

ABSTRAK

Kepopuleran *Spotify* di berbagai belahan dunia sebagai aplikasi *streaming* musik, mengakibatkan sebagian besar masyarakat sering mendengar bahkan menggunakan *Spotify*, sehingga memicu argumentasi/pendapat/sentimen masyarakat terhadap keberadaan dan fungsi *Spotify*. Penelitian ini melakukan analisis sentimen terhadap 61594 data sentimen penduduk dunia terkait tanggapannya terhadap aplikasi *Spotify*. Setelah data diinput dan dibersihkan melalui *text preprocessing*, data sentimen yang sebelumnya tidak diketahui kelas labelnya, diberi label melalui metode pelabelan *VADER Lexicon*. Dari hasil pelabelan, terlihat bahwa data tidak seimbang, sehingga sebagian data perlu diseimbangkan menggunakan teknik *Balancing Random Under-Sampling*. Kemudian, data direpresentasikan secara numerik dan dibagi menjadi data uji dan latih. Selanjutnya, menggunakan kernel *Multinomial* dan *Bernoulli Naïve Bayes* dilakukan pelatihan dan pengujian model baik tanpa dan dengan *balancing* untuk memprediksi kelas label dari data uji. Hasil akurasi optimal pada pengujian model tanpa *balancing* untuk kernel *Multinomial Naïve Bayes* adalah 75.48%, sedangkan kernel *Bernoulli* mencapai puncak akurasi dengan nilai 75.17%. Hasil akurasi optimal pada pengujian model dengan *Balancing Random Under-Sampling* untuk kernel *Multinomial Naïve Bayes* adalah 72.79%, sedangkan kernel *Bernoulli* mencapai puncak akurasi dengan nilai 65.16%. Dengan demikian, dapat disimpulkan bahwa model tanpa *Balancing Random Under-Sampling* lebih unggul dibandingkan ketika model diberi implementasi teknik *balancing* pada data latih.

Kata kunci: *Spotify*, Analisis Sentimen, *Multinomial Naïve Bayes*, *Bernoulli Naïve Bayes*, *Balancing Random Under-Sampling*

ABSTRACT

The popularity of Spotify in various parts of the world as a music streaming application, resulting in most people often hearing and even using Spotify, thus triggering arguments / opinions / public sentiment towards the existence and function of Spotify. This research conducts sentiment analysis on 61594 sentiment data of the world's population regarding their responses to the Spotify application. After the data is inputted and cleaned through text preprocessing, sentiment data that was previously unknown to the label class, is labeled through the VADER Lexicon labeling method. From the labeling results, it can be seen that the data is not balanced, so some data needs to be balanced using the Balancing Random Under-Sampling technique. Then, the data is represented numerically and divided into test and training data. Next, using the Multinomial kernel and Bernoulli Naïve Bayes, training and testing of the model both without and with balancing are carried out to predict the label class of the test data. The optimal accuracy result on model testing without balancing for the Multinomial Naïve Bayes kernel is 75.48%, while the Bernoulli kernel reaches the peak accuracy with a value of 75.17%. The optimal accuracy result in model testing with Random Under-Sampling Balancing for the Multinomial Naïve Bayes kernel is 72.79%, while the Bernoulli kernel reaches peak accuracy with a value of 65.16%. Thus, it can be concluded that the model without Balancing Random Under-Sampling is superior to the model without Balancing Random Under-Sampling.

Keyword: Spotify, Sentiment Analysis, Multinomial Naïve Bayes, Bernoulli Naïve Bayes, Balancing Random Under-Sampling