

ABSTRAK

Penelitian ini mengkaji perbandingan performa algoritma *Support Vector Machine* (SVM) dengan dan tanpa penerapan metode *Synthetic Minority Over-sampling Technique* (SMOTE) dalam tugas analisis sentimen terhadap komentar warganet terkait pembatalan Piala Dunia U-20 di Indonesia. Sumber data berasal dari 1.601 komentar pada video di kanal MetroTV di YouTube yang berjudul “Piala Dunia U20 Indonesia Batal, Rp1,4 T Melayang!”. Data dianalisis melalui tahapan preprocessing teks, pembobotan kata menggunakan TF-IDF, hingga pengujian model klasifikasi.

Evaluasi dilakukan menggunakan *confusion matrix* untuk menilai metrik performa seperti akurasi, presisi, recall, dan F1-score. Hasil menunjukkan bahwa SVM tanpa SMOTE mencatat akurasi lebih tinggi, yaitu 75,38%. Namun, metode ini kurang efektif dalam mengenali sentimen dari kelas minoritas. Sebaliknya, SVM dengan SMOTE memberikan peningkatan nilai recall menjadi 51,63% dan F1-score menjadi 50,32%, meskipun akurasinya lebih rendah, yakni 66,04%.

Temuan ini menunjukkan bahwa penggunaan SMOTE membantu menghasilkan distribusi klasifikasi yang lebih seimbang antar kelas, walaupun sedikit menurunkan akurasi keseluruhan. Oleh karena itu, pemilihan pendekatan yang digunakan sebaiknya disesuaikan dengan tujuan analisis—apakah berfokus pada ketepatan total atau keadilan dalam deteksi semua jenis sentimen..

Kata kunci: **Analisis Sentimen, SVM, SMOTE, TF-IDF, Komentar YouTube, Pembatalan Piala Dunia U-20.**

ABSTRACT

This study explores the comparative performance of the Support Vector Machine (SVM) algorithm with and without the application of the Synthetic Minority Over-sampling Technique (SMOTE) in sentiment analysis of public comments concerning the cancellation of the U-20 World Cup in Indonesia. The dataset, consisting of 1,601 YouTube comments from the MetroTV channel titled "Piala Dunia U20 Indonesia Batal, Rp1.4 T Melayang!" ("Indonesia's U20 World Cup Cancelled, IDR 1.4 Trillion Lost!"), underwent a series of preprocessing stages including text cleaning, TF-IDF-based feature weighting, and classification modeling.

Evaluation using a confusion matrix was conducted to assess key metrics such as accuracy, precision, recall, and F1-score. The results revealed that the SVM model without SMOTE achieved a higher overall accuracy of 75.38%, yet showed limitations in identifying minority classes. In contrast, the SVM model enhanced with SMOTE demonstrated improved recall (51.63%) and F1-score (50.32%), despite a lower accuracy of 66.04%.

These findings suggest that while SMOTE may reduce total accuracy, it enhances class balance and minority class recognition. Therefore, the choice between using SMOTE or not should be aligned with the primary objective of the analysis—whether to prioritize overall accuracy or balanced sentiment classification across all categories.

Keywords: Sentiment Analysis, Support Vector Machine, SMOTE, TF-IDF, YouTube Comments, U-20 World Cup Cancellation