

Analisis Tren dan Prediksi Konsumsi Energi Terbarukan Rumah Tangga Menggunakan Metode XGBoost

Achmad Rizqy Pranata¹⁾, Agustian Ardiansya¹⁾, Zamzam Muazam¹⁾, Salvadore Fremont Ouby Saleh Putra¹⁾, Harry Chandra¹⁾, Muhammad Ibadurrahman Arrasyid Supriyanto^{1*}, Novianti Puspitasari²⁾

¹⁾ Fakultas Teknik, Sistem Informasi, Universitas Mulawarman

²⁾ Fakultas Teknik, Informatika, Universitas Mulawarman

E-mail: ibadurrahman@ft.unmul.ac.id

ABSTRAK

Peningkatan pemanfaatan energi terbarukan oleh rumah tangga menghadirkan tantangan baru dalam memahami pola konsumsi serta faktor-faktor yang memengaruhinya. Kompleksitas tersebut menuntut adanya pendekatan analitis berbasis data untuk menghasilkan prediksi yang akurat. Penelitian ini bertujuan untuk menganalisis tren adopsi energi terbarukan rumah tangga sekaligus membangun model prediksi konsumsi energi menggunakan metode *Extreme Gradient Boosting* (XGBoost). Data penelitian berasal dari dataset *Global Renewable Energy Usage 2020–2024* yang mencakup variabel sosial-ekonomi, demografis, jenis sumber energi, hingga dukungan subsidi pemerintah. Tahapan penelitian meliputi preprocessing data (penanganan *missing value*, duplikasi, *outlier*, dan normalisasi), *feature engineering*, pemisahan data menjadi *training* dan *testing*, serta penerapan XGBoost untuk pemodelan prediksi. Evaluasi model dilakukan menggunakan metrik R^2 , MAE, dan RMSE. Hasil analisis menunjukkan bahwa rumah tangga berpendapatan menengah mendominasi adopsi energi terbarukan, dengan energi angin sebagai sumber yang paling populer, sementara energi surya menunjukkan tren peningkatan yang stabil dari tahun ke tahun. Dari sisi performa, model XGBoost berhasil mencapai nilai R^2 sebesar 0,9881, dengan MAE sebesar 29,42 dan RMSE sebesar 46,42, menandakan tingkat akurasi yang sangat tinggi serta kesalahan prediksi yang rendah. Analisis *feature importance* mengungkap bahwa penghematan biaya per kWh merupakan faktor utama yang memengaruhi konsumsi energi rumah tangga, diikuti oleh total biaya penghematan. Temuan ini menegaskan bahwa metode XGBoost mampu memberikan prediksi yang andal dan dapat dijadikan acuan dalam perencanaan serta kebijakan energi berkelanjutan.

Kata Kunci: Energi Terbarukan, Konsumsi Rumah Tangga, XGBoost, Prediksi Energi

ABSTRACT

The increasing adoption of renewable energy in households presents new challenges in understanding consumption patterns and the factors influencing them. This complexity requires data-driven analytical approaches to generate accurate predictions. This study aims to analyze trends in household renewable energy adoption and develop a predictive model using the Extreme Gradient Boosting (XGBoost) method. The research utilized the Global Renewable Energy Usage 2020–2024 dataset, which includes socio-economic and demographic variables, energy source types, and government subsidy support. The research stages involved data preprocessing (handling missing values, duplicates, outliers, and normalization), feature engineering, splitting the data into training and testing sets, and applying XGBoost for predictive modeling. Model evaluation was conducted using R^2 , MAE, and RMSE metrics. The results show that middle-income households dominate renewable energy adoption, with wind energy as the most widely used source, while solar energy demonstrates a steady upward trend over the years. In terms of performance, the XGBoost model achieved an R^2 of 0.9881, with an MAE of 29.42 and RMSE of 46.42, indicating very high accuracy and low prediction error. Feature importance analysis revealed that cost savings per kWh is the most influential factor affecting household energy consumption, followed by total cost savings. These findings confirm that XGBoost can deliver reliable predictions and provide valuable insights to support planning and policymaking for sustainable energy development.

Keyword: Renewable Energy, Household Consumption, XGBoost, Energy Prediction

1. Pendahuluan

Energi terbarukan (*Renewable Energy*) merupakan energi yang berasal dari sumber aliran energi yang alami serta berkelanjutan seperti Sinar Matahari, angin, air, biomassa, serta panas bumi (*Geothermal*). Energi ini berasal dari sumber daya alam yang dapat diperbarui (*Renewable*) secara alami dalam rentang waktu yang relatif singkat (Brandon et al., 2003). Pemanfaatan energi terbarukan menjadi salah satu fokus utama dalam upaya mengurangi ketergantungan terhadap energi fosil dan dampak perubahan iklim global. Dalam beberapa tahun terakhir, khususnya pada periode 2020 hingga 2024, tingkat adopsi energi terbarukan oleh rumah tangga di seluruh dunia mengalami peningkatan yang signifikan. Namun, seiring meningkatnya penggunaan energi terbarukan, pola konsumsi energi rumah tangga juga menjadi semakin kompleks. Berbagai faktor, seperti jenis energi terbarukan, pendapatan rumah tangga, ukuran rumah tangga, dan kebijakan subsidi, memengaruhi besarnya konsumsi energi bulanan.

