



Rancang Bangun dan Evaluasi Sistem Pipeline Asinkron untuk Pengelolaan dan Analitik Data Meteorologi dan Panel Surya

Ibrahim Frosly Alesandro, Alisha Deana Tabina, Muhammad Raihan Fasya,
Atik Wintarti, Ibnu Febry Kurniawan

Departemen Sains Data, Fakultas Matematika dan Ilmu Pengetahuan, Universitas Negeri Surabaya
{ibrahim.23021,alishadeana.23244,muhammadraihan.23037}@mhs.unesa.ac.id, atikwintarti@unesa.ac.id
Corresponding author email: ibnufebry@unesa.ac.id

Abstract: This research addresses the challenge of processing large volumes of meteorological and solar radiation data for photovoltaic (PV) systems through cost-effective Extract-Transform-Load (ETL) workflows. While scalable data processing infrastructure is often inaccessible to institutions with budget constraints, this study proposes a solution using affordable asynchronous processing technologies. Two distinct ETL workflows are designed: one for raw data ingestion and transformation through systematic feature engineering, and another for applying machine learning algorithms for PV electricity production regression and prediction. The system utilizes RabbitMQ and Celery for sophisticated job scheduling, accommodating varying task specifications and costs. The XGBoost model demonstrated the best performance in predicting solar panel output, with an RMSE of 0.0289 and MAE of 0.0200. K-Means Clustering grouped weather, cloud, and temperature conditions into several clusters, optimized through Silhouette Score evaluation. These clusters were categorized into "Clear," "Cloudy," and "Rainy/Overcast" weather conditions, facilitating more adaptive decision-making. The study develops a proof-of-concept ETL system using open-source technologies, offering scalable and cost-effective solutions for processing meteorological data. This research shows that the system supports data-driven decision-making and provides a solid foundation for the development of sustainable solar energy applications in Indonesia, while being a practical and flexible solution for the national green energy transition.

Keywords: data pipeline, data telemetry, data interpretation, distributed processing, photovoltaic estimates

Abstrak: Penelitian ini menghadirkan solusi terjangkau dengan mengembangkan sistem Data Pipeline ETL asinkron yang mampu memproses data cuaca dan radiasi matahari untuk menganalisis serta memprediksi kinerja panel surya secara *real-time*. Sistem ini menggunakan teknologi *open-source* seperti RabbitMQ, Celery, dan PostgreSQL yang hemat biaya dan mudah diterapkan. Pipeline ini diuji sebagai *proof-of-concept* dan berhasil memproses sekitar 35.000 entri data *timeseries* secara modular dan paralel, mendukung pemrosesan data besar dengan kecepatan dan akurasi tinggi. Model XGBoost yang digunakan menunjukkan performa terbaik dalam prediksi kinerja panel surya, dengan nilai RMSE 0.0289 dan MAE 0.0200. *Clustering* menggunakan K-Means berhasil mengelompokkan kondisi cuaca, awan, dan suhu ke dalam beberapa kluster utama, dengan optimalisasi jumlah cluster berdasarkan evaluasi *Silhouette Score*. Hasil klasterisasi ini terbagi menjadi kategori utama seperti "Cerah", "Berawan", dan "Hujan/Mendung", yang memungkinkan pengambilan keputusan berbasis kondisi lingkungan yang lebih tepat dan adaptif. Pipeline ini dirancang sebagai solusi *scalable*, memungkinkan ekspansi sistem dan integrasi data IoT di masa depan. Penelitian ini menunjukkan bahwa sistem ini mendukung pengambilan keputusan berbasis data dan memberikan dasar kuat untuk pengembangan aplikasi energi surya berkelanjutan di Indonesia, sekaligus menjadi solusi praktis dan fleksibel untuk transisi energi hijau nasional.

Kata kunci: data pipeline, data telemetry, data interpretation, distributed processing, photovoltaic estimates

I. PENDAHULUAN

Indonesia sebagai negara tropis memiliki potensi energi surya sebesar 207,8 GWp, namun realisasinya hanya mencapai 0,15 GWp pada tahun 2020, sedangkan target untuk 2025 sebesar 0,87 GWp [1]. Kontribusi pembangkit tenaga surya *photovoltaic* (PLTS) yang terhitung rendah pada tahun 2023, yakni hanya sekitar 0,05%, meskipun Indonesia memiliki intensitas radiasi matahari yang cukup tinggi, yakni 4,8—5,1 kWh/m²/hari, yang seharusnya berpotensi untuk pengembangan energi terbarukan [2]. Tantangan yang dihadapi adalah ketidakpastian kondisi cuaca sehingga menyebabkan fluktuasi intensitas radiasi matahari, serta pemanfaatannya yang masih rendah pada daerah dengan potensi tinggi, yang menghambat pencapaian SDG 7 tentang penggunaan energi terbarukan dan SDG 13 mengenai mitigasi perubahan iklim [3] [4]. Pengoptimalan energi surya memiliki potensi besar untuk mendukung pertumbuhan ekonomi hijau dan meningkatkan ketahanan energi nasional dalam



menghadapi perubahan iklim [5]. Menurut Kementerian Perekonomian, pengembangan sektor energi terbarukan dapat meningkatkan PDB Indonesia rata-rata sebesar 6,3% pada periode 2025-2045 dan menciptakan 1,7 juta lapangan kerja hijau pada 2045. Namun, meskipun potensi tersebut sangat besar, masih terdapat kekurangan inovasi dalam pengelolaan data meteorologi dan radiasi matahari yang dapat meningkatkan efisiensi sistem energi surya.

