

ABSTRAK

Huda, Septi Putri Rahma. 2026. Klasifikasi Risiko Penyakit Hipertensi Menggunakan Algoritma *Random Forest*. Tugas Akhir. Program Sarjana. Program Studi Teknik Informatika. Universitas Muhammadiyah Jember.

Pembimbing: Agung Nilogiri, S. T., M. Kom.; Taufiq Timur W, S. Kom., M. Kom.

Hipertensi merupakan salah satu permasalahan kesehatan global yang memerlukan upaya deteksi dini untuk mencegah terjadinya komplikasi yang lebih serius. Penelitian ini bertujuan mengembangkan model klasifikasi risiko hipertensi menggunakan algoritma *random forest* dengan fokus pada penanganan ketidakseimbangan data dan optimasi model. Proses penelitian dilakukan berdasarkan metodologi *Cross-Industry Standard Process for Data Mining*. Kebaruan penelitian ini terletak pada penerapan metode SMOTE-NC (*Synthetic Minority Over-sampling Technique-Nominal Continuous*) untuk menangani *dataset* yang terdiri atas fitur numerik dan kategorikal, serta optimasi *hyperparameter* yang mencakup *n_estimators*, *max_depth*, dan *max_features*. Hasil penelitian menunjukkan bahwa kombinasi *random forest*, SMOTE-NC, dan optimasi *hyperparameter* menghasilkan model terbaik dengan konfigurasi *n_estimators* sebanyak 50, *max_depth* sebesar 6, dan *max_features* bernilai *sqrt*. Model tersebut memperoleh nilai *accuracy* sebesar 89,47%. Pada kelas Berisiko, model menghasilkan nilai *precision* sebesar 76,52%, *recall* sebesar 95,45%, dan *F1-score* sebesar 84,94%. Selain itu, model terbaik diimplementasikan ke dalam aplikasi berbasis web menggunakan *framework Flask* untuk memudahkan pengguna melakukan deteksi dini secara mandiri.

Kata Kunci: Hipertensi, ketidakseimbangan data, Optimasi *Hyperparameter*, *Random Forest*, SMOTE-NC

ABSTRACT

Huda, Septi Putri Rahma. 2026. *Classification of Hypertension Risk Using the Random Forest Algorithm. Undergraduate Thesis. Bachelor's Program in Informatics Engineering. University of Muhammadiyah Jember.*

Advisors: Agung Nilogiri, S. T., M. Kom.; Taufiq Timur W, S. Kom., M. Kom.

Hypertension is one of the global health problems that requires early detection to prevent more serious complications. This study aims to develop a hypertension risk classification model using the random forest algorithm, with a focus on addressing imbalanced data and optimizing model performance. The research process was conducted based on the Cross-Industry Standard Process for Data Mining methodology. The novelty of this study lies in the application of the Synthetic Minority Over-sampling Technique-Nominal Continuous (SMOTE-NC) to handle datasets consisting of both numerical and categorical features, as well as hyperparameter optimization involving `n_estimators`, `max_depth`, and `max_features`. The results show that the combination of random forest, SMOTE-NC, and hyperparameter optimization produced the best-performing model with a configuration of 50 `n_estimators`, a `max_depth` of 6, and `max_features` set to `sqrt`. The model achieved an accuracy of 89.47%. For the At-risk class, the model obtained a precision of 76.52%, a recall of 95.45%, and an F1-score of 84.94%. Furthermore, the best-performing model was implemented in a web-based application using the Flask framework to facilitate independent early detection by users.

Keywords: *Hypertension, Imbalanced Data, Hyperparameter Optimization, Random Forest, SMOTE-NC*