

CHAPTER I INTRODUCTION

This chapter introduces the research topic, the problem statement arising from it, the research objective, the scope and limitations, and the potential significance and benefits that can be derived from this research.

I.1 Background

Mental health has emerged as a critical sector in public safety and well-being worldwide, especially in Indonesia, the world's fourth most populated country. According to the Indonesia-National Mental Health Survey (I-NAMHS), the 2022 report suggests that one in three adolescents had a mental health problem, amassing over 15 million Indonesian young adults, as shown in Figure I-1. (Wahdi et al., 2022)

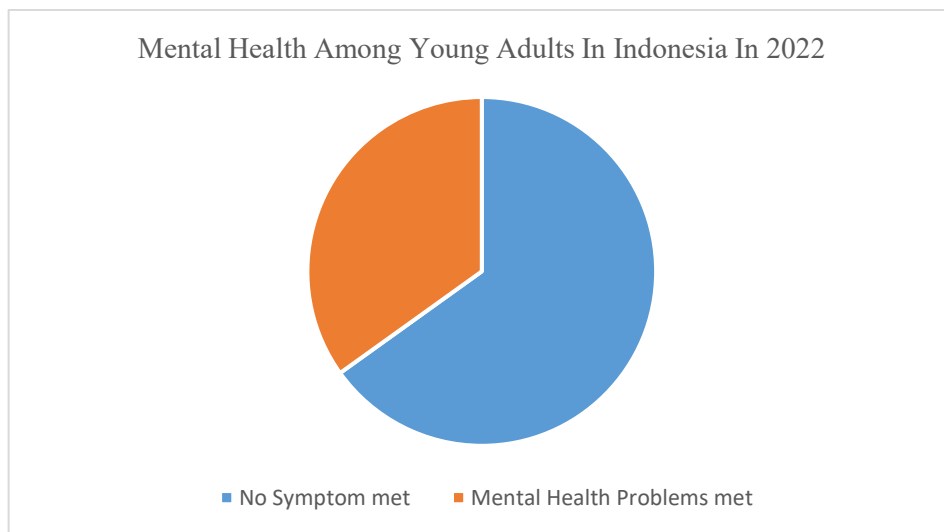


Figure I-1. Mental health in Indonesia (Wahdi et al., 2022)

While the data from Figure I-1 highlights the magnitude of the mental health problems faced by Indonesian youth, it is crucial to understand the specific challenges that they met. The report reveals a more detailed picture of the struggles of Indonesian youth. The most common mental health issues are anxiety, hyperactivity, depression, and others, as shown in Figure I-2.