Energi terbarukan memiliki potensi yang sangat besar, terutama karena keberlanjutan dan ketersediaannya. Sumber energi berbasis fosil seperti batubara dan minyak bumi yang pada saat ini telah menjadi tulang punggung dalam sumber energi telah terus menerus makin menipis, karena alasan tersebut energi fosil ini menjadi rentan terhadap fluktuasi harga di pasar global. Sebaliknya, sumber energi terbarukan seperti matahari, angin, air, biomassa, dsb tersedia secara melimpah serta dapat diperbarui (Pambudi et al., 2023).

Permasalahan utama dalam pemanfaatan energi terbarukan pada rumah tangga global adalah keragaman pola konsumsi yang dipengaruhi oleh berbagai faktor, seperti jenis energi yang digunakan, tingkat pendapatan, ukuran rumah tangga, serta kebijakan subsidi dari pemerintah. Kompleksitas ini membuat prediksi kebutuhan energi rumah tangga menjadi tidak sederhana. Meskipun tren adopsi energi terbarukan terus meningkat dari tahun 2020 hingga 2024, masih terdapat ketidakpastian terkait keberlanjutan pola konsumsi dan efektivitas faktor-faktor ekonomi maupun kebijakan dalam mendorong transisi energi. Kondisi ini menuntut adanya analisis berbasis data besar untuk memahami tren, mengidentifikasi faktor dominan, dan memprediksi pola penggunaan energi terbarukan secara lebih akurat (Abubakar et al., 2024).

Penelitian ini menggunakan metode *Extreme Gradient Boost* (XGBoost) yang banyak digunakan dalam melakukan analisis dan prediksi. *Extreme Gradient Boost* (XGBoost) merupakan algoritma machine learning yang bekerja dengan menggunakan metode ensemble learning. XGBoost akan secara bertahap membangun model prediksi dengan menambahkan model pohon keputusan (*Decision Tree*) secara berurutan. Setiap model akan mencoba memperbaiki kesalahan prediksi dari model sebelumnya. Berkat kinerja dan fleksibilitas yang tinggi, algoritma XGBoost seringkali menjadi pilihan dalam kompetisi data dan aplikasi di berbagai industri (Salsabil et al., 2024).

Penelitian ini bertujuan untuk menganalisis pola/tren konsumsi energi terbarukan di sektor rumah tangga dengan melakukan identifikasi faktor-faktor yang mempengaruhi pola tersebut dalam data yang ada. Selain itu, penelitian ini juga mengembangkan model *machine learning* menggunakan algoritma XGBoost.

Diharapkan hasil penelitian ini dapat memberikan pemahaman lebih komprehensif mengenai perkembangan adopsi energi terbarukan, khususnya dalam sektor rumah tangga, termasuk faktor-faktor yang mempengaruhi tingkat konsumsi energi, seperti jenis sumber energi, kondisi sosial-ekonomi, hingga subsidi pemerintah. Selain itu, penelitian ini juga diharapkan dapat membantu dalam perencanaan penggunaan energi melalui penerapan machine learning untuk mendukung pemecahan tantangan di masa depan (Powroźnik & Szcześniak, 2024).

2. Metode Penelitian

Penelitian ini dilakukan melalui beberapa tahapan yang meliputi proses pengumpulan, pengolahan, dan analisis data untuk membangun model prediksi konsumsi energi terbarukan rumah tangga menggunakan algoritma Extreme Gradient Boosting (XGBoost). Adapun tahapan-tahapan yang dilakukan selama penelitian adalah sebagai berikut:

A. Pengumpulan Data

Data yang digunakan dalam analisis ini berasal dari dataset yang tersedia di platform Kaggle. Dataset ini berisi informasi mengenai adopsi dan penggunaan energi terbarukan oleh rumah tangga di berbagai negara di dunia dalam kurun waktu 2020 hingga 2024. Dataset mencakup berbagai

variabel seperti sumber energi yang digunakan (tenaga surya, angin, hidro, atau panas bumi), lokasi geografis (region, negara, urban/rural), ukuran rumah tangga, tingkat pendapatan, serta adanya subsidi dari pemerintah. Selain itu, data juga mencatat konsumsi energi bulanan dalam kilowatt-jam (kWh) dan penghematan biaya yang diperoleh dari penggunaan energi terbarukan. Dataset ini dapat dimanfaatkan untuk menganalisis tren adopsi energi terbarukan di berbagai wilayah, memahami faktor-faktor sosial-ekonomi yang mempengaruhi konsumsi energi, serta mengeksplorasi peran subsidi dalam mempercepat transisi energi (Li et al., 2024). Dataset *Renewable Energy Usage 2020–2024* yang digunakan dalam studi ini memiliki total 1.000 baris data dan 12 kolom. Setiap baris merepresentasikan informasi dari satu rumah tangga terkait penggunaan energi terbarukan, sedangkan kolom-kolomnya memuat detail penting yang diantaranya:

