

Perbandingan Kinerja SVR dan XGBoost untuk Peramalan Emisi CO₂ Global berbasis Machine Learning

Alrafiqul Rahman^{1*}, Valentinus Paramarta², Agnes Novita Ida³, Mohammad Harits Akbar⁴,
Vica Sonya M. Simanjuntak⁵

^{1,2,3,4,5} Analitika Data Bisnis/ Fakultas Teknologi Informasi, Perbanas Institute

*email: alrafiqul.rahman@perbanas.id

DOI: <https://doi.org/10.31603/komtika.v9i1.13449>

Received: 27-04-2025, Revised: 30-04-2025, Accepted: 05-05-2025

ABSTRACT

The increase in global carbon dioxide (CO₂) emissions has become a major concern in climate change studies. This research aims to compare the performance of two machine learning algorithms, namely Support Vector Regression (SVR) and XGBoost, in predicting global CO₂ emission trends based on historical data. The dataset used includes related variables such as energy consumption, gross domestic product (GDP), and population, obtained from open data sources like Our World in Data. SVR is optimized through grid search to obtain the best parameters, while XGBoost is used as the main comparator due to its ability to handle non-linear relationships and feature interactions. Model evaluation was conducted using the MAE, RMSE, and R² metrics. The results show that XGBoost is significantly superior with an MAE of 1745.70 and an RMSE of 2663.18, as well as an R² value of 0.93, compared to SVR which has an MAE of 5476.54, an RMSE of 8153.37, and an R² value of 0.82. The visualization of the prediction results also indicates that XGBoost is more accurate in following the fluctuation patterns of actual data, especially in detecting sharp changes. These findings suggest that XGBoost is a more suitable method for forecasting CO₂ emissions in complex and non-linear global contexts.

Keywords: CO₂ emissions; XGBoost; SVR; machine learning; forecasting climate change; prediction.

ABSTRAK

Peningkatan emisi karbon dioksida (CO₂) secara global telah menjadi perhatian utama dalam studi perubahan iklim. Penelitian ini bertujuan untuk membandingkan kinerja dua algoritma machine learning, yaitu Support Vector Regression (SVR) dan XGBoost, dalam memprediksi tren emisi CO₂ global berdasarkan data historis. Dataset yang digunakan mencakup variabel terkait seperti konsumsi energi, produk domestik bruto (PDB), dan populasi, yang diperoleh dari sumber data terbuka seperti Our World in Data. SVR dioptimasi melalui pencarian grid (Grid Search) untuk mendapatkan parameter terbaik, sementara XGBoost digunakan sebagai pembanding utama karena kemampuannya dalam menangani relasi non-linear dan interaksi antar fitur. Evaluasi model dilakukan menggunakan metrik MAE, RMSE, dan R². Hasil menunjukkan bahwa XGBoost secara signifikan lebih unggul dengan MAE sebesar 1745,70 dan RMSE sebesar 2663,18 serta nilai R² sebesar 0,93, dibandingkan SVR yang memiliki MAE sebesar 5476,54, RMSE 8153,37 dan R² sebesar 0,82. Visualisasi hasil prediksi juga mengindikasikan bahwa XGBoost lebih akurat dalam mengikuti pola fluktuasi data aktual, terutama dalam mendeteksi perubahan tajam. Temuan ini menyarankan bahwa XGBoost merupakan metode yang lebih tepat untuk digunakan dalam peramalan emisi CO₂ pada konteks global yang kompleks dan non-linear.

Keywords: Emisi CO₂; XGBoost; SVR; machine learning; peramalan perubahan iklim; prediksi.

PENDAHULUAN

Dalam beberapa dekade terakhir, peningkatan emisi gas rumah kaca, khususnya karbon dioksida (CO₂), telah menjadikan perubahan iklim sebagai masalah lingkungan yang semakin mendesak [1]. Untuk memahami dan mengelola bahaya yang ditimbulkan oleh perubahan iklim, diperlukan pendekatan prediktif yang akurat terhadap tren emisi CO₂ di masa depan [2]. Metode pembelajaran mesin, atau ML, mulai banyak digunakan dalam konteks ini karena kemampuan untuk menangani data besar dan menemukan pola tersembunyi yang sulit ditemukan dengan pendekatan statistik konvensional [3].

