

PENGGUNAAN ALGORITMA RANDOM OVER SAMPLING UNTUK MENGATASI MASALAH IMBALANCE DATA PADA KLASIFIKASI GIZI BALITA

Syam Suryo Utomo¹, Triawan Adi Cahyanto², Bakhtiyar Hadi Prakoso³

Program Studi Informatika Fakultas Teknik Universitas Muhammadiyah Jember

Syamsuryo15@gmail.com, triawanac@unmuhjember.ac.id, bahtiyar.hp@gmail.com

ABSTRAK

Puskesmas Tegaldlimo adalah salah satu unit pelaksana kesehatan yang ada di daerah kecamatan dan mempunyai fungsi salah satunya melakukan upaya pemenuhan akan gizi dan kesehatan anak usia bawah lima tahun (balita). Di Puskesmas Tegaldlimo sampai 2018 tercatat ada sebanyak 1186 balita, disini ada empat status gizi yang menurut penulis sangat jauh perbedaannya yaitu “gizi baik”, “gizi kurang”, “gizi lebih”, “gizi buruk”, dengan rincian gizi baik sebanyak 1101 balita, gizi kurang sebanyak 39 balita, gizi lebih sebanyak 39 balita dan gizi buruk sebanyak 7 balita. Untuk mencari solusi masalah ketidakseimbangan kelas pada dataset, diusulkan model dengan pendekatan level data. Perancangan model yang diusulkan meliputi perancangan pendekatan level data ROS, dan algoritma pengklasifikasian berbasis *Naive Bayes*. Hasilnya menunjukkan bahwa model *Naive Bayes* menghasilkan akurasi *Excellent* (Sangat Baik) dengan rata – rata akurasi yang dihasilkan adalah 0,957844. Sedangkan akurasi dengan model ROS + NB cenderung lebih rendah atau bisa dikatakan belum maksimal untuk meningkatkan nilai akurasi dengan memiliki rata – rata yaitu 0,528716 *Fail* (Gagal) dalam mengklasifikasi *dataset* Puskesmas, hal ini dikarenakan jumlah data yang sangat signifikan banyak dari data “gizi baik” dan akan mempengaruhi hasil Akurasi dari *Random Over Sampling + Naive Bayes*.

Kata Kunci : *Random Over Sampling, Ketidakseimbangan Kelas, Naive Bayes, Akurasi.*

PENGGUNAAN ALGORITMA RANDOM OVER SAMPLING UNTUK MENGATASI MASALAH IMBALANCE DATA PADA KLASIFIKASI GIZI BALITA

Syam Suryo Utomo¹, Triawan Adi Cahyanto², Bakhtiyar Hadi Prakoso³

Program Studi Informatika Fakultas Teknik Universitas Muhammadiyah Jember

Syamsuryo15@gmail.com, triawanac@unmuhjember.ac.id, bahtiyar.hp@gmail.com

ABSTRACT

Tegaldlimo Health Center is one of the health implementing units in the sub-district and has the function of one of which is to make efforts to fulfill the nutrition and health of children under the age of five (toddlers). In Tegaldlimo Health Center up to 2018 there were 1186 children under five, here there are four nutritional status according to the authors which are very far apart, namely "good nutrition", "undernutrition", "over nutrition", "poor nutrition", with details of good nutrition as many as 1101 toddlers, malnutrition as many as 39 toddlers, over nutrition as many as 39 toddlers and malnutrition as many as 7 toddlers. To find solutions to class imbalance problems in the dataset, a model with a data level approach is proposed. The proposed model design includes the design of ROS level data approaches, and Naive Bayes based classification algorithms. The results show that the Naive Bayes model produces Excellent accuracy with an average accuracy of 0.957844. While the accuracy with the ROS + NB model tends to be lower or can be said to be not optimal to increase the accuracy value by having an average of 0.528716 Fail (failing) in classifying the Puskesmas dataset, this is because the amount of data is very significant from many "nutritional data" good "and will affect the Accuracy results of Random Over Sampling + Naive Bayes.

Keywords: Random Over Sampling, Class Imbalance, Naive Bayes, Accuracy.