

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1. Kesimpulan

Berdasarkan hasil dari keseluruhan proses yang telah dilakukan sebelumnya maka dapat disimpulkan beberapa hal, diantaranya adalah:

1. Ketimpangan data yang diperoleh pada penelitian ini membuat penerapan skenario penanganan kelas tidak seimbang dengan ADASYN dilakukan. Sebanyak 12 skenario dilakukan dengan menggunakan ADASYN yang terbagi menjadi beberapa skenario pembagian kelas dan skenario penyetelan hyperparameter. Hasil pengujian memperlihatkan bahwa skor tertinggi diperoleh 3 model dengan skenario tanpa penggunaan ADASYN, yaitu model 4, 16, dan 10. Namun, setelah melihat pada fase evaluasi kurva pembelajaran didapati bahwa ketiga model tersebut terindikasi *overfit*. Model 2 dan 3 dengan penggunaan ADASYN mampu memberikan performa yang lebih stabil pada terutama pada skenario proporsi 70:30.
2. Pengujian *cpu usage* dan *processing time* memberikan hasil yang saling berkebalikan. Model dengan algoritma *random forest* unggul pada penggunaan cpu namun kalah dalam waktu pemrosesan, sementara *lightgbm* unggul dalam waktu pemrosesan namun kalah dalam penggunaan cpu. Model 2 diputuskan untuk digunakan sebagai model *deployment* karena menghasilkan performa yang lebih stabil dan memiliki penggunaan cpu yang lebih rendah, yaitu 1%. Sedangkan model 3 menggunakan cpu sebanyak 3,9%.

3. Hasil evaluasi menggunakan *confusion matrix* menampilkan bahwa 5 model teratas di dominasi oleh basis algoritma *lightgbm* dengan model 4, 10 dan 16. Sedangkan, algoritma *random forest* menempati sisanya dengan urutan performa ke 4 dan ke 5. Performa *lightgbm* secara *default* mampu mengungguli kemampuan *random forest* dalam melakukan klasifikasi yang dibuktikan dengan ketidakmampuan *random forest* dalam memprediksi data positif. Hal ini dibuktikan dengan nilai model berbasis *random forest* secara *default* mencatatkan angka 0% pada indikator penilaian *precision*, *recall*, dan *f1* di seluruh skenario pembagian data. Sehingga, memerlukan bantuan ADASYN dan penyetelan *hyperparameter*.
4. Model 2 digunakan sebagai model *deployment*, yaitu model *random forest* default dengan ADASYN dan ukuran test size 30%. Model dimuat dalam website menggunakan pickle yang sebelumnya telah dibuat. Sehingga, sistem dapat melakukan prediksi dari data yang dimasukkan. Hasil pengujian data validasi pada website menghasilkan nilai *accuracy* sebesar 85% dimana terdapat 3 data *true positive*, 4 data *false positive*, 0 data *false negative*, dan 20 data *true negative*.

## 5.2. Saran

1. Penambahan data terutama pada label positif untuk meningkatkan keseimbangan dataset yang digunakan. Hal ini dapat dilakukan dengan memperluas lokasi pengambilan data ataupun mengambil data dengan rentang waktu yang lebih panjang.
2. Penambahan variabel yang digunakan mungkin dapat meningkatkan keragaman model dan memperbaiki hasil pemodelan, terutama variabel

yang berhubungan dengan pemeriksaan laboratorium ataupun pola hidup pasien.

3. Pengembangan model dengan hasil evaluasi yang lebih baik lagi, terutama pada segi *precision* dan *recall* karena dipergunakan pada dataset medis. Hal ini karena dataset medis cenderung lebih memerlukan kemampuan prediksi positif dimana penyakit yang berisiko mungkin dapat mengancam nyawa pasien.