Salah satu pendekatan machine learning yang menjanjikan dalam peramalan (*forecasting*) adalah Support Vector Regression (SVR) dan XGBoost (eXtreme Gradient Boosting), yang dikenal dengan kemampuan generalisasi tinggi dan efektivitasnya dalam menangani data nonlinear [4]. SVR dan XGBoost telah digunakan secara luas dalam berbagai studi peramalan, termasuk prediksi kualitas udara dan emisi polutan, dengan hasil yang kompetitif dibandingkan metode lain seperti regresi linier atau decision tree [5]. Selain itu, SVR dan XGBoost memiliki fleksibilitas dalam pemilihan kernel, yang memungkinkan adaptasi terhadap berbagai pola temporal dalam data emisi CO₂ [6].

Penelitian ini bertujuan untuk mengevaluasi kinerja dan XGBoost dalam memodelkan dan meramalkan tren emisi CO₂ global berdasarkan data historis; selanjutnya, penelitian ini mencari pola-pola utama yang mendasari dinamika emisi dari waktu ke waktu [7]. Selain itu, Penelitian yang dikerjakan oleh (Naseri dkk, 2020) menemukan bahwa penggunaan SVR dapat memberikan kontribusi signifikan dalam pengambilan kebijakan berbasis data untuk mitigasi perubahan iklim [8].

Pembelajaran mesin untuk memodelkan dan meramalkan emisi karbon dioksida di seluruh dunia telah meningkat seiring dengan tersedianya data lingkungan berskala besar dan peningkatan kemampuan komputasi dalam beberapa tahun terakhir [9]. Regresi linier, forest random, artificial neural network (ANN), dan long short-term memory (LSTM) adalah beberapa metode yang telah digunakan [10].

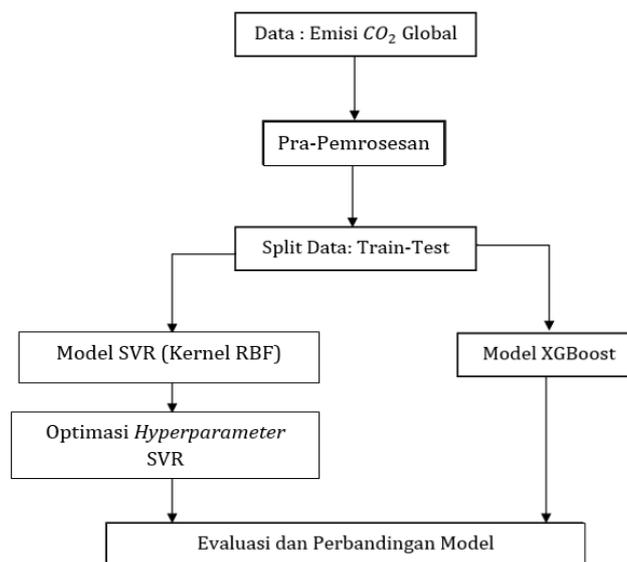
Masing-masing metode, bagaimanapun, memiliki keterbatasan. Misalnya, sementara ANN dan LSTM membutuhkan banyak data pelatihan dan tuning parameter yang kompleks, regresi linier tidak dapat menangani hubungan non-linear dalam data emisi [11]. Karena kemampuan mereka untuk menangani data nonlinear dan ketahanan terhadap overfitting, Support Vector Regression (SVR) muncul sebagai alternatif yang kompetitif dalam konteks ini. Metode yang menggunakan *Support Vector Regression* (SVR) dalam domain regresi dan peramalan berbasis waktu (*time series forecasting*) [12].

Studi yang dilakukan oleh Lu., dkk menunjukkan bahwa SVR dengan kernel radial basis function (RBF) dapat dengan akurat memprediksi emisi CO₂ regional bahkan melebihi model ANN dalam kondisi data terbatas [13]. Studi tambahan oleh Kumar & Srinivas (2022) menemukan bahwa SVR lebih baik untuk memodelkan emisi karbon jangka panjang dalam industri dan sektor transportasi [14]. Kemudian penelitian yang dilakukan alvian, dkk menggunakan metode SVR yang dimana pengembangan dari SVM jauh lebih baik dalam mengidentifikasi tutupan lahan [15]. Penelitian yang lain Suwarno dan Kusnadi menjelaskan bahwa Ini juga menunjukkan betapa fleksibel metode ini untuk mengidentifikasi tren musiman dan siklikal.

Namun, penelitian masih terbatas tentang penggunaan SVR dan XGBoost untuk peramalan emisi karbon dioksida secara global, terutama dalam konteks temporal dan spasial yang kompleks. Sebagian besar studi berkonsentrasi pada wilayah atau negara tertentu, menggunakan berbagai asumsi dan preprocessing. Dengan menggunakan dataset historis skala global yang mencakup berbagai variabel penentu emisi, penelitian ini bertujuan untuk mengisi celah tersebut dengan mempelajari bagaimana model SVR berfungsi di lingkungan global.

