BAB V

PENUTUP

5.1 KESIMPULAN

Kesimpulan yang dapat diambil dari hasil penelitian yang telah dibahas pada sebelumnya adalah sebagai berikut:

1. Model *Random Forest* menunjukkan akurasi tinggi dengan MSE sebesar 0.0192, yang menunjukkan kesalahan prediksi yang sangat kecil, serta R^2 sebesar 0.9986, yang berarti model dapat menjelaskan 99.86% variasi dalam data. Keunggulan *Random Forest* terletak pada kemampuannya menangani dataset besar dan kompleks dengan stabilitas tinggi, mengurangi overfitting, serta efektif dalam memilih variabel yang relevan, menjadikannya pilihan yang sangat baik untuk aplikasi prediksi yang memerlukan akurasi tinggi.
2. Model *Gradient Boosting* menunjukkan akurasi yang sangat baik dengan MSE sebesar 0.0183, yang lebih kecil dibandingkan dengan *Random Forest*, serta R^2 sebesar 0.9986, yang menandakan bahwa model ini mampu menjelaskan 99.86% variasi dalam data dengan sangat baik. Meskipun menghasilkan kesalahan prediksi yang lebih kecil, *Gradient Boosting* lebih sensitif terhadap penyetelan hyperparameter dan memerlukan penyesuaian yang lebih hati-hati untuk mencapai performa optimal. Algoritma ini sangat efektif dalam meningkatkan akurasi model, terutama pada dataset yang lebih kompleks, tetapi perlu perhatian lebih dalam proses tuning-nya.
3. Berdasarkan hasil analisis, *Gradient Boosting* lebih unggul dibandingkan *Random Forest* karena memiliki MSE yang lebih kecil (0.0183

dibandingkan 0.0191), menunjukkan kemampuan yang lebih baik dalam meminimalkan kesalahan prediksi, dengan R^2 yang sama tinggi di angka 0.9986, yang berarti kedua algoritma sama-sama mampu menjelaskan 99.86% variabilitas data. Hal ini menjadikan Gradient Boosting pilihan yang lebih baik untuk tugas dengan fokus utama pada akurasi prediksi.

4. Dalam komparasi antara algoritma *Random Forest* dan *Gradient Boosting*, terdapat beberapa perbedaan yang mencolok. Dari segi akurasi, *Gradient Boosting* memiliki MSE yang lebih rendah (0.0183) dibandingkan dengan *Random Forest* yang memiliki MSE sebesar 0.0192, menunjukkan bahwa *Gradient Boosting* lebih akurat dalam memprediksi nilai. Namun, *Random Forest* lebih stabil dan efisien dalam menangani dataset besar dengan variasi yang tinggi tanpa memerlukan banyak penyetelan hyperparameter, menjadikannya lebih cocok untuk aplikasi dengan kebutuhan pelatihan yang cepat dan dataset yang lebih besar. Di sisi lain, meskipun Gradient Boosting lebih sensitif terhadap penyetelan hyperparameter dan membutuhkan waktu pelatihan yang lebih lama, ia unggul dalam mengidentifikasi pola-pola kompleks dalam data, membuatnya lebih sesuai untuk aplikasi yang memerlukan analisis lebih mendalam pada data yang memiliki hubungan yang lebih rumit.

5.2 SARAN

Penulis menyadari bahwa penelitian ini masih jauh dari kata sempurna. Oleh karena itu, penulis memiliki saran yang dapat digunakan untuk mengembangkan

penelitian mendatang agar mendapatkan hasil yang lebih baik. Adapun beberapa saran sebagai berikut:

1. Penelitian lebih lanjut dapat mencoba membandingkan hasil prediksi dengan algoritma machine learning lain, seperti *Support Vector Machines* (SVM), *Neural Networks*, atau *XGBoost*. Menggunakan teknik ensemble atau stacking dari berbagai model ini dapat membantu memperoleh model yang lebih kuat dan meningkatkan akurasi prediksi energi molekul. Dengan demikian, dapat ditemukan pendekatan yang lebih optimal untuk dataset yang lebih besar dan lebih kompleks.
2. Penelitian selanjutnya dapat mengimplementasikan teknik cross-validation untuk mendapatkan estimasi yang lebih akurat mengenai kinerja model dan mengurangi bias yang dihasilkan oleh pembagian dataset. Selain itu, eksperimen dengan dataset yang lebih besar dan lebih beragam akan memberikan gambaran lebih jelas mengenai kemampuan generalisasi kedua algoritma dalam memprediksi energi molekul, serta menguji stabilitas model pada skala yang lebih besar.