

## INTISARI

### PENINGKATAN AKURASI PREDIKSI ENERGI ATOMISASI MOLEKUL MENGGUNAKAN PENDEKATAN ALGORITMA GENERALISASI BERTUMPUK

Oleh

Yusyri Dwi Kurniawan

17/414602/PA/18102

Pembelajaran mesin (*machine Learning*) dapat menjadi jalan pintas untuk menyelesaikan permasalahan fisika yang rumit dengan tidak membutuhkan banyak daya komputasi. Pada penelitian ini dilakukan penerapan *Extreme Gradien Boosted (XGB) Regression Tree*, *Light Gradien Boosting Machine (LGBM) Regression Tree* dan *Random Forest Regression Tree* dalam memprediksi energi atomisasi molekul. Fitur yang digunakan untuk melatih model dikonstruksi dari definisi Matriks Coulomb. *Dataset* yang digunakan diperoleh dari database PubChem yang memuat 16,242 molekul dengan unsur penyusun C, H, N, O, P, dan S. *Dataset* dibagi menjadi dua yaitu data *training* sebanyak 70% dan data *testing* sebanyak 30%. Model dievaluasi dengan metrik MAE. Hasil terbaik yang diperoleh adalah model XGB, MAE = 16.6038 kcal/mol dengan durasi pelatihan 90.89 detik. Diterapkan pula metode Generalisasi Bertumpuk (*Stacked Generalization*) yang dapat menggabungkan beberapa algoritma tunggal sekaligus yang berbeda. Hasil yang diperoleh *Stacked Generalization*, MAE = 15.6918 kcal/mol, mengalami peningkatan akurasi prediksi sebesar 5.49% dari hasil terbaik model pembangunnya (XGB).

Kata-kata kunci : matriks Coulomb, *machine learning*, *stacked generalization*, energi atomisasi

## ABSTRACT

### IMPROVING THE ACCURACY OF MOLECULAR ATOMIZATION ENERGY PREDICTION USING A STACKED GENERALIZATION ALGORITHM APPROACH

By

Yusyri Dwi Kurniawan

17/414602/PA/18102

Machine Learning can be a bypass to solving complex physics problems without requiring a lot of computational power. In this research, machine learning is implemented using Extreme Gradient Boosted (XGB) Regression Tree, Light Gradient Boosting Machine (LGBM) Regression Tree and Random Forest Regression Tree to predict molecular atomization energy. The features used to train the model are constructed from the definition of the Coulomb Matrix. The dataset used was obtained from the PubChem database which contains 16,242 molecules with the constituent elements C, H, N, O, P, and S. Dataset is divided into two, namely data training as much as 70% and 30% testing data. The model is evaluated with the MAE metric. The best results obtained is the XGB model, MAE = 16.6038 kcal/mol with a training duration of 90.89 seconds. The Stacked Generalization is also applied which can combine several different single algorithms at once. In this study, the results obtained for Stacked Generalization is MAE = 15.6918 kcal/mol, increased by 5.49% in prediction accuracy from the best results of the model that built it (XGB).

Keywords : Coulomb Matrix, machine learning, stacked generalization, atomization energy