

## Optimalisasi *Grid Search* pada *Extreme Gradient Boosting* (XGBoost) untuk Prediksi Pendapatan Asli Daerah (PAD) dari Pajak Kendaraan Bermotor (PKB)

Bintang Maulidana Muhammad<sup>1</sup>, Abdulloh Hamid<sup>2\*</sup>, Ahmad Hanif Asyhar<sup>3</sup>, Retna Fetty Idamayanti<sup>4</sup>

<sup>123</sup>Universitas Islam Negeri Sunan Ampel Surabaya, Jl. Dr. Ir. H. Soekarno No. 682, Surabaya, Jawa Timur, Indonesia

<sup>4</sup>UPT Badan Pendapatan Surabaya Utara, Jl. Rajawali No. 6-8, Surabaya, Jawa Timur, Indonesia

Email: <sup>1</sup>muhammadmaulidanabintang@gmail.com, <sup>2</sup>doelhamid@uinsa.ac.id, <sup>3</sup>hanif@uinsa.ac.id,

<sup>4</sup>mayadipendajatim@gmail.com

\*Corresponding Author

### ABSTRACT

*Local Revenue from Motor Vehicle Tax is one of the key sources of local government revenue used to fund development and public services. This study aims to predict Local Revenue (PAD) from the Motor Vehicle Tax (PKB) sector in North Surabaya for the period January – April 2025. The model used in this study is the XGBoost algorithm optimized with Grid Search. The dataset consists of 1,826 daily data points from January 2, 2020, to December 31, 2024, divided into five parts using the Time Series Split technique. Model performance evaluation was conducted using the Mean Absolute Percentage Error (MAPE) metric. The parameter tuning results indicate that the optimal hyperparameter— combination learning\_rate of 0.1, max\_depth of 3, n\_estimators of 100, and subsample of 0.8 yields a MAPE value of 0.4231%. This very low MAPE value (<1%) demonstrates the model's high accuracy in capturing seasonal patterns and daily trends, making it highly suitable for use as a strategic consideration in local financial policy.*

**Keywords:** *Grid Search, Motor Vehicle Tax, Prediction, XGBoost*

### ABSTRAK

Pendapatan Asli Daerah dari Pajak Kendaraan Bermotor merupakan salah satu sumber penerimaan penting daerah yang digunakan untuk mendanai pembangunan serta pelayanan publik. Penelitian ini bertujuan memprediksi Pendapatan Asli Daerah (PAD) dari sektor Pajak Kendaraan Bermotor (PKB) di Surabaya Utara periode Januari–April 2025. Model yang digunakan dalam penelitian ini adalah algoritma XGBoost yang dioptimalkan dengan Grid Search. Dataset terdiri dari 1.826 data harian periode 2 Januari 2020 hingga 31 Desember 2024 yang dibagi menggunakan teknik *Time Series Split* sebanyak lima kali. Penilaian performa model dilakukan menggunakan metrik *Mean Absolute Percentage Error* (MAPE). Hasil *tuning* parameter menunjukkan bahwa kombinasi hyperparameter terbaik *learning\_rate* sebesar 0.1, *max\_depth* 3, *n\_estimators* 100, dan *subsample* 0.8, menghasilkan nilai MAPE sebesar 0.4231%. Nilai MAPE yang sangat rendah (<1%) menunjukkan akurasi model yang sangat tinggi dalam mengikuti pola musiman dan tren harian, sehingga sangat layak digunakan sebagai bahan pertimbangan strategis kebijakan keuangan daerah.

**Kata kunci:** Grid Search, Pajak Kendaraan Bermotor, Prediksi, Xgboost

Dikirim: Desember 2025; Diterima: Januari 2026; Dipublikasikan: Maret 2026

Cara sitasi: Muhammad, B. M. Hamid, A. Asyhar, A. H. Idamayanti, R. F. (2026). Optimalisasi *Grid Search* pada *Extreme Gradient Boosting* (XGBoost) untuk Prediksi Pendapatan Asli Daerah (PAD) dari Pajak Kendaraan Bermotor (PKB). *Teorema: Teori dan Riset Matematika*, 11(1), 45–56. DOI: <https://dx.doi.org/10.25157/teorema.v10i1.22512>.

This is an open access article under the [CC-BY-SA](#) license



## PENDAHULUAN

Surabaya adalah kota terbesar kedua di Indonesia setelah Daerah Khusus Jakarta (DKJ). Surabaya memiliki luas  $\pm 326,81$  km<sup>2</sup>. Menurut Badan Pusat Statistik (BPS), jumlah penduduk Kota Surabaya sekitar 3,02 juta per tahun 2024. Surabaya terdiri dari 31 kecamatan dan 153 kelurahan dengan batas wilayah yang terbagi menjadi 4, yaitu Surabaya Timur, Surabaya Selatan, Surabaya Barat, Surabaya Utara (Surabaya, 2025). Surabaya Utara terdiri dari 8 kecamatan dan 38 kelurahan yang teridentifikasi sebagai wilayah dengan atraksi wisata alam pantai dan permukiman yang padat (Pemerintah Surabaya, 2025). Potensi wisata alam yang tersedia tentunya mendukung pembangunan ekonomi di Surabaya Utara (Pratama & Tucunan, 2021). Untuk mencapai tujuan pembangunan ekonomi di Surabaya Utara, pemerintah Kota Surabaya melalui pemerintah daerah Surabaya Utara melakukan strategi pembangunan dan kebijakan untuk mewujudkan kemakmuran bagi seluruh masyarakat (Hadi et al., 2024). Guna mencapai pembangunan, terdapat instansi pemerintah daerah yang mengelola dan mengumpulkan pajak, yaitu Unit Pelaksana Teknis Badan Pendapatan Daerah (UPT Bapenda) Surabaya Utara.