Tabel 1. Fitur/Kolom dalam Dataset

Kolom/Fitur	Keterangan
Household_ID	Unique Identifier tiap Rumah (Household)
Region	Lokasi Geografis (Benua)
Country	Lokasi Geografis (Negara)
Energy_Source	Tipe Energi Terbarukan
Monthly_Usage_kwh	Penggunaan Energi Terbarukan bulanan (kwh)
Year	Tahun data direkam
Household Size	Jumlah penghuni Rumah (Household)
Income Level	Kategori Penghasilan Rumah (Household)
Urban_Rural	Lokasi Rumah di perkotaan atau pedesaan (Urban, Rural)
Adoption_Year	Tahun mulai adopsi energi terbarukan
Subsidy_Received	Apakah menerima subsidi pemerintah
Cost_Savings_USD	Dana yang dihemat dari penggunaan energi terbarukan dalam USD

Ukuran dan cakupan data ini memungkinkan dilakukan analisis tren adopsi energi terbarukan berdasarkan wilayah, jenis energi, faktor ekonomi, dan kondisi demografis rumah tangga.

B. Preprocessing Data

a. Handling Missing Value

Tahapan awal yang dilakukan pada Preprocessing data yaitu melakukan *handling missing value*. *Missing Value* dapat terjadi apabila terjadi kesalahan penginputan pada saat pengumpulan data atau data tersebut ada/terekam sejak awal pengumpulan. Keberadaan *missing value* dapat mengganggu dalam kinerja dan hasil pada saat pemrosesan data sehingga diperlukan penanganan pada *missing value* (Popoola et al., 2024).

b. Handling Duplicated Data

Pada Tahapan ini pada *preprocessing data* dilakukan penanganan terhadap data yang sama dengan data lainnya (duplikasi). Data duplikasi ini mengacu pada salinan yang identik maupun hampir identik dari data yang sama yang tersimpan di berbagai lokasi di dalam kumpulan data (dataset). Hal ini biasanya terjadi karena berbagai hal seperti kesalahan saat input data, terjadi gangguan sistem, atau integrasi data dari berbagai sumber. Duplikasi data dapat menyebabkan inkonsistensi data, inefisiensi operasional yang dapat mempengaruhi kualitas analisis sehingga

perlu ditangani (dihapus/diisi dengan nilai tertentu) agar data yang diproses berkualitas, tidak berulang (Hasibuan et al., 2022).

c. *Handling Outliers*

Tahapan ini dilakukan dengan melakukan penanganan terhadap Data *Outliers* yang ada di dataset. Data *Outliers* adalah data yang memiliki nilai data yang memiliki perbedaan drastis dari data lainnya atau data memiliki nilai yang sangat berbeda data pola tipikal data yang ada (Al Samara et al., 2022). Data *Outliers* dapat menyebabkan penurunan kinerja pada model *machine learning* atau mempengaruhi hasil akhir dari model algoritma machine learning.

d. *Handling Inconsistent Value/noise*

Pada tahap ini, difokuskan untuk mengidentifikasi adanya nilai-nilai yang tidak konsisten, tidak sesuai format, atau mengandung *noise* dalam dataset. *Inconsistent values* biasanya muncul dalam bentuk kesalahan penulisan (*typo*), perbedaan format antar entri, penggunaan kapitalisasi yang berbeda, atau nilai kategori yang seharusnya sama namun ditulis berbeda. Tahapan ini penting dilakukan untuk memastikan bahwa setiap kategori memiliki representasi yang seragam, sehingga tidak mengganggu proses analisis, agregasi, atau pemodelan di tahap selanjutnya. Tahapan ini juga akan membantu dalam proses *data cleaning*, seperti normalisasi teks atau penggabungan kategori yang redundan.

e. *Normalisasi Data*

Pada tahap ini dilakukan normalisasi pada data. Normalisasi data adalah proses mengubah data ke agar berada dalam skala yang seragam (skala 0 hingga 1). Hal ini bertujuan untuk memastikan semua fitur/kolom memiliki kontribusi yang seimbang dan tidak ada yang mendominasi dalam proses analisis, hal ini terutama pada algoritma *machine learning* yang sensitif terhadap perbedaan skala. Pada penelitian kali ini *MinMax Scaling* digunakan dalam melakukan normalisasi data. Teknik *scaling* ini dilakukan dengan merubah rentang nilai data menjadi antara 0 dan 1 (Izonin et al., 2022).

$$x' = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (1)$$

Keterangan:

- a) x_i = nilai tertentu yang akan dinormalisasi
- b) x' = nilai hasil normalisasi
- c) $\min(x)$ = nilai minimal dari suatu atribut
- d) $\max(x)$ = nilai maksimal dari suatu atribut

f. *Feature Engineering*

Pada tahap ini, dilakukan proses *Feature Engineering*, yaitu proses membuat fitur-fitur baru dari data yang sudah ada untuk memperkaya analisis dan mendapatkan wawasan yang lebih mendalam. Tujuan dari tahap ini adalah mengekstraksi informasi tambahan yang tidak secara langsung tersedia di dataset awal, namun dapat membantu menjelaskan pola, tren, dan hubungan antar variabel. Dalam konteks analisis tren adopsi energi terbarukan pada rumah tangga global, dilakukan penambahan dua fitur baru. Fitur pertama adalah (*SavingsperkWh*), yang menghitung efisiensi penghematan biaya per kWh energi terbarukan yang digunakan, sehingga dapat diketahui rumah tangga mana yang paling hemat biaya relatif terhadap energi yang mereka konsumsi. Fitur kedua adalah (*UsageCategory*), yang mengelompokkan rumah tangga ke dalam tiga kategori konsumsi energi, yaitu *Low*, *Medium*, dan *High*. Pengelompokan ini memudahkan analisis pola konsumsi energi terbarukan di berbagai tingkat pemakaian dan membantu menemukan hubungan antara tingkat konsumsi, pendapatan, dan potensi penghematan. Dengan penambahan fitur-fitur ini, analisis selanjutnya dapat dilakukan dengan lebih komprehensif karena informasi yang diperoleh menjadi lebih detail dan tersegmentasi.

C. Analisis Tren

Pada tahapan ini, analisis tren/pola dilakukan dengan menggunakan visualisasi data seperti grafik dan diagram sebagai alat bantu dalam menampilkan pola dan karakteristik yang terdapat pada data agar

lebih jelas dan mudah dipahami. Hal ini dilakukan untuk mempermudah dalam menganalisa demi mendapatkan informasi mengenai pola – pola seperti perkembangan penggunaan energi terbarukan, perbandingan yang terjadi antara rumah tangga, serta faktor – faktor yang menjadi alasan dari hal – hal tersebut.

D. Modelling

Pada tahap ini algoritma XGBoost digunakan dalam melakukan prediksi konsumsi Energi terbarukan dalam rumah tangga secara global. Hal – hal yang dilakukan pada tahap ini yaitu:

a. *Label Encoding*

Sebelum melakukan modelling dilakukan proses *Label Encoding*. Proses ini dilakukan untuk mengubah struktur data agar sesuai dan dapat diproses oleh algoritma *machine learning*. *Label encoding* merupakan teknik mengubah data kategorikal numerik representatif yang dapat dimengerti komputer sehingga data dapat diproses oleh algoritma *machine learning* dengan baik, Misalnya “*low*” > 0, “*medium*” > 1 dan “*high*” > 2 (Kumar & Bhardwaj, 2025).

b. *Splitting Data*

Pada tahap ini dilakukan pembagian dataset menjadi *Data Training* dan *Data Testing*. Dataset dibagi dengan Rasio 8:2 yang dimana 80% dari Dataset digunakan sebagai *Data Training* yang berfungsi untuk melatih model *machine learning* untuk melakukan pengenalan pola dan karakteristik yang berada di Dataset. Sedangkan 20% sisanya akan menjadi *Data Testing* yang digunakan dalam proses pengujian model yang telah dilatih untuk melihat dan mengevaluasi kinerja dari model *machine learning* yang digunakan (Nguyen et al., 2021).

c. *Algoritma XGBoost*

Algoritma XGBoost bekerja dengan menggunakan *Decision Tree* secara bertahap. Model *Decision Tree* akan dibuat secara berulang yang dimana ‘pohon’ yang baru akan berusaha memperbaiki kesalahan yang dilakukan oleh ‘pohon’ sebelumnya. Algoritma ini memiliki sistem optimalisasi dan regularisasi yang membantu meminimalisir kesalahan dalam prediksi dan juga mencegah terjadinya *overfitting*. Pada Tahap ini, model XGBoost dilatih menggunakan data training hasil dari splitting data sebelumnya dengan menggunakan parameter :

Tabel 2. Parameter Modelling XGBoost

Parameter	Nilai
max_depth	[3, 6]
learning_rate	[0.01, 0.1]
n_estimators	[100, 200]
subsample	[0.8, 1.0]

Keterangan :

- 1) max_depth : kedalaman dari ‘pohon’ Decision tree
- 2) learning_rate : kecepatan dalam proses pembelajaran model
- 3) n_estimators : jumlah ‘pohon’ yang dihasilkan oleh model
- 4) subsample : data yang digunakan untuk melatih model

