



Klasterisasi K-Means untuk Karakterisasi Efisiensi Turbin Angin berdasarkan Karakteristik Operasi

Triuli Novianti^{1*}, Ridho Akbar², Iwan Santosa³, Ainur Romzy⁴

^{1,2,4}Program Studi Teknik Elektro, Universitas Muhammadiyah Surabaya, Indonesia

³Program Studi Teknik Informatika, Universitas Trunojoyo Madura, Indonesia

^{1,2,4}Jl. Raya Sutorejo No. 59, Dukuh Sutorejo, Kec. Mulyorejo, Kota Surabaya, Jawa Timur, 60113

³Jl. Raya Telang PO Box 2, Kec. Kamal, Kab. Bangkalan, Madura, Jawa Timur 69162

E-Mail *: triuli.novianti@ft.um-surabaya.ac.id

ABSTRACT

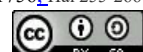
Wind turbine efficiency is an important indicator of a system's ability to convert wind energy into electricity. This study develops a K-Means-based machine learning approach for efficiency characterization of wind turbine operating conditions using Wind Speed, Active Power, and Efficiency. The dataset consists of 100 observations from the Wind Turbine Power Generation Dataset. The research stages include data preprocessing, Min-Max normalization, determination of the optimal number of clusters using the Elbow Method, K-Means clustering, and cluster-quality evaluation using the Silhouette Score. The results indicate that three clusters ($K=3$) are optimal, with WCSS values of 27.4509, 6.7612, 2.9752, 1.7142, and 1.1369 for $K=1-5$. The clusters are interpreted as high-efficiency, transition, and low-efficiency regimes. Cluster 1 has Centroid values of 577.1 kW for Active Power, 6.4 m/s for Wind Speed, and 73.6% for Efficiency; Cluster 2 has 2,482 kW, 12.4 m/s, and 63.7%; while Cluster 3 has 6,122 kW, 18.0 m/s, and 54.7%, respectively. The clusters contain 44, 16, and 40 observations. A Silhouette Score of 0.5863 indicates good clustering quality. The main contribution is the interpretation of K-Means clusters as efficiency regimes rather than merely mathematical groups. The results show that higher wind speed is associated with higher active power but does not necessarily indicate higher efficiency.

Keywords: Machine Learning, K-Means Clustering, Wind Turbine, Efficiency, Operating Characteristics

ABSTRAK

Efisiensi turbin angin merupakan indikator penting dalam mengevaluasi kemampuan sistem mengonversi energi angin menjadi energi listrik. Penelitian ini mengembangkan pendekatan *machine learning* berbasis K-Means untuk karakterisasi efisiensi kondisi operasi turbin angin menggunakan *Wind Speed*, *Active Power*, dan *Efficiency*. Dataset terdiri atas 100 observasi dari *Wind Turbine Power Generation Dataset*. Tahapan penelitian meliputi praproses data, normalisasi menggunakan *Min-Max Scaling*, penentuan jumlah kluster optimal dengan *Elbow Method*, *K-Means clustering*, serta evaluasi kualitas kluster menggunakan *Silhouette Score*. Hasil menunjukkan bahwa jumlah kluster optimal adalah tiga ($K=3$), dengan nilai WCSS sebesar 27,4509; 6,7612; 2,9752; 1,7142; dan 1,1369 untuk $K=1$ hingga $K=5$. Ketiga kluster diinterpretasikan sebagai *high-efficiency regime*, *transition regime*, dan *low-efficiency regime*. Kluster 1 memiliki *Centroid Active Power* 577,1 kW, *Wind Speed* 6,4 m/s, dan *Efficiency* 73,6%; Kluster 2 sebesar 2.482 kW, 12,4 m/s, dan 63,7%; sedangkan Kluster 3 sebesar 6.122 kW, 18,0 m/s, dan 54,7%. Komposisi kluster masing-masing terdiri atas 44, 16, dan 40 observasi. Nilai *Silhouette Score* sebesar 0,5863 menunjukkan kualitas pengelompokan yang baik. Kontribusi utama penelitian terletak pada interpretasi hasil K-Means sebagai rezim efisiensi sehingga memberikan karakterisasi kondisi operasi berbasis data. Hasil menunjukkan bahwa peningkatan kecepatan angin berkaitan dengan peningkatan daya aktif, tetapi tidak selalu menghasilkan efisiensi yang lebih tinggi.

