

BAB V PENUTUP

5.1 Kesimpulan

Berdasarkan percobaan yang telah dilakukan, maka dapat disimpulkan bahwa algoritma *logistic regression* dapat digunakan untuk memprediksi hasil penyakit paru-paru:

1. Berdasarkan hasil penelitian, variabel paling signifikan dalam memprediksi penyakit paru-paru adalah merokok dengan nilai tertinggi, disusul lagi usia, aktivitas begadang, aktivitas olahraga, penyakit bawaan, dan jenis kelamin berada di posisi bawah dengan nilai terendah. Hal ini dapat ditunjukkan melalui analisis *feature importance* dan hasil evaluasi model.
2. Dari hasil model menggunakan algoritma *logistic regression*, akurasi mencapai 89%, pada kelas 0 *precision* berjumlah 100%, *recall* 78%, dan *f1-score* sebesar 88%. Sedangkan pada kelas 1 *precision* memiliki jumlah sebesar 83%, *recall* 100%, dan *f1-score* sebesar 91%. Dan dapat artikan model menunjukkan performa yang cukup baik dalam mengklasifikasi data. Hal ini dapat ditunjukkan pada analisis *classification report*.
3. Penyakit bawaan memiliki kontribusi yang signifikan terhadap risiko terkena penyakit paru-paru. Pasien yang memiliki riwayat penyakit bawaan cenderung memiliki kemungkinan lebih tinggi menderita penyakit paru-paru, hal ini terlihat dari nilai koefisien positif pada hasil model regresi yang digunakan.

5.2 Saran

Berdasarkan hasil dari penelitian yang diperoleh, tentunya penulis berharap bahwa hasil dari penelitian ini dapat menjadi referensi bagi pembaca yang ingin mengembangkan penelitian dengan topik serupa. Penulis juga menyadari bahwa penelitian ini masih memiliki keterbatasan, sehingga penulis menyarankan beberapa hal sebagai bahan pertimbangan untuk penelitian selanjutnya:

1. Dataset yang digunakan dalam penelitian ini masih berskala terbatas. Agar hasil prediksi lebih akurat dan dapat digeneralisasikan dengan lebih baik,

disarankan untuk menggunakan dataset yang lebih besar dan beragam, serta mencakup data dari sumber medis yang kredibel.

2. Penelitian selanjutnya disarankan untuk menggunakan data usia dalam bentuk rentang umur (range), bukan diklasifikasikan ke dalam kategori seperti “Muda” dan “Tua”. Dengan menggunakan data usia dalam bentuk range memungkinkan model memahami pola dengan baik dan meningkatkan ketepatan hasil prediksi.
3. Penelitian ini hanya memanfaatkan algoritma *Logistic regression*. Akan lebih baik dilakukan perbandingan dengan metode lain seperti *Random Forest*, *SVM*, atau *Neural Network* guna mengetahui metode mana yang memberikan hasil paling optimal dalam memprediksi penyakit paru-paru.
4. Untuk meningkatkan akurasi prediksi, fitur dalam dataset sebaiknya diperluas dengan menyertakan informasi klinis seperti hasil rontgen dada, kadar oksigen, maupun data dari pemeriksaan laboratorium.
5. Aplikasi prediksi pada penelitian ini dapat dikembangkan menjadi *platform* berbasis web atau perangkat *mobile* sehingga dapat diakses dengan mudah oleh masyarakat umum maupun dimanfaatkan dalam institusi layanan kesehatan.