority who experienced online-blended learning expressed negative impressions due to slow internet connectivity.[8] In addition, most students displayed anxiety during the entire period of lockdown. 62.64% were worried about food and financial resources, and about 54-56% of the students avoid social contact, large meetings, and gatherings. Although little is known about the effects that online services have on mental health, cross-sectional studies reported that an average of 35% of users facilitated seeking help services from a professional source online.[9] Across studies, young people were satisfied with online mental health resources.

Methodology

In identifying factors that affect the mental health of Filipinos, we underwent several wrangling and natural language processing methods, as described in the next few subsections.

Data Acquisition

Data was collected from Reddit under the r/mentalhealthph thread. The r/mentalhealthph thread contains stories and discussion of Filipinos in relation to mental health issues. The data was scraped using the Python Reddit API Wrapper (PRAW) library. A total of 2396 posts were scraped from June 24, 2018 to December 2, 2020.

The dataset has the following features:

- **author**: creator of the post
- **title**: brief/initial description of post
- **upvotes**: number of positive votes on the post
- **id**: unique identifier of post
- **num_comments**: total number of comments on the post
- **body**: main content of the post. **NaN** if there is no content
- **all_comments**: comments in list format
- **created**: UTC format of the time origin of the post
- **time_stamp**: human readable format of **created**

A sample of the dataset is shown in **Figure 1**.

Data Pre-processing

Data Cleaning and Formatting

After selecting the appropriate data and features, we then cleaned the data to improve its quality. In cleaning the data, we converted all texts to lowercase and removed plenty of stopwords, some of which were in Filipino. Afterwards, we tokenized and applied some lematization methods on the preprocessed dataset.

Figure 1: First 5 Entries of the Dataset

	author	title	upvotes	id	num_comments	body	all_comments	created	time_stamp
0	Fentany1223	FINALLY GOT HIRED AFTER BEING UNEMPLOYED FOR 3...	163	j8t46u	40	Even if I'm super anxious to accept the job be...	['Thank you for posting in r/MentalHealthPH. P...	1.602114e+09	2020-10-07 23:39:28
1	bam03_	Laban us, mga beshyl Meme is sometimes the ans...	140	kh4svm	3	NaN	['Thank you for posting in r/MentalHealthPH. P...	1.608535e+09	2020-12-21 07:20:33
2	BoredAndCurious_	Am I the only one? (ctto)	115	iq45yj	16	NaN	['Thank you for posting in r/MentalHealthPH. P...	1.599776e+09	2020-09-10 22:18:46
3	bam03_	More patience pls	105	io61r4	2	NaN	['Thank you for posting in r/MentalHealthPH. P...	1.599508e+09	2020-09-07 19:49:55
4	AtoySporkchop69	free consultation	90	jvub45	5	NaN	['Thank you for posting in r/MentalHealthPH. P...	1.605654e+09	2020-11-17 22:54:09

Tokenization and Lemmatization

Us humans, when we read text, we immediately understand the meaning and context behind it. Computers however don't understand natural language as much as we do. In order for the machine to better understand the texts we want analyze, we perform natural language processing methods prior to making any computations and analysis. As such we perform two important methods in NLP, which are tokenization and lemmatization. **Tokenization** refers to somehow chopping a sentence into individual words, excluding punctuation and stopwords. This is done so we can identify important, as well as common words that needs to be removed in the dataset, and see how this affects the results of our analysis later on. On the other hand, **Lemmatization** reduces the words in a sentence to their main root word by reducing or reverting the inflectional forms of the word to their original form. Lemmatization allows us to understand and have a context of what the text, in this case posts, are really saying, at a high level. Doing these methods aides the machine in understanding texts better.

Feature Extraction

Bag-of-Words

We then extract information from it by getting its bag of words representation. This is also performed so that the text can be represented as vectors, which can be parsed by the machine.

