Text Mining and Determinants of Sentiments towards the COVID-19 Vaccine Booster of Twitter Users in Malaysia

ABSTRACT

Vaccination is the primary preventive measure against the COVID-19 infection, and an additional vaccine dosage is crucial to increase the immunity level of the community. However, public bias, as reflected on social media, may have a significant impact on the vaccination program. We aim to investigate the attitudes to the COVID-19 vaccination booster in Malaysia by using sentiment analysis. We retrieved 788 tweets containing COVID-19 vaccine booster keywords and identified the common topics discussed in tweets that related to the booster by using latent Dirichlet allocation (LDA) and performed sentiment analysis to understand the determinants for the sentiments to receiving the vaccination booster in Malaysia. We identified three important LDA topics: (1) type of vaccination booster; (2) effects of vaccination booster; (3) vaccination program operation. The type of vaccination further transformed into attributes of "az", "pfizer", "sinovac", and "mix" for determinants' assessments. Effect and type of vaccine booster associated stronger than program operation topic for the sentiments, and "pfizer" and "mix" were the strongest determinants of the tweet's sentiments after the Boruta feature selection and validated from the performance of regression analysis. This study provided a comprehensive workflow to retrieve and identify important healthcare topic from social media.