INTISARI

OPTIMALISASI MODEL *MACHINE LEARNING* PADA SISTEM HIDUNG ELEKTRONIK UNTUK MENDETEKSI *NEONATAL SEPSIS* DENGAN KASUS *DATA IMBALANCE*

IMAN RAHMAN
23/525463/PPA/06604

Penelitian ini bertujuan mengoptimalisasi sistem Electronic Nose (E-Nose) berbasis sensor gas Metal Oxide Semiconductor (MOS) untuk deteksi sepsis neonatal dengan menentukan metode sampling terbaik yang memberikan nilai *balanced accuracy* tertinggi sebagai tujuan optimalisasi yang dilakukan. Menggunakan 1.254 data latih (900 negatif, 354 positif) dan 446 data validasi eksternal (386 negatif, 60 positif), penelitian ini mengevaluasi berbagai metode komputasi untuk menangani bias kelas mayoritas. Hasil komparasi menunjukkan bahwa *Near Miss* 2 (NM2) adalah metode paling optimal. Berbeda dengan metode yang menciptakan data sintetik, NM2 bekerja dengan menyeleksi data mayoritas (negatif) yang paling informatif berdasarkan jarak rata-rata ke data minoritas terjauh, sehingga mempertegas batas keputusan model. Implementasi NM2 pada model SVM menghasilkan *balanced accuracy* 86,01% pada validasi internal, yang mana pada dataset tanpa sampling menghasilkan *balanced accuracy* 50,00%. Pada validasi eksternal, NM2 terbukti meningkatkan performa dibandingkan kondisi tanpa sampling yang mana *balanced accuracy* meningkat dari 50,00% menjadi 91,00%, membuktikan ketahanan model terhadap variabilitas data klinis di lapangan.

Kata Kunci: *Electronic Nose, Machine Learning,* Sepsis Neonatal, Sensor Gas MOS, *Data Imbalance, Near Miss 2.*

ABSTRACT

OPTIMIZATION OF MACHINE LEARNING MODELS IN ELECTRONIC NOSE SYSTEMS FOR NEONATAL SEPSIS DETECTION WITH IMBALANCED DATA

IMAN RAHMAN
23/525463/PPA/06604

This study aims to optimize a Metal Oxide Semiconductor (MOS) gas sensor-based Electronic Nose (E-Nose) system for neonatal sepsis detection by determining the best sampling method that yields the highest balanced accuracy as the primary optimization objective. Using 1,254 training samples (900 negative, 354 positive) and 446 external validation samples (386 negative, 60 positive), this research evaluated various computational methods to handle majority class bias. Comparative results indicate that Near Miss 2 (NM2) is the most optimal method. Unlike methods that create synthetic data, NM2 works by selecting the most informative majority (negative) samples based on the average distance to the farthest minority samples, thereby sharpening the model's decision boundary. The implementation of NM2 on the SVM model achieved a balanced accuracy of 86.01% in internal validation, whereas the unsampled dataset yielded a balanced accuracy of 50.00%. In external validation, NM2 proved to enhance performance compared to the unsampled condition, where balanced accuracy increased from 50.00% to 91.00%, demonstrating the model's robustness against clinical data variability in the field.

Keywords: Electronic Nose, Machine Learning, Neonatal Sepsis, MOS Gas Sensor, Data Imbalance, Near Miss 2