

Abstrak

Penelitian ini bertujuan untuk menyelesaikan nilai dari energi atomisasi molekul tanpa menggunakan metode analitik maupun numerik, dengan berkembang pesatnya teknologi memungkinkan mengetahui nilai energi atomisasi molekul secara cepat dan efisien dengan algoritma *Machine Learning* dan *Matrix Coulomb* sebagai deskriptornya dengan membandingkan dataset asli dari *database PubChem* sebanyak 16242 molekul dari unsur penyusun C, H, N, O, P dan S dengan hasil prediksi masing – masing algoritma. Penelitian ini dilakukan di Laboratorium Komputer Jurusan Fisika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Medan yang dilaksanakan pada bulan Juni hingga Juli 2022. Algoritma yang digunakan ialah *K-Nearest Neighbors*, *Random Forest*, dan *XGBoost* dengan bantuan *library* dari Scikit-Learn. Variabel yang diuji ialah nilai RMSE (*Root Mean Square Error*), R^2 (Koefisien Determinasi), lama waktu, penggunaan *resource* dan *hyperparameter* masing – masing algoritma. Hasil penelitian menunjukkan bahwa algoritma *K-Nearest Neighbors* buruk dalam memprediksi dengan nilai RMSE 2,0226, nilai R^2 0,6989 dan lama waktu yang dibutuhkan selama 5,7 detik. Untuk algoritma *Random Forest* menghasilkan nilai RMSE 0,0089, nilai R^2 0,9999 dan lama waktu yang dibutuhkan selama 343 detik, sedangkan algoritma *XGBoost* menghasilkan nilai RMSE 0,0647, R^2 0,9996 dan lama waktu yang dibutuhkan selama 99 detik. Penggunaan *resource* dari ketiga algoritma ini menyentuh 90 % dari CPU dan RAM nya.

Kata Kunci : *Machine Learning, Matrix Colomb, RMSE, R^2 , K-Nearest Neighbors, Random Forest, XGBoost, hyperparameter, Scikit-Learn*



Abstract

This study aims to solve the value of the atomization energy of a molecule without using analytical or numerical methods, with the rapid development of technology its possible to find out the value of the atomization energy of a molecule quickly and efficiently with Machine Learning algorithms and the Coulomb Matrix as descriptors by comparing the original dataset from the *PubChem database* of 16242 molecules of the constituent elements C, H, N, O, P and S with the prediction results of each algorithm. This research was conducted at the Computer Laboratory of the Physics Department, Faculty of Mathematics and Natural Sciences, Medan State University which was carried out from June to July 2022. The algorithm used was *K-Nearest Neighbors*, *Random Forest*, and *XGBoost* with the help of a library from Scikit-Learn. The variables tested are the value of RMSE (*Root Mean Square Error*), R2 (*Coefficient of Determination*), length of time, resource usage and hyperparameters of each algorithm. The results showed that the K-Nearest Neighbors algorithm was bad at predicting with an RMSE value of 2.0226, an R2 value of 0.6989 and the length of time required for 5.7 seconds. The Random Forest algorithm produces an RMSE value of 0.0089, an R2 value of 0.9999 and the time required is 343 seconds, while the XGBoost algorithm produces an RMSE value of 0.0647, R2 0.9996 and the length of time it takes is 99 seconds. The resource usage of these three algorithms touches 90% of the CPU and RAM.

Keywords : *Machine Learning, Matrix Colomb, RMSE, R², K-Nearest Neighbors, Random Forest, XGBoost, hyperparameter, Scikit-Learn*