Pajak dan retribusi merupakan komponen utama terbentuknya Pendapatan Asli Daerah (PAD) (Irmawati et al., 2024). Salah satu kontributor utama dalam penerimaan kas yaitu Pajak Kendaraan Bermotor (PKB). Menurut data di UPT Bapenda Surabaya Utara pada bulan Januari-April 2025, penerimaan kas mengalami penurunan yang signifikan dibandingkan dengan tahun-tahun sebelumnya. Penurunan PAD dari PKB ini akan menghambat pembangunan infrastruktur dan ekonomi. Untuk mengetahui perolehan tahun selanjutnya UPT Bapenda Surabaya Utara memprediksi target dari PAD dengan kebijakan tertentu. Prediksi target PAD dari PKB menjadi aspek krusial dalam merencanakan dan mengelola keuangan daerah. Prediksi tersebut bertujuan agar pemerintah dapat menyusun strategi kebijakan yang lebih efektif dalam meningkatkan pendapatan serta mengoptimalkan alokasi anggaran (Lubis, 2024).

Untuk meningkatkan dan mengoptimalkan alokasi anggaran, diperlukan solusi berupa memprediksi berbagai aspek untuk memperkirakan suatu nilai masa depan. Salah satu pendekatan yang efektif yaitu menggunakan metode *machine learning*. *Machine learning* mampu mengolah data *time series* karena mampu menangkap pola yang kompleks. Pendekatan ini mampu memproses analisis data menjadi lebih akurat dan dapat digunakan untuk memprediksi tren target PAD di masa mendatang. Salah satu metode *machine learning* yang dapat digunakan untuk memprediksi adalah *Extreme Gradient Boosting* (XGBoost), yaitu algoritma berbasis pohon keputusan yang unggul dalam menangani data tabular serta memiliki performa tinggi pada data kompleks (Zhang et al., 2022). XGBoost memiliki keunggulan dalam menangani data yang bersifat *nonlinear* dan memberikan hasil prediksi yang akurat dengan waktu komputasi yang singkat. Selain itu, penambahan Teknik *Grid Search* dilakukan untuk mencari kombinasi yang optimal dari *hyperparameter* untuk digunakan dalam model XGBoost. Kebaruan penelitian ini terletak pada penerapan optimasi Grid Search khusus pada model XGBoost untuk data PAD di lingkup UPT Bapenda Surabaya Utara guna menangani fluktuasi data harian yang belum banyak dieksplorasi dalam penelitian sebelumnya.

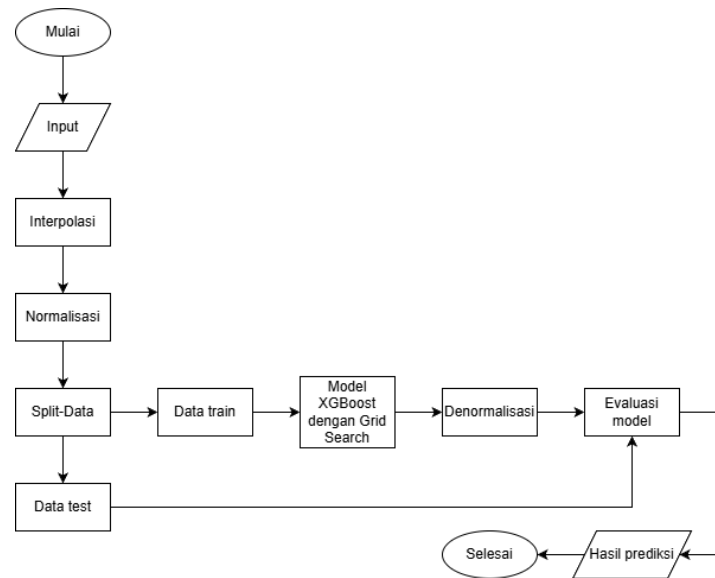
Penelitian terdahulu yang dilakukan oleh Tarwidi et al. (2023) tentang prediksi gelombang naik di pantai miring menggunakan XGBoost yang dioptimalkan menghasilkan akurasi R<sup>2</sup>, MAPE, dan RMSE sebesar 0.98675, 6.635%, dan 0.03902. Penelitian yang dilakukan oleh Islam et al. (2021). tentang metode XGBoost dalam pembuatan aplikasi peramalan dan analisis nilai tukar USD terhadap rupiah menghasilkan akurasi RMSE dan MAPE sebesar 0.23577 dan 0.11643. Perbandingan lima metode yaitu XGBoost, *Random Forest*, *AdaBoost*, SVM, dan *Decision Tree*, pada penelitian oleh Amjad et al. (2022) tentang prediksi kapasitas daya dukung tiang pancang menyatakan bahwa kinerja XGBoost lebih baik dengan R<sup>2</sup>, MAE, dan RMSE sebesar 0.955, 59.929, dan 80.653. Meskipun penelitian-penelitian tersebut menunjukkan hasil yang sangat baik pada fenomena alam dan ekonomi makro, terdapat perbedaan signifikan dengan penelitian ini. Jika studi sebelumnya berfokus pada prediksi gelombang laut atau nilai tukar yang fluktuasinya dipengaruhi oleh faktor fisik dan global, penelitian ini secara khusus menyasar data fiskal daerah. Kompleksitas dalam penelitian ini terletak

pada pola harian pendapatan Pajak Kendaraan Bermotor (PKB) di Surabaya Utara yang sangat bergantung pada kebijakan lokal serta perilaku wajib pajak, sehingga memerlukan pendekatan optimasi yang lebih spesifik.

Oleh karena itu, tujuan utama penelitian ini adalah mengimplementasikan dan mengoptimalkan model XGBoost menggunakan Grid Search untuk meningkatkan akurasi prediksi target PAD dari PKB di UPT Bapenda Surabaya Utara secara harian. Dengan adanya prediksi ini, diharapkan dapat mempersiapkan strategi untuk meningkatkan pendapatan daerah dan pembangunan ekonomi.

## METODE PENELITIAN

Berikut adalah Langkah-langkah dalam penelitian yang akan ditampilkan dalam bentuk diagram alir atau *flowchart*:



**Gambar 1.** Diagram Alur Penelitian

Prosedur penelitian ini diawali dengan tahap pengumpulan data yang bersumber dari UPT Bapenda Surabaya Utara, berupa data harian historis Pajak Kendaraan Bermotor (PKB) periode 2 Januari 2020 hingga 31 Desember 2024 dengan total 1.826 sampel. Data tersebut kemudian melalui tahapan *pre-processing*, meliputi interpolasi linier untuk imputasi data yang hilang dan normalisasi *Min-Max* guna menyeragamkan skala data agar siap untuk pemodelan. Selanjutnya, data dibagi (*Split-Data*) menjadi *training set* dan *testing set*. *Training set* digunakan untuk melatih model utama, yaitu *Extreme Gradient Boosting* (XGBoost), yang dioptimalkan secara sistematis melalui algoritma *Grid Search* guna menentukan kombinasi *hyperparameter* yang memberikan kinerja terbaik. Setelah pelatihan, hasil prediksi model menjalani denormalisasi untuk mengembalikan nilai ke satuan rupiah yang sebenarnya, diikuti dengan evaluasi model menggunakan *Mean Absolute Percentage Error* (MAPE), yang berfungsi mengukur akurasi model sebelum menghasilkan luaran akhir berupa hasil prediksi target pendapatan di masa depan.

