ABSTRAK

Agil Dwi Saputra. 2025. Pengaruh Metode *Random Under Sampling Random Over Sampling* Dan *Smote* Terhadap Kinerja *Support Vector Machine* Untuk Analisis Sentimen. Universitas Muhammadiyah Jember.

Pembimbing: Deni Arifianto, M.Kom; Dr. Reni Umilasari, S.Pd., M.Si.

Ketidakseimbangan data merupakan tantangan dalam analisis sentimen karena dapat menyebabkan model *bias* terhadap kelas mayoritas. Penelitian ini mengevaluasi pengaruh metode *resampling*, yaitu *Random Under Sampling* (*RUS*), *Random Over Sampling* (*ROS*), dan *Synthetic Minority Oversampling Technique* (*SMOTE*), terhadap performa algoritma *Support Vector Machine* (*SVM)* pada data yang tidak seimbang. Data diperoleh dari media sosial *X* (*Twitter*) dengan topik naturalisasi pemain sepak bola di Indonesia. Tahapan penelitian meliputi *preprocessing, TF-IDF*, pengujian model menggunakan *K-Fold Cross Validation* K 2, 3, 5, 6, 9 dan 10, dengan evaluasi *matrix F1-score, Recall, Precision*, dan *Accuracy*. Hasil evaluasi hanya ditampilkan nilai tertinggi dari masing-masing *label* pada setiap metode, Model *SVM* tanpa *Balancing* bekerja baik pada *label negatif* dan *positif*, dengan *F1-score* 91% dan 90%, namun lemah pada *label netral* dengan *F1-score* 44% dan *Recall* 29%. Metode *RUS* meningkatkan *Recall label netral* menjadi 81% dan *F1-score* 68%, meskipun *Accuracy* turun menjadi 84%. *SMOTE* menghasilkan *Precision* tinggi pada *label negatif* dan *netral*, masing-masing 100% dan 90%, namun *Recall label netral* tetap rendah yaitu 41%. *ROS* memberikan hasil terbaik dan paling seimbang, dengan *F1-score* tertinggi pada semua *label*, serta *Accuracy* keseluruhan tertinggi 92%. *ROS* terbukti paling efektif dalam meningkatkan kinerja *SVM* pada data sentimen yang tidak seimbang.

Kata Kunci: *data imbalance*, analisis sentimen, *RUS, ROS, SMOTE, SVM*

ABSTRACT

Agil Dwi Saputra. 2025. Effect Of Random Under Sampling Random Over Sampling And Smote Method On Support Vector Machine Performance For Sentiment Analysis. Muhammadiyah University of Jember.

Advisors: *Deni Arifianto, M.Kom; Dr. Reni Umilasari, S.Pd., M.Si.*

Data imbalance is a challenge in sentiment analysis because it can cause the model to be biased towards the majority class. This study evaluates the effect of resampling methods, namely Random Under Sampling (RUS), Random Over Sampling (ROS), and Synthetic Minority Oversampling Technique (SMOTE), on the performance of the Support Vector Machine (SVM) algorithm on unbalanced data. Data was obtained from X social media (Twitter) with the topic of naturalization of soccer players in Indonesia. The research stages include preprocessing, TF-IDF, model testing using K-Fold Cross Validation K 2, 3, 5, 6, 9 and 10, with evaluation of F1-score matrix, Recall, Precision, and Accuracy. The evaluation results only show the highest value of each label in each method, SVM model without Balancing works well on negative and positive labels, with F1-score 91% and 90%, but weak on neutral labels with F1-score 44% and Recall 29%. The RUS method improved neutral label Recall to 81% and F1-score 68%, although Accuracy dropped to 84%. SMOTE produced high Precision on negative and neutral labels, 100% and 90% respectively, but neutral label Recall remained low at 41%. ROS gave the best and most balanced results, with the highest F1-score on all labels, and the highest overall Accuracy of 92%. ROS proved to be the most effective in improving SVM performance on unbalanced sentiment data.

Keywords*: data imbalance, sentiment analysis, RUS, ROS, SMOTE, SVM*