Comparison of Mutual Information and Recursive Feature Elimination in Hypertension Prediction Using K-Nearest Neighbor (Case Study: Puskesmas Mlati II)

Joko Susilo
Program Studi Informatika Medis, Fakultas Sains dan Teknologi
Universitas Teknologi Yogykarta
Jl. Ringroad Utara Jombor Sleman Yogyakarta
E-mail : jokos0946@gmail.com

ABSTRACT

Hypertension is a significant global health issue and a leading cause of various severe complications. This study aims to develop a hypertension prediction model using the K-Nearest Neighbor (K-NN) algorithm and compare the effectiveness of two feature selection methods — Mutual Information (MI) and Recursive Feature Elimination (RFE) — in improving the model's performance. The research methodology comprises several stages: problem identification, collection of clinical hypertension patient data from Puskesmas Mlati II, data analysis, development of the K-NN model using both feature selection methods, model implementation, and system evaluation. The results show that the hypertension diagnosis prediction model using the K-NN algorithm provides strong outcomes. The feature selection method of Mutual Information, with selected features including TG, HDL, LDL, and Creat, yielded the best performance with an accuracy of 92.59%. On the other hand, the Recursive Feature Elimination method, with selected features such as Age, cholesterol, HDL, Urea, and Creatinine, achieved an accuracy of 87.03%. Both methods were found to effectively reduce data dimensionality without significantly compromising accuracy, and the developed model exhibited high precision, recall, and F1-score values. The developed system was successfully implemented, enabling users to make predictions for hypertension and view their prediction history.

Keywords: Hypertension, K-Nearest Neighbor, Feature Selection, Mutual Information, Recursive Feature Elimination.