Interpolasi linier adalah metode yang digunakan untuk memperkirakan atau mengisi nilai yang hilang (*missing value*) suatu variabel di antara dua titik data yang diketahui. Pendekatan ini secara matematis mengikuti pola Persamaan Garis Lurus (PGL) (Noertjahjani *et al.*, 2024).

Rumus interpolasi linier adalah sebagai berikut:

$$y = y_0 + \frac{y_1 - y_0}{x_1 - x_0} (x - x_0)$$

Dimana:

$y$  : Nilai data hasil interpolasi  
 $x$  : Nilai data kosong  
 $x_0$  dan  $y_0$  : Nilai atas  
 $x_1$  dan  $y_1$  : Nilai bawah

Normalisasi data merupakan proses penting dalam *pre-processing* yang bertujuan untuk menskalakan nilai suatu variabel agar berada dalam rentang yang sama, sehingga mencegah dominasi variabel dengan nilai ekstrem (terlalu besar atau kecil) dalam proses pelatihan model. Teknik yang umum digunakan dalam penelitian ini adalah Normalisasi *Min-Max*, di mana nilai-nilai data PKB mentah diubah dan dipetakan ke dalam rentang interval 0 dan 1. Penerapan ini sangat penting untuk mempermudah analisis statistik lanjutan dan meningkatkan efisiensi komputasi model *machine learning* (Wenny, 2024). Rumus Normalisasi *Min-Max* adalah sebagai berikut:

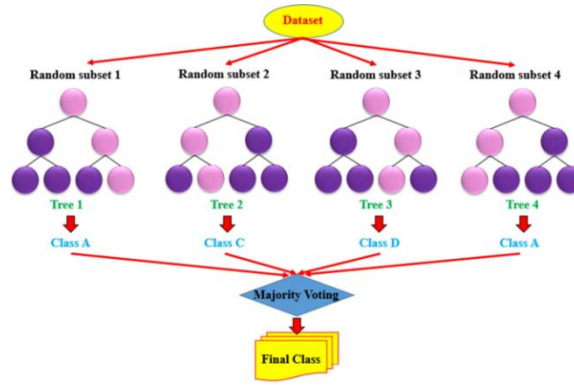
$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Dimana:

$x$  : Nilai data awal  
 $x'$  : Nilai data akhir  
 $x_{min}$  : Nilai minimum dari data  
 $x_{max}$  : Nilai maximum dari data

*Grid Search* adalah teknik optimasi *hyperparameter* yang sistematis dan komprehensif, bertujuan untuk menemukan kombinasi parameter eksternal paling optimal bagi model *machine learning*, khususnya XGBoost dalam penelitian ini. Prinsip kerjanya adalah mengevaluasi secara menyeluruh setiap kemungkinan kombinasi dari rentang nilai yang telah ditentukan sebelum model dilatih. Dalam penelitian ini, rentang parameter (*parameter grid*) yang diuji meliputi *learning\_rate* (0.1), *max\_depth* (3), *n\_estimators* (100), dan *subsample* (0.8). Penentuan rentang parameter tersebut didasarkan pada kebutuhan untuk menyeimbangkan kompleksitas model serta menghindari risiko *overfitting*, sehingga model tetap memiliki kemampuan generalisasi yang baik terhadap data baru. Proses pengujian ini dilakukan untuk memastikan bahwa konfigurasi terbaik yang dipilih menghasilkan kinerja prediksi paling unggul, yang kemudian diverifikasi berdasarkan metrik evaluasi *Mean Absolute Percentage Error* (MAPE) (Shams *et al.*, 2024).

*Extreme Gradient Boosting* (XGBoost) adalah metode *machine learning* berbasis *gradient boosting* yang dirancang guna meningkatkan akurasi dan efisiensi dalam memprediksi. XGBoost bekerja dengan membangun pohon keputusan secara bertahap guna memperbaiki kesalahan prediksi tahap sebelumnya. Prediksi terakhir dihitung sebagai penjumlahan atau penggabungan dari pembelajar lemah untuk menciptakan pembelajar tangguh dari semua pohon keputusan (Mehdary *et al.*, 2024). XGBoost memanfaatkan struktur pohon keputusan dengan Teknik *boosting* untuk meningkatkan akurasi prediksi. Algoritma ini dilengkapi fitur regularisasi, pemrosesan paralel, dan penanganan data hilang secara otomatis. Efisiensi akses data dan kompresi juga dioptimalkan, sehingga XGBoost dapat mengolah data berukuran besar secara efektif dengan sumber daya yang minimal. Pendekatan ini membuat XGBoost unggul dalam pemodelan prediktif yang menuntut kinerja tinggi (Momeni *et al.*, 2023).



**Gambar 2.** Struktur pohon keputusan XGBoost (Abdi *et al.*, 2021)

Rumus penjumlahan semua pohon sebagai berikut:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i)$$

Dimana:

$\hat{y}_i$  : Hasil prediksi

$K$  : Banyaknya pohon

$f_k(x_i)$  : nilai pada pohon ke- $k$  dan daun ke- $i$

XGBoost juga menggunakan *Objective function* untuk meminimalkan kesalahan prediksi. Penambahan fungsi regularisasi kedalam *Objective function* untuk menghindari overfitting. Rumus umum *Objective function*:

$$Obj = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k)$$

Dimana:

$Obj$  : *Objective function* yang akan diminimalkan

$l(y_i, \hat{y}_i)$  : Kesalahan prediksi dari 2 titik

$\Omega(f_k)$  : Penalti kompleksitas pohon

$n$  : Banyaknya data

$K$  : Banyaknya pohon

Denormalisasi data adalah proses kebalikan dari normalisasi data, yaitu pengubahan hasil nilai data Kembali ke nilai real. Cara kerja metode ini dengan mengembalikan nilai data yang memiliki rentang 0 dan 1 menjadi rentang data yang sebenarnya (Riatma *et al.*, 2025).

