

ABSTRACT

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. The Indonesia-National Mental Health Survey (I-NAMHS) suggests that one in three young adults in Indonesia suffers from some form of mental health issues, combined with the rapid digitalization of Indonesia, making mental health issues a concern that needs to be addressed. This research will utilize a clustering algorithm and sentiment analysis to effectively predict and identify potential mental health topics among Indonesian users. Additionally, this research will utilize the CRISP-DM approach, which has six crucial steps: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. In the modeling phase, this research will utilize the K-means clustering algorithm, combined with the Elbow method and Word2Vec embeddings, to ensure the algorithm's results are satisfactory. The evaluation phase will assess the results of the findings using the Silhouette Index and the Calinski-Harabasz Index to validate the performance of the generated clusters. The findings showed that the cluster generated produced a Silhouette Index score of 0.5058, indicating well-structured clusters and a Calinski-Harabasz Index score of 11895.70, indicating a good separation between clusters. The research found four key topics: (1) psychosomatics and early treatment of stress, (2) meaning-seeking and alternative coping strategies, (3) religious relationships and social expectations, and (4) emotional expression and relationship difficulties.

Keywords: K Means clustering, Machine Learning, Mental Health, Sentiment Analysis