Emerging Technologies in Healthcare and Medicine, Vol. 116, 2023, 264–274
https://doi.org/10.54941/ahfe1004380

Generation of User Requirements for a Mental Health Mobile Application From an Online Public Forum: A Topic Modelling Approach

Raymond Freth A. Lagria, Lorelie C. Grepo-Jalao, and Joy Ann N. Malapit

Department of Industrial Engineering and Operations Research, College of Engineering, University of the Philippines, Diliman, Quezon City, Philippines

ABSTRACT

This research paper explores the application of topic modelling algorithms to extract user requirements for a mental health-related mobile application. Specifically, the objective is to generate themes efficiently and effectively from Reddit posts related to mental health narratives, stories, calls for help, and knowledge sharing among others. Particularly, this research examines Latent Dirichlet Allocation algorithm to generate themes coming from the posts and validate using a thematic analysis process to check similarities in generated outputs. The output will be used to establish user requirements for a mental wellbeing app to be developed for the academic community. Hence, the significance of this research. The research findings demonstrate utilizing topic modelling has promising results and categorized thematic terms from the Reddit posts. By leveraging the extracted themes, the research team can gain valuable insights into the needs and preferences of their target audience. The results offer practical implications for the design and development of mobile apps that are guided by a user-centered design process that meets the needs and expectations of the target users. The qualitative analysis further validated the relevance of the generated themes.

Keywords: Mental health, Topic modelling, User requirements definition, Application user requirements

INTRODUCTION

Mental health can be defined as the state of well-being of a person who realizes his or her own abilities and how he/she copes up with the different stressors of life (Senate of the Philippines, 2018). In recent years, the number of Filipinos experiencing mental health conditions has increased. Filipino working employees are third in rank in the Southeast Asia region when it comes to feeling uncomfortable in sharing their mental health state especially with their managers and supervisors (Desiderio, 2023). Perceived stigma is shown to be a factor that worries the Filipinos in sharing about the stressors they experience everyday. Moreover, it was also reported that 14% of

Filipinos with disabilities also experience mental health disorders (Lally et al., 2019).

Particularly, during the height of the COVID-19 pandemic, the Department of Health (DOH) estimated that at least 3.6 million Filipinos experienced mental health issues in various scenarios. It was found out that 1 out of 3 COVID-19 Filipino patients eventually were diagnosed with a certain mental health condition. In response to this, the DOH launched its mobile application (Lusog-Isip) that provides access to self-help and self-care resources that aim to address mental health concerns (University Research Co. (URC), 2021).

Mental health care in the Philippines is characterized by an extreme shortage of mental health professionals in the Philippines (Lally et al., 2019). Around 2400 mental health professionals (e.g., psychiatrists, psychologists, other mental health professionals) comprise the whole human resources for mental health in the Philippines (World Health Organization, 2020). This translates to 2.3 professionals per 100,000 Filipinos. In addition, a total of only 79 mental health facilities supported by the government are available in the country.

In terms of awareness to available mental health care, a survey reported that only about 1% of Filipino households are aware of local government unit (LGU) level programs for mental health. Of the 1%, only 0.1% have availed the various mental health programs (Statistics & City, 2023). A separate survey conducted by the FWD Group supported the fact that there are numerous mental health challenges in the Philippines. According to its survey, 63% of Filipinos believe that mental health is a critical issue in 2023. Moreover, of these 63%, only 39% are willing to seek external support due to financial issues (FWD Group, 2023). This further aggravates the situation for the mental health situation in the country.

To alleviate and address mental health challenges in the Philippines, the government ratified its Mental Health Act in 2018. The Mental Health Act establishes a framework for the comprehensive and integrated access to mental health care services and personnel as well as to protect the rights of people with mental health conditions (Lally et al., 2019). This framework provides easier access to multidisciplinary services, enhanced treatment efforts, and provide more opportunities to increase mental health professionals not only at the national level but also at the community and LGU levels.

