

ABSTRAK

Wardana, Kadek Kusuma (2026), Perbandingan Kinerja Naive Bayes Dan SVM Berbasis SMOTE Dalam Analisis Sentimen Komentar Youtube Mengenai Wacana Redenominasi Rupiah. Tesis, Ilmu Komputer, Program Pascasarjana, Universitas Pendidikan Ganesha

Tesis ini sudah disetujui dan diperiksa oleh Pembimbing I : Dr. I Gede Aris Gunadi, S.Si., M.Kom. dan Pembimbing II: Prof. Dr. Luh Joni Erawati Dewi, S.T., M.Pd.

Kata kunci: Analisis Sentimen, Redenominasi Rupiah, Naïve Bayes, SVM, SMOTE, K-Fold

Wacana redenominasi rupiah merupakan salah satu isu ekonomi yang memunculkan beragam tanggapan masyarakat di media sosial, khususnya YouTube. Analisis sentimen terhadap komentar masyarakat diperlukan untuk mengetahui kecenderungan opini publik terhadap isu tersebut. Namun, ketidakseimbangan distribusi data dapat memengaruhi kinerja model klasifikasi sehingga diperlukan teknik penyeimbangan data. Penelitian ini bertujuan untuk: (1) menganalisis penerapan algoritma Naïve Bayes Classifier (NBC) dan Support Vector Machine (SVM) berbasis Synthetic Minority Oversampling Technique (SMOTE) pada analisis sentimen komentar YouTube mengenai wacana redenominasi rupiah, serta (2) menentukan model klasifikasi dengan tingkat performansi terbaik. Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen. Data penelitian berupa komentar YouTube dari kanal CNBC Indonesia, Kompas Pontianak, dan Kumparan. Tahapan penelitian meliputi *case folding*, *cleaning*, tokenisasi, normalisasi, *stopword removal*, *stemming*, pembobotan TF-IDF, serta penyeimbangan data menggunakan SMOTE pada beberapa rasio distribusi kelas. Evaluasi model dilakukan menggunakan *confusion matrix*, *accuracy*, *precision*, *recall*, *F1-score*, *ROC-AUC*, *K-Fold Cross Validation*, dan analisis *overfitting*. Hasil penelitian menunjukkan bahwa SMOTE meningkatkan performa kedua algoritma dengan hasil terbaik pada rasio 50:50. Peningkatan performa relatif terbatas, sekitar 1–2%, yang menunjukkan bahwa ketidakseimbangan kelas bukan merupakan faktor utama karena representasi fitur TF-IDF telah mampu membedakan sentimen pada data asli. Pada rasio tersebut, SVM memperoleh *accuracy* 76,34%, *precision* 76,33%, *recall* 76,34%, *F1-score* 75,46%, dan *AUC* 81,44%, sedangkan NBC memperoleh *accuracy* 72,35%, *precision* 73,86%, *recall* 72,35%, *F1-score* 72,68%, dan *AUC* 80,46%. Hasil *K-Fold Cross Validation* menunjukkan bahwa SVM lebih stabil dibandingkan NBC, sedangkan analisis *overfitting* menunjukkan SVM memiliki *generalization gap* lebih besar tetapi tetap memberikan performa pengujian terbaik. Analisis homogenitas menunjukkan panjang komentar pada data training dan testing tidak berbeda signifikan ($p\text{-value} = 0,3437$), sehingga representasi TF-IDF tidak dipengaruhi oleh perbedaan karakteristik dokumen. Dengan demikian, kombinasi SVM dan SMOTE pada rasio 50:50 merupakan model terbaik untuk analisis sentimen komentar YouTube mengenai wacana redenominasi rupiah.

ABSTRACT

Wardana, Kadek Kusuma (2026), *Comparison of the Performance of Naive Bayes and SMOTE-Based SVM in Sentiment Analysis of YouTube Comments Regarding the Rupiah Redenomination Discourse*. Thesis, Computer Science, Master Program, Ganesha University of Education.

This thesis has been approved and reviewed by Supervisor I: Dr. I Gede Aris Gunadi, S.Si., M.Kom. and Supervisor II: Prof. Dr. Luh Joni Erawati Dewi, S.T., M.Pd.

Keywords: Sentiment Analysis, Redenomination, Rupiah, Naïve Bayes, SVM, SMOTE

The rupiah redenomination discourse is an economic issue that has generated diverse public responses on social media, especially YouTube. Sentiment analysis of public comments is needed to determine the tendency of public opinion towards the issue. However, imbalanced data distribution can affect the performance of classification models, so data balancing techniques are needed. This study aims to: (1) analyze the application of the Naïve Bayes Classifier (NBC) and Support Vector Machine (SVM) algorithms based on Synthetic Minority Oversampling Technique (SMOTE) in sentiment analysis of YouTube comments regarding the rupiah redenomination discourse, and (2) determine the classification model with the best performance level. This study uses a quantitative approach with an experimental method. The research data consists of YouTube comments from the channels CNBC Indonesia, Kompas Pontianak, and Kumparan. The research stages include case folding, cleaning, tokenization, normalization, stopword removal, stemming, TF-IDF weighting, and data balancing using SMOTE at several class distribution ratios. Model evaluation was conducted using confusion matrix, accuracy, precision, recall, F1-score, ROC-AUC, K-Fold Cross Validation, and overfitting analysis. The results showed that SMOTE improved the performance of both algorithms with the best results at a 50:50 ratio. The performance improvement was relatively limited, around 1–2%, indicating that class imbalance was not a major factor because the TF-IDF feature representation was already able to distinguish sentiments in the original data. At this ratio, SVM obtained an accuracy of 76.34%, a precision of 76.33%, a recall of 76.34%, an F1-score of 75.46%, and an AUC of 81.44%, while NBC obtained an accuracy of 72.35%, a precision of 73.86%, a recall of 72.35%, an F1-score of 72.68%, and an AUC of 80.46%. The results of K-Fold Cross Validation show that SVM is more stable than NBC, while overfitting analysis shows that SVM has a larger generalization gap but still provides the best testing performance. Homogeneity analysis shows that the length of comments in the training and testing data is not significantly different ($p\text{-value} = 0.3437$), so the TF-IDF representation is not affected by differences in document characteristics. Thus, the combination of SVM and SMOTE at a 50:50 ratio is the best model for sentiment analysis of YouTube comments regarding the rupiah redenomination discourse.