

ABSTRAK

Hipertensi merupakan penyakit tidak menular yang menjadi penyebab utama kematian di Indonesia. Di Puskesmas Anggadita, prevalensi hipertensi cukup tinggi dengan kurang lebih dari 1.000 pasien terdiagnosis setiap tahun. Permasalahan yang dihadapi menurut kepala puskesmas yaitu, belum tersedia klasifikasi data pasien hipertensi dan normal yang terstruktur. Karena datanya masih tercampur antara pasien hipertensi, normal, dan diabetes. Penelitian ini bertujuan mengembangkan model klasifikasi hipertensi menggunakan algoritma *Support Vector Machine (SVM)* dan *Random Forest (RF)*. Data sebanyak 3.230 pasien digunakan, melalui tahapan *preprocessing* seperti penanganan *missing values*, penghapusan duplikat, *label encoding*, normalisasi, dan pemilihan fitur dengan analisis korelasi, serta pembagian data menjadi *training* dan *testing*. Model SVM menggunakan kernel *Radial Basis Function (RBF)*, sementara RF menggunakan 100 pohon keputusan. Evaluasi dilakukan menggunakan *confusion matrix* dengan metrik akurasi presisi, *recall*, dan *f1-score*. Hasil menunjukkan model SVM memperoleh akurasi 97%, presisi 0.98 (Normal) dan 0.96 (Hipertensi), serta *f1-score* 0.98 dan 0.96. Sedangkan model RF menunjukkan performa lebih baik dengan akurasi 98%, presisi 1.00 (Normal) dan 0.97 (Hipertensi), serta *f1-score* masing-masing 0.99. Dengan demikian, algoritma *Random Forest* dinilai lebih unggul dan berpotensi diimplementasikan sebagai alat bantu dalam pengelolaan data pasien hipertensi di fasilitas kesehatan.

Kata Kunci: Hipertensi, Klasifikasi, *Random Forest*, *Support Vector Machine*



KARAWANG

ABSTRACT

Hypertension is a non-communicable disease and one of the leading causes of death in Indonesia. At Puskesmas Anggadita, the prevalence of hypertension is relatively high, with approximately over 1,000 patients diagnosed each year. According to the head of the health center, the main issue is the absence of a structured classification system for differentiating between hypertensive and non-hypertensive patients, as the data is still mixed with records of hypertension, normal, and diabetes patients. This study aims to develop a hypertension classification model using the Support Vector Machine (SVM) and Random Forest (RF) algorithms. A dataset of 3,230 patient records was used and underwent preprocessing steps, including handling missing values, removing duplicates, label encoding, normalization, feature selection through correlation analysis, and data splitting into training and testing sets. The SVM model employed the Radial Basis Function (RBF) kernel, while the RF model used 100 decision trees. Evaluation was conducted using a confusion matrix with accuracy, precision, recall, and F1-score as performance metrics. Results showed that the SVM model achieved an accuracy of 97%, with precision scores of 0.98 (Normal) and 0.96 (Hypertension), and F1-scores of 0.98 and 0.96, respectively. Meanwhile, the RF model demonstrated superior performance with 98% accuracy, 1.00 precision (Normal), 0.97 precision (Hypertension), and F1-scores of 0.99 for both classes. Therefore, the Random Forest algorithm is considered more effective and has strong potential for implementation as a decision-support tool in managing hypertension patient data at healthcare facilities.

Keywords: *Classification, Hypertension, Random Forest, Support Vector Machine*



KARAWANG