Berbagai inovasi telah dikembangkan di berbagai negara untuk mendukung optimalisasi energi surya. Abbas et al. dengan membangun sistem *pipeline* paralel untuk memproses data meteorologi secara *real-time* dan *batch* data satelit, namun terkendala keterbatasan cakupan dan resolusi data di area terpencil serta kebutuhan infrastruktur data yang stabil [6]. Di Indonesia, Wahyudi et al. mengembangkan *pipeline* otomatis dengan *Apache Airflow* untuk mengakuisisi data prediksi curah hujan secara periodik dari BMKG, namun *pipeline* ini sangat bergantung pada ketersediaan server BMKG dan belum mengintegrasikan analitik lanjutan berbasis *machine learning* [7]. Nadeem et al. menunjukkan peningkatan akurasi prediksi energi surya melalui *deep learning* berbasis citra satelit dan *sky camera* dalam *pipeline*, tetapi pendekatan ini bergantung pada kualitas input dan sumber daya komputasi tinggi [8]. Berbagai inovasi tersebut menunjukkan bahwa sistem prediksi yang lebih adaptif, efisien, dan terdesentralisasi, terutama untuk menghadapi tantangan volume dan variasi data panel surya dan cuaca diperlukan.

Sistem yang dirancang dalam penelitian ini menggunakan teknologi terjangkau dan dapat diimplementasikan secara berkelanjutan (*scalable*) sesuai ketersediaan dana dan kebutuhan *data processing*. Penelitian ini berperan sebagai *proof of concept* untuk mengevaluasi kelayakan dan efektivitas pendekatan sistem *asynchronous data pipeline* yang diusulkan. Sistem ini dirancang untuk menganalisis dan memprediksi kinerja panel surya berdasarkan data meteorologi dan radiasi matahari, dengan kemampuan menangani aliran data dalam skala besar secara paralel, termasuk data waktu nyata (*real-time*) yang berasal dari sensor IoT dan sumber data cuaca. Pendekatan *pipeline* yang bersifat asinkron memungkinkan percepatan proses akuisisi dan pemrosesan data, sekaligus meningkatkan ketepatan estimasi keluaran energi dari panel surya. Dalam jangka panjang, sistem ini diharapkan dapat memperkuat ekosistem energi terbarukan nasional dengan meningkatkan efisiensi operasional, khususnya bagi pembangkit skala kecil seperti rumah tangga dan UMKM. Selain itu, ketersediaan data yang terstruktur dan dapat diakses secara kontinu memungkinkan penerapan kebijakan energi berbasis data (*evidence-based policy*), terutama di tingkat lokal dan regional. Dengan demikian, sistem ini tidak hanya mendukung pengambilan keputusan teknis dan operasional secara cepat dan efisien, tetapi juga memberikan kontribusi terhadap peningkatan kinerja pembangkit listrik tenaga surya (PLTS) di Indonesia serta mendorong akselerasi transisi energi menuju keberlanjutan.

II. METODE PENELITIAN

2.1. Jenis penelitian

Penelitian ini merupakan penelitian rekayasa sistem (*engineering research*) dengan pendekatan kuantitatif dan eksperimental. Fokus utama dari penelitian ini adalah mengembangkan dan mengimplementasikan sistem *asynchronous data pipeline* untuk memproses data cuaca dan radiasi matahari dalam rangka menganalisis serta memprediksi kinerja panel surya. Proses rekayasa sistem dilakukan secara berkala dengan tujuan untuk menghasilkan sistem yang efisien, terukur, dan bisa diterapkan dalam kejadian nyata.



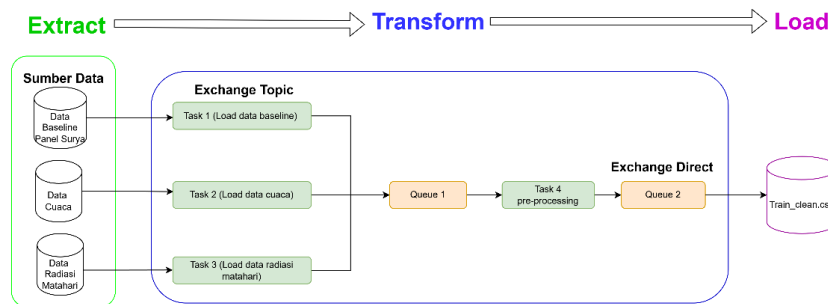
2.2. Data dan Sumber Data

Penelitian ini menggunakan data deret waktu (*timeseries*) yang diperoleh dari situs publik Solcast. Terdapat tiga jenis dataset utama yang digunakan: (1) Dataset kinerja panel surya, yang memuat informasi waktu dan persentase kinerja panel pada setiap periode pengamatan; (2) Dataset cuaca, yang mencakup data meteorologis per jam seperti suhu maksimum dan minimum, kelembaban, kecepatan angin, curah hujan, serta durasi penyinaran matahari; dan (3) Dataset radiasi matahari, yang berisi parameter seperti *Diffuse Horizontal Irradiance* (DHI), *Direct Normal Irradiance* (DNI), *Global Horizontal Irradiance* (GHI), serta informasi tambahan seperti jenis awan dan sudut zenith matahari.