METODE

Penelitian ini dilakukan dengan tahapan-tahapan sistematis yang digambarkan dalam diagram alir, yang bertujuan untuk membangun dan membandingkan model SVR dan XGBoost dalam memprediksi emisi CO₂ global. Tahapan metode yang digunakan termasuk:



Gambar 1. Bagan Alir Metode Penelitian

Penelitian ini diawali dengan proses pengumpulan data yang diperoleh dari sumber-sumber publik seperti Our World in Data dan Global Carbon Project. Data yang dikumpulkan mencakup emisi karbon dioksida tahunan dalam satuan metrik ton dengan jumlah berisi 50.191 baris data dan 79 kolom, baik per negara maupun secara global, serta variabel lain yang relevan seperti konsumsi energi, Produk Domestik Bruto (PDB), dan jumlah populasi [16].

Tahap pra-pemrosesan data dilakukan untuk menyiapkan data mentah agar siap digunakan dalam proses pemodelan machine learning. Pada tahap ini, dilakukan pembersihan data dengan mengatasi nilai-nilai yang hilang, menghapus data duplikat, serta memastikan konsistensi format antar kolom, seperti satuan emisi karbon dan format penulisan tahun. Selanjutnya, dilakukan seleksi fitur dengan memilih variabel-variabel yang relevan seperti emisi karbon dioksida, konsumsi energi, Produk Domestik Bruto (PDB), dan jumlah populasi. Data dari berbagai sumber seperti Our World in Data dan Global Carbon Project juga disinkronisasi dan digabungkan berdasarkan entitas yang sama, seperti negara dan tahun. Untuk memastikan setiap fitur berada dalam skala yang seragam, data numerik kemudian dinormalisasi sebelum digunakan dalam pelatihan model. Jika terdapat variabel kategorikal,

maka dilakukan pengkodean agar bisa dikenali oleh algoritma. Setelah data dinyatakan bersih dan siap digunakan, dilakukan pembagian data menjadi dua bagian, yaitu data latih dan data uji, dengan rasio 80:20. Pembagian ini bertujuan agar model dapat dievaluasi secara objektif menggunakan data yang belum pernah dilatih sebelumnya.

SVR dengan kernel *Radial Basis Function* (RBF) digunakan dalam penelitian ini untuk memodelkan hubungan non-linear antara emisi karbon dioksida (CO₂) dan variabel lain seperti konsumsi energi, PDB, dan populasi. SVR bekerja dengan mencari fungsi regresi yang dapat memprediksi nilai emisi CO₂ dalam batas toleransi kesalahan tertentu (epsilon), sambil meminimalkan kompleksitas model. Kernel RBF digunakan untuk mentransformasikan data input ke ruang berdimensi lebih tinggi, sehingga pola yang tidak linear dapat dipelajari secara efektif. Parameter penting seperti C (penalti kesalahan), gamma (ukuran pengaruh data), dan epsilon disesuaikan menggunakan Grid Search untuk menghasilkan model dengan akurasi terbaik. Model ini dievaluasi menggunakan metrik RMSE dan R² untuk mengukur seberapa baik prediksi emisi CO₂ dibandingkan dengan data aktual.

Dalam proses pemodelan dengan SVR, dilakukan penyesuaian terhadap beberapa *hyperparameter* penting untuk meningkatkan akurasi dan performa model. Salah satu *hyperparameter* utama adalah C (parameter regulasi), yang mengontrol trade-off antara kompleksitas model dan toleransi terhadap kesalahan. Nilai C yang besar akan memaksa model untuk meminimalkan kesalahan pada data pelatihan, namun berisiko menyebabkan overfitting, sedangkan nilai C yang kecil memungkinkan adanya beberapa kesalahan tetapi menghasilkan model yang lebih sederhana dan general. *Hyperparameter* penting lainnya adalah gamma, yang mengatur seberapa jauh pengaruh dari satu data latih dalam ruang fitur ketika menggunakan kernel RBF. Gamma yang kecil menghasilkan model yang lebih halus dengan pengaruh data yang lebih luas, sementara gamma besar membuat model lebih kompleks dengan pengaruh data yang lebih sempit dan lokal. Selain itu, digunakan pula parameter epsilon (ϵ) yang mendefinisikan margin kesalahan yang masih ditoleransi tanpa dikenakan penalti. Artinya, model tidak akan berusaha melakukan prediksi yang sangat presisi jika kesalahan masih berada dalam batas epsilon, sehingga membantu menghindari sensitivitas berlebih terhadap noise. Ketiga *hyperparameter* ini disesuaikan menggunakan teknik Grid Search untuk menemukan kombinasi terbaik yang memberikan hasil prediksi paling optimal.

