

ABSTRAK

ADITYO KHORI RAMADHAN. Analisis Sentimen Komentar *YouTube* terkait Peluncuran DANANTARA Menggunakan Metode *Naïve Bayes* dengan Pendekatan *Random Undersampling* dan *Tomek Link* untuk Menangani Ketidakseimbangan Data. Dibimbing oleh NURFALINDA dan FORTIA MAGFIRA.

Peluncuran DANANTARA di Indonesia memicu beragam respons publik yang tercermin melalui komentar *YouTube*. Penelitian ini bertujuan menganalisis pengaruh kombinasi metode *Random Undersampling* dan *Tomek Link* terhadap kinerja algoritma *Naïve Bayes* dalam mengklasifikasikan sentimen komentar menjadi positif, negatif, dan netral. Data penelitian dikumpulkan melalui *YouTube API v3* sebanyak 12.232 komentar dari dua video terkait. Tahapan *Preprocessing* meliputi *case Folding*, *cleaning*, translasi, tokenisasi, *stopword removal*, normalisasi, dan *stemming*, dilanjutkan dengan pelabelan menggunakan *Lexicon InSet* dan pembobotan *TF-IDF*. Evaluasi dilakukan menggunakan *10-Fold Cross Validation* pada *Dataset* asli (*imbalanced*) dan *Dataset* hasil *balancing*. Hasil menunjukkan bahwa kombinasi *Random Undersampling* dan *Tomek Link* meningkatkan *F1-Score* dari 58,58% menjadi 64,97% (+6,39%), *Recall* dari 59,51% menjadi 65,08% (+5,57%), dan akurasi dari 65,26% menjadi 67,11% (+1,85%). Meskipun *Precision* menurun sebesar 3,71%, model *balanced* menunjukkan performa yang lebih seimbang dalam mengklasifikasikan kelas minoritas, terlihat dari peningkatan prediksi yang lebih merata pada *confusion matrix*.

Kata kunci: Analisis Sentimen, DANANTARA, *Naïve Bayes*, *Random Undersampling*, *Tomek Link*.

ABSTRACT

ADITYO KHORI RAMADHAN. Sentiment Analysis of YouTube Comments on the Launch of DANANTARA Using the Naïve Bayes Method with Random Undersampling and Tomek Link Approach to Handle Imbalanced Data. Supervised by NURFALINDA and FORTIA MAGFIRA.

The launch of DANANTARA in Indonesia sparked a variety of public responses reflected through YouTube comments. This study aims to analyze the effect of the combination of Random Undersampling and Tomek Link methods on the performance of the Naïve Bayes algorithm in classifying comment sentiment into positive, negative, and neutral. Research data was collected through YouTube API v3 totaling 12,232 comments from two related videos. Preprocessing stages include case Folding, cleaning, translation, tokenization, stopword removal, normalization, and stemming, followed by labeling using Lexicon InSet and TF-IDF weighting. Evaluation was carried out using 10-Fold Cross Validation on the original (imbalanced) dataset and the balanced dataset. The results show that the combination of Random Undersampling and Tomek Link improves F1-Score from 58.58% to 64.97% (+6.39%), Recall from 59.51% to 65.08% (+5.57%), and accuracy from 65.26% to 67.11% (+1.85%). Although Precision decreased by 3.71%, the balanced model showed a more balanced performance in classifying minority classes, as seen from the more evenly distributed predictions in the confusion matrix.

Keywords: DANANTARA, Naïve Bayes, Random Undersampling, Sentiment Analysis, Tomek Link.