ABSTRAK

Ferdy Ilham Maulana. 2025. Pengaruh Metode *Imbalance Data Sampling* Terhadap Kinerja *Support Vector Machine* Untuk Sentimen Analisis Naturalisasi Pemain Sepak Bola. Tugas Akhir. Program Sarjana. Program Studi Teknik Informatika. Universitas Muhammadiyah Jember.

Pembimbing: (1) Deni Arifianto, M.Kom; (2) Dr. Reni Umilasari, S.Pd, M.Si

Media sosial kini menjadi wadah penting bagi masyarakat untuk menyampaikan pandangan terhadap berbagai isu publik, termasuk wacana naturalisasi pemain sepak bola. Untuk mengidentifikasi arah opini tersebut, digunakan pendekatan analisis sentimen berbasis algoritma *Support Vector Machine (SVM).* Namun, distribusi kelas data yang tidak seimbang menjadi kendala utama karena menyebabkan model lebih condong ke kelas mayoritas. Penelitian ini bertujuan untuk mengukur efektivitas tiga teknik *imbalance data sampling*, yaitu *Borderline-SMOTE, SVM-SMOTE,* dan *ADASYN,* dalam meningkatkan kinerja *SVM*. Data yang digunakan diperoleh dari platform X (Twitter), kemudian diklasifikasikan ke dalam sentimen positif, netral, dan negatif. Evaluasi dilakukan menggunakan teknik *K-Fold Cross Validation* dengan variasi nilai K = 2, 3, 5, 6, 9, dan 10, serta metrik evaluasi berupa akurasi, presisi, recall, dan F1-score. Hasil eksperimen menunjukkan bahwa model tanpa penyeimbangan data memperoleh akurasi 86%, presisi 91%, *recall* 85%, dan *F1-score* 91%. Penerapan *Borderline-SMOTE* meningkatkan akurasi menjadi 87%, presisi 91%, *recall* 86%, dan *F1-score* tetap 91%. Metode *SVM-SMOTE* menghasilkan akurasi tertinggi yaitu 91%, disertai presisi 91%, *recall* 86%, dan *F1-score* 91%. Sementara itu, *ADASYN* menghasilkan *F1-score* tertinggi sebesar 92%, dengan akurasi 87%, presisi 91%, dan *recall* 87%. *Borderline-SMOTE* dan SVM-SMOTE juga menunjukkan peningkatan signifikan pada label netral dengan *F1-score* 69%, sedangkan *ADASYN* unggul dalam klasifikasi kelas negatif dengan *F1-score* 92%. Dari temuan tersebut, dapat disimpulkan bahwa meskipun *SVM-SMOTE* unggul secara akurasi, *ADASYN* memberikan hasil klasifikasi yang lebih seimbang dan andal, terutama dalam menangani ketimpangan distribusi kelas minoritas pada data sentimen.

Kata Kunci: *Imbalance Data, Borderline-SMOTE, SVM-SMOTE, ADASYN, SVM, Analisis Sentimen.*

ABSTRACT

Ferdy Ilham Maulana. 2025. The Effect of Imbalanced Data Sampling Methods on the Performance of Support Vector Machines for Sentiment Analysis of Soccer Player Naturalization.. Final Project. Undergraduate Program. Informatics Engineering Study Program. Universitas Muhammadiyah Jember.

Advisors: (1) Deni Arifianto, M.Kom; (2) Dr. Reni Umilasari, S.Pd, M.Si

Social media has become an important platform for the public to express their views on various public issues, including the discourse on the naturalization of soccer players. To identify the direction of these opinions, a sentiment analysis approach based on the Support Vector Machine (SVM) algorithm was used. However, the imbalance in the distribution of data classes posed a major challenge, as it caused the model to favor the majority class. This study aims to measure the effectiveness of three data imbalance sampling techniques—Borderline-SMOTE, SVM-SMOTE, and ADASYN—in improving SVM performance. The data used was obtained from platform X (Twitter) and classified into positive, neutral, and negative sentiments. Evaluation was conducted using the K-Fold Cross Validation technique with varying values of K = 2, 3, 5, 6, 9, and 10, and evaluation metrics including accuracy, precision, recall, and F1-score. The experimental results show that the model without data balancing achieved an accuracy of 86%, precision of 91%, recall of 85%, and an F1-score of 91%. The application of Borderline-SMOTE improved accuracy to 87%, precision to 91%, recall to 86%, and the F1-score remained at 91%. The SVM-SMOTE method achieved the highest accuracy of 91%, accompanied by precision of 91%, recall of 86%, and an F1-score of 91%. Meanwhile, ADASYN produced the highest F1-score of 92%, with an accuracy of 87%, precision of 91%, and recall of 87%. Borderline-SMOTE and SVM-SMOTE also showed significant improvements in neutral labels with an F1-score of 69%, while ADASYN excelled in negative class classification with an F1-score of 92%. From these findings, it can be concluded that while SVM-SMOTE excels in accuracy, ADASYN provides more balanced and reliable classification results, particularly in addressing the imbalance in the distribution of minority classes in sentiment data.

Keywords**:** *Imbalanced Data, Borderline-SMOTE, SVM-SMOTE, ADASYN, SVM, Sentiment Analysis.*