2.3. Pengembangan Sistem

Sistem dikembangkan secara modular dan *asynchronous* dengan pendekatan ETL (*Extract, Transform, Load*) yang terdiri dari tiga tahap utama:

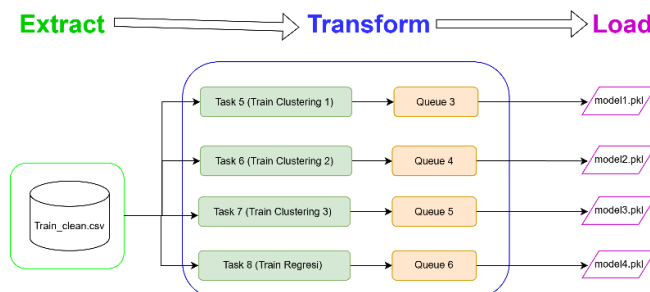
2.3.1. ETL1 : Pre-Processing dan merging data



Gambar 1. Alur ETL 1

- *Extract*: data mentah dalam format CSV dibaca secara bertahap dan dikirim ke sistem antrian RabbitMQ untuk diproses lebih lanjut.
- *Transform*: Penggantian format waktu, penggabungan data, *encoding* fitur kategorikal, penanganan *missing value*, penyesuaian tipe data dan inisiasi indeks.
- *Load*: Menyimpan hasil transformasi ke *buffer* untuk dilanjutkan ke tahap ETL 2.

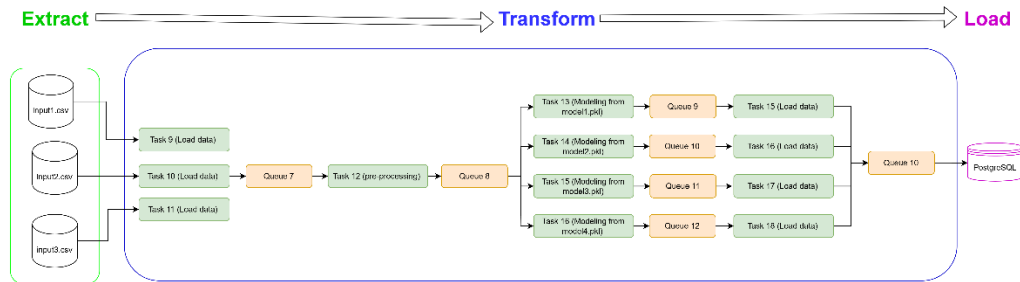
2.3.2. ETL 2 – Train Model



Gambar 2. Alur ETL 2

- *Extract*: data yang telah bersih pada ETL 1 dibaca dan dikirim ke system RabbitMQ.
- *Transform*: *Feature Engineering*, pelatihan model regresi, penerapan teknik *clustering*.
- *Load*: Simpan model terlatih kedalam format .pkl untuk dilanjutkan ke tahap selanjutnya.

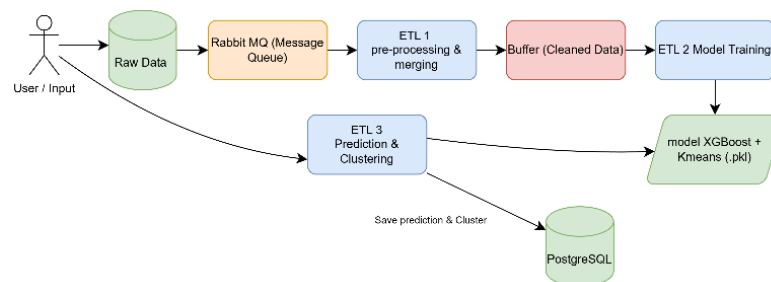
2.3.3. ETL 3 – Test Model



Gambar 2. Alur ETL 3

- *Extract*: baca data input satu persatu lalu dikirim ke server RabbitMQ.
- *Transform*: *Data Pre-Processing*: *timestamp to posix* dan Melakukan inferensi menggunakan model yang telah dilatih di ETL 2.
- *Load*: menyimpan hasil *cluster* dan prediksi ke database PostgreSQL

Untuk memperjelas alur sistem yang dibangun, berikut ditampilkan diagram arsitektur ETL secara keseluruhan yang memvisualisasikan bagaimana data diproses mulai dari data mentah hingga hasil akhir disimpan di basis data.