Sebagai pembanding, dikembangkan pula model regresi menggunakan algoritma XGBoost, sebuah metode boosting berbasis pohon keputusan yang dikenal efektif untuk menangani relasi non-linear dan robust terhadap multikolinearitas antar variabel. Setelah kedua model dibangun, dilakukan proses evaluasi dan perbandingan performa pada data uji menggunakan metrik RMSE dan R². Tujuan dari evaluasi ini adalah untuk mengetahui model mana yang memiliki kinerja lebih unggul dalam memprediksi tren emisi karbon dioksida secara global.

HASIL DAN PEMBAHASAN

Penelitian ini membandingkan dua model machine learning untuk peramalan emisi CO₂ global, yaitu Support Vector Regression (SVR) dan XGBoost Regressor. Masing-masing model dievaluasi berdasarkan tiga metrik utama:

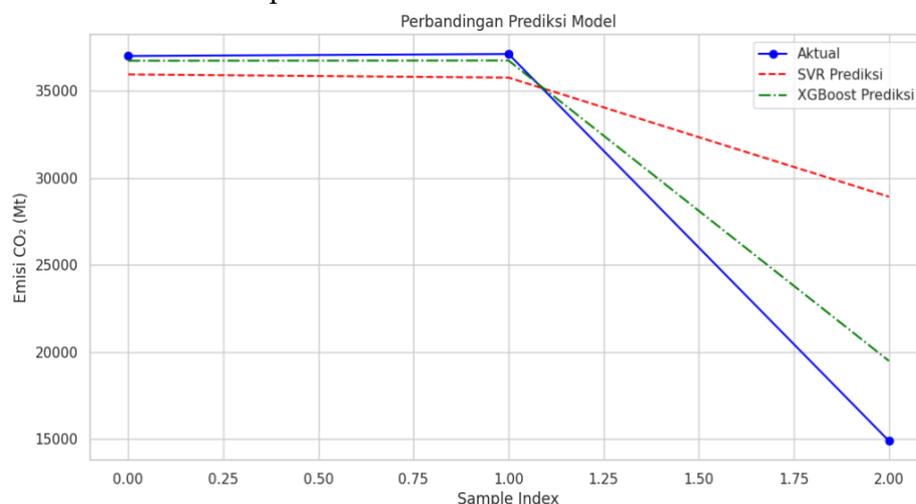
Tabel 1. Hasil Evaluasi Kinerja Model SVR dan XGBoost

Metrik	SVR (Tuning)	XGBoost
MAE	5476.54	1745.70
RMSE	8153.37	2663.18
R ² Score	0.82	0.93

Evaluasi performa model dalam penelitian ini menggunakan tiga metrik utama, yaitu Mean Absolute Error (MAE), Root Mean Square Error (RMSE), dan Koefisien Determinasi (R² Score). Metrik MAE mengukur rata-rata kesalahan absolut antara nilai prediksi dan nilai aktual. MAE memberikan gambaran seberapa jauh, secara rata-rata, hasil prediksi menyimpang dari nilai sebenarnya, tanpa mempertimbangkan arah kesalahan. Oleh karena itu, semakin kecil nilai MAE yang diperoleh, semakin baik performa model dalam menghasilkan prediksi yang mendekati nilai aktual.

Sementara itu, metrik RMSE digunakan untuk mengukur akar dari rata-rata kuadrat selisih antara nilai prediksi dan nilai aktual. Berbeda dengan MAE, RMSE memberikan penalti yang lebih besar terhadap kesalahan prediksi yang besar, sehingga metrik ini lebih sensitif terhadap outlier. Dalam hasil evaluasi, model XGBoost menunjukkan performa yang lebih baik dibandingkan model SVR, dengan menghasilkan nilai RMSE yang jauh lebih kecil. Hal ini menunjukkan bahwa prediksi XGBoost cenderung lebih stabil dan konsisten, terutama dalam menghindari kesalahan prediksi yang ekstrem.