Kata Kunci: Machine Learning, K-Means Clustering, Turbin Angin, Efisiensi, Karakteristik Operasi



1. PENDAHULUAN

Energi angin merupakan salah satu sumber energi terbarukan yang dapat dikonversi menjadi energi listrik melalui turbin angin. Kinerja turbin sangat dipengaruhi oleh perubahan kecepatan angin yang menyebabkan variasi daya keluaran dan efisiensi pada kondisi operasi yang berbeda. Oleh karena itu, evaluasi performa turbin tidak cukup hanya berdasarkan daya yang dihasilkan, tetapi perlu mempertimbangkan hubungan antara kecepatan angin, daya aktif, dan efisiensi secara bersamaan. Karakteristik dan prediksi kecepatan angin telah dikaji menggunakan pendekatan statistik maupun *deep learning* untuk memahami pola perubahan kondisi angin (Tyass dkk., 2023). Studi karakteristik turbin juga menunjukkan bahwa perubahan kondisi angin dan kurva daya berpengaruh terhadap kemampuan turbin dalam menghasilkan energi listrik (Yandri dkk., 2024).

Variasi kondisi operasi juga menyebabkan tingkat efisiensi turbin tidak selalu meningkat seiring dengan bertambahnya daya keluaran. Penelitian mengenai turbin angin menunjukkan bahwa perubahan konfigurasi dan kecepatan angin dapat menghasilkan karakteristik daya dan efisiensi yang berbeda (Sugeng dkk., 2024; Chalimah dkk., 2024). Selain itu, perkembangan sistem monitoring memungkinkan parameter operasi turbin diperoleh secara kontinu. Arief dkk. (2023) mengembangkan monitoring arus dan tegangan PLTB berbasis ESP8266 dan Node-RED, sedangkan Baihaqi (2023) mengembangkan sistem pemantauan efisiensi daya DC pada turbin angin dengan akuisisi data secara periodik. Pengembangan sistem monitoring turbin angin berbasis IoT juga memungkinkan pengiriman parameter operasi secara nirkabel dari lokasi turbin yang berada pada lingkungan rural (Kurniawati dkk., 2026). Kondisi tersebut menghasilkan data operasional yang dapat dimanfaatkan lebih lanjut untuk analisis berbasis data.

Ketersediaan data operasional membuka peluang penerapan *machine learning* untuk mengidentifikasi karakteristik kondisi turbin. Wiranata dkk. (2021) menerapkan *machine learning* dengan metode *Support Vector Regression* (SVR) untuk analisis prognostik kerusakan bantalan turbin angin. Meskipun demikian, penelitian tersebut berfokus pada prediksi kerusakan komponen, sedangkan penelitian mengenai performa turbin masih banyak diarahkan pada pengaruh parameter desain, karakteristik angin, monitoring, dan evaluasi daya atau efisiensi (Sugeng dkk., 2024; Yandri dkk., 2024). Pendekatan *machine learning* telah digunakan untuk mendeteksi gangguan, mengidentifikasi false alarm, serta meningkatkan kemampuan diagnosis kondisi turbin angin (Allal dkk., 2024; Peco Chacón dkk., 2023; Sun dkk., 2023; Tian dkk., 2022; Zhang dkk., 2023). Pendekatan *machine learning* untuk mengelompokkan kondisi operasi berdasarkan karakteristik efisiensi masih relatif terbatas.

Berdasarkan kondisi tersebut, diperlukan pendekatan yang mampu mengidentifikasi pola

kondisi operasi tanpa bergantung pada label kelas yang telah ditentukan sebelumnya. *Unsupervised learning* dengan metode K-Means sesuai digunakan karena mampu mengelompokkan observasi berdasarkan kemiripan karakteristik. Penelitian K-Means pada data SCADA turbin angin telah menunjukkan manfaat *clustering* untuk eksplorasi kondisi operasi dan kebutuhan pemeliharaan prediktif (Rodriguez dkk., 2023). Pendekatan *clustering* dinamis juga telah digunakan untuk mengelompokkan turbin berdasarkan pola sinyal SCADA dan mengidentifikasi perilaku yang menyimpang (Marti-Puig dan Núñez-Vilaplana, 2024). Analisis data SCADA juga telah banyak digunakan untuk mengevaluasi performa dan kondisi operasi turbin angin secara berkelanjutan (Astolfi dkk., 2022; Wu dkk., 2024). Pendekatan *clustering* pada data SCADA juga dapat digunakan untuk mengidentifikasi berbagai mode operasi turbin berdasarkan karakteristik data berdimensi tinggi (Yang dkk., 2024). Namun, pendekatan tersebut berfokus pada pengelompokan karakteristik performa dan belum secara khusus menginterpretasikan kelompok sebagai rezim efisiensi berdasarkan kombinasi *Wind Speed*, *Active Power*, dan *Efficiency*. Dengan demikian, celah penelitian terletak pada kebutuhan karakterisasi kondisi operasi yang menghubungkan hasil *clustering* dengan tingkat efisiensi secara eksplisit.