Rumus denormalisasi *min-max* sebagai berikut:

$$x = x' (x_{max} - x_{min}) + x_{min}$$

Dimana:

$x$  : Nilai data setelah denormalisasi

$x'$  : Nilai data sebelum di denormalisasi

$x_{min}$  : Nilai minimum dari data

$x_{max}$  : Nilai maximum dari data

Setelah prediksi model dilakukan, kinerja model dievaluasi secara kuantitatif berdasarkan indikator yang relevan. Dalam penelitian ini, MAPE dipilih sebagai metrik evaluasi utama 1 karena beberapa alasan krusial. Pertama, MAPE adalah metrik standar yang mendefinisikan akurasi metode forecasting, yang sangat relevan untuk prediksi data time series seperti PAD dari PKB. Kedua, MAPE menghitung rata-rata kesalahan dari selisih persentase antara nilai prediksi dan nilai aktual, sehingga hasilnya mudah diinterpretasikan oleh pengambil kebijakan dan manajemen. Kemampuan untuk mengukur kesalahan relatif dalam persentase ini menjadikan MAPE ideal untuk membandingkan kinerja model prediksi target pendapatan (Wan *et al.*, 2024). Rumus MAPE adalah sebagai berikut:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\%$$

- $n$  : Jumlah sampel dalam data
- $y_i$  : Nilai aktual
- $\hat{y}_i$  : Nilai prediksi

Penelitian ini menggunakan data penerimaan dari Pajak Kendaraan bermotor (PKB) di UPT Bapenda Surabaya Utara. Data ini mencakup selama data harian dari 2 Januari 2020 – 31 Desember 2024 sebanyak 1826 data. Berikut sampel data penelitian:

**Tabel 1.** Sampel Data Penerimaan PKB di UPT Bapenda Surabaya Utara

| No    | Tanggal          | PKB              |
|-------|------------------|------------------|
| 1.    | 2 Januari 2020   | Rp 1.194.257.400 |
| 2.    | 3 Januari 2020   | Rp 1.089.673.800 |
| 3.    | 4 Januari 2020   | Rp 910.648.900   |
| 4.    | 5 Januari 2020   | Rp 92.953.350    |
| 5.    | 6 Januari 2020   | Rp 1.529.653.700 |
| ⋮     | ⋮                | ⋮                |
| 1822. | 27 Desember 2024 | Rp 1.829.162.383 |
| 1823. | 28 Desember 2024 | Rp 1.109.972.725 |
| 1824. | 29 Desember 2024 | Rp 114.528.700   |
| 1825. | 30 Desember 2024 | Rp 1.257.271.100 |
| 1826. | 31 Desember 2024 | Rp 0             |

**HASIL DAN PEMBAHASAN**

**Hasil Interpolasi Data**

**Tabel 2.** Sampel Hasil Interpolasi Data Penerimaan PKB di UPT Bapenda Surabaya Utara

| No    | Tanggal          | PKB              | Interpolasi      |
|-------|------------------|------------------|------------------|
| 1.    | 2 Januari 2020   | Rp 1.194.257.400 | Rp 1.194.257.400 |
| 2.    | 3 Januari 2020   | Rp 1.089.673.800 | Rp 1.194.257.400 |
| 3.    | 4 Januari 2020   | Rp 910.648.900   | Rp 1.089.673.800 |
| 4.    | 5 Januari 2020   | Rp 92.953.350    | Rp 910.648.900   |
| 5.    | 6 Januari 2020   | Rp 1.529.653.700 | Rp 1.529.653.700 |
| 1822. | 27 Desember 2024 | Rp 1.829.162.383 | Rp 1.829.162.383 |
| 1823. | 28 Desember 2024 | Rp 1.109.972.725 | Rp 1.109.972.725 |
| 1824. | 29 Desember 2024 | Rp 114.528.700   | Rp 114.528.700   |
| 1825. | 30 Desember 2024 | Rp 1.257.271.100 | Rp 1.257.271.100 |
| 1826. | 31 Desember 2024 | Rp 0             | Rp 1.264.863.422 |

Penerapan interpolasi linier pada dataset ini merupakan langkah strategis untuk mengatasi anomali data pada data harian tanpa transaksi (nilai nol) yang teridentifikasi muncul akibat hari libur nasional maupun kendala operasional teknis. Secara analitis, keberadaan nilai nol yang ekstrem pada data *time series* pendapatan dapat menciptakan *stochastic noise* yang berisiko menyesatkan algoritma XGBoost dalam menangkap pola ekonomi yang sebenarnya. Dengan menggunakan interpolasi linier, kontinuitas varians data tetap terjaga sehingga model mampu memahami sinyal musiman (*seasonality*) secara lebih halus dan konsisten. Hal ini memastikan hasil prediksi yang dihasilkan menjadi lebih reliabel dan akurat sebagai instrumen bagi pihak manajemen dalam menyusun perencanaan target penerimaan pajak kendaraan bermotor di masa mendatang.

### Hasil Normalisasi Data

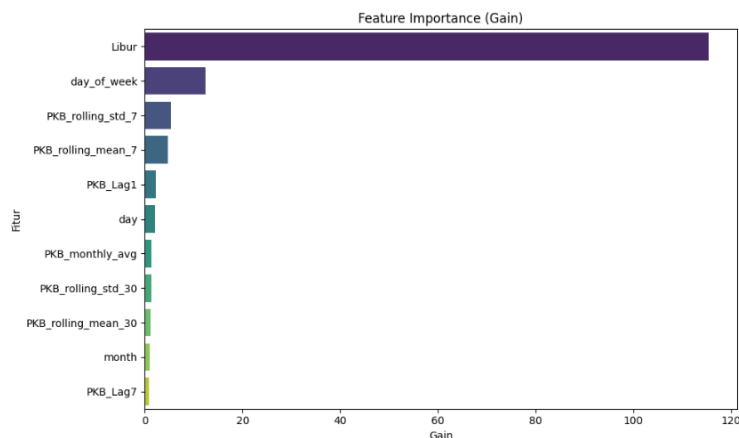
**Tabel 3.** Sampel Hasil Normalisasi Data Penerimaan PKB di UPT Bapenda Surabaya Utara