Selanjutnya, metrik R² Score digunakan untuk mengukur seberapa besar proporsi variasi dalam data target yang dapat dijelaskan oleh model. Nilai R² berkisar antara 0 hingga 1, di mana nilai yang mendekati 1 menunjukkan bahwa model mampu menjelaskan sebagian besar variasi dalam data, sehingga memiliki kemampuan prediktif yang baik. Pada evaluasi ini, model XGBoost memperoleh nilai R² sebesar 0,93, yang menunjukkan bahwa 93% variasi dalam data target berhasil dijelaskan oleh model. Sebaliknya, model SVR hanya memperoleh R² sebesar 0,82. Dengan demikian, berdasarkan ketiga metrik evaluasi ini, dapat disimpulkan bahwa model XGBoost memiliki performa prediktif yang lebih unggul dibandingkan dengan SVR dalam memodelkan dan memprediksi tren emisi karbon dioksida.



Gambar 2. Perbandingan Prediksi Model SVR dan XGBoost

Visualisasi hasil prediksi model terhadap data emisi CO₂ disajikan dalam bentuk grafik garis untuk mempermudah analisis komparatif. Pada grafik tersebut, garis biru merepresentasikan nilai aktual dari emisi CO₂, yang menjadi acuan utama dalam evaluasi model. Sementara itu, garis merah putus-putus menunjukkan hasil prediksi dari model Support Vector Regression (SVR), dan garis hijau putus-putus merepresentasikan hasil prediksi dari model XGBoost.

Dari grafik ini, terlihat bahwa model XGBoost mampu mengikuti pola tren data aktual dengan lebih baik, terutama dalam menangkap penurunan tajam yang terjadi pada sampel ke-2. Hal ini menunjukkan sensitivitas XGBoost terhadap perubahan mendadak dalam data, yang menjadi keunggulan penting dalam konteks prediksi emisi CO₂ yang dapat dipengaruhi oleh faktor dinamis seperti kebijakan energi atau gangguan ekonomi. Sebaliknya, model SVR tampak menghasilkan prediksi yang lebih halus dan cenderung konservatif. Karakteristik ini menyebabkan SVR kurang responsif terhadap perubahan drastis, sehingga tidak mampu menangkap pola penurunan tajam dengan akurasi yang memadai. Dengan demikian, berdasarkan visualisasi ini, XGBoost menunjukkan performa yang lebih adaptif dan akurat dibandingkan SVR dalam merepresentasikan dinamika data aktual.

KESIMPULAN

Penelitian ini menunjukkan bahwa model XGBoost memiliki performa yang lebih baik dibandingkan Support Vector Regression (SVR) dalam meramalkan emisi CO₂ global. Hal ini dibuktikan dari nilai MAE dan RMSE yang lebih rendah serta R² yang lebih tinggi, yang mengindikasikan akurasi dan kemampuan generalisasi XGBoost yang lebih unggul. Selain itu, XGBoost mampu mengikuti pola tren dan fluktuasi data aktual dengan lebih baik, sementara SVR cenderung kurang responsif terhadap perubahan tajam meskipun telah dioptimasi dengan Grid Search. Mengingat kompleksitas dan sifat non-linear data emisi CO₂, XGBoost direkomendasikan sebagai metode yang lebih efektif dan efisien untuk peramalan dalam penelitian ini.

Dengan mempertimbangkan kompleksitas data emisi CO₂ yang bersifat non-linear dan melibatkan banyak variabel, model berbasis pohon keputusan seperti XGBoost dinilai lebih sesuai. Selain kemampuannya dalam menangani hubungan non-linear dan interaksi antar fitur tanpa perlu transformasi data yang rumit, XGBoost juga lebih efisien dalam implementasi. Oleh karena itu, XGBoost direkomendasikan sebagai metode yang lebih efektif untuk peramalan emisi CO₂ global, dan berpotensi untuk dikembangkan lebih lanjut melalui pendekatan seperti ensemble atau deep learning.

DAFTAR PUSTAKA

- [1] I. P. on C. C. (IPCC), "The Intergovernmental Panel on Climate Change (IPCC) is the United Nations body for assessing the science related to climate change.," 2022. [Daring]. Tersedia pada: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.ipcc.ch/report/ar6/wg3/downloads/report/IPCC_AR6_WGIII_FullReport.pdf
- [2] C. Le Quéré *et al.*, "Global Carbon Budget 2018," *Earth Syst. Sci. Data Discuss.*, vol.