While the Mental Health Act has been considered as a means to improve the mental health situation in the Philippines, still, the mental health care suffers from lack of resources even to this day. It is often described that the state of the mental health care services of the National Center for Mental Health (NCMH) is a tragic and heart-breaking situation (Dela Peña, 2023). It has been cited that families of mental health patients are discouraged from seeking help from such institutions and programs because of such conditions. Moreover, back in 2020, for the NCMH alone, the budget attributed per Filipino was only valued at Php 350.00 per person. This is considered to be insufficient to a week's worth of medication (Dela Peña, 2023). Lastly, the mental healthcare expenditure is pegged only at

5% of the total healthcare budget under the Universal Health Care Law (Maravilla & Tan, 2021).

In this regard, the lack of resources and facilities leads to mental health patients to try coping up with a normal life environment inside their homes. It is the hope of this research that during this age, technology would be able to help in potentially reduce the challenges to access, and obstacles to an improved overall mental health state in the Philippines. As such, it is the ultimate goal of this paper to aid in the development of a mental well-being mobile application for Filipinos which is deemed necessary at this time.

A group of researchers from the University of the Philippines (UP) proposed the development of a mental wellbeing app funded by the UP Office of the Vice President for Academic Affairs (OVPAA). The development process implements a user-centered design process to extract and solicit critical features and design elements from its intended users and experts within the academic community. To extract features, user requirements elicitation was performed using data gathered from (1) interviews, (2) focus group discussions and (3) content analysis from public forum content.

As part of the software development process, it is the goal of this paper to aid in building the mental wellbeing app through the user requirements elicitation step. This explores the application of a topic modelling algorithm to extract user requirements using social media data from the public forum platform, Reddit.

Specifically, the purpose of this research paper tries to answer the following research questions:

- RQ1 – What are the themes generally discussed in public forum posts related to mental health?
- RQ2 – How can these themes aid in the generation of user requirements for a mobile application?

The generic requirements gathering process for building software applications includes requirements elicitation, documentation and confirmation. Ideally, user requirements elicitation involves examining and collecting high level requirements from all stakeholders. Methods such as interviews, focus group discussions, surveys, narratives, and stories fall under elicitation. Information gathered from these methods are then documented using different requirements modelling tools. These tools include but are not limited to the following: use case diagrams, activity diagrams, data flow diagrams, and class diagrams. The third step in the requirements gathering process involves confirmation from the project team members as well as the stakeholders to achieve a common agreement and understanding on the application to be built.

It is difficult to identify user roles in Reddit especially in the case of public posts because speakers or users are not directly asked of user requirements as opposed to structured requirements gathering like interviews. Hence, this paper focuses only on the aspect of what or elicitation of user requirements and emphasizes on extracting use cases from publicly available data.

The significance of this study includes supporting and enhancing user requirements elicitation through additional points of view. In addition, such methods of using publicly available data are considered automatic which can have a more efficient and faster generation of insights from large crowd-produced datasets (Gulle et al., 2020). This attempts to level up and to accelerate the direct requirements definition process by using an analytics approach.

LITERATURE REVIEW

Several studies have utilized different data science and analytics methods in generating user requirements for different systems in different application areas.

Over the past decade, an increased trend was observed in using machine learning (ML) techniques to automate the elicitation of requirements. A review on several studies was done to characterize and determine requirements elicitation activities that use ML as well to identify these ML tools. Results showed that data sources can be categorized into textual documents, user-generated content (UFC) and existing requirement datasets. Furthermore, it was discovered that the most popular data mining algorithms used in general purpose analytics are also used in requirements elicitation such as Naïve Bayes, Support Vector Machines, Decision Trees and Neural Networks (Cheligeer et al., 2022).

A study by Siahaan et al. (2023), capitalized on using online news as its primary source for gathering user requirements. Specifically, several Natural Language Processing (NLP) tools were used such as Parts-of-Speech (POS) chunking and named entity recognition (NER). Their study was able to determine that requirements generated from online news are geared towards hard-goals or soft-goals (Siahaan et al., 2023).

Jiang et al. (2014) proposed a systematic approach in generating user requirements from online reviews data. They adopted opinion mining techniques to obtain expressions on software application features which were later on used to generate evolutionary requirements using user satisfaction analysis. The study showed acceptable performance of their proposed method.

