

BAB V

PENUTUP

5.1 KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan mengenai klasifikasi risiko penyakit stroke menggunakan algoritma *Support Vector Machine* (SVM), maka dapat disimpulkan beberapa hal sebagai berikut:

1. Penerapan algoritma SVM berhasil digunakan untuk mengklasifikasikan risiko penyakit stroke berdasarkan data pasien yang diambil dari *Kaggle*. Dataset yang digunakan mencakup 5111 data dengan 11 fitur yang relevan, seperti *id*, *gender*, *age*, *hypertension*, *heart_disease*, *ever_married*, *work_type*, *residence_type*, *avg_glucose_level*, *bmi*, *smoking_status*, *stroke*.
2. Proses seleksi fitur menggunakan metode *Recursive Feature Elimination* (RFE) menunjukkan bahwa pemilihan fitur yang tepat dapat meningkatkan kinerja model. Hasil seleksi terhadap 5 fitur terbaik justru memberikan akurasi lebih tinggi dibandingkan dengan penggunaan semua fitur.
3. Evaluasi model SVM dilakukan sebanyak 10 kali pengujian menghasilkan rata-rata performa yang baik dengan:
 - Rata-rata akurasi: 93%
 - Precision kelas positif (stroke): 91%
 - Recall kelas positif (stroke): 96%
 - F1-score kelas positif (stroke): 93%

4. Berdasarkan hasil *confusion matrix*, model mampu memprediksi stroke dengan tingkat kesalahan yang rendah, yaitu:

- *True Positive* (TP): 932
- *True Negative* (TN): 881
- *False Positive* (FP): 94
- *False Negative* (FN): 38

Hal ini menunjukkan bahwa model cukup andal untuk digunakan dalam deteksi awal risiko stroke.

5. Model dengan 5 fitur terbaik memiliki akurasi 95.02%, sedikit lebih tinggi dari model dengan semua fitur (95.00%), yang menunjukkan bahwa model sederhana namun tepat guna dapat memberikan hasil yang lebih efisien tanpa mengorbankan akurasi.

5.2 SARAN

Berdasarkan hasil penelitian yang telah dilakukan, beberapa saran yang dapat diberikan adalah sebagai berikut:

1. Penggunaan data yang lebih beragam dan besar sangat disarankan untuk meningkatkan generalisasi model. Penelitian ini menggunakan 5111 data, namun untuk penerapan lebih luas di dunia nyata, dibutuhkan lebih banyak data dari berbagai sumber dan latar belakang demografis yang berbeda.
2. Perlu pengujian dengan algoritma lain seperti *Random Forest*, *K-Nearest Neighbor* (KNN), atau *Deep Learning* untuk membandingkan performa dan memilih model yang paling optimal dalam klasifikasi penyakit stroke.
3. Penggunaan teknik balancing data seperti SMOTE atau ADASYN dapat dilakukan untuk menangani ketidakseimbangan data kelas stroke dan non-stroke agar performa model pada kelas minoritas menjadi lebih baik.
4. Integrasi sistem ke dalam aplikasi berbasis web atau mobile akan sangat membantu tenaga medis atau masyarakat umum dalam melakukan prediksi dini risiko stroke secara praktis dan real time.
5. Penelitian selanjutnya disarankan untuk menggabungkan metode seleksi fitur yang berbeda dan mengamati pengaruhnya terhadap performa model guna mendapatkan hasil yang lebih optimal.