Abstract

Access to reliable health information remains a significant challenge in Indonesia, particularly in regions with diverse linguistic and cultural backgrounds. This research aims to develop an Indonesian language chatbot that enhances the accessibility and reliability of health information through the application of Natural Language Processing (NLP) techniques. The primary purpose of this study is to address the communication barriers posed by linguistic diversity in Indonesia's healthcare system by providing a scalable and efficient chatbot solution. The chatbot incorporates essential NLP preprocessing steps, including case folding, tokenization, stopword removal, and stemming, ensuring effective understanding of user input. To process queries and generate responses, the system uses the Term Frequency-Inverse Document Frequency (TF-IDF) and Cosine Similarity algorithms, which were evaluated with a health-related dataset from the Ministry of Health Service Unit (UPK Kemenkes). The chatbot achieved a precision of 0.9333, recall of 0.7962, and F1-score of 0.8421, with an average response time of 0.32 seconds, demonstrating strong performance in real-time environments. Deployment on the Telegram platform further emphasizes the system's practical application. The results show that the chatbot can overcome linguistic barriers and provide reliable health information, making it a valuable tool for improving access to healthcare services, especially in linguistically diverse regions. This study lays the foundation for future research in multilingual healthcare chatbots, offering a scalable solution to address regional language variations and enhance health communication in Indonesia.

Keywords

Chatbot, Health Information, Indonesian Language, Natural Language Processing (NLP), TF-IDF, Cosine Similarity