Similarly, a topic modelling approach was utilized in generating topics from a publicly available user stories dataset (i.e., CrowdRE). The user stories were based on crowd-based requirements for smart home applications. This study compared three topic modelling methods namely, Latent Dirichlet Allocation (LDA) algorithm, combination of word embeddings and principal component analysis (PCA), and combination of word embeddings and Word Mover's Distance. According to this research, the combination of Word Mover's Distance and word embeddings proved to be the most effective (Gulle et al., 2020).

These related works cited above further support the significance of this study in aiding the generation of requirements for a product, specifically, mobile application.

METHODOLOGY

In order to extract user requirements from online forum data, this study employs a systematic approach. To this end, Figure 1 shows the methodology used in this paper consisting of five key phases.

The data collection process involves extracting reddit posts from the subreddit r/MentalHealthPH. Posts were gathered from March 1, 2022 to June 1, 2022 to cover a period of four (4) months. The period of collection was limited by the extraction limitations used in python. All attributes available by the python package psaw were collected. Examples of attributes are flair, author, selftext, and permalink to the post. Comments and replies were excluded in the data collection to simplify the extraction. Data collected were then stored in a spreadsheet.

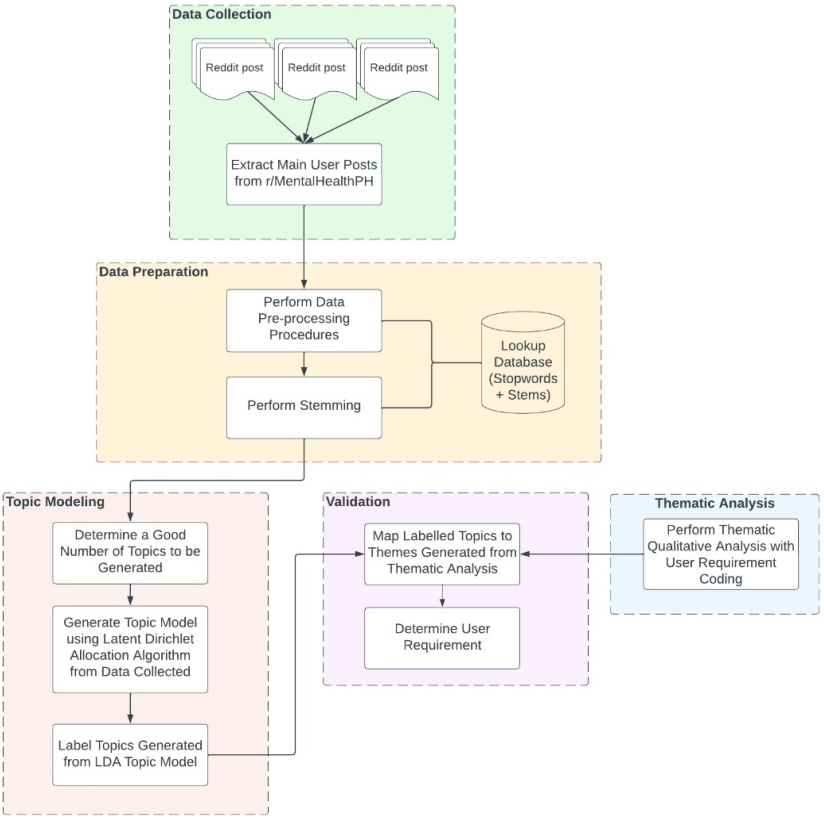

Figure 1: Methodology framework for generating user requirements.

In the data preparation phase, steps were performed to ensure that each post is prepared for the topic modelling process and to be free from undesired stopwords, punctuations, and whitespaces. The following steps comprise the pre-processing phase and were run using R and R packages:

1. Convert crazy characters to text
2. Replace emoticons with whitespaces
3. Remove special characters, numbers, and links
4. Remove extra whitespaces
5. Remove stopwords
6. Perform stemming

To elaborate on the stemming process, it features the use of lookup tables to replace words in the posts into its word stem. Word stems are word forms before adding inflectional and derivational affixes (e.g., fish is the stem for fisher, fishing, fishers). The lookup tables used for the stemming process as well as to remove the stopwords are adopted from another study conducted by the principal author. The lookup tables included in the study are English stems, Tagalog stems, and some Slang Tagalog stems.

