ABSTRAK

Penyakit jantung merupakan penyebab kematian utama di dunia dengan diagnosis dini yang penting namun sering terkendala akurasi interpretasi data rekam medis kompleks. Algoritma klasifikasi tradisional seperti K-NN memiliki kelemahan dalam menangani *noise* dan *outliers* dalam data medis. Penelitian ini bertujuan mengimplementasikan algoritma *Local Mean K-Nearest Neighbor* untuk mengklasifikasikan penyakit jantung berdasarkan data rekam medis dengan akurasi yang lebih baik. *Dataset* terdiri dari 403 observasi dengan 10 variabel meliputi jenis kelamin, umur, tekanan darah, *heart rate*, *respiratory rate*, hasil elektrokardiogram, kondisi nyeri dada, dan klasifikasi diagnosis. Metode *Local Mean* K-NN mengadaptasi konsep K-NN tradisional dengan pendekatan local mean calculation untuk mengatasi *noise* dan *outliers*. Tahapan penelitian mencakup *preprocessing* data, *feature encoding*, *feature scaling*, *hyperparameter tuning*, dan evaluasi menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil menunjukkan algoritma *Local Mean* K-NN dengan nilai K optimal 11 mampu mengklasifikasikan penyakit jantung dengan *accuracy* 71.60%, *precision* 69.21%, *recall* 71.60%, dan *F1-score* 70.27%. Model menunjukkan performa sangat baik dalam mendeteksi Penyakit Jantung Koroner dengan *precision* 91.89% dan *recall* 97.14%. Analisis *feature importance* mengidentifikasi nyeri dada sebagai indikator terpenting (73.79%), diikuti *heart rate* (36.40%) dan *respiratory rate* (25.25%). Penelitian membuktikan efektivitas *Local Mean* K-NN sebagai *clinical decision support tool* dalam klasifikasi penyakit kardiovaskular meskipun terdapat *tantangan class imbalance* pada kelas minoritas.
.

Kata Kunci: *Data Mining*, Jantung, Klasifikasi, *Local Mean K-Nearest Neighbor*, *Machine Learning*

ABSTRACT

Heart disease is a leading cause of death worldwide, with early diagnosis being crucial but often hindered by accuracy challenges in interpreting complex medical record data. Traditional classification algorithms such as K-NN have limitations in handling noise and outliers commonly present in medical data. This study aims to implement the Local Mean K-Nearest Neighbor algorithm to classify heart disease based on medical record data with improved accuracy. The Dataset consists of 403 observations with 10 variables including gender, age, blood pressure, heart rate, respiratory rate, electrocardiogram results, chest pain conditions, and diagnosis classification. The Local Mean K-NN method adapts the traditional K-NN concept with a local mean calculation approach to address noise and outliers. Research stages include data preprocessing, feature encoding, feature scaling, hyperparameter tuning, and evaluation using accuracy, precision, recall, and F1-score metrics. Results show that the Local Mean K-NN algorithm with optimal K value of 11 successfully classified heart disease with accuracy of 71.60%, precision of 69.21%, recall of 71.60%, and F1-score of 70.27%. The model demonstrated excellent performance in detecting Coronary Heart Disease with precision of 91.89% and recall of 97.14%. Feature importance analysis identified chest pain as the most important indicator (73.79%), followed by heart rate (36.40%) and respiratory rate (25.25%). The study proves the effectiveness of Local Mean K-NN as a clinical decision support tool in cardiovascular disease classification despite challenges with class imbalance in minority classes.

Keywords: *Classification, Data Mining, Heart, Local Mean K-Nearest Neighbor, Machine Learning*