

KLASIFIKASI TIPE PENYAKIT DIABETES MENGUNAKAN ALGORITMA *SUPPORT VECTOR MACHINE (SVM)*

Nur Afifah Az-Zahra

ABSTRAK

Diabetes melitus adalah penyakit kronis yang menjadi salah satu penyebab utama kematian di Indonesia. Pada tahun 2019, diabetes merupakan penyebab kematian tertinggi ketiga di Indonesia dengan sekitar 57,42 kematian per 100.000 penduduk. Prevalensi kasus diabetes di Indonesia meningkat signifikan, termasuk peningkatan prevalensi pada anak-anak sebanyak 70 kali lipat sejak tahun 2010. Mengingat dampak serius terhadap kesehatan masyarakat, klasifikasi yang akurat antara diabetes tipe 1 dan tipe 2 sangat diperlukan untuk mencegah dampak yang merugikan. Penggunaan algoritma Support Vector Machine (SVM) dalam machine learning menawarkan solusi untuk klasifikasi cepat dan akurat berdasarkan data klinis. Penelitian ini bertujuan mengetahui kinerja algoritma SVM dalam klasifikasi diabetes berdasarkan data pasien dari Puskesmas Pekayon Jaya, Bekasi, Jawa Barat. Hasil menunjukkan bahwa penggunaan SMOTE dan *hyperparameter tuning* C, Gamma, dan Kernel secara signifikan meningkatkan akurasi model. Percobaan terbaik menunjukkan akurasi data latih 99%, validasi 89%, dan uji 91% dengan parameter optimal $C = 1000$, $\text{Gamma} = \text{scale}$, dan kernel RBF. Hasil validasi yang lebih mendekati performa test memberikan indikasi bahwa model tidak overfit pada data validasi dan kemungkinan besar akan memberikan performa yang serupa pada data yang benar-benar baru. Kesimpulannya, metode SMOTE dan *hyperparameter tuning* efektif dalam menangani ketidakseimbangan data dan meningkatkan akurasi model. Pembagian data yang tepat juga berpengaruh positif terhadap akurasi.

Kata Kunci: Support Vector Machine, Diabetes, Klasifikasi, SMOTE, Imbalance Dataset

Classification of Diabetes Types Using Support Vector Machine (SVM) Algorithm

Nur Afiifah Az-Zahra

ABSTRACT

Diabetes mellitus is a chronic disease that has become one of the leading causes of death in Indonesia. In 2019, diabetes was the third highest cause of death in Indonesia, with approximately 57.42 deaths per 100,000 population. The prevalence of diabetes cases in Indonesia has increased significantly, including a seventy-fold increase in prevalence among children since 2010. Given the serious impact on public health, accurate classification between type 1 and type 2 diabetes is essential to prevent adverse effects. The use of Support Vector Machine (SVM) algorithms in machine learning offers a solution for fast and accurate classification based on clinical data. This study aims to determine the performance of the SVM algorithm in classifying diabetes based on patient data from Puskesmas Pekayon Jaya, Bekasi, West Java. The results show that using SMOTE and hyperparameter tuning significantly improves model accuracy. The best experiment showed training data accuracy of 99%, validation accuracy of 89%, and test accuracy of 91%, with optimal parameters $C = 1000$, $\text{Gamma} = \text{scale}$, and kernel RBF. In conclusion, the SMOTE method and hyperparameter tuning are effective in handling data imbalance and improving model accuracy. Proper data partitioning also positively impacts accuracy.

Keywords: *Support Vector Machine, Diabetes, Classification, SMOTE, Imbalance Dataset*