The resulting dataset coming from the stemming process is converted into a document-term matrix (DTM) which is composed of term features. These features are selected based on terms with minimum frequency of 5 instances and the DTM results to a sparsity of 97%. This would ensure that most commonly used words without the stopwords are included in the topic modelling process.

Before subjecting the subsequent DTM into the LDA algorithm, a number of topics to be generated must be determined first. While there is no standard or formula for an optimal number of topics to be produced by LDA, there are several metrics proposed which can suggest a good number of topics based on the dataset the LDA will process. For this study, the metrics proposed by CaoJuan2009 (Cao et al., 2009) and Deveaud2014 (Deveaud et al., 2014)will be used to generate the number of topics. The objective is to minimize the topics generated by CaoJuan2009 and maximize the number of topics by Deveaud2014. These metrics were separately used to validate LDA modelling in 2009 and 2014, respectively.

Afterwards, the DTM is processed using the LDA algorithm to generate topics via the topicmodels package in R. LDA is a statistical tool used in text mining and natural language processing to produce topics from a collection of documents. In this paper, each forum post is considered a text document and the content of each post are the text data that LDA processes. To simplify, LDA examines how terms are used inside these forum posts and supposes how these words relate to several topics. LDA does not generate labelled topics, rather it assumes that there is an existing distribution of topics over the corpora (e.g., 25% is about calling for help). This in turn finds words that are related to the assumed topics due to its probability (based on frequency) of appearing in a particular post. Then, for each forum post, depending on the words contained on each post, a probability is assigned to that particular post (document) to which topic it belongs to. For example, if words such as “help”, “call”, and “hotline” are in a particular forum post, it should have a high probability that the post is about calling for help. Finally, after the assigned topic distributions, these forum posts are then clustered according to their similarity. Hence, the resulting output is composed of groups (topics/clusters) of words.

Following the generation of topics, the researchers inspect and label each topic related to mental health environment. Next, the labelled topics are mapped to the themes generated by a thematic analysis of sampled set of forum posts from the same dataset. It has to be noted that the analysts must have prior knowledge in mental health environment as well as user requirements determination to be able to properly label each topic and discern user requirements from these topics. Finally, a thematic analysis of sampled forum post is conducted to validate if there are similar user requirements generated from both methods.

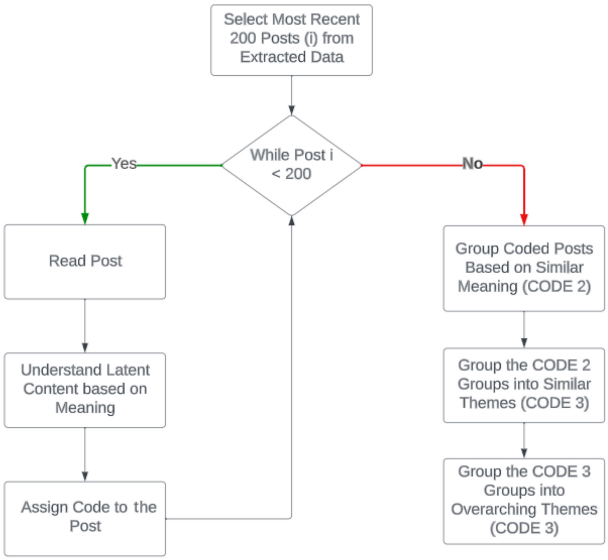

Figure 2: Thematic analysis performed for validation.

To validate the themes or topics generated by the topic modelling algorithm, a thematic analysis was performed. Thematic analysis is a technique for assessing qualitative data and is conducted by a researcher by reading through a collection of data (e.g., text data) and attempts to search for patterns or meaning to identify themes. Thematic analysis further emphasizes on interpreting qualitative patterns in the data. The thematic analysis was conducted by an expert in the field of qualitative research. In addition, a workshop was also conducted to perceive user requirements from themes generated. Figure 2 above shows the simple process of thematic analysis performed by the researchers.

