

BAB 5

PENUTUP

5.1 Kesimpulan

Sebelum mengaplikasikan metode optimasi seperti Particle Swarm Optimization (PSO), memastikan bahwa data bersih dan lengkap adalah fondasi yang kuat untuk mendapatkan hasil yang optimal. PSO, sebagai algoritma optimasi, dapat bekerja lebih efektif dan efisien ketika diterapkan pada data yang telah melewati tahap penanganan masalah seperti *noisy data* dan *missing values*. Dalam pengembangan model prediktif, penanganan *noisy data* dan *missing values* memiliki peran krusial dalam memastikan keandalan dan ketepatan model. *Noisy data* dapat memengaruhi interpretasi model dan konsistensi prediksi, sehingga dalam mengidentifikasi dan membersihkan *outlier* serta nilai yang tidak wajar menjadi langkah esensial. Selain itu juga, keberadaan *missing values* bisa mengakibatkan bias dan ketidakakuratan dalam hasil prediksi, maka proses imputasi yang tepat dapat membantu mengisi kekosongan tersebut dan memastikan setiap *instance data* dapat memberikan kontribusi yang maksimal pada pembentukan model.

Setelah *preprocessing*, ketiga algoritma klasifikasi: k-Nearest Neighbor, Random Forest, dan Deep Learning, menunjukkan kinerja yang baik. Algoritma k-Nearest Neighbor memiliki akurasi sekitar 96.84%, Random Forest sekitar 95.79%, dan Deep Learning sekitar 93.68%. Kemudian setelah diterapkan Particle Swarm Optimization untuk optimalisasi, terjadi peningkatan yang signifikan pada semua algoritma ini. k-Nearest Neighbor dan Random Forest mencapai performa yang sempurna dengan akurasi, presisi, *recall*, dan AUC mencapai 100%. Sedangkan Deep Learning menunjukkan peningkatan pada akurasi menjadi 98.95%, presisi 100%, dan *recall* 96.30%. Peningkatan performa ini setelah dioptimasi oleh PSO menunjukkan bahwa algoritma-algoritma tersebut dapat lebih baik memahami pola-pola kompleks dalam data, memberikan prediksi yang lebih akurat, dan memiliki kehandalan yang lebih tinggi dalam mengklasifikasikan *instance*. Dengan demikian, kombinasi penanganan data yang baik dan juga optimasi menggunakan

PSO menjadi kunci keberhasilan dalam mencapai kinerja model yang optimal. Keseluruhan proses ini menunjukkan pentingnya tahapan-tahapan ini dalam memaksimalkan potensi dari algoritma klasifikasi.

Tetapi meski algoritma k-Nearest Neighbor dan Random Forest memiliki akurasi yang lebih tinggi dibandingkan dengan Deep Learning, akan tetapi secara statistik *inference* ketiganya adalah sama. Hasil uji t-Test yang disajikan dalam Tabel 4.3 menunjukkan bahwa ketiga algoritma ini tidak memiliki perbedaan yang signifikan, sehingga dalam penggunaannya mereka akan memberikan *output* yang sama baiknya.

5.2 Saran

Terakhir, untuk penelitian selanjutnya, beberapa hal yang dapat penulis sarankan diantaranya:

- Untuk mendalami strategi pengisian *missing values* dengan mengintegrasikan Algoritma Imputasi Khusus seperti kNN atau menerapkan Teknik Interpolasi, selain pendekatan umum menggunakan nilai *mean*. Variasi dalam metode pengisian *missing values* ini dapat memberikan wawasan lebih lanjut tentang cara menangani ketidaklengkapan data atau pun struktur data yang mungkin lebih kompleks dengan lebih efektif lagi.
- Selain itu, eksplorasi terhadap berbagai teknik algoritma klasifikasi lainnya dapat membuka peluang untuk menemukan model yang dapat memberikan hasil lebih baik atau memiliki keunggulan pada jenis *dataset* tertentu.
- Penggunaan *dataset* yang berbeda juga penting untuk menguji generalitas dari model yang dikembangkan.
- Penelitian lain selanjutnya juga bisa mempertimbangkan “variabilitas” dalam penggunaan parameter PSO untuk mengeksplorasi pengaturan yang optimal dan memahami sejauh mana performa model dapat ditingkatkan.