

INTISARI

Penelitian ini berjudul Optimasi Analisis Sentimen Program Makan Bergizi Gratis Menggunakan Algoritma *Support Vector Machine* dan *Naive Bayes* Berbasis Teknik SMOTE. Penelitian ini dilakukan karena studi terdahulu pada sentimen MBG umumnya melaporkan akurasi tinggi, tetapi masih menyisakan kelemahan pada deteksi kelas minoritas akibat ketidakseimbangan kelas, sehingga evaluasi berisiko terjebak pada *accuracy paradox*. Penelitian ini bertujuan membandingkan dan mengevaluasi kinerja *Support Vector Machine* (SVM) dan *Multinomial Naive Bayes* (MNB). Variabel penelitian dibatasi pada 3.359 data dari Kaggle yang telah diseleksi dari 10.000 data awal, dipraolah, dan diberi label positif, negatif, dan netral. Metode analisis menggunakan TF-IDF 5.000 fitur, stratified train-test split 80:20, SMOTE hanya pada data latih, serta evaluasi *accuracy*, *precision*, *recall*, *macro F1-score*, dan *overfitting*.

Hasil penelitian menunjukkan bahwa SVM memiliki kinerja lebih tinggi daripada MNB. Pada data uji, SVM memperoleh akurasi 99,70 dan macro F1-score 99,75, sedangkan MNB memperoleh akurasi 95,86 dan macro F1-score 95,33. Kebaruan penelitian ini terletak pada pengujian terkontrol pengaruh SMOTE pada komparasi SVM dan MNB untuk klasifikasi sentimen tiga kelas MBG menggunakan dataset yang telah melalui seleksi kualitas, dengan SMOTE diterapkan hanya pada data latih agar evaluasi tetap objektif. Hasil tersebut menunjukkan bahwa penyeimbangan kelas dan evaluasi berbasis metrik per kelas memberikan penilaian model yang lebih sahih daripada hanya mengandalkan akurasi.

Kesimpulan penelitian ini menunjukkan bahwa SVM merupakan model terbaik pada konfigurasi penelitian ini. Penerapan SMOTE pada data latih efektif meningkatkan representasi kelas minoritas, sedangkan gap macro F1-score sebesar 0,24 pada SVM dan 2,10 pada MNB menunjukkan generalisasi model yang tetap stabil. Kontribusi ilmiah penelitian ini adalah penegasan bahwa pada korpus sentimen kebijakan publik yang sangat timpang, validitas evaluasi model lebih ditentukan oleh penanganan ketidakseimbangan kelas, metrik per kelas, dan kestabilan generalisasi daripada akurasi tunggal.

Kata kunci : Analisis Sentimen, Makan Bergizi Gratis, SMOTE, SVM, *Naive Bayes*

ABSTRACT

This thesis is entitled Optimization of Sentiment Analysis of the Free Nutritious Meal Program Using Support Vector Machine and Naive Bayes Algorithms Based on the SMOTE Technique. This study was conducted because previous studies on MBG sentiment analysis generally reported high accuracy, but still showed weaknesses in detecting minority classes due to class imbalance, creating a risk of the accuracy paradox. The objective of this study was to compare and evaluate the performance of Support Vector Machine (SVM) and Multinomial Naive Bayes (MNB). The research variables were limited to 3,359 Kaggle records selected from 10,000 initial records, preprocessed, and labeled into positive, negative, and neutral sentiments. The analysis used TF-IDF with 5,000 features, an 80:20 stratified train-test split, SMOTE applied only to the training data, and evaluation based on accuracy, precision, recall, macro F1-score, and overfitting.

The results showed that SVM outperformed MNB. On the test data, SVM achieved 99.70 accuracy and 99.75 macro F1-score, while MNB achieved 95.86 accuracy and 95.33 macro F1-score. The novelty of this study lies in the controlled testing of SMOTE in the comparison of SVM and MNB for three-class MBG sentiment classification using a quality-filtered dataset, with SMOTE applied only to the training set to preserve objective evaluation. The findings indicate that class balancing and class-wise evaluation provide a more valid assessment of model performance than relying on accuracy alone.

This study concludes that SVM was the best-performing model under the applied experimental setting. Applying SMOTE to the training data effectively improved minority-class representation, while the macro F1-score gaps of 0.24 for SVM and 2.10 for MNB indicated stable generalization. The scientific contribution of this study is the confirmation that, in highly imbalanced public policy sentiment corpora, the validity of sentiment classification is determined more by imbalance handling, class-wise metrics, and generalization stability than by single aggregate accuracy

Keywords: Sentiment Analysis, Free Nutritious Meal, SMOTE, SVM, Naive Bayes.