Latent Dirichlet Allocation

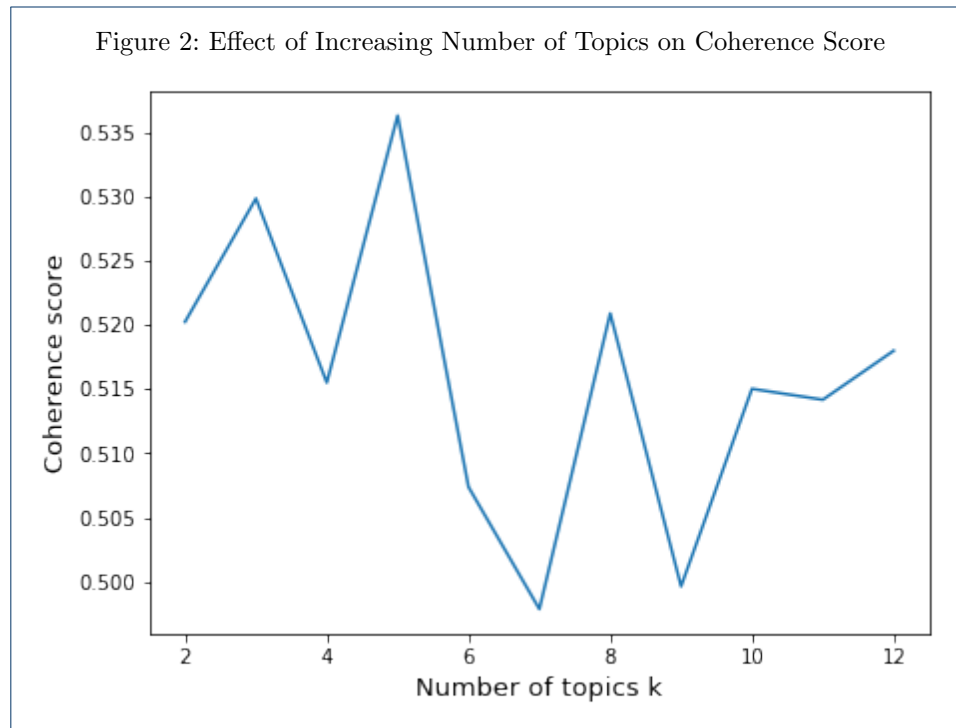
After feature extraction, we implemented a topic modeling using the Latent Dirichlet Allocation to identify the topics of the documents. LDA is an unsupervised generative probabilistic model of a corpus which works under the assumption that a document is made up of several topics, and these topics are made up of several words.[10]

Selecting the number of topics

For selecting the number of topics, we used topic coherence as a metric instead of the more common metric, which is the perplexity. This is because recent studies have shown that predictive likelihood (or equivalently, perplexity) and human judgment are often not correlated. This implies that optimizing the number of topics based on perplexity may not yield interpretable results [11].

Such limitations are addressed by topic coherence. By intuition, a statement or a topic can be considered as coherent if the sentences or words are logical and

support one another. This is basically the intuition that is supposed to be captured by topic coherence. In the context of topic modeling, this is measured based on the degree of semantic similarity between high scoring words in the topic. These measurements help distinguish between semantically interpretable topics and topics that are artifacts of statistical inference [11].



From the figure, the coherence score is maximum at 5 topics with a coherence score of 0.54. We therefore perform the topic modeling using 5 topics, and each document were then assigned a topic.

Keyword Extraction with TextRank Algorithm

After identifying the topics (i.e. the mental health issues), we perform keyword extraction for all documents belonging to a certain topic. The idea is that using keyword extraction, we may be able to identify the factors that affect the mental health of Filipinos. For the keyword extraction, we employ the TextRank algorithm [12].

The TextRank algorithm, essentially works similar to the Pagerank algorithm. Unlike Pagerank however which ranks webpages, TextRank uses texts. Furthermore, TextRank treats texts as nodes and edges as similarity scores.

