

ABSTRAK

Media sosial X menjadi ruang utama masyarakat dalam menyampaikan opini terhadap kebijakan pemerintah, termasuk isu kenaikan tunjangan Dewan Perwakilan Rakyat (DPR) yang memicu perdebatan luas. Penelitian ini bertujuan mengklasifikasikan sentimen publik terhadap kebijakan tersebut ke dalam kategori positif dan negatif menggunakan algoritma *Decision Tree* CART. Data yang digunakan berupa 4.682 komentar dari media sosial X yang dikumpulkan melalui kata kunci “tuntutan 17+8”, “tunjangan rumah DPR”, dan “jarah rumah” pada periode Agustus hingga Oktober 2025, diberi label secara manual, dan divalidasi oleh ahli bahasa. Distribusi awal menunjukkan ketidakseimbangan kelas dengan 73,8% sentimen negatif dan 26,2% sentimen positif. Untuk mengatasinya, diterapkan tiga teknik *text augmentation* yaitu *Synonym Replacement*, *Back Translation* berbasis Google Translate API dengan perantara Bahasa Mandarin, dan kombinasi keduanya untuk menyeimbangkan distribusi kelas menjadi 50:50. Data diproses melalui *pre-processing* meliputi *cleaning*, *case folding*, tokenisasi, normalisasi, *stopword removal*, dan *stemming*, lalu dibobot menggunakan TF-IDF. Pengujian dilakukan dalam lima skenario dengan pembagian data 80:20 dan optimasi *hyperparameter* menggunakan GridSearch 5-fold *cross-validation*. Hasil terbaik diperoleh pada skenario kombinasi *Synonym Replacement* dan *Back Translation* dengan akurasi 84,56%, *precision* 85,00%, *recall* 84,56%, dan f1-score 84,67%, meningkat 10,07% dibandingkan *baseline* tanpa augmentasi sebesar 74,49%. Hasil ini membuktikan bahwa teknik *text augmentation* efektif mengatasi ketidakseimbangan data dan meningkatkan kemampuan model dalam mengenali kelas minoritas pada analisis sentimen berbahasa Indonesia.

Kata kunci: Analisis Sentimen, *Text Augmentation*, *Decision Tree* CART, TF-IDF, *Imbalance Dataset*.

ABSTRACT

Social media platform X has become a primary space for the public to express opinions on government policies, including the controversial issue of the House of Representatives (DPR) allowance increase that triggered widespread debate. This study aims to classify public sentiment regarding this policy into positive and negative categories using the Decision Tree CART algorithm. The data used consists of 4,682 comments collected from social media X through the keywords "tuntutan 17+8", "tunjangan rumah DPR", and "jarah rumah" during the period of August to October 2025, which were manually labeled and validated by a language expert. The initial distribution revealed a class imbalance with 73.8% negative sentiment and 26.2% positive sentiment. To address this, three text augmentation techniques were applied, namely Synonym Replacement, Back Translation using the Google Translate API via Mandarin as an intermediate language, and a combination of both, in order to balance the class distribution to 50:50. The data was processed through pre-processing stages including cleaning, case folding, tokenization, normalization, stopword removal, and stemming, then weighted using TF-IDF. Testing was conducted in five scenarios with an 80:20 data split and hyperparameter optimization using GridSearch 5-fold cross-validation. The best result was achieved in the combination of Synonym Replacement and Back Translation scenario with an accuracy of 84.56%, precision of 85.00%, recall of 84.56%, and f1-score of 84.67%, an improvement of 10.07% compared to the baseline without augmentation at 74.49%. These results demonstrate that text augmentation techniques are effective in overcoming data imbalance and improving the model's ability to recognize the minority class in Indonesian sentiment analysis.

Keywords: *Sentiment Analysis, Text Augmentation, Decision Tree CART, TF-IDF, Imbalanced Dataset.*