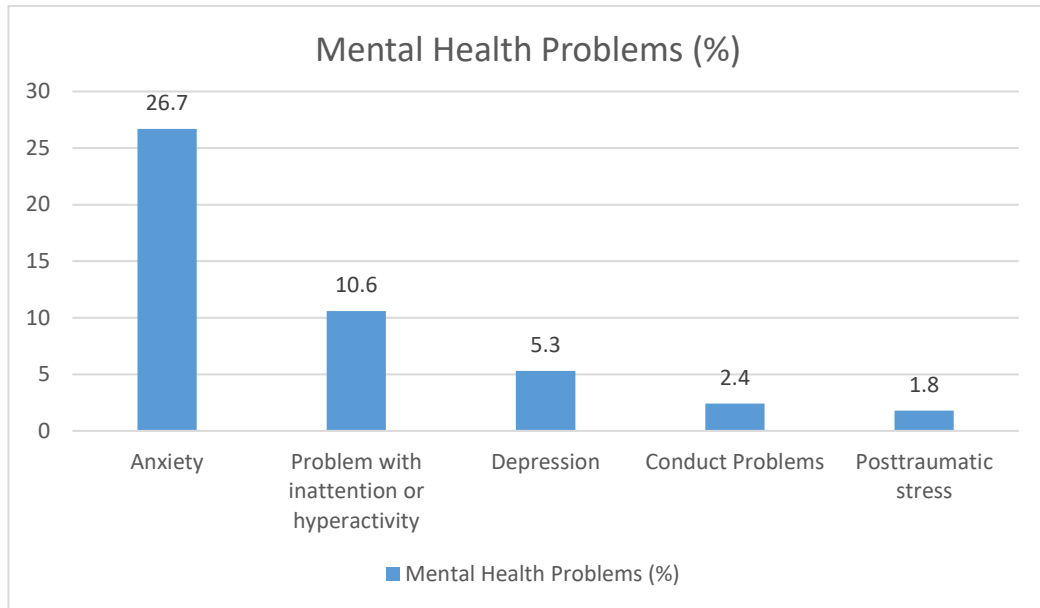


Figure I-2. Mental health problems in Indonesia (Wahdi et al., 2022)

While the number suggests in Figure I-2 the significance of this problem, mental health issues often go undetected and untreated due to stigma, lack of awareness, and limited access to mental health services (Ghafari et al., 2022).

The rapid digitalization in Indonesia presents both unique opportunities and challenges for a new approach to mental health detection and prevention. The We Are Social global overview report in 2024 (We Are Social & Meltwater, 2024) suggests that nearly 70% of Indonesians have adopted the internet. With an increasing number of people using social media, individuals have become more vocal in expressing their thoughts, fostering a sense of connectedness (Luo & Hancock, 2020), which provides an opportunity to gain insight into their mental health and well-being.

The sentiment analysis technique effectively detects human emotions and feelings by analyzing words and sentences. This technique has proven effective in classifying human emotions and mental health (Benrouba & Boudour, 2022).

Saifullah et al. (2023) focused on utilizing sentiment analysis with machine learning to identify anxiety among Indonesian users in YouTube comments. The research focused on YouTube comments about the Indonesian government's COVID-19 program, utilizing a dataset of approximately 9,700 comments. The research found

that using the RF-CV method, a combination of Random Forest and Count Vectorization yield the result reached 98.4% accuracy compared to other methods.

In separate research, Liu (2022) employed the K-means clustering method to examine the mental health of college students. The research involved approximately 1000 students and successfully categorized them into three groups, each comprising 20.6%, 31.9%, and 47.1% of the population. The research demonstrates that utilizing a clustering algorithm effectively identifies distinct mental health states among college students.

Banerjee et al. (2024) investigated the impact of online learning on the mental health of Indian college students during the COVID-19 Pandemic, using datasets comprising approximately 500 undergraduate students and covering their stress levels, learning experiences, and psychological effects. The research found that using autoencoders and clustering techniques in online learning can lead to social isolation among students, as these methods can segregate and categorize students into distinct groups, potentially exacerbating existing social divisions.

The intricacy of Indonesian literature, which can have many ways to convey emotion to its readers, combined with the mental health problems that Indonesians face, makes these challenges call for the need to use multi-label sentiment analysis to analyze further and predict mental health problems that might be faced by Indonesian people with more points of detection. Additionally, there are ethical concerns regarding the collection and analysis of personal data, which can be sensitive if not handled appropriately.

This study aims to bridge the gap by developing a sentiment analysis tool for mental health prediction, specifically designed to predict the mental health of the Indonesian people. The research combines machine learning and utilizes clustering techniques to create a comprehensive and nuanced linguistic profile of the Indonesian language, aiming to predict mental health problems through Indonesian social media use.

I.2 Problem Formulation

Based on the background that has been laid out, the formulation of this study's problem can be described as follows:

1. How can this research develop a mental health topic analysis through clustering using the K-Means clustering algorithm to identify mental health topics from social media used by Indonesians?
2. How does the K-Means clustering algorithm perform when being evaluated by the standard method used to evaluate clustering results?
3. What mental health topics and insights can be found from the K-Means clustering algorithm method?

