

# Analisis Sentimen Vaksin COVID-19 Di Indonesia Menggunakan Model Bahasa XLMR Dan Teknik Augmentasi Data = Sentiment Analysis of COVID-19 Vaccine in Indonesia Using Pre-trained Language Model XLMR and Augmentation Data Methods

Putri Rizqiyah, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920547371&lokasi=lokal>

---

## Abstrak

Vaksinasi COVID-19 merupakan salah satu solusi jangka panjang untuk mengatasi pandemi COVID-19 di Indonesia. Topik vaksinasi COVID-19 menjadi perbincangan yang hangat, khususnya di media sosial. Berbagai macam pro dan kontra mengenai program vaksinasi terus bermunculan sehingga penelitian mengenai analisis publik terhadap program vaksinasi COVID-19 sangat berguna untuk komunikasi publik. Penelitian ini berfokus kepada lima jenis vaksin yang banyak digunakan di Indonesia yaitu, AstraZeneca, Moderna, Pfizer, Sinopharm dan Sinovac. Sebanyak 252,805 data dikumpulkan melalui media sosial twitter menggunakan Twitter API di tahun 2021. Lalu sebanyak 11,361 dipilih secara acak untuk dianotasi secara manual. Selanjutnya, proses klasifikasi dilakukan menggunakan model bahasa XLMR dan beberapa metode baseline berbasis pre-trained language model, deep learning, machine learning dan lexicon. Augmentasi data seperti Easy Data Augmentation (EDA), An Easier Data Augmentation (AEDA) dan Seqgan juga dilakukan untuk menyeimbangkan jumlah kelas data minoritas. Pembagian data latih dan data uji dilakukan dengan menggunakan dua metode yaitu simple random sampling dan stratified sampling untuk mengetahui performa model yang dilatih. Hasil penelitian menunjukkan bahwa metode yang diusulkan yaitu XLMR, memiliki performa yang tinggi dibandingkan metode baseline lainnya, dengan akurasi sebesar 71.91% sebelum dilakukan augmentasi dan 72.19% setelah dilakukan augmentasi menggunakan Seqgan menggunakan metode pembagian data simple random sampling. Lalu, dengan menggunakan metode pembagian data stratified, XLMR juga memiliki performa terbaik dengan akurasi 59.96% sebelum dilakukan augmentasi dan 74.37% setelah dilakukan augmentasi menggunakan EDA. Penelitian ini akan sangat bermanfaat untuk komunikasi publik dengan kasus serupa. Di masa yang akan datang, penelitian ini bisa dilanjutkan dengan melakukan domain transfer untuk meningkatkan performa model.

.....COVID-19 vaccination is one of the long-term solutions to address the COVID-19 pandemic in Indonesia. The topic of COVID-19 vaccination has become a hot discussion, especially on social media. Various pros and cons regarding the vaccination program continue to emerge, making research on public analysis of the COVID-19 vaccination program very useful for public communication. This study focuses on five types of vaccines widely used in Indonesia, namely AstraZeneca, Moderna, Pfizer, Sinopharm, and Sinovac. A total of 252,805 data were collected through social media Twitter using the Twitter API in 2021. Then, 11,361 were randomly selected to be manually annotated. Subsequently, the classification process was performed using the XLMR language model and several baseline methods based on pre-trained language models, deep learning, machine learning, and lexicon. Data augmentation such as Easy Data Augmentation (EDA), An Easier Data Augmentation (AEDA), and Seqgan was also carried out to balance the number of minority class data. The division of training data and test data was done using two methods, namely simple random sampling and stratified sampling, to determine the performance of the trained model. The results of the study show that the proposed method, XLMR, has high performance compared to other baseline

methods, with an accuracy of 71.91% before augmentation and 72.19% after augmentation using Seqgan with the simple random sampling data splitting method. Then, using the stratified data splitting method, XLMR also had the best performance with an accuracy of 59.96% before augmentation and 74.37% after augmentation using EDA. This research will be very useful for public communication with similar cases. In the future, this research can be continued by conducting domain transfer to improve model performance.