E. Evaluasi Model

Tahapan ini dilakukan untuk mengukur kinerja/performa dari model *machine learning* yang digunakan. Pada Penelitian ini ada 3 metrik evaluasi yang digunakan dalam melakukan Evaluasi model yaitu diantaranya :

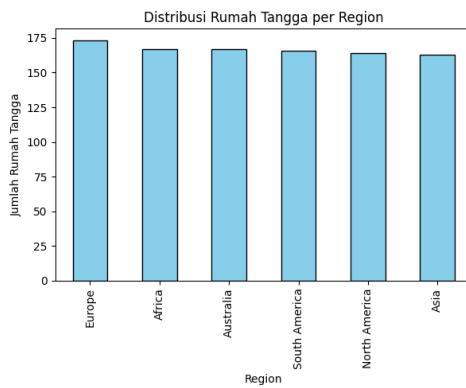
- a. *R²* Score menunjukkan seberapa baik model dapat menjelaskan variasi data. Nilai *R²* berada dalam rentang $0 > R^2 > 1$ yang menandakan lebih dekat nilai *R²* score terhadap 1 maka model dapat menjelaskan pola dari data dengan lebih baik. Contohnya, jika nilai *R²* score adalah 0,90 berarti model dapat menjelaskan 90% variasi data yang ada (Chaabani & Alfadl, 2024).
- b. *Mean Absolute Error* (MAE) menunjukkan rata – rata selisih absolut antara hasil prediksi dengan data aktualnya. Contohnya, pada suatu model memiliki nilai MAE = 5 artinya model tersebut memiliki rata – rata perbedaan nilai prediksi 5 dari nilai aktualnya tanpa melihat apakah perbedaan nilai tersebut positif atau negatif (Atalan et al., 2025).

- c. *Root Mean Squared Error* (RMSE) mirip dengan MAE namun sebelum dicari rata – ratanya, selisih nilai prediksi dari aktual di kuadratkan terlebih dahulu lalu diakarkan. Hal ini membuat RMSE memiliki nilai yang lebih besar untuk setiap kesalahan dalam prediksi data. Metrik ini biasa digunakan pada kasus dimana kesalahan prediksi merupakan hal yang serius dan memberikan penekanan yang lebih berat terhadap kesalahan dibandingkan oleh MAE (Hodson, 2022).

3. Hasil dan Pembahasan

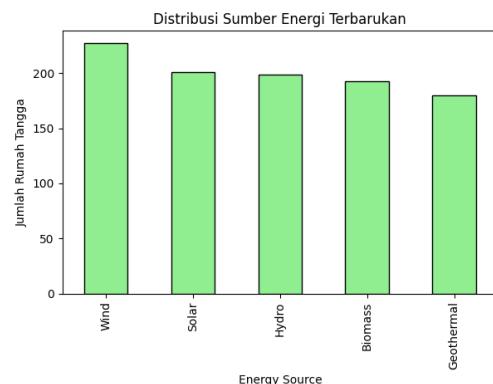
A. Analisis Tren

Melalui Visualisasi data, kita dapat melihat secara lebih jelas untuk melihat struktur dan pola tren yang terjadi, hal ini terkait pada perilaku atau kecenderungan yang terjadi dalam penggunaan sumber energi terbarukan dalam sektor rumah tangga. Dengan grafik dan diagram yang ditampilkan dalam visualisasi kita dapat melakukan analisis untuk memahami informasi mengenai pola atau tren yang ada dalam data. Hasil analisis ini nantinya dapat digunakan untuk mendorong pemanfaatan energi terbarukan dengan lebih optimal.



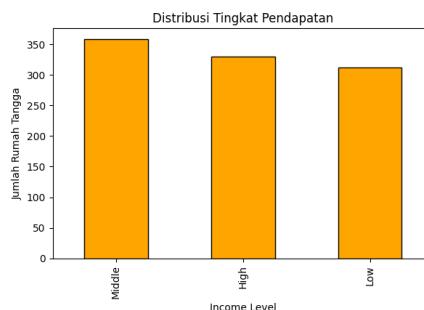
Gambar 1. Diagram Distribusi Rumah Tangga tiap Region

Diagram pada Gambar 1 di atas menunjukkan distribusi rumah tangga per region yang mengadopsi energi terbarukan. Dari hasil visualisasi, dapat diketahui bahwa Eropa memiliki jumlah rumah tangga terbanyak yang menggunakan energi terbarukan, menggambarkan bahwa wilayah ini sudah lebih maju dalam penerapan sumber energi alternatif yang ramah lingkungan. Posisi berikutnya ditempati oleh Afrika dan Asia, yang juga menunjukkan tingkat adopsi cukup tinggi meskipun masih dipengaruhi oleh faktor ketersediaan infrastruktur dan kebijakan energi di masing-masing negara. Sementara itu, Amerika Utara dan Amerika Selatan berada pada kategori menengah, dengan tren adopsi yang stabil namun tidak sebesar tiga region sebelumnya. Adapun Australia tercatat sebagai wilayah dengan jumlah adopsi energi terbarukan paling rendah dibandingkan region lainnya, kemungkinan dipengaruhi oleh jumlah populasi rumah tangga yang lebih sedikit serta kondisi geografis. Visualisasi ini membantu kita memahami adanya perbedaan tingkat adopsi energi terbarukan antarwilayah serta mengidentifikasi potensi pengembangan yang dapat difokuskan pada setiap region yang ada dengan tingkat adopsi yang masih rendah namun memiliki peluang besar untuk ditingkatkan melalui dukungan kebijakan dan inovasi teknologi.