Gambar 4. Arsitektur sistem ETL *asynchronous* untuk prediksi kinerja panel surya

Pada Gambar 4, ditunjukkan alur dimulai dari pembacaan tiga data input yang kemudian dikirim melalui RabbitMQ sebagai *message queue*. Proses ETL 1 melakukan pembersihan dan penggabungan data sebelum disimpan sementara di *buffer*. Tahap berikutnya, ETL 2 memproses data bersih untuk pelatihan model menggunakan XGBoost dan segmentasi menggunakan K-Means. Model yang telah dilatih disimpan dalam format *.pkl*. Kemudian, pada ETL 3, sistem menerima data input berupa waktu (*timestamp*), memuat model terlatih, dan menghasilkan *output* prediksi serta hasil *clustering* yang akhirnya disimpan dalam PostgreSQL.

2.4. Komponen teknologi

- RabbitMQ: Digunakan sebagai *message broker* untuk mengirim data mentah dalam format CSV ke berbagai antrian (*queue*), seperti petugas pos yang mengirim data mentah ke jalur pemrosesan yang tepat.
- Celery: Merupakan *job scheduler* yang mengatur proses kerja secara paralel dan tidak harus menunggu satu sama lain selesai (*asynchronous*), seperti dapur dengan banyak koki yang memasak secara bersamaan.
- PostgreSQL: Sebagai sistem basis data relasional untuk menyimpan hasil akhir.
- Python: Bahasa pemrograman utama yang menjalankan logika (*core logic system*) dan menghubungkan semua komponen.



2.5. Validasi Sistem

Validasi sistem dilakukan menggunakan data baru berupa *timestamp* dari tanggal 1 Oktober hingga 31 Desember, yang tidak digunakan selama proses pelatihan. Pengujian dilaksanakan secara otomatis melalui *pipeline* ETL 3, di mana sistem hanya menerima *input* berupa waktu. Model akan memproses *input* tersebut untuk menghasilkan prediksi keluaran panel surya, karena seluruh informasi cuaca telah dipelajari pada tahap pelatihan. Evaluasi performa model regresi dilakukan dengan tiga metrik utama, yaitu:

- *Root Mean Square Error* (RMSE) untuk mengukur deviasi prediksi terhadap nilai aktual
- *Mean Absolute Error* (MAE) untuk menghitung rata-rata selisih absolut
- *Mean Absolute Percentage Error* (MAPE) untuk menunjukkan tingkat kesalahan dalam bentuk persentase

Sementara itu, evaluasi hasil *clustering* dilakukan menggunakan *Silhouette Score* untuk menilai kualitas pemisahan data berdasarkan kondisi cuaca, awan, dan suhu. Sistem diuji menggunakan data riil dari *platform* Solcast selama tiga bulan, mencakup berbagai kondisi cuaca. Selain akurasi prediksi, validasi juga mempertimbangkan kecepatan proses otomatis serta ketahanan *pipeline* dalam menangani aliran data secara *asynchronous*.

III. HASIL DAN PEMBAHASAN

3.1. Deskripsi Sistem dan Implementasi Pipeline

Sistem *asynchronous data pipeline* yang dikembangkan berhasil memproses sekitar 35.000 entri data *time series* dari gabungan data cuaca, data radiasi matahari, dan data kinerja panel surya. Sistem berhasil dibangun secara modular menggunakan pendekatan ETL yang berjalan secara terpisah dan asinkron melalui middleware RabbitMQ dan *task queue* Celery. Setiap tahap ETL telah diuji secara unit untuk memastikan keberhasilan proses pengambilan (*extract*), transformasi (*transform*), dan penyimpanan data (*load*) pada masing-masing tahapan.

Tahap *ETL 1 (Pre-processing dan Merging)*, data dari berbagai sumber diolah untuk menyamakan format waktu menjadi POSIX dan digabungkan berdasarkan kolom *posix_time*. Kolom kategorikal diubah menjadi numerik, sementara *missing value* ditangani menggunakan metode *forward fill* dan *back fill*. Tipe data disesuaikan agar konsisten lalu setelah semuanya selesai data disimpan dalam format csv. Pengubahan format waktu juga berguna supaya data input bisa sebagai detik, jam, harian, bulanan, bahkan tahunan.

Tahap *ETL 2 (Train Model)* menggunakan model regresi XGBoost yang menghasilkan performa paling optimal. *Clustering* menggunakan K-Means berhasil mengelompokkan kondisi cuaca ke dalam 4 cluster utama berdasarkan fitur suhu, kelembaban, kecepatan angin, dan radiasi matahari. Model yang telah dilatih disimpan dalam format (.pkl) untuk inferensi pada tahap berikutnya. Sistem ini juga dapat menjadi pelatihan model berulang dengan menambah beberapa input data baru lagi untuk dilatih.

Tahap *ETL 3 (Test Model)* untuk menangani *streaming data* secara *asynchronous*, sehingga sangat cocok diterapkan pada sistem monitoring panel surya yang membutuhkan *update* prediksi secara berkala atau *real-time*. Dengan adanya *input data* baru dengan waktu yang bervariasi, kemampuan prediksi otomatis dan penyimpanan terstruktur, sistem ini dapat mendukung pengambilan keputusan berbasis data (*data-driven decisions*) dalam efektivitas panel surya. Hasil prediksi dan cluster disimpan secara otomatis ke dalam database PostgreSQL.



