

## Abstrak

Proses klasifikasi dengan berbagai algoritma *machine learning* bertujuan untuk mendapatkan target kelas yang akurat. Namun pada kenyataannya muncul permasalahan dalam proses klasifikasi tersebut ketika salah satu kelas memiliki jumlah data yang jauh lebih kecil pada *training dataset* nya. Permasalahan tersebut disebut juga dengan *imbalance dataset problem*.

Untuk menyelesaikan permasalahan tersebut salah satu pendekatan yang sangat populer adalah dengan metode *sampling*. Metode *sampling* bekerja dengan cara menyeimbangkan data *imbalance*. Dalam penerapannya akan digunakan tiga metode *sampling*, yaitu *oversampling*, *undersampling* dan *Combine*. Pengukuran data dilakukan dengan menentukan nilai *precision*, *recall*, *Fmeasure*, dan *HammLoss* yang merupakan hasil dari proses klasifikasi dengan menggunakan *SVM*, *kNN* dan *NaiveBayes*.

Yang akan dianalisis pada tugas akhir ini adalah membandingkan pengukuran data hasil klasifikasi sebelum dan setelah dilakukan metode *sampling*. Selain itu analisis akan melihat pula metode *sampling* mana yang menghasilkan metode terbaik untuk *resample* data.

**Kata kunci :** *imbalance dataset problem*, *sampling*, klasifikasi, *precision*, *recall*, dan *Fmeasure*.