I.3 Research Objective

Based on the identified problem that has been identified, here are the objectives of this research to solve the problem:

1. This research develops a topic analysis using a K-means clustering algorithm to accurately cluster the data, helping to identify the topics that emerge from the clusters of social media data.
2. This research will utilize a standardized evaluation method to assess the performance of the K-means clustering algorithm.
3. The research will determine the topic that emerges through bigram analysis to identify the topics that arise from the clusters of social media data, and validation with a practitioner will ensure accuracy and effectiveness.

I.4 Research Benefits

The benefits that can be extracted from this study can be divided into:

1. For the public:
 - Increases mental health awareness among the Indonesian people and the general public.

- Potential for detecting mental health problems through a combination of human and technological approaches.
 - Increased access to mental health analysis.
2. For industry:
- This study aims to inspire the healthcare industry to develop a tool that helps doctors identify mental health themes.
 - Potential for more collaboration with machine learning technologies to further assist in predicting mental health problems
3. For research:
- To expand its knowledge within the scope of mental health topic analysis, particularly by utilizing K-Means clustering techniques.
4. For further research:
- Advancement in mental health topic analysis using K-means clustering algorithm in the Indonesian language.
 - Foundation for further research within the mental health topic analysis model using the clustering technique in Indonesia, and for cross-cultural and language comparisons using similar studies in other regions.

I.5 Research Limitation

This section outlines the limitations and assumptions of this research. Limitations and assumptions are essential to setting a clear goal and acting as the research guideline to yield the best results. The limitations and assumptions of this research are as follows:

1. The data used for this research are limited to text data in Bahasa Indonesia from Facebook, spanning the period from 2021 to 2024.
2. This research determines the topic based only on the clustering result that has been generated.

I.6 Report Structure

This research is described with the following systematic writing:

1. Chapter I. Introduction

This chapter explains the research problem, utilizing background information, objectives, research scope, study advantages, and the writing structure. It highlights the importance of mental health prediction using clustering algorithms in relation to Indonesian social media data. This chapter also addresses the potential gaps in mental health research using clustering, with Indonesia serving as the primary dataset.

2. Chapter II. Literature Review

This chapter provides a theoretical review of the work by presenting relevant literature and research conducted in the past. Additionally, this chapter also mentions the CRISP-DM methodology, K-means clustering, Word2Vec embeddings, and text preprocessing techniques. This chapter also mentions the evaluation metric used to evaluate the results of the clusters that have been produced.

3. Chapter III. Problem Solving Methodology

This chapter describes the research methodology employed to achieve the study objectives. This study involves thorough research, following the CRISP-DM methodology, and employing a systematic problem-solving approach.

4. Chapter IV. Problem Resolution

This chapter outlines the research process, beginning with the collection of the mental health dataset from social media and proceeding to the data preprocessing stage, where the data is thoroughly processed and prepared for analysis. Additionally, this chapter discusses the implementation of Word2Vec and how the study utilizes K-means clustering to categorize the dataset. This chapter also mentions the evaluation methods used to assess the results of the cluster, mainly the Silhouette Index and Calinski-Harabasz Index.

5. Chapter V. Validation, Result Analysis, and Implication

The chapter covers the evaluation of the cluster itself and the identification of topics from the generated cluster. Furthermore, an analysis of each cluster will be conducted to determine the topic that corresponds to a mental health issue. The analysis of each cluster is conducted by examining the bigram pattern within each cluster, in combination with the cluster contributions, and then validating the results with a health professional to determine their effectiveness.

6. Chapter VI. Conclusion and Suggestion

This chapter presents the conclusions derived from the research accompanied by the findings and analysis discussed previously. It addresses the research questions formulated at the outset of this study and highlights the significant contributions in identifying mental health topics on Indonesian social media. Additionally, suggestions are provided for enhancing text preprocessing for the Indonesian language, as well as avenues for future research in mental health clustering analysis.