3.2. Evaluasi performa model

3.2.1. Regresi

Evaluasi performa model regresi dilakukan menggunakan dua metrik utama: RMSE dan MAE. Hasil perbandingan performa antar model disajikan pada Tabel 1.

Tabel 1. Evaluasi performa model regresi

Model	RMSE	MAE	MAPE (%)
ARIMA	0.11	0.1	254.66%
SARIMA	0.16	0.13	190.36%
SARIMAX	0.1431	0.1229	194.90%
XGBoost (<i>Tuning</i>)	0.0289	0.0200	46.11%

Dari tabel di atas menunjukkan bahwa XGBoost memiliki kemampuan menangkap pola data non-linearitas dan interaksi kompleks antar fitur yang baik. Keunggulan ini menjadikannya lebih unggul untuk prediksi berbasis data multi-variat seperti pada sistem panel surya.

3.2.2. Clustering 1

Pada tahap ini klasterisasi berdasarkan fitur yang digunakan adalah ‘*Cloud Type*’ dan ‘*Cloud Cover*’.

- *Cloud type* memberikan konteks jenis awan dan karakteristik fisiknya (misal cirrus tipis di ketinggian tinggi vs nimbostratus tebal di ketinggian rendah).
- *Cloud cover* memberikan informasi intensitas atau kepadatan awan [9].

Tabel 2. Evaluasi *silhouette score cluster 1*

Jumlah Cluster	Silhouette Score
2	0.427727
3	0.510813
4	0.518809
5	0.482733
6	0.444487
7	0.496355
8	0.513236
9	0.482615
10	0.496582

Setelah mengevaluasi berbagai jumlah *cluster* ($k = 2$ hingga 10), diperoleh nilai *Silhouette Score* tertinggi sebesar 0.518809 saat jumlah *cluster* = 4. *Cluster* tersebut kemudian diberi label sebagai berikut berdasarkan karakteristik awan yang dominan di masing-masing cluster:

- Cluster 0: Cumulus/Stratocumulus
- Cluster 1: Stratus/Alto cumulus
- Cluster 2: Nimbostratus/Cumulonimbus
- Cluster 3: Cirrus/Cirrostratus



Pengelompokan ini mencerminkan tipe-tipe awan utama yang umum dalam meteorologi dan memudahkan interpretasi hasil *clustering* dalam konteks studi cuaca dan radiasi matahari [10]. Dengan 4 cluster ini, analisis dapat dilakukan lebih spesifik terhadap karakteristik awan yang berbeda dan dampaknya terhadap kondisi atmosfer.

3.2.3. Clustering 2

Fitur yang digunakan dalam *clustering* ini adalah GHI, DHI, dan DNI.

- GHI (*Global Horizontal Irradiance*) mengukur total radiasi matahari yang diterima permukaan horizontal, mencakup radiasi langsung dan tersebar, penting untuk sistem panel surya datar.
- DHI (*Diffuse Horizontal Irradiance*) merepresentasikan radiasi matahari yang tersebar di atmosfer dan mencapai permukaan secara tidak langsung, sangat berpengaruh saat kondisi berawan.
- DNI (*Direct Normal Irradiance*) mengukur radiasi matahari langsung dari arah matahari yang jatuh tegak lurus pada permukaan, krusial untuk panel surya dengan sistem tracking dan pembangkit tenaga surya konsentrator.

Tabel 3. Evaluasi *Sillhouette Score Clustering 2*

Jumlah Cluster	<i>Silhouette Score</i>
2	0.420588
3	0.451485
4	0.398134
5	0.429916
6	0.4223
7	0.411651
8	0.41211
9	0.415141
10	0.403118

Untuk segmentasi kondisi radiasi matahari berdasarkan variabel GHI, DHI, dan DNI, menggunakan 3 *cluster* adalah pilihan optimal menurut evaluasi *Silhouette Score*. Hal ini dapat digunakan untuk mengklasifikasikan cuaca seperti "Cerah", "Berawan", dan "Hujan/Mendung" [11], [12].

3.2.4. Clustering 3

Pada tahap ini fitur yang digunakan adalah *temperature*, *maxtempC*, dan *DewpointC*.

- Temperature: Suhu rata-rata udara.
- maxtempC: Puncak suhu panas harian.
- DewPointC: Tingkat kelembapan udara, menunjukkan titik embun.

