

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan, dapat diambil kesimpulan sebagai berikut.

1. Telah dihasilkan 3 model *machine learning* untuk memprediksi energi atomisasi molekul dengan fitur Matriks Coulomb menggunakan algoritma *K-Nearest Neighbors*, *Random Forest* dan *XGBoost* berdasarkan data dari *website* PubChem dengan 16242 molekul dari unsur penyusun C ,H, N ,O, P dan S dalam bentuk visualisasi.
2. Telah dihasilkan metrik koefisien determinasi (R^2) sebesar 0,6989 untuk model *K-Nearest Neighbors*, 0,9999 untuk model *Random Forest* dan 0,9996 untuk model *XGBoost*. Dengan model *XGBoost* menghasilkan tingkat keakuratan tertinggi dengan mengabaikan perbedaan yang tidak signifikan dengan algoritma *Random Forest*.
3. Telah dihasilkan metrik *Root Mean Square Error* (RMSE) sebesar 2,0226 untuk model *K-Nearest Neighbors*, 0,0089 untuk model *Random Forest*, 0,0647 untuk model *XGBoost*. Dengan model *Random Forest* menghasilkan nilai tingkat kesalahan rata – rata kuadratnya paling rendah.
4. Telah dihasilkan waktu untuk menghasilkan distribusi energi atomisasi molekul selama 5,7 detik untuk model *K-Nearest Neighbors*, 343 detik untuk model *Random Forest*, 99 detik untuk model *XGBoost*. Dengan model *K-Nearest Neighbors* paling efektif jika ditinjau dari lama waktu yang dibutuhkan masing – masing model.

5. Untuk biaya komputasi yang digunakan dari ketiga algoritma, sama – sama membutuhkan sumber daya yang tinggi yakni sebesar 90 % dari penggunaan CPU dan RAM untuk menjalankan masing – masing model.

Kesimpulan akhir yang dapat ditarik ialah algoritma *XGBoost* adalah algoritma yang terbaik untuk memprediksi energi atomisasi molekul dibandingkan dengan algoritma *Random Forest* dan model *K-Nearest Neighbors*.

5.2 Saran

Saran yang dapat diberikan untuk penelitian selanjutnya sebagai berikut.

1. Meningkatkan jumlah data yang akan diolah demi meningkatkan variansi dan akurasi model
2. Menggunakan algoritma – algoritma bertipe *ensemble* lainnya.
3. Menggunakan dataset dengan unsur penyusunnya berbeda dengan penelitian ini

