

## INTISARI

### **ANALISIS SENTIMEN TWEET BERBAHASA INDONESIA TENTANG VAKSIN COVID-19 MENGGUNAKAN FASTTEXT EMBEDDING DAN SUPPORT VECTOR MACHINE**

Oleh

Dikih Arif Wibowo

20/466402/PPA/05968

Quartal-I 2020 kasus pandemi Covid-19 di Indonesia mengalami peningkatan. Pemerintah Indonesia melaksanakan vaksinasi sebagai solusi untuk mengatasi pandemi, hal ini menuai pro-kontra dalam masyarakat. Analisis sentimen merupakan salah satu langkah penting untuk mengetahui pandangan masyarakat terhadap vaksinasi. Berbagai penelitian menggunakan *Word2vec* sebagai ekstraksi fitur, *Word2vec* memiliki keunggulan dapat melihat hubungan semantik antar kata, namun *Word2vec* memiliki kelemahan dalam representasi kata yang tidak ada dalam proses *training corpus*.

*FastText* merepresentasikan kata sebagai sekumpulan *n-gram* karakter. Kata yang tidak ada dalam proses *training corpus* dapat ditangani *FastText*. Untuk merepresentasikan vektor kata *FastText* terdiri dari arsitektur *skip-gram* dan *CBoW*. Data tweet yang digunakan pada penelitian ini berjumlah 832 data, terdiri dari 672 sebagai data latih, dan 160 sebagai data uji. Untuk mengenali pola data yang diberikan, digunakan metode klasifikasi *Support Vector Machine*. Hasil klasifikasi sentimen terdiri dari kelas positif, negatif, dan netral.

Berdasarkan hasil pengujian yang dilakukan, *FastText* menghasilkan akurasi sebesar 88.1% dengan arsitektur *skip-gram*, sementara *Word2vec* menghasilkan akurasi sebesar 76.8% dengan arsitektur *CBoW*. *FastText* dapat diandalkan untuk permasalahan *out of vocabulary*. *FastText* unggul saat berhubungan dengan kata tidak baku dan kesalahan penulisan.

**Kata Kunci:** *Vaksin Covid-19, Twitter, Sentimen, FastText, SVM.*

## **ABSTRACT**

### **ANALYSIS OF INDONESIAN LANGUAGE TWEET SENTIMENTS ABOUT COVID-19 VACCINE USING FASTTEXT EMBEDDING AND SUPPORT VECTOR MACHINE**

By

Dikih Arif Wibowo

20/466402/PPA/05968

In the first quarter of 2020, cases of the Covid-19 pandemic in Indonesia have increased. The Indonesian government implemented a vaccination program as a solution to overcome pandemic, this reaped the pros and cons in society. Sentiment analysis is one of the important steps to determine the public's view of vaccination. Various studies use Word2vec as feature extraction, Word2vec has the advantage of being able to see the semantic relationship between words, but Word2vec has a weakness in word representation that is not in training corpus process.

FastText represents a word as a set of n-gram characters. Words that are not in the training corpus can be handled by FastText. To represent word vectors FastText consists of skip-gram and CBoW architecture. The Tweet data used in this study amounted to 832 data, consisting of 672 as training data, and 160 as test data. To recognize the given data pattern, the Support Vector Machine classification method is used. Sentiment classification results consist of positive, negative, and neutral classes.

Based on the results of the tests carried out, FastText produces an accuracy of 88.1% with the skip-gram architecture, while Word2vec produces an accuracy of 76.8% with the CBoW architecture. FastText is reliable for out of vocabulary problems. FastText excels when it comes to non-standard words and spelling errors.

**Keywords : Covid-19 Vaccine, Twitter, Sentiment, FastText, SVM.**