

INTISARI

PENGARUH ENSEMBLE FEATURES PADA SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PUBLIK TERHADAP PEMBATASAN SOSIAL BERSKALA BESAR JILID II DKI JAKARTA

Oleh

Almas Fauzia Wibawa
17/409427/PA/17734

Salah satu kebijakan yang dibuat pemerintah dalam mengatasi dan mencegah wabah COVID-19 adalah Pembatasan Sosial Berskala Besar (PSBB) di Provinsi DKI Jakarta. Berbagai respon diterima pemerintah. Sementara itu, pemimpin perlu memiliki sudut pandang lain sebagai pertimbangan pengambilan kebijakan dan mengetahui keadaan masyarakat sebelum mengumumkan kebijakan agar tidak terjadi kepanikan publik. Instagram merupakan media sosial yang cukup populer di Indonesia. Pemimpin daerah pun mulai menggunakannya untuk mengumpulkan aspirasi dari masyarakat melalui kolom komentar. Sementara itu, analisis sentimen kerap dijadikan sarana penyelesaian untuk merangkum opini publik di berbagai topik dengan data bersumber dari media sosial.

Penelitian ini melakukan analisis sentimen untuk mengetahui respon publik terhadap diberlakukannya kembali PSBB di DKI Jakarta dengan melihat pengaruh fitur tekstual, fitur *parts of speech*, dan fitur *lexicon-based* pada *ensemble features* terhadap metode ekstraksi fitur *Term Frequency – Inverse Document Frequency* (TF-IDF) serta metode klasifikasi *Support Vector Machine* dengan data komentar Instagram berbahasa Indonesia. Data yang digunakan bersumber dari komentar pada akun @aniesbaswedan dan @dkijakarta di Instagram yang mengumumkan akan diberlakukannya PSBB Jilid II.

Setelah dilakukannya penelitian ini, diperoleh kesimpulan bahwa mayoritas respon publik terhadap kebijakan PSBB Jilid II bersentimen negatif. Selain itu, *ensemble features* mendatangkan pengaruh yang baik di sisi akurasi, *recall*, dan *F1-measure*, namun tidak di sisi presisi. Begitu juga dengan fitur *parts of speech* dan *lexicon-based* secara individu, namun tidak dengan fitur tekstual yang tidak berpengaruh baik di sisi akurasi. Model terbaik yang ditemukan merupakan model yang dihasilkan oleh kombinasi fitur *lexicon-based* dan TF-IDF dengan akurasi dan *F1-measure* tertinggi, secara berturut-turut, 79,41 % dan 55,80 %.

Kata kunci: analisis sentimen, Bahasa Indonesia, Instagram, *ensemble features*, TF-IDF, *Support Vector Machine*, COVID-19, Pembatasan Sosial Berskala Besar, PSBB Jilid II, DKI Jakarta.

ABSTRACT

THE EFFECT OF ENSEMBLE FEATURES ON SUPPORT VECTOR MACHINE FOR PUBLIC SENTIMENT ANALYSIS ON THE LARGE SCALE SOCIAL LIMITATIONS, VOLUME II, DKI JAKARTA

by

Almas Fauzia Wibawa
17/409427/PA/17734

One of government's policy to overcome and prevent further damage from COVID-19 pandemic is to enact the Large Scale Social Limitation (LSSL) in DKI Jakarta. The government received all kinds of responses. Meanwhile, a leader should have another perspective as a consideration before making a decision and as a knowledge about public's perceptions before announcing a new policy in order to control public's response. Instagram is one of the most popular social media in Indonesia. Regional governments are already using it to gain public's aspiration through the comment section. Meanwhile, sentiment analysis often used as a solution to summarize public's opinion in a particular topics with social media data.

In this research, public's response's sentiment about the re-enactment of PSBB in DKI Jakarta are analyzed. The effect of textual features, parts of speech features, and lexicon-based features, as a part of ensemble features, on Term Frequency – Inverse Document Frequency (TF-IDF) and Support Vector Machine (SVM) with comments data from Instagram in Bahasa Indonesia are also being researched. Data that are used in this research are gained from public's comments on @aniesbaswedan and @dkijakarta's post, that are also LSSL Volume II enactment announcement, in Instagram.

After the research is done, we can conclude that the majority of public's responses about LSSL Volume II have negative sentiments. Meanwhile, ensemble features make a good effect on the accuracy, recall, and F1-measure of classification model using TF-IDF, SVM, and the data, but bad effect on the precision, as well as parts of speech features and lexicon-based features individually. It is different with the textual features that also have bad effect on accuracy. The combination of lexicon-based features and TF-IDF forms the best model with the best accuracy and F1-measure score 79.41 % and 55.80 %.

Keywords: sentiment analysis, Bahasa Indonesia, Instagram, ensemble features, TF-IDF, Support Vector Machine, COVID-19, *Pembatasan Sosial Berskala Besar, PSBB Jilid II*, DKI Jakarta.