The thematic analysis employed by this research follows selecting the most recent 200 posts from the same dataset collected. It is noted that the selected posts also came from the same dataset subjected through the LDA modelling. For each selected post, the designated researcher reads through each and understand its latent meaning. The meaning is further given context by the assumption that the Reddit user has a specific purpose in publishing his/her post and that there is something he/she needs that is not readily accessible and can only be accessed or found by posting it in Reddit.

After understanding the post, a code or a simple short phrase is assigned to the post to represent the latent content. These steps are performed on all selected posts.

Then, the coded posts are grouped according to similar meaning based on the researcher's understanding. These coded posts are labelled CODE 2 to indicate that this is the second coding level. The CODE 2 groups are then grouped intro similar themes. These themes are labelled CODE 3 to indicate that this is a third coding level. Finally, the CODE 3 themes are further grouped into overarching themes. These themes are the final themes used to validate against those themes generated by the topic model.

RESULTS AND DISCUSSION

The total extracted forum posts using the psaw python library was 1,279 forum posts with 82 different attributes. These attributes were all default to the psaw library. Since this paper is tasked to only generate a topic model, only the sequence number and the *selftext* attribute was used in the next set of processing steps. This paper excludes the inclusion of other attributes other than the post itself to preserve the integrity and quality of the content as well as to mimic the same path as the validation method.

The series of pre-processing steps ultimately reduced the number of data points to 879 forum posts. Further reduction of data points can be due to one of several reasons such as required sparsity (97%) of the DTM because of the minimum number of frequency of words included, posts with pure emoticons were eventually removed and posts that became blank or whitespace after pre-processing. The resulting DTM was then subjected to LDA tuning method using metrics CaoJuan2009 and Deveaud2014. Figure 2 shows the resulting plots for both metrics after running the LDA tuning method.

Based on Figure 2, a good number of topics is $n = 12$ since it can be observed from the plots that it has the second minimum value for the CaoJuan2009 metric while it has a good number for the Deveaud2014 metric showing not too low or too high. Hence, the LDA modelling process will utilize $n = 12$ topics to be generated.

After running the LDA modelling algorithm (with 500 iterations using Gibbs sampling) to the DTM, Figure 3 shows the resulting topic-word clusters. Gibbs sampling is a Markov chain Monte Carlo (MCMC) method that follows a probabilistic algorithm that samples from a probability distribution[1]. The topic-word clusters were then labelled by the researchers to have a better and meaningful descriptions. The labellers have had experience in labelling topic models and user requirements elicitation. Table 1 shows some examples of labelled topics.

As seen in Table 1, forum posts generally exhibit showing care and understanding, seeking professional help, feelings and stressors in life, and medical experiences. These messages are all related to the mental health of Filipinos who posted in the public forum.

[1]Maklin, Cory. Gibbs Sampling. *Towards Data Science*. 2020. https://towardsdatascience.com/gibbs-sampling-8e4844560ae5 (accessed 20 August 2023)

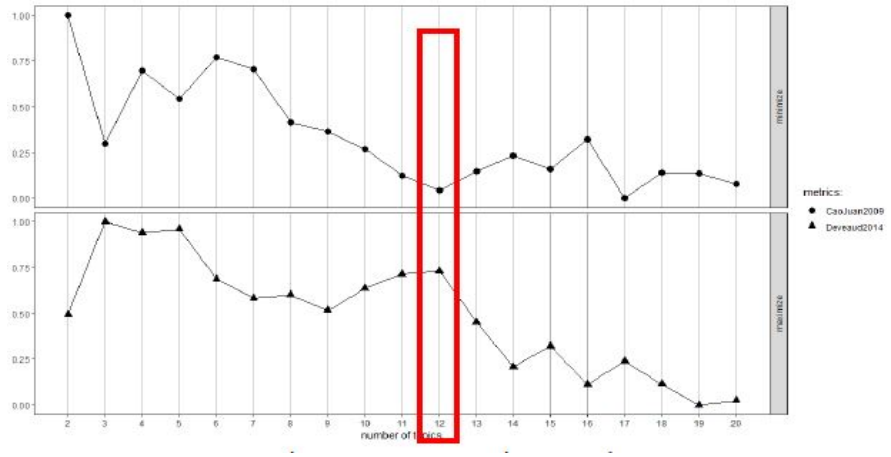

Figure 3: LDA tuning results.

