

## INTISARI

# **METODE *ADAPTIVE SYNTHETIC SAMPLING APPROACH* (ADASYN) DENGAN STRATEGI KLASIFIKASI MULTIKELAS *ONE VS ONE* (OVO) DAN *ONE VS ALL* (OVA) PADA DATA TIDAK SEIMBANG**

Oleh

Mesy Diah Ariani  
20/456586/PA/19773

Ketidakseimbangan kelas masih menjadi tantangan baik dalam kasus klasifikasi biner maupun multikelas. Ketidakseimbangan kelas menyebabkan model yang diperoleh cenderung fokus pada kelas yang jumlahnya banyak dan mengabaikan pentingnya kelas yang jumlahnya sedikit sehingga menyebabkan efek buruk pada ketepatan prediksi kelas minoritas. Salah satu metode penanganan ketidakseimbangan kelas adalah *Adaptive Synthetic Sampling Approach* (ADASYN). Penelitian ini bertujuan untuk melihat efektifitas ADASYN dalam peningkatan performa model berdasarkan karakteristik datasetnya. *Support Vector Machine* digunakan sebagai metode klasifikasi. Dataset utama yang digunakan merupakan data sekunder dari *UCI Machine Learning Repository*, yaitu *glass* dan *dry bean*. ADASYN bekerja efektif pada data *glass*. Berbeda dengan data *dry bean*, sebelum dan sesudah penerapan ADASYN, baik dengan *One versus One* (OVO) dan *One versus All* (OVA) tidak terjadi perubahan signifikan performa pada *kappa score*, *F1-score macro*, dan *F1-score weighted*. Pengujian lebih lanjut dilakukan pada data simulasi dan data *date fruit* untuk melihat performa ADASYN yang cenderung tidak optimal diterapkan pada data *dry bean*. Hasil analisis menunjukkan bahwa karakteristik dataset dengan mayoritas distribusi fitur dengan dua puncak (bimodal) atau lebih dan korelasi antar fitur yang cenderung tergolong tinggi berpeluang menyebabkan ADASYN kurang optimal diterapkan.

Kata kunci: ADASYN, ketidakseimbangan kelas, OVA, OVO, *Support Vector Machine*

## **ABSTRACT**

# **ADAPTIVE SYNTHETIC SAMPLING APPROACH (ADASYN) METHOD WITH ONE VS ONE (OVO) AND ONE VS ALL (OVA) MULTICLASS CLASSIFICATION STRATEGIES ON UNBALANCED DATA**

by

Mesy Diah Ariani  
20/456586/PA/19773

Class imbalance remains a challenge in both binary and multiclass classification cases. Class imbalance causes the obtained model to focus on the large number of classes and ignore the importance of the small number of classes, causing adverse effects on the prediction accuracy of minority classes. One method of handling class imbalance is the Adaptive Synthetic Sampling Approach (ADASYN). This study aims to determine the effectiveness of ADASYN in improving model performance based on the characteristics of the dataset. Support vector machines are used as the classification method. The main datasets used are secondary data from the UCI Machine Learning Repository, namely glass and dry bean. ADASYN works effectively on glass data. Unlike the dry bean data, before and after the application of ADASYN, both with One versus One (OVO) and One versus All (OVA), there was no significant change in performance on the kappa score, F1-score macro, or F1-score weighted. Further testing was carried out on simulated data and date fruit data to see the performance of ADASYN, which tends to be less optimal when applied to dry bean data. The results of the analysis show that the characteristics of datasets with a majority feature distribution with two peaks (bimodal) or more and correlations between features that tend to belong high are likely to cause ADASYN to be less optimally applied.

**Keywords:** ADASYN, Imbalanced Class, OVA, OVO, Support Vector Machine