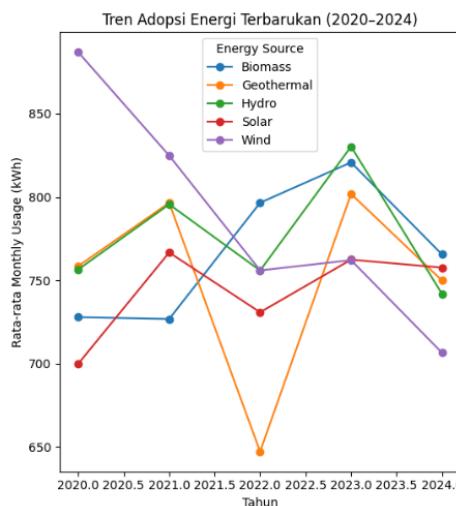
Gambar 2. Diagram Distribusi sumber Energi Terbarukan

Selanjutnya, Gambar 2 memperlihatkan distribusi rumah tangga berdasarkan sumber energi terbarukan yang digunakan. Dari hasil analisis, terlihat bahwa energi angin (*Wind*) menjadi pilihan paling populer dan mendominasi penggunaan di sebagian besar rumah tangga. *Solar* dan *Hydro* berada di posisi berikutnya sebagai alternatif energi yang cukup banyak dimanfaatkan. Sementara itu, *Biomass* dan *Geothermal* memiliki jumlah pengguna yang lebih sedikit, namun tetap berperan penting sebagai opsi pendukung dalam bauran energi terbarukan. Hasil ini memberikan gambaran mengenai preferensi rumah tangga terhadap jenis energi yang digunakan.



Gambar 3. Diagram Distribusi Tingkat Pendapatan Rumah Tangga

Gambar 3 menampilkan distribusi rumah tangga berdasarkan tingkat pendapatan. Hasil visualisasi menunjukkan bahwa rumah tangga dengan pendapatan menengah (*Middle*) menjadi kelompok yang paling banyak mengadopsi energi terbarukan, diikuti oleh kategori pendapatan tinggi (*High*). Sementara itu, rumah tangga dengan pendapatan rendah (*Low*) tercatat sebagai kelompok dengan tingkat adopsi paling rendah.



Gambar 4. Tren Adopsi tiap Tipe Sumber Energi Terbarukan 2020 - 2024

Gambar 4 menampilkan tren adopsi energi terbarukan berdasarkan rata-rata penggunaan energi bulanan (*Monthly Usage kWh*) dari tahun 2020 hingga 2024 untuk setiap jenis sumber energi. Dari hasil visualisasi, terlihat bahwa energi angin (*Wind*) mendominasi pada awal periode dengan tingkat penggunaan tertinggi pada tahun 2020, namun mengalami penurunan signifikan hingga 2024. Sebaliknya, energi biomassa (*Biomass*) dan energi hidro (*Hydro*) menunjukkan tren yang relatif stabil, bahkan mengalami sedikit peningkatan pada tahun 2022–2023 sebelum menurun kembali di 2024. Untuk energi surya (*Solar*), tingkat penggunaannya meningkat secara bertahap sejak 2020 dan mencapai kestabilan pada 2023–2024. Sementara itu, energi panas bumi (*Geothermal*) memperlihatkan fluktuasi paling tajam, dengan penurunan drastis pada 2022 sebelum kembali naik pada 2023 dan stabil menjelang 2024.

Dari tren keseluruhan, dapat disimpulkan bahwa preferensi rumah tangga terhadap jenis energi terbarukan terus mengalami perubahan dari tahun ke tahun. Faktor-faktor seperti biaya, ketersediaan infrastruktur, dan kebijakan subsidi kemungkinan berperan besar dalam memengaruhi pergeseran pola konsumsi energi tersebut.

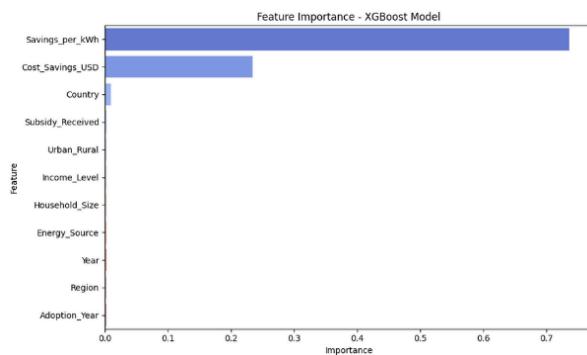
B. Hasil Prediksi

Pada tahap ini, evaluasi model dilakukan menggunakan metrik *R² Score*, *Mean Absolute Error* (MAE), dan *Root Mean Squared Error* (RMSE) untuk menilai akurasi dan kesalahan prediksi model.