		Topic											
	#	1	2	3	4	5	6	7	8	9	10	11	12
	1	just	wala	help	get	feel	iatm	lang	like	work	day	thank	tell
	2	realli	ako	mental	just	dont	say	yung	peopl	year	anxieti	consult	life
	3	friend	alam	need	famili	like	donatt	kasi	autist	time	ive	med	want
	4	talk	naman	health	live	know	like	naman	group	job	take	psychiatrist	start
	5	tri	sobra	issu	mom	think	time	talaga	actual	point	want	diagnos	end
	6	know	pag	therapi	one	peopl	itat	din	person	plan	school	adhd	just
	7	mayb	hirap	profession	thing	want	iatv	parang	parent	month	sleep	doctor	way
T	8	say	isip	studi	know	thought	want	tapo	use	struggl	past	ask	thing
e	9	month	gawa	hope	stop	ive	thought	nung	good	tri	week	experi	person
r	10	ask	mag	sure	money	hate	bad	sana	time	pandem	attack	psych	relationship
m	11	start	sya	servic	stuff	make	happen	yun	autism	week	think	session	love
s	12	one	nag	good	dad	scare	lot	haha	thing	depress	normal	depress	awayin
	13	doesnt	ganito	great	look	better	think	sabi	think	social	high	hello	look
	14	reason	nya	seek	lot	hard	canatt	daw	place	experi	deal	medic	get
	15	tire	ayoko	trauma	leav	self	open	nga	one	meet	everyday	free	anymor
	16	understand	buhay	student	take	kind	just	sakin	support	abl	stress	disord	felt
	17	care	baka	current	problem	idk	didnatt	tao	tell	current	read	effect	didnt
	18	differ	gawin	affect	amp	futur	one	okay	condit	cours	anxious	onlin	sad
	19	sure	nalang	appreci	fight	peac	hear	kaso	filipino	new	hope	kayo	right
	20	messag	ramdam	hello	come	right	breakdown	isa	child	task	advic	recommend	know

Figure 4: Resulting topics from LDA topic modelling.

Table 1. Sample labelled topics.

Topics	1	3	5	10	11
Label / Description	Messages showing or seeking care and understanding	Messages calling for professional help or therapy	Messages about feelings such hate, scare, and future thinking	Messages about anxiety and dealing with everyday stress	Messages about diagnosis and medical experiences

VALIDATION USING THEMATIC ANALYSIS

To compare results between the LDA topics and the themes generated from the thematic analysis, simple comparison of context and meaning between

topics is performed. To limit the scope, only CODE 3 categories are used to compare both results. As seen in Table 2 and comparing these with labelled topics from Table 1, themes generally resemble topics generated from the LDA topic model. The column Theme also refers to the user requirement identified by the workshop performed. These user requirements were considered in the development of the mental wellbeing mobile application.

Table 2. User requirement mapping to CODE 2 topics.

#	Code 2 Topic	Theme/User Requirement
1	Seeking information for self-assessment on how to recognize when to seek professional help	Psychoeducation and Information Dissemination
2	Seeking information how to deal with symptoms or mental health challenges	Psychoeducation and Information Dissemination
3	Seeking information about mental health professional services, treatment and medication	Psychoeducation and Information Dissemination
4	Psychoeducation how to support others	Psychoeducation and Information Dissemination
5	Seeking others with similar experience	Community
6	Seek advice/ subjective opinion of others	Community
7	Seeking connection or support from peers	Community

While there is no exact similarity in terms of actual labels (from LDA topics) and coded topics (from thematic analysis), it has to be emphasized that the Code 2 topics and labelled LDA topics exhibit the same meaning, thought and context. In particular, topics 1, 5 and 10 in LDA relate to topics 2, 5 and 6 in the thematic analysis; topic 3 in the LDA (messages calling for professional help or therapy) corresponds to seeking information about mental health professional services, treatment and medication from the thematic analysis; and topic 11 from LDA is related to topics 1 and 3 from the thematic analysis. This supports the hypothesis of this paper that topic modelling can be used in user requirements determination.

