

DOI: doi.org/10.21009/03.1301.FA10

PERBANDINGAN PERFORMA ALGORITMA *RANDOM FOREST REGRESSOR* DAN *XGBOOST REGRESSOR* DALAM PREDIKSI BAND GAP MATERIAL SEMIKONDUKTOR SILIKON

Muhammad Rizky Anugrah^{a)}, Teguh Budi Prayitno^{b)}, Haris Suhendar^{c)}*Program Studi Fisika, Universitas Negeri Jakarta, Jakarta Timur 13220, Indonesia*Email: ^{a)}muhammadrizkyanugrah_179@gmail.com, ^{b)}haris_suhendar@unj.ac.id, ^{c)}teguh-budi@unj.ac.id

Abstrak

Material semikonduktor memainkan peran krusial dalam berbagai aplikasi teknologi modern, termasuk bidang elektronik, fotovoltaik, dan optoelektronik. Salah satu sifat utama dari material semikonduktor adalah band gap, yang merupakan energi yang diperlukan suatu elektron untuk dapat tereksitasi. Band gap adalah parameter kunci yang memengaruhi perilaku elektronik dan optik material semikonduktor. Dalam penelitian ini, machine learning digunakan untuk memprediksi nilai band gap material semikonduktor yaitu silikon. Silikon merupakan salah satu material semikonduktor yang sangat penting dan banyak digunakan dalam berbagai aplikasi teknologi modern. Dataset yang digunakan diambil dari Materials Project (MP), yang menyediakan beragam data mengenai material yang telah diuji, serta ciri-ciri dan karakteristiknya. MP menyediakan informasi tentang berbagai jenis material, termasuk data sifat-sifat material, struktur kristal, kestabilan termal, dan lainnya yang dapat digunakan untuk membangun model machine learning. Penelitian ini bertujuan untuk mengembangkan dan membandingkan performa dua algoritma machine learning, yaitu Random Forest Regressor dan XGBoost Regressor, dalam memprediksi band gap material silikon dengan akurasi tinggi berdasarkan ciri-ciri dan karakteristik material yang tersedia dalam dataset Materials Project. Evaluasi dilakukan juga terhadap kedua model machine learning tersebut dalam memprediksi band gap material silikon.

Kata-kata kunci: Semikonduktor, Silikon, Material, Machine Learning.

Abstract

Semiconductor materials play a crucial role in various modern technological applications, including electronics, photovoltaics, and optoelectronics. One of the primary properties of semiconductor materials is the band gap, which is the energy required to excite an electron. The band gap is a key parameter influencing the electronic and optical behavior of semiconductor materials. In this study, machine learning is used to predict the band gap values of the semiconductor material silicon. Silicon is one of the most important and widely used semiconductor materials in various modern technological applications. The dataset used is taken from the Materials Project (MP), which provides a wide range of data on tested materials, including their features and characteristics. MP offers information on various types of materials, such as material properties, crystal structures, thermal stability, and others that can be used to build machine learning models. This research aims to develop and compare the performance of two machine learning algorithms, namely Random Forest Regressor and XGBoost Regressor, in predicting the band gap of silicon materials with high accuracy based on the features and characteristics available in the Materials Project dataset. This study also involves a comprehensive evaluation of both machine learning models in predicting the band gap of silicon materials.

Keywords: Semiconductor, Silicon, Material, Machine Learning.

PENDAHULUAN

Dokumen Material semikonduktor adalah tulang punggung teknologi modern, mendukung berbagai aplikasi seperti perangkat elektronik dan energi terbarukan [1]. Semikonduktor memiliki konduktivitas antara isolator dan konduktor, dengan parameter utama band gap, yaitu perbedaan energi antara valence band dan conduction band [2]. Band gap menentukan energi yang dibutuhkan untuk elektron melompat dari valence band ke conduction band [3]. Semikonduktor memiliki band gap antara 0-4 eV, di mana band gap sempit (0,5-1 eV) memudahkan lompatan elektron sehingga menghasilkan arus listrik lebih efisien [4]. Silikon adalah semikonduktor paling banyak digunakan karena melimpah di alam dan memiliki band gap 1,1 eV, ideal untuk sel surya berefisiensi tinggi [5]. Band gap material sangat penting dalam aplikasi seperti fotokatalitik dan fotovoltaik, di mana bahan dengan band gap rendah dapat menyerap lebih banyak foton [6]. Sebagai elemen kedua paling melimpah di kerak bumi, silikon juga dapat membentuk banyak senyawa dengan band gap berbeda, sehingga dapat memperluas aplikasi teknologinya [7].

