

INTISARI

ANALISIS SENTIMEN TERHADAP OBESITAS PADA *TWEET* BERBAHASA INDONESIA MENGGUNAKAN BERT *FINE-TUNING*

Oleh

Evita Claudia
19/439152/PA/18975

Peningkatan obesitas terjadi di Indonesia dari 10,5% pada tahun 2007 menjadi 21.8% pada tahun 2018. Obesitas beresiko besar terhadap perkembangan penyakit lainnya. Twitter digunakan oleh penggunanya untuk menyebarkan opininya terkait obesitas di Indonesia. Dengan menganalisis sentimen pengguna, diharapkan dapat membantu mengerti persepsi pengguna agar dapat mengambil langkah yang preventif untuk mengurangi stigma buruk obesitas di Indonesia.

Penelitian ini membandingkan 2 *pre-trained* model BERT (*Bidirectional Encoder Representations from Transformers Embedding*) dengan menambahkan *fully connected layer* tanpa *hidden layer* dan *softmax*. Data yang digunakan diperoleh dari Twitter menggunakan Twint. Langkah-langkah yang dilakukan dalam tahap *preprocessing* adalah *case folding*, *data cleaning*, tokenisasi, dan penambahan token spesial. Model IndoBERT dan IndoBERTweet digunakan sebagai model analisis sentimen. Performa model tersebut diuji dan dibandingkan dengan menghitung nilai akurasi, presisi, *recall*, *loss* dan *F1 score*.

Hasil pengujian model IndoBERT mempunyai akurasi 0,785, presisi 0,816, *recall* 0,788, dan *F1 score* 0,786, sedangkan model IndoBERTweet mempunyai akurasi 0,803, presisi 0,833, *recall* 0,806, dan *F1 score* 0,803. Hasil perbandingan *F1 score* menunjukan model IndoBERTweet mengungguli model IndoBERT dalam menganalisa sentimen obesitas.

Kata Kunci: Analisis Sentimen, Obesitas, Twitter, Bahasa Indonesia, IndoBERT, IndoBERTweet

ABSTRACT

SENTIMENT ANALYSIS TOWARD OBESITY ON INDONESIAN TWEETS USING BERT FINE-TUNING

By

Evita Claudia
19/439152/PA/18975

An increase in obesity has occurred in Indonesia from 10.5% in 2007 to 21.8% in 2018. Obesity is at great risk for the development of other diseases. Twitter is used by its users to spread their opinions regarding obesity in Indonesia. By analyzing user sentiment, it is hoped that it can help understand user perceptions so that we can take preventive steps to reduce the bad stigma of obesity in Indonesia.

This study compares 2 pre-trained BERT (Bidirectional Encoder Representations from Transformers Embedding) models which are added with a fully connected layer without a hidden layer and softmax. The data used was obtained from Twitter using Twint. The steps taken in the preprocessing stage are case folding, data cleaning, tokenization, and the addition of special tokens. The IndoBERT and IndoBERTweet models are used as sentiment analysis models. The performance of the model is tested and compared by calculating the values for accuracy, precision, recall, loss, and F1 score.

The results of the IndoBERT model have an accuracy of 0.785, precision of 0.816, recall of 0.788, and F1 score of 0.786, while the IndoBERTweet model has an accuracy of 0.803, precision of 0.833, recall of 0.806, and F1 score of 0.803. The results of the comparison of F1 scores show that the IndoBERTweet model outperforms the IndoBERT model in analyzing obesity sentiment.

Keywords: Sentiment Analysis, Obesity, Twitter, Indonesian, IndoBERT, IndoBERTweet