



INTISARI

Prediksi Energi Celah Pita Pada Perovskit Halida Ganda Menggunakan *Machine Learning*

Oleh

Geervan Thanra
21/477331/PA/20650

Penelitian ini mengembangkan model *machine learning* untuk memprediksi energi celah pita pada material perovskit halida ganda berbasis data hasil perhitungan *Density Functional Theory* (DFT). Pendekatan ini menjadi alternatif yang lebih efisien dibandingkan metode komputasi tradisional dalam estimasi sifat elektronik material semikonduktor. Empat algoritma digunakan dalam penelitian ini, yaitu *Gradient Boosted Regression Trees* (GBRT), *Support Vector Regression* (SVR), *Random Forest Regressor* (RF), dan *Multi-Layer Perceptrons* (MLP), dengan dataset yang terdiri dari 540 senyawa hipotetis dan 34 fitur. Proses seleksi fitur dilakukan menggunakan tiga metode: *LASSO*, *SelectKBest*, dan *Permutation Importance*, untuk mengidentifikasi 17 fitur paling relevan terhadap prediksi energi celah pita. Fitur-fitur hasil seleksi kemudian dikompilasi ke dalam subset bernama *selected*, yang digunakan dalam pelatihan model. Hasil evaluasi menunjukkan bahwa model MLP dengan subset *selected* mampu menghasilkan performa tinggi dengan nilai evaluasi $RMSE = 0.255$, $R^2 = 0.946$, dan $RMAE = 0.178$, meskipun menggunakan jumlah fitur yang lebih sedikit. Uji validasi terhadap data eksternal menunjukkan hasil prediksi model mendekati nilai hasil DFT. Selain itu, dibangun antarmuka aplikasi berbasis model terbaik untuk memudahkan implementasi prediksi. Temuan ini menegaskan bahwa *machine learning* dapat menjadi solusi yang cepat dan akurat dalam memperkirakan energi celah pita material perovskit, serta berpotensi mendukung percepatan eksplorasi material baru di bidang energi terbarukan.

Kata kunci : *machine learning*, perovskit halida ganda, seleksi fitur

ABSTRACT

Energy Bandgap Predictions in Double Halide Perovskites Using Machine Learning

By

Geervan Thanra
21/477331/PA/20650

This research develops machine learning models to predict the bandgap energy of double halide perovskite materials using data derived from Density Functional Theory (DFT) calculations. This approach offers an efficient alternative to traditional computational methods for estimating the electronic properties of semiconductor materials. Four machine learning algorithms were utilized: Gradient Boosted Regression Trees (GBRT), Support Vector Regression (SVR), Random Forest Regressor (RF), and Multi-Layer Perceptrons (MLP), with a dataset consisting of 540 hypothetical compounds and 34 features. Feature selection was conducted using LASSO, SelectKBest, and Permutation Importance to identify 17 key features with the highest influence on bandgap prediction. These features were compiled into a new subset named *selected*, which was then used for model training. The evaluation results show that the MLP model using the *selected* subset achieved high performance with $RMSE = 0.255$, $R^2 = 0.946$, and $RMAE = 0.178$, despite using fewer features. Validation testing using external data confirmed that the model predictions closely matched the DFT values. Furthermore, a user interface application was developed based on the best-performing model to support practical implementation. These findings demonstrate that machine learning provides a fast and accurate solution for predicting bandgap energy and has significant potential to accelerate the discovery of new materials in the field of renewable energy.

Keywords : machine learning, double halide perovskite, feature selection