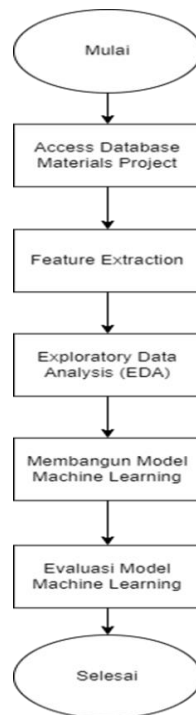
Prediksi yang akurat terhadap band gap semikonduktor telah menjadi tantangan dalam kimia kuantum karena band gap adalah parameter penting yang menentukan kinerja aplikasi elektronik dan optoelektronik [8]. Meskipun metode perhitungan teori fungsional densitas standar (DFT) dan aproksimasi gradien umum (GGA) telah digunakan, hasilnya sering kali terlalu rendah [9]. Teknik yang lebih canggih seperti fungsional densitas hibrida dan aproksimasi GW meningkatkan akurasi prediksi tetapi memerlukan biaya komputasi tinggi [10]. Dalam dekade terakhir, machine learning (ML) telah muncul sebagai metode yang menjanjikan untuk memprediksi band gap dengan akurat melalui model berbasis data. Machine learning adalah bidang dalam Artificial Intelligence (AI) yang menggunakan prinsip-prinsip ilmu komputer dan statistik untuk mengembangkan model yang mampu menangkap pola kompleks dalam data dengan tujuan mengajari mesin untuk membuat prediksi yang akurat. Kemajuan dalam keterjangkauan dan efisiensi komputer, bersama dengan teknik komputasi yang lebih baik, telah memungkinkan penerapan ML yang luas di berbagai bidang, termasuk prediksi band gap. Algoritma ML dilatih menggunakan data training dan diuji dengan data testing untuk meningkatkan performa prediksi. Penggunaan algoritma machine learning seperti Random Forest Regressor dan XGBoost Regressor memungkinkan hasil dengan akurasi maksimum dalam memprediksi band gap material semikonduktor silikon. Random Forest Regressor bekerja dengan membangun banyak pohon keputusan selama pelatihan dan menggabungkan hasilnya untuk meningkatkan akurasi prediksi serta mengurangi risiko overfitting. Algoritma ini berguna karena dapat menangani data yang kompleks dan non-linear, serta memiliki kemampuan untuk menentukan pentingnya fitur. Di sisi lain, XGBoost Regressor adalah algoritma boosting yang meningkatkan kinerja prediksi melalui pendekatan iteratif, di mana model baru terus dibangun untuk memperbaiki kesalahan dari model sebelumnya. XGBoost dikenal karena efisiensinya dalam komputasi dan kemampuannya untuk memberikan hasil prediksi yang sangat akurat [11].

Dalam pemanfaatan machine learning dalam riset material, Materials Project menjadi tonggak penting. Materials project muncul sebagai solusi terobosan dalam mengatasi tantangan klasik dalam pengembangan material. Dengan memanfaatkan komputasi tingkat tinggi dan kecerdasan buatan, platform ini memberikan akses cepat dan terukur terhadap data sifat material yang luas, termasuk material silikon. Dari dataset silikon yang diperoleh dari Materials Project, dilakukan ekstraksi fitur berdasarkan pendekatan composition based, struktur based, dan lainnya. Fitur-fitur ini digunakan untuk membangun model machine learning yang dapat memprediksi band gap dari material silikon [12]. Evaluasi model machine learning dilakukan berdasarkan metrik seperti Mean Squared Error (MSE) dan R-Square untuk mengukur kinerja model tersebut. Dengan pendekatan ini, Materials Project dan pemanfaatannya dalam pengembangan model machine learning memberikan kontribusi signifikan dalam memahami dan memprediksi sifat material secara efisien [13].

Dengan demikian, penelitian ini berkontribusi pada pengembangan model machine learning yang akurat menggunakan algoritma machine learning dalam memprediksi band gap material semikonduktor silikon, serta mengevaluasi model machine learning yang telah dibuat untuk memprediksi band gap material silikon. Penelitian ini juga meningkatkan pemahaman tentang sifat material semikonduktor, mendukung perancangan material yang lebih baik, dan membantu dalam pemilihan algoritma yang efektif.