| No    | Tanggal          | PKB              | Interpolasi      | Normalisasi |
|-------|------------------|------------------|------------------|-------------|
| 1     | 2 Januari 2020   | Rp 1.194.257.400 | Rp 1.194.257.400 | 0,31        |
| 2     | 3 Januari 2020   | Rp 1.089.673.800 | Rp 1.194.257.400 | 0,28        |
| 3     | 4 Januari 2020   | Rp 910.648.900   | Rp 1.089.673.800 | 0,24        |
| 4     | 5 Januari 2020   | Rp 92.953.350    | Rp 910.648.900   | 0,02        |
| 5     | 6 Januari 2020   | Rp 1.529.653.700 | Rp 1.529.653.700 | 0,40        |
| ⋮     | ⋮                | ⋮                | ⋮                | ⋮           |
| 1822. | 27 Desember 2024 | Rp 1.829.162.383 | Rp 1.829.162.383 | 0,47        |
| 1823. | 28 Desember 2024 | Rp 1.109.972.725 | Rp 1.109.972.725 | 0,29        |
| 1824. | 29 Desember 2024 | Rp 114.528.700   | Rp 114.528.700   | 0,03        |
| 1825. | 30 Desember 2024 | Rp 1.257.271.100 | Rp 1.257.271.100 | 0,33        |
| 1826. | 31 Desember 2024 | Rp 0             | Rp 1.264.863.422 | 0,33        |

Proses normalisasi *Min-Max* ke dalam rentang 0 hingga 1 dilakukan bukan sekadar untuk penyederhanaan angka, melainkan sebagai upaya teknis untuk menyeragamkan bobot fitur sehingga model XGBoost dapat memproses data secara lebih efisien tanpa terdistorsi oleh besarnya nilai nominal rupiah. Secara analitis, transformasi ini memastikan bahwa fluktuasi harian seperti pada 2 Januari 2020 dengan nilai indeks 0,31 dapat dibandingkan secara proporsional terhadap titik puncak pendapatan dalam seluruh periode historis. Dengan menghilangkan dominasi nilai absolut yang ekstrem, normalisasi memungkinkan algoritma untuk fokus pada deteksi pola tren dan hubungan antarvariabel secara lebih objektif. Hal ini menjadi landasan krusial bagi model dalam menghasilkan proyeksi yang stabil dan meminimalisir bias selama proses *training*, sehingga visualisasi serta hasil peramalan akhir memiliki akurasi yang lebih konsisten.

### Proses XGBoost dengan *Tunning Hyperparameter Grid search*

#### 1. Feature Engineering



**Gambar 3.** Fitur Terbaik

Feature engineering berisi berbagai variabel yang digunakan untuk meningkatkan kemampuan model dalam memahami pola. Berdasarkan analisis Feature Importance (Gain), fitur Libur dan day of week memiliki kontribusi terbesar dalam meningkatkan prediksi, menunjukkan bahwa pola hari dan libur memainkan peran penting dalam perubahan PKB. Fitur lag memungkinkan model melihat nilai PKB dari hari sebelumnya, seminggu sebelumnya, dan sebulan sebelumnya, sehingga dapat menangkap hubungan antara data historis dan pola perubahan PKB (Surakhi *et al.*, 2021). Dalam analisis ini, lag1 dan lag7 menjadi fitur yang cukup berkontribusi, membantu model memahami dampak langsung dari PKB sebelumnya terhadap nilai terkini. Fitur rolling statistic digunakan karena pada data penelitian terdapat outlier. Fitur seperti rolling mean7 dan rolling mean30 memberikan informasi tentang rata-rata PKB dalam 7 hari dan 30 hari terakhir untuk mengidentifikasi tren stabilisasi atau lonjakan. Selain itu, rolling std7 dan rolling std30 dapat mengukur besar fluktuasi PKB dalam periode mingguan dan bulanan, membantu model memahami volatilitas yang terjadi (Setiawan *et al.*, 2024). Variabel monthly avg, yang menunjukkan rata-rata PKB dalam satu bulan, juga memberikan wawasan terhadap pola musiman dari tahun ke tahun. Dengan kombinasi fitur ini, model dapat menangkap tren, pola musiman, serta volatilitas yang terjadi dalam PKB, menghasilkan prediksi yang lebih akurat dan adaptif terhadap perubahan ekonomi. Struktur ini memberikan pemahaman lebih dalam tentang bagaimana berbagai fitur berkontribusi terhadap model, sesuai dengan analisis yang terlihat dalam Feature Importance (Gain).

**2. Model Hyperparameter**

**Tabel 4.** Parameter

| No. | Parameter            | Nilai/Range | Fungsi Parameter                                                                                                                 |
|-----|----------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------|
| 1.  | <i>n_estimators</i>  | [100, 300]  | Menentukan jumlah pohon keputusan yang akan diterapkan dalam proses klasifikasi atau regresi.                                    |
| 2.  | <i>max_depth</i>     | [3, 6]      | Mengontrol kedalaman maksimum setiap pohon yang dibangun untuk mencegah <i>overfitting</i> .                                     |
| 3.  | <i>learning_rate</i> | [0.01, 0.1] | Mengontrol seberapa banyak model diubah sebagai respons terhadap kesalahan yang diperkirakan setiap kali bobot model diperbarui. |
| 4.  | <i>subsample</i>     | [0.8, 1.0]  | Rasio subset data pelatihan yang diambil secara acak sebelum membangun setiap pohon, membantu mengurangi <i>overfitting</i> .    |