Results

Topic Modeling Results

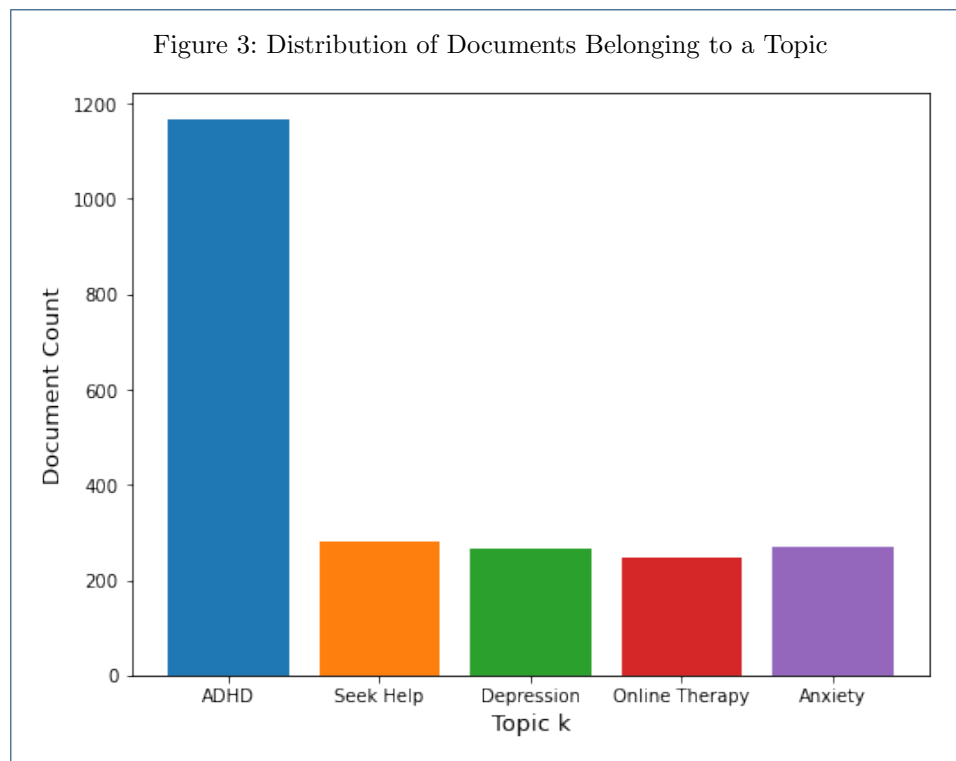
After performing the topic modeling via LDA, we identified the most frequent words for each topic, and from which, the appropriate label were inferred, as shown in table 1 below.

Table 1 shows the corresponding labels for the topics identified using LDA. Based from the most common words for topic 1, we deduced that this topic is about ADHD.

Table 1: Topic model obtained via LDA

Topic #	Topic Label	Top 5 Key Words
1	ADHD	adhd, psychologists, psychiatrist, need, want
2	Seeking Help	need, feel, psychiatrist, friend, know
3	Depression	depression, disorder, life, parent, therapist
4	Online Therapy	online, consultation, therapy, health, first
5	Anxiety	anxiety, support, group, think, thinking