METODE

Pada penelitian ini dilakukan pengumpulan data Data material semikonduktor silikon untuk penelitian ini akan diperoleh melalui akses dan pengunduhan data dari database Materials Project. Database ini menyediakan informasi lengkap terkait properti material, termasuk data eksperimental dan teoretis mengenai band gap. Proses ekstraksi data akan dilakukan menggunakan API atau metode ekstraksi yang disediakan oleh Materials Project yaitu dengan menggunakan library Pymatgen. Berikut merupakan diagram alir penelitiannya :

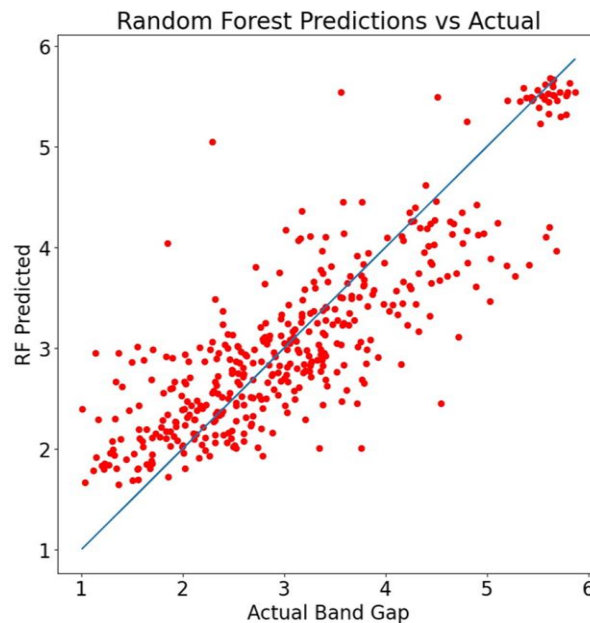


GAMBAR 1 Diagram Alir Penelitian

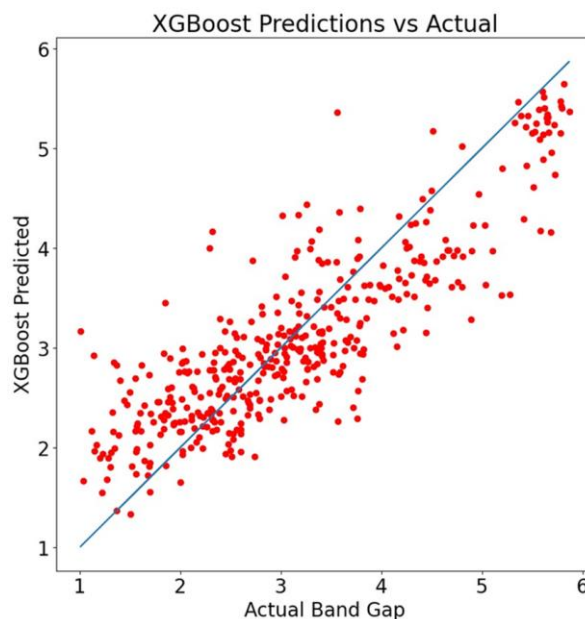
Feature Extraction dilakukan untuk mengidentifikasi dan mengolah data mentah menjadi format yang sesuai untuk pembangunan model. Langkah berikutnya adalah analisis data eksploratif (EDA) untuk memahami dataset dengan lebih baik, termasuk visualisasi distribusi data, mengidentifikasi pola, mendeteksi outlier, dan memahami hubungan antar variabel. Setelah EDA, dilakukan pembangunan model machine learning dengan tujuan memprediksi nilai band gap material silikon. Model-model ini dilatih menggunakan fitur yang telah diekstraksi dan dioptimalkan untuk kinerja terbaik. Kemudian, evaluasi model machine learning dilakukan untuk menilai akurasi prediksi model, termasuk pengujian dengan data baru dan perbandingan performa algoritma, seperti Random Forest Regressor dan XGBoost Regressor. Proses penelitian diakhiri dengan dokumentasi hasil evaluasi model yang telah dilakukan.

HASIL DAN PEMBAHASAN

Penelitian dimulai dengan mengakses database material proyek, di mana data material dikumpulkan dari sumber yang relevan. Selanjutnya, dilakukan ekstraksi fitur untuk mengidentifikasi atribut-atribut yang relevan dari data tersebut. Proses ini diikuti dengan Exploratory Data Analysis (EDA) yang bertujuan memahami karakteristik data, melihat distribusi, dan mengidentifikasi pola. Setelah itu, model machine learning dibangun berdasarkan data material dan fitur-fiturnya. Model ini kemudian dievaluasi menggunakan metrik kinerja seperti Mean Squared Error (MSE) dan R-squared (R^2) dengan hasil yaitu sebagai berikut :



