ABSTRAK

Pratama, Wayan Andre (2026), *Komparasi Metode Random Forest dan Light Gradient Boost Machine dalam Diagnosa Penyakit Nyeri Akut dengan Penerapan SMOTE pada Dataset Pakar*. Tesis, Ilmu Komputer, Program Pascasarjana, Universitas Pendidikan Ganesha.

Tesis ini sudah disetujui dan diperiksa oleh Pembimbing 1: Dr. I Made Gede Sunarya, S.Kom., M.Cs. dan Pembimbing II: Dr. Putu Hendra Suputra, S.Kom., M.Cs.

Kata kunci: Dataset Pakar, LightGBM, Nyeri Akut, Random Forest, SMOTE.

Penelitian ini bertujuan untuk membandingkan performa metode Random Forest dan Light Gradient Boosting Machine (LightGBM) dalam mendiagnosis penyakit nyeri akut, serta mengevaluasi pengaruh penerapan Synthetic Minority Over-sampling Technique (SMOTE) dalam menangani ketidakseimbangan kelas data, mengingat penanganan awal nyeri akut yang kurang optimal seringkali memicu masyarakat melakukan swamedikasi yang berisiko menimbulkan kesalahan penggunaan obat. Penelitian kuantitatif komparatif ini dilaksanakan di Universitas Pendidikan Ganesha, menggunakan dataset berjumlah 2.722 data yang dibangkitkan dari sistem pakar berdasarkan kerangka asesmen nyeri PQRST. Dataset mencakup 36 fitur gejala dan 9 kelas diagnosis nyeri akut, divalidasi secara menyeluruh (100%) oleh pakar medis (Penata Anestesi) sebagai bukti validitas instrumen, meskipun reliabilitas antar-penilai belum dapat diuji karena keterbatasan jumlah pakar yang terlibat. Data dibagi menjadi data latih dan data uji dengan proporsi 80:20 menggunakan stratified sampling, kemudian dianalisis menggunakan metrik akurasi, *precision*, *recall*, *F1-score*, dan AUC-ROC pada skenario tanpa SMOTE serta dengan SMOTE menggunakan parameter *k_neighbors* 1 dan 5. Hasil penelitian menunjukkan LightGBM secara konsisten mengungguli Random Forest pada seluruh skenario, dengan akurasi tertinggi pada kondisi SMOTE *k_neighbors*=5 (80,73%), diikuti tanpa SMOTE (80,37%), sementara Random Forest mencapai akurasi tertinggi 75,23%. Evaluasi SMOTE menunjukkan pengaruh tidak konsisten, dengan *k_neighbors*=1 menurunkan performa kedua model akibat mekanisme interpolasi yang kurang optimal pada fitur kategorikal. Analisis lebih lanjut mengungkapkan mayoritas kesalahan klasifikasi terjadi antar jenis nyeri pada tingkat keparahan yang sama. Disimpulkan bahwa LightGBM merupakan metode yang lebih optimal untuk klasifikasi diagnosis nyeri akut, dan penelitian selanjutnya disarankan melibatkan lebih dari satu pakar guna menguji reliabilitas antar-penilai.

ABSTRACT

Pratama, Wayan Andre (2026). Comparison of Random Forest and Light Gradient Boost Machine Methods in Diagnosing Acute Pain with the Application of SMOTE on an Expert Dataset. Thesis, Computer Science, Postgraduate Program, Ganesha University of Education.

This thesis has been reviewed, examined, and approved by Supervisor I, Dr. I Made Gede Sunarya, S.Kom., M.Cs., and Supervisor II, Dr. Putu Hendra Suputra, S.Kom., M.Cs.

Keywords: Acute Pain, Expert Dataset, LightGBM, Random Forest, SMOTE.

This study aims to compare the performance of the Random Forest and Light Gradient Boosting Machine (LightGBM) methods in diagnosing acute pain, as well as to evaluate the effect of applying the Synthetic Minority Over-sampling Technique (SMOTE) in addressing class imbalance, given that inadequate initial management of acute pain often prompts individuals to resort to self-medication, which risks resulting in medication errors. This comparative quantitative study was conducted at Ganesha University of Education, employing a dataset of 2,722 records generated from an expert system based on the PQRST pain assessment framework. The dataset comprises 36 symptom features and 9 classes of acute pain diagnosis, fully validated (100%) by a medical expert (a certified anesthesia nurse) as evidence of instrument validity, although inter-rater reliability could not be tested due to the limited number of experts involved. The data were divided into training and testing sets with an 80:20 ratio using stratified sampling, then analyzed using *accuracy*, *precision*, *recall*, *F1-score*, and AUC-ROC metrics under scenarios without SMOTE and with SMOTE using *k_neighbors* values of 1 and 5. The results show that LightGBM consistently outperformed Random Forest across all scenarios, achieving its highest *accuracy* under the SMOTE *k_neighbors*=5 condition (80.73%), followed by the condition without SMOTE (80.37%), while Random Forest achieved its highest *accuracy* of 75.23%. Evaluation of SMOTE revealed an inconsistent effect, with *k_neighbors*=1 decreasing the performance of both models due to a suboptimal interpolation mechanism on categorical features. Further analysis revealed that the majority of misclassifications occurred between different pain types within the same severity level. It is concluded that LightGBM is a more optimal method for acute pain diagnosis classification, and future research is recommended to involve more than one expert to test inter-rater reliability.