CONCLUSION

In conclusion, this paper highlights the growing importance of addressing mental health challenges in the Philippines, where a significant portion of the population faces issues with limited access to mental health care. It emphasizes the need for innovative solutions, such as the development of a mental well-being mobile application. The paper's methodology demonstrates the potential of topic modelling, specifically LDA, in extracting user requirements from publicly available data, validating the results against thematic analysis. The findings indicate that LDA-generated topics align with the themes identified through thematic analysis, supporting the feasibility of this approach for user requirements determination in mental health app development. This research contributes to addressing the critical issue of mental health in the

Philippines by providing a data-driven approach to understanding user needs and facilitating the development of relevant solutions.

ACKNOWLEDGMENT

The authors would like to acknowledge the UP Office of the Vice President for Academic Affairs for the support in conducting the project related to this research.

REFERENCES

Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing*, *72*(7–9), 1775–1781. https://doi.org/10.1016/j.neucom.2008.06.011

Cheligeer, C., Huang, J., Wu, G., Bhuiyan, N., Xu, Y., & Zeng, Y. (2022). Machine learning in requirements elicitation: a literature review. In *Artificial Intelligence for Engineering Design, Analysis and Manufacturing: AIEDAM* (Vol. 36). Cambridge University Press. https://doi.org/10.1017/S0890060422000166

Dela Peña, K. (2023, April 11). Mental health care in PH starving from lack of resources. *Inquirer. Net*.

Desiderio, L. (2023, January 16). *Most Pinoy workers stay mum about mental health concerns*.

Deveaud, R., San Juan, É., & Bellot, P. (2014). Accurate and effective latent concept modeling for ad hoc. *Document Numérique*, *17*(1), 61–84.

FWD Group. (2023, May 31). *63% of Filipinos believe mental health as one of the most critical issues for 2023*.

Gulle, K. J., Ford, N., Ebel, P., Brokhausen, F., & Vogelsang, A. (2020). Topic Modeling on User Stories using Word Mover's Distance. *Proceedings - 7th International Workshop on Artificial Intelligence and Requirements Engineering, AIRE 2020*, 52–60. https://doi.org/10.1109/AIRE51212.2020.00015

Lally, J., Samaniego, R. M., & Tully, J. (2019). Mental health legislation in the Philippines: Philippine Mental Health Act. *BJPsych International*, *16*(03), 65–67. https://doi.org/10.1192/bji.2018.33

Maravilla, N. M. A. T., & Tan, M. J. T. (2021). Philippine Mental Health Act: Just an Act? A Call to Look Into the Bi-directionality of Mental Health and Economy. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.706483

Senate of the Philippines. (2018). *An Act Establishing a National Mental Health Policy for the Purpose of Enhancing the Delivery of Integrated Mental Health Services, Promoting and Protecting the Rights of Persons Utilizing Psychosocial Health Services, Appropriating Funds Therefor and Other Purposes. REPUBLIC ACT No. 11036*.. 17th Congress, Congress of the Philippines.

Siahaan, D., Raharjana, I. K., & Fatichah, C. (2023). User story extraction from natural language for requirements elicitation: Identify software-related information from online news. *Information and Software Technology*, *158*, 107195. https://doi.org/10.1016/J. INFSOF.2023.107195

Statistics, P., & City, Q. (2023). *Philippines 2022 Philippine National Demographic and Health Survey (NDHS) Final Report*. www. DHSprogram.com.

University Research Co. (URC). (2021, December 22). *Mental Health on the Move in the Philippines – Meet the Lusog-Isip App*.

World Health Organization. (2020). *Philippines WHO Special Initiative for Mental Health Situational Assessment CONTEXT*.