GAMBAR 2 Hasil Prediksi Random Forest Regressor



GAMBAR 3 Hasil Prediksi XGBoost Regressor

| Evaluasi | Random Forest Regressor | XGBoost Regressor |
|----------|-------------------------|-------------------|
| R-Square | 0,7226 | 0,7272 |
| MSE | 0,3837 | 0,3773 |

TABEL 1 Hasil Evaluasi Performa Prediksi Machine Learning

Berdasarkan data yang telah didapatkan melalui proses pengolahan, diketahui bahwa nilai R-Square yang diperoleh oleh algoritma random forest memiliki nilai sebesar 0,7226 dengan nilai MSE yang diperoleh adalah 0,3837. Sedangkan, nilai R-Square yang diperoleh oleh algoritma XGBoost memiliki nilai sebesar 0,7272 dengan nilai MSE yang diperoleh adalah 0,3773. Pengujian menunjukkan bahwa algoritma XGBoost memiliki performa yang lebih baik untuk meningkatkan akurasi untuk memprediksi band gap material semikonduktor silikon.

KESIMPULAN

Dengan demikian dapat diketahui bahwa algoritma XGBoost Regressor memiliki nilai R^2 sebesar 0,7272 dengan nilai MSE sebesar 0,3773. Nilai tersebut menunjukkan performa yang lebih baik dalam memprediksi band gap material semikonduktor silikon dibandingkan dengan algoritma Random Forest Regressor.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada dosen pembimbing yang telah membimbing dalam penyusunan jurnal ini. Semoga jurnal ini dapat memberikan manfaat para pembaca. Mohon maaf apabila terdapat kekurangan dan kesalahan pada penulisan jurnal ini.

REFERENSI

- [1] A. Huda, C. T. Handoko, M. D. Bustan, B. Yudono, and F. Gulo. (2018). New route in the synthesis of Tin(II) oxide micro-sheets and its thermal transformation. *Materials Letters*, 211, 293–295.
- [2] A. Huda et al. (2019). Visible light-driven photoelectrocatalytic degradation of acid yellow 17 using Sn₃O₄ flower-like thin films supported on Ti substrate (Sn₃O₄/TiO₂/Ti). *Journal of Photochemistry and Photobiology A: Chemistry*.
- [3] A. Hernández-Ramírez and I. Medina-Ramírez, Eds. (2015). *Photocatalytic Semiconductors*. Cham: Springer International Publishing.
- [4] Boylestad, Robert and Nashelsky, Louis. (2002). *Electronic Devices and Circuit Theory*. New Jersey: Prentice Hall.
- [5] Greiner, R.A. (1961), *Semiconductor Devices and Applications*, McGraw-Hill book Co., inc., New York.
- [6] A. Huda et al. (2019). *SC. Journal of Photochemistry and Photobiology A: Chemistry*.
- [7] Arnett, D. (1996). *Supernovae and Nucleosynthesis* (First ed.). Princeton, New Jersey: Princeton University Press.
- [8] Sham, L. J., & Schlüter, M. (1983). Density-Functional Theory of the Energy Gap. *Physical Review Letters*, 51(22), 1888–1891.
- [9] Perdew, J. P., Yang, W., Burke, K., Yang, Z., Gross, E. K. U., Scheffler, M., Scuseria, G. E., Henderson, T. M., Zhang, I. Y., Ruzsinszky, A., et al. (2017). Understanding band gaps of solids in generalized Kohn–Sham theory. *Proceedings of the National Academy of Sciences of the United States of America*, 114(11), 2801–2806.
- [10] Hybertsen, M. S., & Louie, S. G. (1986). Electron correlation in semiconductors and insulators: Band gaps and quasiparticle energies. *Physical Review B: Condensed Matter and Materials Physics*, 34(8), 5390–5413.
- [11] Giarsyani, N., Hidayatullah, A. F., & Rahmadi, R. (2020). Komparasi Algoritma Machine Learning dan Deep Learning untuk Named Entity Recognition: Studi Kasus Data Kebencanaan. *JIRE (Jurnal Informatika & Rekayasa Elektronika)*, 3(1), 48-57.
- [12] Ward, L., Dunn, A., Faghaninia, A., Zimmermann, N. E. R., Bajaj, S., Wang, Q., Montoya, J. H., Chen, J., Bystrom, K., Dylla, M., Chard, K., Asta, M., Persson, K., Snyder, G. J., Foster, I., Jain, A. (2018). Matminer: An open source toolkit for materials data mining. *Computational Materials Science*, 152, 60-69.
- [13] S. Sherif. (2019). Evaluating machine learning models: a beginner’s guide to Mean Squared Error (MSE) and R-Squared. *Journal of Data Science*, 15(4), 215-228.