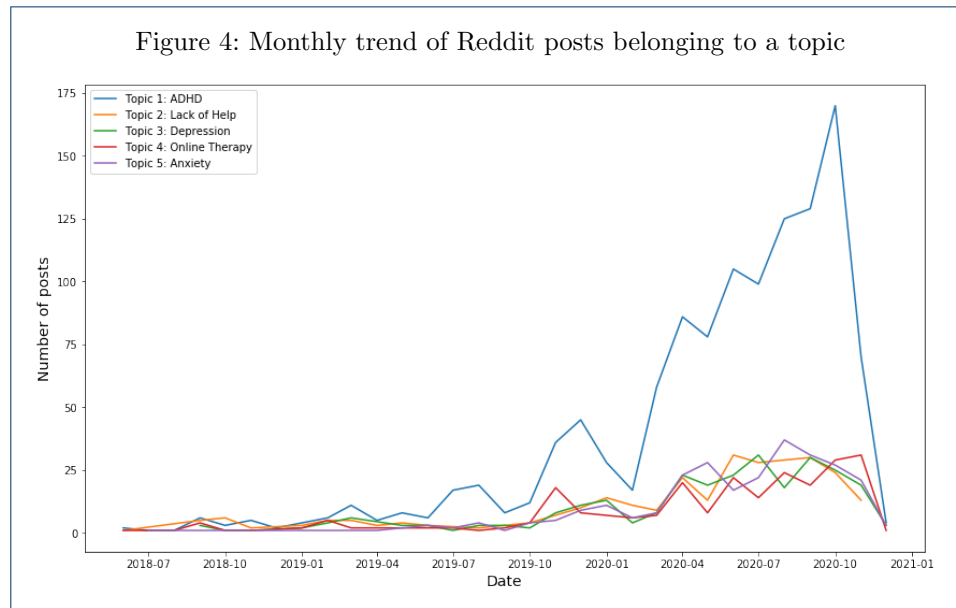
Topic 2 is about seeking help as evident from common words need, feel, psychiatrist, friend. Topic 3 is about depression and other mental disorders as evident from the dominant words for the topic. Topic 4 is about seeking other avenues of help, for example, online consultation and therapy. Lastly, and as such, was labeled as online therapy. Lastly, topic 5, was labeled as anxiety based from the most common words. Furthermore, we also identified the distribution of the topics in the corpus by assigning the topic to all documents, as shown in figure 3.



Majority of the posts (52%) relates to ADHD as shown above. This likely suggests that a number of Filipinos are affected by ADHD. In fact, a study from the University of Iowa have shown that the number of ADHD cases for children aged 4 to 17, have increased from 6.1% to 10.2% for 1997 to 2015 and the reason is still unknown [13]. This can be problematic as it has been shown that ADHD can be a precursor to other mental health issues such as bipolar disorder.[14] The remaining topics on the other hand, are almost equally distributed, with approximately, 12% each.

Time Series Trends in the Number of Posts

To uncover insights into how the posts per topic has progressed since the online forum was established up to the present, we identified the monthly number of post per topic as shown in figure 4:



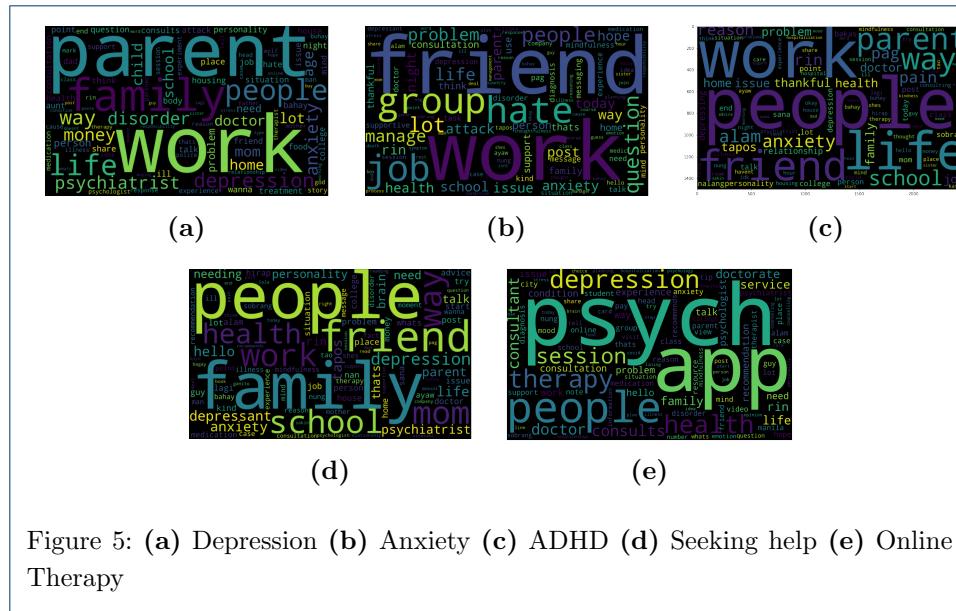
It can be observed that the number of posts are relatively flat from 2018-2019. This could be due to lack of awareness into the existence of the forum. It can also be observed that there is a surge in the number of posts around April, 2020, which is 2-3 weeks since the enhanced community quarantine was imposed. This suggests that mental health issues was aggravated by the lockdown, and as such, has caused more people to resort to online forums in alleviating their mental health problems.

