

BAB V PENUTUP

5.1 Kesimpulan

Dari penelitian yang telah dilakukan mengenai perbandingan algoritma *Support Vector Machine* (SVM), *Random Forest*, *K-Nearest Neighbors* (KNN), dan *Gradient Boosting* dengan menggunakan *hyperparameter tuning* dalam klasifikasi kanker payudara, dapat disimpulkan bahwa:

1. Algoritma *Random Forest* memberikan performa terbaik dalam klasifikasi kanker payudara dibandingkan algoritma lainnya (SVM, KNN, dan *Gradient Boosting*) setelah dilakukan *hyperparameter tuning*, dengan nilai akurasi tertinggi dan keseimbangan metrik evaluasi seperti *precision*, *recall*, dan *F1-score*. *Hyperparameter tuning* menggunakan *Random Search* terbukti dapat meningkatkan performa semua algoritma yang diuji. Model *Random Forest* mengalami peningkatan akurasi menjadi 95%, *Gradient Boosting* menjadi 94%, KNN menjadi 91%, dan SVM meningkat menjadi 86%. Dengan demikian, *tuning* berkontribusi pada peningkatan akurasi dan efektivitas model dalam klasifikasi kanker payudara.
2. Penerapan *hyperparameter tuning* secara signifikan meningkatkan performa model, terutama pada algoritma *Random Forest* dan *Gradient Boosting*. Berdasarkan hasil evaluasi, algoritma *Random Forest* menjadi yang paling optimal setelah dilakukan *tuning*, dengan akurasi tertinggi dan performa yang stabil. Dari dua metode *tuning* yang digunakan, yaitu *GridSearchCV* dan *RandomizedSearchCV*, *GridSearchCV* memberikan hasil yang lebih konsisten dan optimal dalam pencarian kombinasi parameter terbaik.
3. Integrasi *preprocessing data* (*cleaning*, *encoding*, normalisasi, dan *SMOTE*) dengan seleksi fitur yang relevan menggunakan *SelectKBest* berkontribusi besar terhadap peningkatan kinerja model. Tahapan ini memastikan data

yang digunakan dalam pelatihan model lebih bersih, seimbang, dan informatif.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan mengenai perbandingan machine learning terhadap analisis kanker payudara, masih terdapat kekurangan. Peneliti memberikan beberapa saran untuk pengembangan penelitian selanjutnya, yaitu :

1. Penelitian selanjutnya disarankan untuk menambahkan lebih banyak algoritma dan teknik ensemble seperti XGBoost atau LightGBM guna memperluas cakupan perbandingan dan menguji stabilitas model pada dataset yang lebih kompleks. Menerapkan metode balancing data yang lebih baik untuk menangani ketidakseimbangan kelas dalam dataset sehingga model dapat bekerja lebih optimal.
2. Penggunaan dataset dari berbagai sumber dengan karakteristik berbeda juga dapat memperkuat validitas model serta mengevaluasi performa algoritma dalam konteks dunia nyata yang lebih bervariasi, seperti data dari citra medis atau rekam medis elektronik.