*Hyperparameter tuning* dalam XGBoost memainkan peran krusial dalam menentukan kinerja dan akurasi prediksi. *Number of Estimators* dengan rentang [100, 300] mengontrol jumlah pohon keputusan dalam *ensemble*, di mana jumlah pohon yang lebih besar memungkinkan model menangkap pola lebih kompleks tetapi berisiko menyebabkan *overfitting*. *Maximum Depth* dengan rentang [3, 6] membatasi kedalaman maksimum pohon keputusan untuk menjaga keseimbangan antara fleksibilitas dalam menangkap hubungan antarfitur dan kompleksitas berlebihan. *Learning Rate* dalam rentang [0.01, 0.1] mengatur seberapa besar kontribusi setiap pohon terhadap model akhir, dengan nilai kecil menghasilkan pembelajaran yang lebih stabil dan terkendali, sementara nilai besar mempercepat konvergensi tetapi bisa menyebabkan *overfitting* jika tidak dikontrol dengan baik. *Subsample Ratio* dengan rentang [0.8, 1.0] menentukan persentase sampel data yang digunakan untuk membangun setiap pohon, di mana nilai lebih kecil membantu dalam mengurangi risiko *overfitting* dengan memberikan variasi lebih besar dalam data pelatihan. Kombinasi parameter ini mencerminkan strategi optimal untuk menemukan keseimbangan terbaik antara kompleksitas, kecepatan, dan kemampuan model dalam menangkap pola yang ada dalam data, sehingga menghasilkan prediksi yang lebih akurat dan adaptif terhadap perubahan ekonomi.



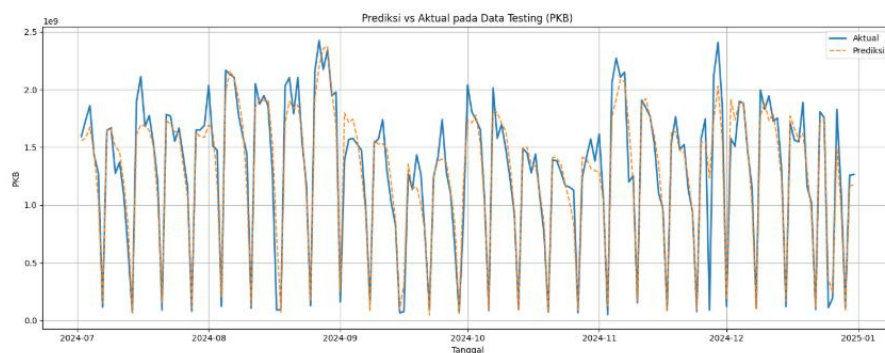
### 3. Training Model dan MAPE

|    | learning_rate | max_depth | n_estimators | subsample | MAPE     | Uji_ke |
|----|---------------|-----------|--------------|-----------|----------|--------|
| 0  | 0.010000      | 3         | 100          | 0.800000  | 0.593682 | 1      |
| 1  | 0.010000      | 3         | 100          | 1.000000  | 0.596247 | 2      |
| 2  | 0.010000      | 3         | 300          | 0.800000  | 0.436162 | 3      |
| 3  | 0.010000      | 3         | 300          | 1.000000  | 0.440633 | 4      |
| 4  | 0.010000      | 6         | 100          | 0.800000  | 0.572427 | 5      |
| 5  | 0.010000      | 6         | 100          | 1.000000  | 0.570640 | 6      |
| 6  | 0.010000      | 6         | 300          | 0.800000  | 0.435138 | 7      |
| 7  | 0.010000      | 6         | 300          | 1.000000  | 0.446135 | 8      |
| 8  | 0.100000      | 3         | 100          | 0.800000  | 0.423101 | 9      |
| 9  | 0.100000      | 3         | 100          | 1.000000  | 0.428010 | 10     |
| 10 | 0.100000      | 3         | 300          | 0.800000  | 0.435194 | 11     |
| 11 | 0.100000      | 3         | 300          | 1.000000  | 0.439408 | 12     |
| 12 | 0.100000      | 6         | 100          | 0.800000  | 0.436681 | 13     |
| 13 | 0.100000      | 6         | 100          | 1.000000  | 0.451395 | 14     |
| 14 | 0.100000      | 6         | 300          | 0.800000  | 0.439409 | 15     |
| 15 | 0.100000      | 6         | 300          | 1.000000  | 0.454273 | 16     |

**Gambar 4.** Uji Hyperparameters Optimal

Berdasarkan hasil uji *hyperparameter tuning* menggunakan *Grid Search*, diperoleh bahwa kombinasi parameter terbaik terdapat pada uji ke-9. Kombinasi tersebut terdiri dari *n\_estimators* sebesar 100, *max\_depth* sebesar 3, *learning\_rate* sebesar 0.1, dan *subsample* sebesar 0.8. Kombinasi ini menghasilkan nilai MAPE (*Mean Absolute Percentage Error*) sebesar 0.4231, yang merupakan nilai MAPE terendah dari seluruh 16 percobaan. Hal ini menunjukkan bahwa model yang dihasilkan memiliki tingkat kesalahan rata-rata hanya sekitar 0,42% dari nilai aktual. Jika diklasifikasikan berdasarkan standar evaluasi kemampuan peramalan menurut Lewis (1982), nilai MAPE di bawah 10% termasuk dalam kategori sangat akurat. Dengan tingkat *error* yang sangat rendah (di bawah 1%), model ini tergolong sangat andal dalam memprediksi nilai PKB. Hasil ini juga menunjukkan performa yang sangat kompetitif dibandingkan dengan penelitian terdahulu, seperti penelitian Tarwidi *et al* (2023) pada prediksi gelombang laut dengan MAPE 6,635% atau Islam *et al* (2021) pada peramalan nilai tukar mata uang dengan MAPE 0,11643%. Hal ini membuktikan bahwa penggunaan XGBoost yang dioptimalkan dengan *Grid Search* memiliki performa yang optimal untuk digunakan dalam peramalan target PAD dari Pajak Kendaraan Bermotor di UPT Bapenda Surabaya Utara.