Tabel 4. Evaluasi *Sillhouette Score Clustering 3*

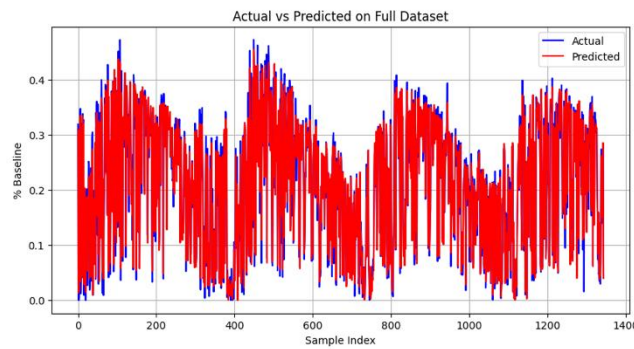
Jumlah Cluster	<i>Silhouette Score</i>
2	0.571011
3	0.496958
4	0.425242
5	0.392296
6	0.346764



7	0.304777
8	0.290719
9	0.263996
10	0.256378

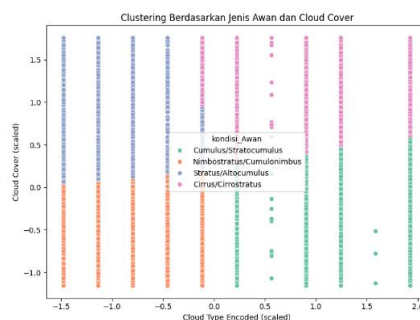
Meskipun hasil analisis *Silhouette Score* menunjukkan bahwa pemisahan paling optimal secara statistik dicapai dengan 2 kluster (nilai *Silhouette Score* = 0.587), pemilihan jumlah kluster sebanyak 3 dilakukan atas pertimbangan konseptual dan interpretatif. Hal ini dilakukan agar hasil *clustering* dapat merepresentasikan kondisi suhu yang lebih bermakna secara praktis untuk analisis iklim dan pengelolaan energi, yaitu membagi data menjadi kategori "Panas", "Lembap", dan "Sejuk"[13] [14].

3.3. Visualisasi



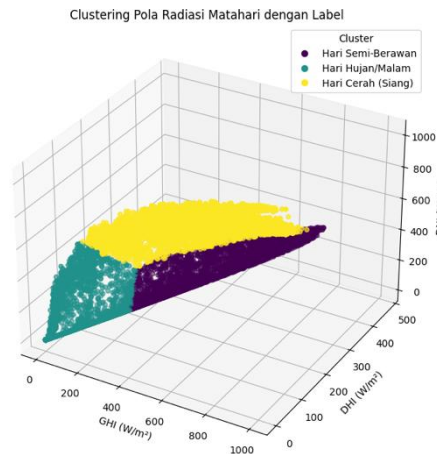
Gambar 3. Visualisasi *actual vs. predicted* regresi *XGBoost*

Grafik pada Gambar 4 menunjukkan bahwa prediksi model (garis merah) mengikuti pola nilai aktual (garis biru) dengan baik, terutama dalam menangkap fluktuasi naik-turun pada data keluaran panel surya. Walaupun terlihat ada beberapa *noise*, *XGBoost* secara garis besar bahwa tidak hanya menghafal rata-rata, tetapi mampu memahami dinamika data dengan baik. Selain itu, *XGBoost* dipilih karena kemampuannya dalam memanfaatkan berbagai fitur input, seperti parameter cuaca dan radiasi matahari [15], yang tidak dapat ditangani secara langsung oleh model *time-series* klasik seperti ARIMA atau SARIMA.



Gambar 5. Visualisasi *Scatter plot Clustering 1*

Gambar 5 menunjukkan hasil *clustering* berdasarkan kombinasi jenis awan (*cloud type*) dan tutupan awan (*cloud cover*). Setiap warna mewakili kelompok awan yang berbeda, yaitu Cumulus/Stratocumulus, Nimbostratus/Cumulonimbus, Stratus/Altostratus, dan Cirrus/Cirrostratus. Visualisasi ini memudahkan identifikasi pola hubungan antara tipe awan dan tingkat tutupan awan untuk mengenali kondisi langit yang paling berdampak terhadap penurunan *output* panel surya, misalnya Nimbostratus/Cumulonimbus umumnya berkorelasi dengan penurunan tajam produksi energi.

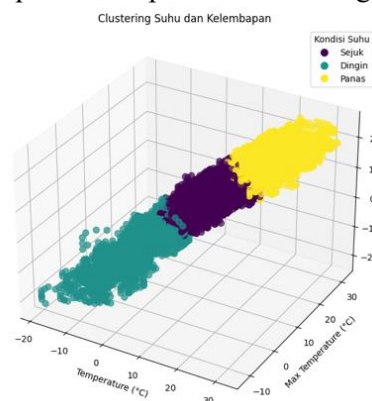


Gambar 4. Visualisasi Scatter Plot Clustering 2

Pada gambar 6 ditunjukkan hasil *clustering dengan* tiga variable utama yaitu GHI, DHI, dan DNI yang masing-masing direpresentasikan menjadi sumbu x, y, dan z. Distribusi titik data hasil klasterisasi memuat pola.

- Hari Cerah (Siang) (kuning) cenderung memiliki nilai GHI, DHI, dan DNI yang tinggi, menandakan intensitas radiasi matahari maksimal.
- Hari Semi-Berawan (ungu) berada di area tengah, dengan nilai radiasi yang sedang.
- Hari Hujan/Malam (biru) terkonsentrasi pada nilai GHI, DHI, dan DNI yang rendah, sesuai dengan minimnya radiasi matahari pada kondisi tersebut.

