

ABSTRACT

Social media offers a real-time means of monitoring the emotional needs of pregnant women, yet quantitative evidence from Indonesia remains limited. This study designs a sentiment–topic analysis framework based on IndoBERT, enriched with self-training pseudo-labeling and BERTopic, to extract insights from more than 10,000 public TikTok comments posted between 2022 and 2024. After expanding 4,000 manually annotated labels with 6,789 pseudo-labels, the model achieved a macro-F1 score of 0.94 under five-fold cross-validation, while sentiment analysis revealed that roughly 65 % of the comments carried a negative tone. Positive sentiment primarily highlighted coping strategies, spirituality, and social support, whereas negative sentiment centred on emotional distress, pregnancy-related health issues, and partner conflict. These coherent topic patterns were translated into five proposed mental-health app features: interactive coping modules, culturally grounded spiritual content, a peer-support forum, a partner-engagement toolkit, and real-time sentiment-triggered micro-CBT. The findings confirm the synergy of IndoBERT, pseudo-labeling, and BERTopic as an effective method for capturing emotional context in short Indonesian texts and provide a scientific foundation for more targeted and personalised digital interventions to enhance the mental well-being of pregnant women in Indonesia.

Keywords: *BERTopic, IndoBERT, Mental health, Pregnant women, Sentiment analysis, Topic modeling*