#### Hasil Prediksi vs Aktual dari Data Testing

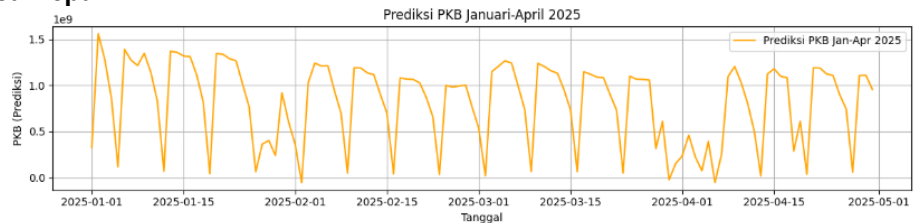


**Gambar 5.** Grafik Prediksi vs Aktual Data Testing

Visualisasi perbandingan antara nilai aktual dan prediksi pada periode Juli hingga Desember 2024 menunjukkan kemampuan model dalam mereplikasi dinamika pendapatan harian secara presisi. Secara analitis, keselarasan garis prediksi yang konsisten mengikuti fluktuasi tajam pada data aktual membuktikan bahwa model XGBoost berhasil mengekstraksi ketergantungan temporal, terutama dalam membedakan karakteristik *volume* transaksi antara hari kerja dan hari libur. Munculnya *deviasi*

*minor* pada periode Desember tidak mengurangi validitas model, melainkan mencerminkan kemampuan algoritma dalam menjaga stabilitas tren di tengah volatilitas akhir tahun yang biasanya dipengaruhi oleh anomali perilaku wajib pajak. Makna dari temuan ini mengonfirmasi bahwa nilai MAPE yang rendah bukan sekadar hasil optimasi matematis, melainkan representasi dari keandalan model dalam menangkap pola periodik yang kompleks. Dengan demikian, sinkronisasi visual ini memberikan jaminan bahwa model memiliki tingkat generalisasi yang kuat dan sangat layak digunakan sebagai instrumen proyeksi pendapatan PKB di masa depan.

**Prediksi Masa Depan**



**Gambar 6.** Grafik Prediksi PKB (Januari-April 2025)

Proyeksi pendapatan PKB untuk periode Januari hingga April 2025 mengungkapkan persistensi pola musiman yang sangat terstruktur, ditandai dengan ritme mingguan yang konsisten antara puncak hari kerja dan lembah akhir pekan. Secara analitis, kestabilan fluktuasi dalam rentang 0 hingga 1,5 miliar rupiah per hari ini menunjukkan bahwa model berhasil mengidentifikasi *baseline* pendapatan daerah yang tangguh terhadap perubahan waktu. Penurunan ketajaman yang terdeteksi pada awal April 2025 memberikan sinyal analitis mengenai adanya potensi interupsi aktivitas transaksi, yang kemungkinan besar berkaitan dengan periode libur panjang atau penyesuaian kalender administratif. Makna dari temuan ini memberikan nilai strategis bagi UPT Bapenda Surabaya Utara: bahwa meskipun tidak terlihat tren penurunan jangka panjang yang mengkhawatirkan, pemahaman atas titik-titik rendah (lembah) dalam grafik tersebut dapat digunakan sebagai dasar perancangan intervensi kebijakan atau optimalisasi layanan digital guna menjaga stabilitas likuiditas daerah. Dengan demikian, hasil prediksi ini bukan sekadar estimasi angka, melainkan instrumen navigasi fiskal yang realistis untuk memitigasi risiko ketidaktercapaian target di awal tahun 2025.

**Prediksi Pendapatan Asli Daerah**

**Tabel 5.** Prediksi PAD per Hari

| Tanggal        | Prediksi PKB     |
|----------------|------------------|
| 1 Januari 2025 | Rp 324.983.695   |
| 2 Januari 2025 | Rp 1.560.762.475 |
| 3 Januari 2025 | Rp 1.276.802.347 |
| 4 Januari 2025 | Rp 862.874.785   |
| 27 April 2025  | Rp 57.645.249    |
| 28 April 2025  | Rp 1.106.533.984 |
| 29 April 2025  | Rp 1.110.697.595 |
| 30 April 2025  | Rp 956.106.692   |

**Tabel 6.** Total PAD per Bulan

| Bulan | Total Prediksi PKB Bulanan |
|-------|----------------------------|
| 1     | Rp 28.359.761.183          |
| 2     | Rp 23.503.672.463          |
| 3     | Rp 25.159.411.255          |
| 4     | Rp 21.135.092.742          |

Akumulasi proyeksi bulanan mengungkap dinamika kapasitas fiskal daerah yang diprediksi mencapai puncaknya pada Januari 2025 dengan total Rp 28,36 miliar. Secara analitis, dominasi

pendapatan di awal tahun ini mengindikasikan adanya efek *momentum-driven*, di mana kepatuhan wajib pajak cenderung menguat pasca pergantian tahun administratif. Sebaliknya, fluktuasi pada bulan-bulan berikutnya termasuk kontraksi pada Februari dan April menandakan adanya pengaruh siklus musiman dan hari libur nasional yang secara periodik menekan frekuensi transaksi harian. Makna dari temuan ini menunjukkan bahwa model tidak hanya menghitung angka kumulatif, tetapi juga sensitif terhadap variasi jumlah hari aktif per bulan. Bagi UPT Bapenda, disparitas nilai antar bulan ini merupakan landasan krusial dalam melakukan kalibrasi target operasional, di mana strategi intensifikasi penagihan dapat difokuskan pada bulan-bulan dengan prediksi rendah (seperti April) guna menjaga stabilitas arus kas PAD. Dengan demikian, tabel prediksi ini berfungsi sebagai peta navigasi fiskal yang memungkinkan perencanaan anggaran dilakukan secara lebih presisi, terukur, dan berbasis data harian yang valid.

## KESIMPULAN

Penelitian ini menyimpulkan bahwa implementasi model XGBoost yang dioptimalkan dengan *Grid Search* sangat efektif untuk memprediksi Pendapatan Asli Daerah (PAD) dari Pajak Kendaraan Bermotor (PKB) secara harian. Penggunaan parameter terbaik (*learning\_rate* 0.1, *max\_depth* 3, *n\_estimators* 100, dan *subsample* 0.8) terbukti menghasilkan tingkat presisi tinggi dengan nilai MAPE sebesar 0,4231%. Secara metodologis, penelitian ini memberikan kontribusi berupa prosedur optimasi sistematis pada data fiskal yang fluktuatif, sementara secara praktis, hasil prediksi ini menjadi instrumen navigasi bagi pemerintah dalam memitigasi risiko penurunan pendapatan yang diproyeksikan terjadi pada April 2025. Implikasi kebijakan dari temuan ini menuntut Pemerintah Kota Surabaya perlu merespons dengan strategi kebijakan yang tepat guna menjaga kestabilan pendapatan daerah dan mendukung pertumbuhan ekonomi.

