

INTISARI

ANALISIS SENTIMEN TWEET BERBAHASA INDONESIA TENTANG VAKSIN COVID-19 MENGGUNAKAN NAIVE BAYES DAN FASTTEXT EMBEDDING

Oleh

Risman Yorza

20/466432/PPA/05998

Penanganan COVID-19 menjadi salah satu permasalahan kompleks yang dihadapi oleh seluruh negara di dunia termasuk Indonesia. Pemerintah telah menerapkan berbagai cara untuk mencegah penularan virus COVID-19 ini, salah satunya dengan adanya vaksin yang dapat memberikan kekebalan pada tubuh. Namun sejak dikabarkan bahwa vaksin COVID-19 datang ke Indonesia, muncul berbagai isu dan tanggapan yang membuat masyarakat merasa ragu tentang keamanannya dan banyak dari pengguna twitter di Indonesia yang menulis *tweet* tentang vaksin COVID-19 dengan berbagai opini yang beragam. Tanggapan dari *tweet* inilah yang kemudian akan dilakukan analisis untuk mengetahui polaritasnya menggunakan analisis sentimen yang dapat mengklasifikasikan tanggapan masyarakat ke dalam kelas positif, negatif, dan netral.

Penelitian ini mengusulkan model analisis sentimen masyarakat terhadap aspirasi yang disampaikan melalui media sosial twitter tentang vaksin COVID-19 menggunakan metode *Naive Bayes* dan memanfaatkan representasi kata *fastText*. *FastText* adalah sebuah pustaka yang dikeluarkan oleh *Facebook AI Research* yang dapat digunakan sebagai *word embedding* yang merupakan pengembangan dari metode sebelumnya yaitu *word2vec* yang dapat menangani kata yang tidak pernah dijumpai sebelumnya atau *out of vocabulary*.

Penelitian ini menggunakan dataset berupa data *tweet* tentang COVID-19 yang berasal dari media sosial twitter. Ada beberapa pengujian yang dilakukan untuk menilai kinerja dari model yang digunakan, yakni dengan menghitung akurasi, *precision*, *recall*, dan *f-measure*. Hasil penelitian menunjukkan bahwa penggunaan ekstraksi fitur *fastText* dapat meningkatkan akurasi lebih baik dari pada metode *word2vec*, percobaan yang telah dilakukan menggunakan algoritma *naive bayes* dengan *fastText* mendapatkan nilai akurasi tertinggi sebesar 87,0%, *precision* 87%, *recall* 87%, dan *f-measure* 87% sedangkan pada algoritma *naive bayes* dengan *word2vec* memiliki nilai akurasi sebesar 78,7%, *precision* 80%, *recall* 79%, dan *f-measure* 79%.

Kata Kunci: *Twitter, COVID-19, Vaksin, Sentimen, FastText, Naive Bayes*

ABSTRACT

SENTIMENT ANALYSIS OF INDONESIAN TWITTER ABOUT COVID-19 VACCINE USING NAIVE BAYES AND FASTTEXT EMBEDDING

By

Risman Yorza

20/466432/PPA/05998

Handling COVID-19 is one of the world's most complicated problems, and Indonesia is no exception. The government has implemented a number of measures to combat the spread of the COVID-19 virus, one of which is the availability of a vaccination that can confer bodily immunity. However, after it was reported that the COVID-19 vaccination had arrived in Indonesia, a slew of difficulties and comments have surfaced, casting doubt on the vaccine's safety. As a result, many Indonesian twitter users have written tweets concerning the COVID-19 vaccine, expressing a variety of viewpoints. The sentiment analysis will be used to classify public replies into positive, negative, and neutral categories, and the responses from these tweets will be studied to assess polarity.

Using the Naive Bayes approach and fastText word representation, this study presents a model for evaluating public sentiment regarding ambitions given by social media tweeting about the COVID-19 vaccination. FastText is a Facebook AI Research library that may be used for word embedding. It is a development of the previous method, word2vec, and can handle words that have never been encountered before or are out of lexicon.

This study makes use of a COVID-19 tweet dataset obtained from the social media site Twitter. Several tests are used to evaluate the performance of the system under construction, including calculating accuracy, precision, recall, and f-measure. The results suggest that utilizing fastText feature extraction improves accuracy over using the word2vec approach. Experiments using the Naive Bayes algorithm with fastText yielded the greatest accuracy, precision, recall, and f-measure values of 87.0%, 87%, 87%, and 87%, respectively. The accuracy, precision, recall, and f-measure of the naive bayes algorithm with word2vec are all 78.7%, 80%, 79%, and 79% respectively.

Keywords: Twitter, COVID-19, Vaccine, Sentiment, FastText, Naive Bayes