Informasi ini diharapkan berguna untuk mengelola ekspektasi energi yang dihasilkan serta menjadi dasar dalam strategi penyimpanan atau pemanfaatan energi cadangan.



Gambar 5. Visualisasi Scatter Plot Clustering 3

Scatter Plot pada Gambar 7 merupakan visualisasi hasil *clustering* 3 dimensi berdasarkan tiga variabel utama, yaitu *Temperature (°C)*, *Max Temperature (°C)*, dan *Dew Point (°C)*, yang masing-masing direpresentasikan pada sumbu X, Y, dan Z. Data dikelompokkan menjadi tiga kategori suhu: Dingin, Sejuk, dan Panas.

- Kelompok Dingin didominasi oleh suhu dan kelembapan rendah.
- Kelompok Sejuk berada di rentang menengah antara kelompok dingin dan panas.
- Kelompok Panas terkonsentrasi di area dengan suhu dan kelembapan tinggi.

Klaster ini membantu mengenali kondisi suhu ekstrem yang dapat mempengaruhi efisiensi panel surya. Kelompok suhu panas, misalnya, mengindikasikan potensi penurunan efisiensi meskipun radiasi tinggi, sehingga penting dalam perencanaan pendinginan pasif atau aktif.



Secara keseluruhan, hasil proses klasterisasi memiliki peran signifikan dalam mendukung sistem prediksi untuk mengenali pola-pola lingkungan yang kompleks. Identifikasi karakteristik masing-masing klaster memungkinkan sistem untuk menyesuaikan model prediktif secara lebih konsisten terhadap kondisi-kondisi tertentu, sehingga dapat meningkatkan akurasi prediksi keluaran panel surya. Selain itu, informasi yang dihasilkan dari klasterisasi juga dapat dimanfaatkan sebagai dasar dalam pengambilan keputusan operasional di lapangan, seperti penjadwalan pemeliharaan panel surya secara preventif maupun strategi manajemen beban energi berdasarkan proyeksi kinerja sistem dalam kondisi cuaca yang beragam.

Tabel 5. Performa System

Task / Modul	ID Task	Durasi (s)	Metrik/Output
etl_train	f0d01986-3e26-4d41-99d9-3ab959dc6365	2 -	
etl_weather	f2a7ca83-27d5-4195-b584-6a38e465742e	2.03 -	
etl_solar	6149a02b-3739-45a9-853f-2f4ebc23214b	2.03 -	
etl_pra	7ef5ffbc-b4e4-4381-9e47-dd44ff7978d	1.02	Posix time range: 13.8e9 - 15.0e9
modelregresi (Regresi GHI)	5508b13a-749c-462e-9089-0eb674fd51d4	18.38	RMSE: 0.0289, MAE: 0.0200
Clustering_kondisicuaca	db78408b-015c-44b6-a28c-859fd8f143c0	0.75	Silhouette Score: 0.451485
Clustering_kondisiawan	2b2f5594-c354-405a-acbc-4dffca9681f8	29.08	Silhouette Score: 0.5121
Clustering_kondisisuhu	95db44d9-d6c6-4e5f-a6af-a1eff9a345f5	30.52	Silhouette Score (k=3): 0.497
predict_regression	2x ID	0.61 -	
predict_cluster_* (awan, cuaca, suhu)	6x ID	0.03 - 0.30	Semua task berhasil dalam waktu < 1 detik
load_to_postgres	5x ID	0.5 - 0.7	File hasil dimuat ke database

KESIMPULAN

Penelitian ini berhasil mengembangkan sistem *asynchronous data pipeline* berbasis ETL dengan dukungan RabbitMQ dan Celery untuk prediksi kinerja panel surya menggunakan data cuaca dan radiasi matahari. Sistem ini beroperasi dengan efisien, akurat, dan mampu berjalan otomatis dalam waktu nyata. Model XGBoost menunjukkan kemampuan tinggi dalam memetakan pola non-linear, sementara metode K-Means efektif dalam melakukan segmentasi kondisi lingkungan. Validasi sistem menunjukkan potensi penerapannya di dunia nyata, khususnya untuk pemantauan panel surya secara berkala.

Sistem ini dapat diperluas melalui integrasi dengan *Internet of Things* (IoT) dan *platform cloud*, yang mendukung skalabilitas dan fleksibilitas lebih baik. Integrasi ini memungkinkan pemantauan dan pengelolaan data energi surya secara lebih efisien dan *real-time*. Selain itu, penerapan *blockchain* dapat memperkuat keamanan dan transparansi data, serta memberikan solusi terdesentralisasi untuk sistem pemantauan energi surya. Namun, tantangan seperti keterbatasan data dan infrastruktur yang ada perlu



diperhatikan dalam implementasi agar sistem ini dapat beroperasi secara optimal dan mendukung pengambilan keputusan yang lebih efektif.