Moreover, we also see a huge spike in the number of ADHD related posts in October. This huge spike also coincides with the fact that October is also the ADHD awareness month [15]. This gives us insight that events such as awareness months are effective in raising awareness towards various issues. As such, health agencies could look into other mental health problems that lack awareness from the public, and have awareness events.

Factors that affect Mental Health Issues with Keyword Extraction

After identifying the prevalent mental health issues among Filipinos with LDA topic modeling, we then identify the factors or potential triggers of these mental health issues through keyword extraction. As mentioned in the previous section, we perform this through the TextRank Algorithm. Using the said algorithm, we were able to obtain the following results:

Figure 5 shows the keywords extracted from the documents belonging to each topic. Figure 4a are the keywords extracted for posts related to depression. Keywords such as parents, family, work are seen. It is possible that these are factors that induce depression among Filipinos. Figure 4b, which is related to anxiety, has keywords such as friend, work, group, hate. We suspect that anxiety among Filipinos



are induced because they hate their work or jobs. Additionally, it could also entail their need for friend or a group which could help them overcome such. Figure 4c is for ADHD. While the root cause of ADHD is known, scientists have confirmed that it is not something developed, and is likely to be due to genetic factors. Hence, the keywords extracted may not be interpreted as factors that induces ADHD. However, we may interpret it as how it affects aspects of our lives. For example, ADHD can perhaps affect relationships with people, friends and work and school performance as evident from the keywords extracted.

Moreover, the word "work" is shown to be one of the most important words for depression and anxiety related topics. This is indicative of a burn-out culture found in Philippine workplace. In fact, a recent study which shows Manila as 5th among world cities with the highest burn out rate, reinforces our result [16].

Figure 4d and Figure 4e, are related to seeking help and online therapy, respectively. Seeking help consists of keywords such as people, friend and family suggesting the importance of support from love ones. Lastly, for posts related to online therapy, a common word shown is "app", which may be indicative of the need for a mobile application that Filipinos can readily use to seek help for addressing mental health problems.

Conclusion and Future Works

In this study, we successfully implemented topic modeling and keyword extraction to identify some of the factors that affect the mental health of Filipinos. Using Topic Modeling, we identified the prevalent mental health issues to be ADHD, Depression and Anxiety. Among the three mental health disorder, ADHD is shown to be the most prevalent as 52% of the posts are related to such. This is concerning given that ADHD can be a precursor to other mental health issues such as bipolar disorder. Additionally, we found out that Filipinos are using online platform such as Reddit to seek help from professionals whether physically or online.

Using TextRank algorithm, we also identified what could potentially be factors that drive or affect these mental health issues. We discovered that "work" is one of the most important words for topics related to anxiety and depression. This is likely indicative of a burn out culture in the Philippine work place, and in fact, agrees with a study that shows Manila, Philippines which ranked 5th among world cities with the highest burnout rate. Also, for depression, aside from work, parents, family and money were found to be most important words, which suggests that poverty and family problems are also factors that affect depression. Our results also emphasize the importance of family, friends and surrounding oneself around people, as it was shown to one of the most important word for all five topics.

To improve the results of the study, we propose to explore more NLP techniques in identifying triggers or factors that affect mental health issues. TextRank algorithm may not fully capture the semantic relationship between words, and other state of the art methods can be utilized. Filipino words were not also considered in the analysis, due to lack of pre-processing techniques that support the language. This however, may lead to loss of valuable insights. As such, we may opt to use language translation techniques so that there is minimal loss of information and a better understanding into the mental health issues of Filipinos can be obtained. Lastly, we can also explore other forums so we can have more data.