- 10, no. 4, hal. 2141–2194, 2018, doi: 10.5194/essd-10-2141-2018.
- [3] M. I. Jordan and T. M. Mitchell dan D. Mulligan, “Machine learning: Trends, perspectives, and prospects,” *Science (80-.)*, vol. 349, no. 6245, hal. 255–260, 2015, doi: 10.1126/science.aaa8415.
- [4] O. Oktafia dan R. S. A. Nugroho, “Comparison of Support Vector Machine(Svm), Xgboost and Random Forest for Sentiment Analysis of Bumble App User Comments,” *Proxies J. Inform.*, vol. 6, no. 1, hal. 32–46, 2024, doi: 10.24167/proxies.v6i1.12453.
- [5] R. Wahyudi, S. Annas, dan Z. Rais, “Analisis Support Vector Regression (Svr) Untuk Meramalkan Indeks Kualitas Udara Di Kota Makassar,” *Variansi J. Stat. Its Appl. Teach. Res.*, vol. 5, no. 3, hal. 104–117, 2023, doi: 10.35580/variansiunm107.
- [6] A. A. Ajala, O. L. Adeoye, O. M. Salami, dan A. Y. Jimoh, “An examination of daily CO2 emissions prediction through a comparative analysis of machine learning, deep learning, and statistical models,” *Environ. Sci. Pollut. Res.*, vol. 32, no. 5, hal. 2510–2535, 2025, doi: 10.1007/s11356-024-35764-8.
- [7] J. Rogelj *et al.*, “Zero emission targets as long-term global goals for climate protection,” *Environ. Res. Lett.*, vol. 10, no. 10, hal. 105007, 2015, doi: 10.1088/1748-9326/10/10/105007.
- [8] H. Naseri, H. Jahanbakhsh, P. Hosseini, dan F. Moghadas Nejad, “Designing sustainable concrete mixture by developing a new machine learning technique,” *J. Clean. Prod.*, vol. 258, hal. 120578, 2020, doi: <https://doi.org/10.1016/j.jclepro.2020.120578>.
- [9] T. Janhuaton *et al.*, “Data-Driven Forecasting of CO2 Emissions in Thailand’s Transportation Sector Using Nonlinear Autoregressive Neural Networks,” *Big Data Cogn. Comput.*, vol. 9, no. 3, 2025, doi: 10.3390/bdcc9030071.
- [10] A. Rahman dan A. Bustamam, “Deep learning with concatenate model to detect COVID-19 lung disease with CT scan images,” in *AIP Conference Proceedings*, AIP Publishing, 2022. doi: 10.1063/5.0072411.
- [11] A. Rahman, L. S. Istiyowati, V. Valentinus, I. Ivan, dan Z. Azis, “Implementasi Data Mining Dalam Prediksi Harga Saham BBNI Dengan Pemodelan Matematika Menggunakan Metode LSTM Dengan Optimasi Adam,” *JUTECH J. Educ. Technol.*, vol. 5, no. 2, hal. 427–439, 2024, doi: 10.31932/jutech.v5i2.4137.
- [12] D. Armiady, “Analisis Algoritma Logistic Regression Dan Support Vector Machine Pada Kasus Klasifikasi Citra Hewan Rawa Dengan Dataset Yang Tidak Seimbang,” vol. 4, no. 1, hal. 69–77, 2024, doi: 10.47709/dsi.v4i1.4433.
- [13] Y. Inoue, “Winner-Takes-All or Co-Evolution among Platform Ecosystems: A Look at the Competitive and Symbiotic Actions of Complementors,” *Sustainability*, vol. 11, no. 3, 2019, doi: 10.3390/su11030726.
- [14] M. Azadian, H. Ghanadzadeh Gilani, H. Rajabi-Kuyakhi, dan N. Gharibipour, “Kinetic, isotherm, thermodynamic, and adsorption capacity studies of magnetic *Spirulina* microalgae onto zinc (II),” *Environ. Prog. & Sustain. Energy*, vol. 41, no. 2, hal. e13751, 2022, doi: <https://doi.org/10.1002/ep.13751>.

- [15] A. Aji Purboyo *et al.*, “Identifikasi Tutupan Lahan Menggunakan Support Vector Machine di Kawasan Perkotaan Cekungan Bandung,” *J. Komtika (Komputasi dan Inform.*, vol. 8, no. 1, hal. 54–64, 2024, [Daring]. Tersedia pada: <https://doi.org/10.31603/komtika.v8i1.11140>
- [16] O. W. in D. (OWID), “CO2 Data.” 2024. [Daring]. Tersedia pada: <https://github.com/owid/co2-data?tab=readme-ov-file>



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)