Secara sosial-ekonomi, sistem ini berpotensi mendorong pemanfaatan energi surya pada skala kecil, seperti untuk rumah tangga atau UMKM, yang pada gilirannya akan mempercepat adopsi energi terbarukan, mengurangi ketergantungan pada energi fosil, serta mendukung kebijakan energi lokal yang lebih hijau dan berkelanjutan. Dengan demikian, sistem ini turut berkontribusi pada pencapaian SDG 7 dan SDG 11.

REFERENSI

- [1] F. Afif dan A. Martin, "Tinjauan Potensi dan Kebijakan Energi Surya di Indonesia," *J. Engine Energi Manufaktur Dan Mater.*, vol. 6, no. 1, hlm. 43, Mei 2022, doi: 10.30588/jeemm.v6i1.997.
- [2] Wijaya Andy dan Widoatmodjo Sawiji, "Analisa Potensi Dan Hambatan Penerapan Solar Photovoltaic Terhadap Program Net-Zero Emission Di Indonesia," 2023.
- [3] P. R. Ansyah, G. R. Cahyono, A. Nihin, Z. Aqli, dan N. H. Lala, "Studi Kinerja Ground Heat Exchanger (GHE) sebagai Media Pendingin Temperatur Permukaan Panel Surya dengan Variasi Kecepatan Udara," *Bul. Profesi Ins.*, vol. 6, no. 1, hlm. 20–25, 2023, doi: 10.20527/bpi.v6i1.167.
- [4] K. T. N. Ihsan, T. S. Anggraini, M. Adrian, P. Rohayani, dan A. D. Sakti, "SPATIAL MODELING of MULTI-SCENARIO OPTIMAL SOLAR PV POWER PLANT DISTRIBUTION to SUPPORT INDONESIA'S CLEAN ENERGY ACHIEVEMENT TARGETS," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.*, vol. 46, no. M-2–2022, hlm. 119–126, 2022, doi: 10.5194/isprs-Archives-XLVI-M-2-2022-119-2022.
- [5] M. I. El-afifi, "Solar Energy: Our Future for Sustainable Energy," *Nile J. Commun. Comput. Sci.*, vol. 7, no. 1, hlm. 67–83, 2024, doi: 10.21608/njccs.2024.276627.1031.
- [6] M. A. Abbas, B. Nadjia, dan B. N. El-Islam, "Parallel Data Collection Pipelines for Decision-Support Systems: Addressing Data Gaps in Solar Energy Assessment".
- [7] E. E. Wahyudi *dkk.*, "Akuisisi Data Prediksi Curah Hujan Secara Periodik Menggunakan Apache Airflow," *J. Inform. Inf. Syst. Softw. Eng. Appl. INISTA*, vol. 4, no. 2, hlm. 1–12, Mei 2022, doi: 10.20895/inista.v4i2.574.
- [8] A. Nadeem *dkk.*, "AI-Driven precision in solar forecasting: Breakthroughs in machine learning and deep learning," *AIMS Geosci.*, vol. 10, no. 4, hlm. 684–734, 2024, doi: 10.3934/geosci.2024035.
- [9] A. E. Yuchechen, S. Gabriela Lakkis, A. Caferri, P. O. Canziani, dan J. P. Muszkats, "A cluster approach to cloud cover classification over South America and adjacent oceans using a k-means/k-means++ unsupervised algorithm on GOES IR imagery," *Remote Sens.*, vol. 12, no. 18, 2020, doi: 10.3390/RS12182991.
- [10] A. Scholar, "Identifying Four Year Average Cloud Field Regimes from World Wide Merged Cloud Analysis Dataset By Way of K-Means Clustering," 2021.
- [11] L. Song *dkk.*, "The quantile-matching approach to improving radar quantitative precipitation estimation in South China," *Remote Sens.*, vol. 13, no. 23, hlm. 1–18, 2021, doi: 10.3390/rs13234956.
- [12] A. Jalil *dkk.*, "Perancangan PLTS Off Grid di Pulau Wawonii dengan PVLib Python," vol. 09, no. 03, hlm. 203–210, 2024.
- [13] M. Gao dan C. L. E. Franzke, "Quantile regression-based spatiotemporal analysis of extreme temperature change in China," *J. Clim.*, vol. 30, no. 24, hlm. 9897–9914, 2017, doi: 10.1175/JCLI-D-17-0356.1.
- [14] D. R. Gergel *dkk.*, "Global Downscaled Projections for Climate Impacts Research (GDPCIR): Preserving quantile trends for modeling future climate impacts," *Geosci. Model Dev.*, vol. 17, no. 1, hlm. 191–227, 2024, doi: 10.5194/gmd-17-191-2024.
- [15] D. P. Mishra, S. Jena, R. Senapati, A. Panigrahi, dan S. R. Salkuti, "Global solar radiation forecast using an ensemble learning approach," *Int. J. Power Electron. Drive Syst. IJPEDS*, vol. 14, no. 1, hlm. 496, Mar 2023, doi: 10.11591/ijpeds.v14.i1.pp496-505.



SENADA

Seminar Nasional Sains Data 2025 (SENADA 2025)
UPN “Veteran” Jawa Timur

E-ISSN 2808-5841
P-ISSN 2808-7283