Tabel 3. Hasil Evaluasi Model

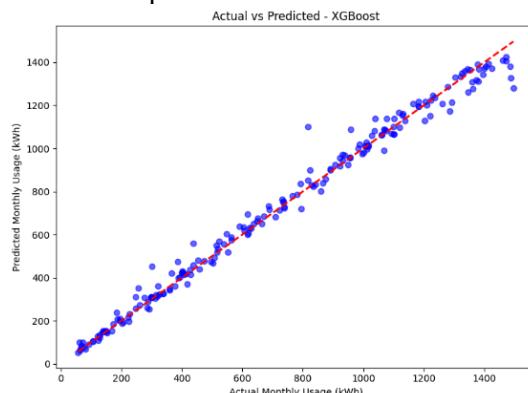
Metrik	Nilai
R ² Score	0,9881
Mean Absolute Error	29,4154
Mean Squared Error	46,4214

Hasil evaluasi pada Tabel 3 menunjukkan bahwa model XGBoost memperoleh nilai $R^2 = 0.9782$, dengan $MAE = 39.49$ dan $RMSE = 62.67$. Setelah melakukan evaluasi, langkah selanjutnya adalah melakukan visualisasi hasil prediksi. Visualisasi hasil prediksi menggunakan metode *Extreme Gradient Boosting* (XGBoost) bertujuan untuk memberikan pemahaman yang lebih mendalam mengenai kinerja model dalam melakukan prediksi. Melalui analisis *Feature Importance*, dapat diketahui variabel mana yang paling berpengaruh dalam menentukan hasil prediksi. Sementara itu, grafik *Actual vs Predicted* membantu dalam mengevaluasi sejauh mana hasil prediksi mendekati nilai aktual, sehingga dapat menilai akurasi serta keandalan model.



Gambar 5. Visualisasi Feature Importance pada Model XGBoost

Gambar 5 menunjukkan bahwa fitur (SavingsperWh) memiliki pengaruh paling besar dengan tingkat kepentingan mencapai lebih dari 70%, artinya penghematan biaya per kWh menjadi faktor utama yang memengaruhi keputusan rumah tangga untuk beralih ke energi terbarukan. Fitur berikutnya yang juga memiliki peran signifikan adalah (CostSavingsUSD) dengan tingkat kepentingan sekitar 23%, menunjukkan bahwa besarnya total penghematan biaya secara keseluruhan juga menjadi pertimbangan penting dalam proses adopsi. Sedangkan fitur lainnya, seperti Country, (SubsidyReceived), (UrbanRural), (IncomeLevel), dan faktor demografis lainnya, memberikan kontribusi yang jauh lebih kecil terhadap model.



Gambar 6. Grafik Aktual Dibandingkan dengan Prediksi pada Model XGBoost

Grafik Aktual dibandingkan dengan Prediksi pada Gambar 6 menggambarkan kinerja model XGBoost dalam memprediksi penggunaan energi bulanan rumah tangga (kWh). Sumbu X merepresentasikan nilai aktual, sementara sumbu Y menunjukkan nilai prediksi yang dihasilkan model. Garis merah putus-putus menggambarkan garis ideal (*perfect prediction*), yaitu kondisi di mana prediksi model sepenuhnya sama dengan nilai aktual. Sebagian besar titik biru pada grafik berada sangat dekat dengan garis merah, menandakan bahwa prediksi model hampir mendekati kondisi ideal. Hal ini mengindikasikan bahwa model XGBoost memiliki tingkat akurasi yang tinggi dalam memprediksi konsumsi energi rumah tangga.

4. Kesimpulan

Berdasarkan hasil penelitian, dapat disimpulkan bahwa rumah tangga berpendapatan menengah merupakan kelompok yang paling dominan dalam mengadopsi energi terbarukan, dengan energi angin menjadi sumber yang paling banyak digunakan. Sementara itu, energi surya menunjukkan tren peningkatan yang stabil dari tahun 2020 hingga 2024. Faktor utama yang memengaruhi konsumsi energi terbarukan adalah penghematan biaya per kWh, diikuti oleh total biaya penghematan secara keseluruhan. Dari sisi performa model, Extreme Gradient Boosting (XGBoost) menghasilkan nilai R^2 sebesar 0,9881 dengan tingkat kesalahan prediksi MAE sebesar 29,4154 dan RMSE sebesar 46,4214 menunjukkan kinerja yang lebih baik sehingga prediksi lebih mendekati kondisi ideal. Dengan performa akurasi yang tinggi dan kesalahan prediksi yang relatif rendah, metode XGBoost terbukti unggul dalam memodelkan

pola konsumsi energi terbarukan rumah tangga. Temuan ini dapat menjadi dasar bagi perencanaan kebijakan energi berkelanjutan dan strategi peningkatan adopsi energi terbarukan di masa depan.

