

ABSTRAK

Diabetes mellitus merupakan penyakit metabolisme kronis yang semakin umum di Indonesia, diperkirakan akan memengaruhi pada tahun 2020, tercatat lebih dari 10,8 juta jiwa. Penyakit ini perlu dikenali sejak dini untuk mencegah komplikasi berat yang dapat meningkatkan angka kesakitan dan kematian. Dengan membandingkan kedua metode, penelitian ini dilakukan untuk mengetahui apakah salah satu pendekatan menunjukkan tingkat akurasi yang lebih baik serta untuk mengembangkan model klasifikasi berdasarkan data pasien. Data penelitian ini disediakan oleh Puskesmas Anggadita yang meliputi data demografi, gaya hidup, dan hasil penilaian kesehatan dari 1001 pasien. Salah satu langkah penelitian adalah pra-pemrosesan data sampai dengan evaluasi. Pemodelan SVM dan RF, dapat mengevaluasi model yang menggunakan metrik akurasi, presisi, *recall*, dan *f1-score*. Berdasarkan hasil pengujian, algoritma *Random Forest* menunjukkan performa terbaik dengan akurasi sebesar 99%, presisi 99%, *recall* 100%, dan *f1-score* 99%, sedangkan SVM mendapatkan akurasi 91%, presisi 93%, *recall* 0,91%, dan *f1-score* 92%. Hal ini menunjukkan seberapa baik *Random Forest* memisahkan pasien dengan dan tanpa diabetes. Penelitian ini diharapkan mampu menjadi salah satu rujukan dalam memperoleh informasi pembuatan sistem pendukung keputusan medis sehingga para tenaga kesehatan dapat lebih cepat dan akurat dalam mendiagnosis penyakit diabetes mellitus.

Kata Kunci: Diagnosis, Klasifikasi, *Machine Learning*, *Random Forest*, *Support Vector Machine*

ABSTRACT

Diabetes mellitus is a chronic metabolic disease that is increasingly common in Indonesia, estimated to affect more than 10.8 million people in 2020. This disease needs to be recognized early to prevent serious complications that can increase morbidity and mortality. By comparing the two methods, this study was conducted to determine whether one approach shows a better level of accuracy and to develop a classification model based on patient data. The research data was provided by the Anggadita Health Center which includes demographic data, lifestyle, and health assessment results from 1001 patients. One of the research steps is data pre-processing to evaluation. SVM and RF modeling can evaluate models using accuracy, precision, recall, and f1-score metrics. Based on the test results, the Random Forest algorithm showed the best performance with an accuracy of 99%, precision of 99%, recall of 100%, and f1-score of 99%, while SVM got an accuracy of 91%, precision of 93%, recall of 91%, and f1-score of 92%. This shows how well Random Forest separates patients with and without diabetes. This study is expected to be one of the references in obtaining information for making medical decision support systems so that health workers can be faster and more accurate in diagnosing diabetes mellitus.

Keywords: *Diagnosis, Classification, Machine Learning, Random Forest, Support Vector Machine*