

ABSTRAK

Penelitian mengenai pelabelan data tanaman Keladi Tikus menghasilkan data yang tidak seimbang dengan distribusi kelas mayoritas dan minoritas mencapai 653398 dan 9830 data. Ketidakseimbangan data merupakan masalah yang dapat menurunkan akurasi model prediktif. Penelitian ini bertujuan untuk menguji bagaimana pengaruh dari teknik penyeimbangan data *Random Oversampling* dan *Random Undersampling* terhadap algoritma klasifikasi ID3 pada dataset Keladi Tikus. Eksperimen dilakukan dengan melakukan preprocessing, balancing menggunakan *Random Oversampling* dan *Random Undersampling* dilanjutkan dengan klasifikasi dengan menggunakan ID3. Hasil klasifikasi menunjukkan rata-rata *running time* 1.10 detik pada dataset *imbalance*, 2.02 detik pada dataset *random oversampling*, dan 0.04 detik pada dataset *random undersampling*. Rata-rata akurasi yang didapatkan adalah 99.98% pada dataset *imbalance*, 99.99% pada dataset *random oversampling*, dan 99.16% pada dataset *random undersampling*. Teknik *balancing Random Oversampling* dan *Random Undersampling* tidak terlalu berpengaruh terhadap performa klasifikasi dan *running time* ID3 pada dataset Keladi Tikus.

Kata Kunci: *Keladi Tikus, Data Tidak Seimbang, Random Oversampling, Random Undersampling, ID3, Klasifikasi*

ABSTRACT

Research on the labeling of Rodent Tuber plant data produced an imbalanced dataset with a majority class distribution of 653398 instances and a minority class of 9830 instances. Data imbalance is a problem that can reduce the accuracy of predictive models. This study aims to test how the influence of the data balancing techniques Random Oversampling and Random Undersampling is on the ID3 classification algorithm on the Rodent Tuber dataset. The experiments were conducted by performing preprocessing, applying Random Oversampling and Random Undersampling for balancing, and then classifying the data using the ID3 algorithm. The results of the classification show an average running time of 1.10 seconds on the imbalance dataset, 2.02 seconds on the random oversampling dataset, and 0.04 seconds on the random undersampling dataset. The average accuracy obtained is 99.98% on the imbalance dataset, 99.99% on the random oversampling dataset, and 99.16% on the random undersampling dataset. Balancing techniques using Random Oversampling and Random Undersampling do not significantly affect the classification performance and running time of ID3 on the Rodent Tuber dataset.

Keywords: *Rodent Tuber, Imbalance Data, Random Oversampling, Random Undersampling, ID3, Classification*