5. Daftar Pustaka

- Al Samara, M., Bennis, I., Abouaissa, A., & Lorenz, P. (2022). A Survey of Outlier Detection Techniques in IoT: Review and Classification. *Journal of Sensor and Actuator Networks*, 11(1). <https://doi.org/10.3390/jsan11010004>
- Brandon, P. S., Lombardi, P. L., & Bentivegna, V. (2003). Evaluation of the built environment for sustainability. In *Evaluation of the Built Environment for Sustainability*. <https://doi.org/10.4324/9780203362426>
- Hasibuan, E., Informasi, S., Ilmu, F., Informasi, T., Gunadarma, U., Margonda, J., No, R., Cina, P., & Jawa, D. (2022). Implementasi Machine Learning untuk Prediksi Harga Mobil Bekas dengan Algoritma Regresi Linear berbasis Web. *Jurnal Ilmiah Komputasi*, 21(4), 595–602. <https://doi.org/10.32409/jikstik.21.4.3327>
- Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. *Geoscientific Model Development*, 15(14), 5481–5487. <https://doi.org/10.5194/gmd-15-5481-2022>
- Izonin, I., Tkachenko, R., Shakhovska, N., Ilchyshyn, B., & Singh, K. K. (2022). A Two-Step Data Normalization Approach for Improving Classification Accuracy in the Medical Diagnosis Domain. *Mathematics*, 10(11), 1–18. <https://doi.org/10.3390/math10111942>
- Kumar, M., & Bhardwaj, V. (2025). Evaluating Label Encoding and Preprocessing Techniques for Breast Cancer Prediction Using Machine Learning Algorithms. In *International Journal of Computational Intelligence Systems* (Vol. 18, Issue 1). Springer Netherlands. <https://doi.org/10.1007/s44196-025-00957-7>
- Pambudi, N. A., Firdaus, R. A., Rizkiana, R., Ulfa, D. K., Salsabila, M. S., Suharno, & Sukatiman. (2023). Renewable Energy in Indonesia: Current Status, Potential, and Future Development. *Sustainability (Switzerland)*, 15(3). <https://doi.org/10.3390/su15032342>
- Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random Forest Algorithm Overview. *Babylonian Journal of Machine Learning*, 2024, 69–79. <https://doi.org/10.58496/bjml/2024/007>
- Salsabil, M., Lutvi, N., & Eviyanti, A. (2024). Implementasi Data Mining Dalam Melakukan Prediksi Penyakit Diabetes Menggunakan Metode Random Forest Dan Xgboost. *Jurnal Ilmiah Komputasi*, 23(1), 51–58. <https://doi.org/10.32409/jikstik.23.1.3507>
- Abubakar, I. R., Alola, A. A., Bekun, F. V., & Onifade, S. T. (2024). Investigating the determinants of household energy consumption in Nigeria: insights and implications. *Energy Sustainability and Society*, 14(1). <https://doi.org/10.1186/s13705-024-00451-6>
- Powroźnik, P., & Szcześniak, P. (2024). Predictive Analytics for energy efficiency: Leveraging machine learning to optimize household energy consumption. *Energies*, 17(23), 5866. <https://doi.org/10.3390/en17235866>
- Li, W., Zhu, J., Li, Y., Li, Y., & Ding, Z. (2024). Determinants of Solar Photovoltaic Adoption Intention among Households: A Meta-Analysis. *Sustainability*, 16(18), 8204. <https://doi.org/10.3390/su16188204>
- Atalan, Y. A., Şahin, H., Keskin, A., & Atalan, A. (2025). Strategic forecasting of renewable energy production for sustainable electricity supply: A machine learning approach considering environmental, economic, and oil factors in Türkiye. *PLoS ONE*, 20(8), e0328290. <https://doi.org/10.1371/journal.pone.0328290>
- Chaabani, A. K., & Alfadl, N. (2024). A comparative study of machine learning approaches for an accurate predictive modeling of solar energy generation. *Energy Reports*, 12, 1293–1302. <https://doi.org/10.1016/j.egyr.2024.07.010>
- Nguyen, Q. H., Ly, H., Ho, L. S., Al-Ansari, N., Van Le, H., Tran, V. Q., Prakash, I., & Pham, B. T. (2021). Influence of data splitting on performance of machine learning models in prediction of shear strength of soil. *Mathematical Problems in Engineering*, 2021, 1–15. <https://doi.org/10.1155/2021/4832864>
- Popoola, P. A., Tapamo, J., & Assounga, A. G. H. (2024). Effective and efficient handling of missing data in supervised machine learning. *Data Science and Management*. <https://doi.org/10.1016/j.dsm.2024.12.002>