## DAFTAR PUSTAKA

- Abdi, J., Hadavimoghaddam, F., Hadipoor, M., & Hemmati-Sarapardeh, A. (2021). Modeling of CO<sub>2</sub> adsorption capacity by porous metal organic frameworks using advanced decision tree-based models. *Scientific Reports*, 11(1), 1–14. <https://doi.org/10.1038/s41598-021-04168-w>
- Amjad, M., Ahmad, I., Ahmad, M., Wróblewski, P., Kamiński, P., & Amjad, U. (2022). Prediction of Pile Bearing Capacity Using XGBoost Algorithm: Modeling and Performance Evaluation. *Applied Sciences (Switzerland)*, 12(4). <https://doi.org/10.3390/app12042126>
- Hadi, A. F., Enantya, F. B. P., & Khakimah, H. (2024). Analysis of the Success of Economic Development in Surabaya in the Era of Mayor Tri Rismaharini'S Government. *Journal of Finance, Economics and Business*, 3(1), 41–52. <https://doi.org/10.59827/jfeb.v3i1.111>
- Irmawati, Syam, A. B., Baharuddin, & Sudirman. (2024). Optimalisasi Pendapatan Asli Daerah melalui Pajak dan Retribusi. *Jurnal Ilmiah Administrasita*, 15(01), 74–82. <https://doi.org/https://doi.org/10.47030/administrasita.v15i1.752>
- Islam, S. F. N., Sholahuddin, A., & Abdullah, A. S. (2021). Extreme gradient boosting (XGBoost) method in making forecasting application and analysis of USD exchange rates against rupiah. *Journal of Physics: Conference Series*, 1722(1). <https://doi.org/10.1088/1742-6596/1722/1/012016>
- Lubis, D. B. (2024). *Manajemen Keuangan Sektor Publik* (D. I. M. G. Efgivia (ed.)). Widina Media Utama.
- Mehdary, A., Chehri, A., Jakimi, A., & Saadane, R. (2024). Hyperparameter Optimization with Genetic Algorithms and XGBoost: A Step Forward in Smart Grid Fraud Detection. *Sensors*, 24(4). <https://doi.org/10.3390/s24041230>
- Momeni, E., He, B., Abdi, Y., & Armaghani, D. J. (2023). Novel Hybrid XGBoost Model to Forecast Soil Shear Strength Based on Some Soil Index Tests. *CMES - Computer Modeling in Engineering and Sciences*, 136(3), 2527–2550. <https://doi.org/10.32604/cmes.2023.026531>

- Noertjahjani, S., Zaki, S. A., & Kiswanto, A. (2024). Analysis of the Application of Linear Interpolation and Quadratic Interpolation in Electrical Distribution Performance. *International Journal of Informatics and Computation (IJICOM)*, 6(2). <https://doi.org/https://doi.org/10.35842/ijicom.v6i2.84>
- Pemerintah Surabaya. (2025). *Kelurahan Di Surabaya*. <https://www.surabaya.go.id/id/page/0/8169/kelurahan>
- Pratama, R. B., & Tucunan, K. P. (2021). Analisis Rekognisi Citra Ruang Kota Surabaya Berdasarkan Persepsi Masyarakat Melalui Lensa Sosial Media. *Jurnal Teknik ITS*, 10(2). <https://doi.org/10.12962/j23373539.v10i2.63408>
- Riatma, D. L., Rahman, Y. F., Roshinta, T. A., Masbahah, Sani, A. F., Khoirunisa, R., & Haqimi, N. A. (2025). Model Prediksi Manajemen Stok Produk Berbasis Deep Learning Gated Recurrent Unit untuk Optimalisasi Rantai Pasok E-Commerce. *Jurnal Pendidikan Teknologi Informasi*, 5(1), 314–323. <https://doi.org/http://dx.doi.org/10.51454/decode.v5i1.1130>
- Setiawan, A., Arnita, Yusuf, D., Syafira, N., & Tania, T. (2024). Analisis Deskriptif Ump (Upah Minimum Provinsi) Seindonesia (2002-2022) Menggunakan Metode Fuzzy C Means Clustering. *Jurnal Ilmu Komputer Revolusioner*, 08(12), 1–13.
- Shams, M. Y., Elshewey, A. M., El-kenawy, E. S. M., Ibrahim, A., Talaat, F. M., & Tarek, Z. (2024). Water quality prediction using machine learning models based on grid search method. *Multimedia Tools and Applications*, 83(12), 35307–35334. <https://doi.org/10.1007/s11042-023-16737-4>
- Surabaya, B. P. S. K. (2025). Kota Surabaya Dalam Angka Surabaya Municipality In Figures. *Buku*, 37(1), 1–341. <https://surabayakota.bps.go.id/id/publication/2025/02/28/bd1f25e59ae790cc8a7c0c07/kota-surabaya-dalam-angka-2025.html>
- Surakhi, O., Zaidan, M. A., Fung, P. L., Motlagh, N. H., Serhan, S., Alkhanafseh, M., Ghoniem, R. M., & Hussein, T. (2021). Time-lag selection for time-series forecasting using neural network and heuristic algorithm. *Electronics (Switzerland)*, 10(20), 1–22. <https://doi.org/10.3390/electronics10202518>
- Tarwidi, D., Pudjaprasetya, S. R., Adytia, D., & Apri, M. (2023). An optimized XGBoost-based machine learning method for predicting wave run-up on a sloping beach. *MethodsX*, 10(March), 102119. <https://doi.org/10.1016/j.mex.2023.102119>
- Wan, A., Du, C., Gong, W., Wei, C., AL-Bukhaiti, K., Ji, Y., Ma, S., Yao, F., & Ao, L. (2024). Using Transfer Learning and XGBoost for Early Detection of Fires in Offshore Wind Turbine Units. *Energies*, 17(10), 1–20. <https://doi.org/10.3390/en17102330>
- Wenny. (2024). Normalisasi Data Kependudukan Dengan Model Min Max Dan Algoritma K-Means Untuk Pengelompokkan Tingkat Ekonomi Masyarakat. *Bulletin of Information System Research (BIOS)*, 2(2), 53–63. <https://doi.org/https://doi.org/10.62866/bios.v2i2.141>
- Zhang, P., Jia, Y., & Shang, Y. (2022). Research and application of XGBoost in imbalanced data. *International Journal of Distributed Sensor Networks*, 18(6). <https://doi.org/10.1177/15501329221106935>