Data Availability

The implementation and datasets used in this paper are available from the corresponding author upon request.

Competing Interests

The authors declare that they have no competing interests.

Author's contributions

The authors of this study are the first to utilize topic modeling and keyword extraction in identifying factors that affect mental health of Filipinos. Furthermore, the authors are the first to use Reddit as the source of Data, as opposed to existing studies which used Twitter or Facebook for sentiment analysis only.

Acknowledgements

The completion of this study would not be possible without the guidance of our Natural Language Processing professor, Prof. Madhavi Devaraj, PhD.

References

1. Mental Health Atlas 2017. World Health Organization
2. Tuliao, A.P.: Mental health help seeking among filipinos: a review of the literature. *Asia Pacific Journal of Counseling and Psychotherapy* **5**(2) (2014)
3. Conde, B.: Philippines mental health country profile. *International Review of Psychiatry* **16**(1-2), 159–166 (2004)
4. Lagaya, A.T.i.G.B., Ong, C.K., Grundy, C., Burford, G., Shein, K.E.: Who global atlas of traditional, complementary, and alternative medicine. Kobe: World Health Organization Centre for Health Development, 199–204 (2005)
5. Taylor, C.B., Ruzek, J.I., Fitzsimmons-Craft, E.E., Sadeh-Sharvit, S., Topooco, N., Weissman, R.S., Eisenberg, D., Mohr, D., Graham, A., Jacobi, O.B. C.: Using digital technology to reduce the prevalence of mental health disorders in populations: Time for a new approach. *Journal of Medical Internet Research* (22(7):e17493) (2020)
6. Carron-Arthur, B., Reynolds, J., Bennett, K.e.a.: What's all the talk about? topic modelling in a mental health internet support group. *BMC Psychiatry* **16**(367) (2016)
7. Sato, K., Kuroda, S., Owan, H.: Mental health effects of long work hours, night and weekend work, and short rest periods. *Social Science Medicine* **112**774 (2019)
8. Ja'ashan, M.: Perceptions and attitudes towards blended learning for english courses: A case study of students at university of bisha mohammed. *English Language Teaching* **8**(9), 40–50 (2015)
9. Kauer, S., Mangan, C., Sanci, L.: Do online mental health services improve help-seeking for young people? a systematic review. *Journal of medical Internet research* **16**(e66.10.2196/jmir.3103) (2014)
10. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. *Journal of Machine Learning Research* **3**(Jan), 993–1022 (2003)
11. Kapadia, S.: Evaluate Topic Models: Latent Dirichlet Allocation (LDA). <https://towardsdatascience.com/evaluate-topic-model-in-python-latent-dirichlet-allocation-lda-7d57484bb5d0>
12. Liang, X.: Understand TextRank for Keyword Extraction by Python. <https://towardsdatascience.com/textrank-for-keyword-extraction-by-python-c0bae21bcec0>

13. Bluth, R.: ADHD Numbers Are Rising, and Scientists Are Trying to Understand Why. https://www.washingtonpost.com/national/health-science/adhd-numbers-are-rising-and-scientists-are-trying-to-understand-why/2018/09/07/a918d0f4-b07e-11e8-a20b-5f4f84429666_story.html
14. Meier, e.a. Sandra M: Attention-deficit hyperactivity disorder and anxiety disorders as precursors of bipolar disorder onset in adulthood. *British Journal of Psychiatry* **213**(3), 555–560 (2018)
15. CHADD: ADHD Awareness Month October 2020. <https://chadd.org/awareness-month/>
16. Gonzales, C.: Manila Ranks 5th Among World Cities with Highest Burnout Rate. <https://newsinfo.inquirer.net/1297777/manila-ranks-5th-in-the-world-with-highest-employee-burnout-rate-think-tank>