Kontribusi utama penelitian ini adalah pengembangan pendekatan *K-Means-based efficiency characterization* untuk mengidentifikasi *efficiency regime* turbin angin berdasarkan kombinasi *Wind Speed*, *Active Power*, dan *Efficiency*. Metode *clustering* juga telah diterapkan pada data turbin angin untuk mengelompokkan pola operasi, mengidentifikasi *outlier* pada kurva daya, dan menganalisis pola alarm turbin (Paik dkk., 2023; Wei dkk., 2023). Namun, penelitian-penelitian tersebut belum secara khusus menginterpretasikan hasil *clustering* berdasarkan tingkat efisiensi turbin. Berbeda dengan penelitian terdahulu yang berfokus pada pengaruh parameter, monitoring, diagnosis, atau prediksi kondisi komponen, penelitian ini menginterpretasikan hasil *clustering* menjadi tiga karakteristik operasi, yaitu *high-efficiency regime*, *transition regime*, dan *low-efficiency regime*. Dengan demikian, K-Means tidak hanya digunakan untuk menghasilkan kelompok secara matematis, tetapi juga untuk memberikan interpretasi teknis terhadap pola efisiensi pada kondisi operasi yang berbeda.

Penelitian ini menggunakan 100 data pengamatan dari *Wind Turbine Power Generation Dataset*. Data dilakukan *preprocessing* dan normalisasi menggunakan *Min-Max Scaling*, kemudian dikelompokkan menggunakan K-Means. Penelitian bertujuan menentukan jumlah cluster optimal, mengevaluasi kualitas pengelompokan, serta mengidentifikasi karakteristik setiap *efficiency regime* berdasarkan *Wind Speed*, *Active Power*, dan *Efficiency*. Hasil penelitian diharapkan dapat memberikan pendekatan yang sederhana dan

interpretatif untuk mendukung analisis kondisi operasi dan evaluasi efisiensi turbin angin.

2. METODE

2.1. Konsep Karakterisasi Efisiensi Berbasis *Machine Learning*

Turbin angin merupakan sistem konversi energi yang memanfaatkan energi kinetik angin untuk menghasilkan energi listrik melalui turbin dan generator. Kinerja sistem sangat dipengaruhi oleh kondisi angin dan karakteristik operasi turbin. Perubahan kecepatan angin dapat menyebabkan perubahan daya keluaran dan karakteristik efisiensi sehingga evaluasi kinerja turbin perlu mempertimbangkan beberapa parameter operasi secara bersamaan. Hubungan antara daya turbin, daya generator, dan efisiensi menjadi salah satu aspek penting dalam mengevaluasi kemampuan sistem menghasilkan energi listrik pada kondisi operasi yang berbeda (Chalimah dkk., 2024).

Efisiensi digunakan untuk menggambarkan kemampuan sistem dalam mengonversi energi yang tersedia menjadi energi keluaran. Pada sistem turbin angin, perubahan daya masukan dan daya keluaran dapat menyebabkan perubahan nilai efisiensi. Hasil penelitian menunjukkan bahwa variasi daya turbin dapat menghasilkan perubahan daya generator dan nilai efisiensi pada kondisi operasi tertentu, sehingga analisis kinerja tidak cukup dilakukan hanya berdasarkan nilai daya keluaran (Chalimah dkk., 2024). Oleh karena itu, hubungan antara *Wind Speed*, *Active Power*, dan *Efficiency* perlu dianalisis secara simultan untuk memperoleh karakteristik kondisi operasi yang lebih representatif.

Ketersediaan sistem monitoring memungkinkan parameter operasi turbin diperoleh secara lebih terstruktur. Parameter seperti kecepatan angin, kecepatan putaran, arus, tegangan, dan daya dapat diperoleh melalui sistem monitoring dan digunakan untuk mengevaluasi kondisi operasi pembangkit listrik tenaga angin (Fajri & Yuhendri, 2023). Sistem monitoring tersebut menunjukkan bahwa data operasional turbin dapat dikumpulkan secara periodik dan digunakan sebagai dasar untuk melakukan analisis terhadap perubahan kondisi operasi.

Pemanfaatan data operasional selanjutnya dapat dikembangkan melalui pendekatan *machine learning*. Pendekatan berbasis data memungkinkan pola atau kelompok kondisi operasi diidentifikasi tanpa harus menentukan kategori kondisi secara manual sejak awal. Penerapan K-Means pada data SCADA turbin angin menunjukkan bahwa metode *unsupervised learning* dapat digunakan untuk mengeksplorasi pola operasi dan mengelompokkan data berdasarkan karakteristik yang memiliki kemiripan (Rodriguez dkk., 2023). Pendekatan tersebut menjadi dasar penggunaan clustering dalam penelitian ini.

Konsep *Machine Learning-Based Efficiency Characterization* pada penelitian ini diarahkan untuk mengubah data operasi turbin menjadi informasi mengenai karakteristik kondisi efisiensi. Berbeda

dengan penggunaan K-Means yang hanya menghasilkan kelompok data, penelitian ini menginterpretasikan karakteristik setiap kelompok berdasarkan nilai *Wind Speed*, *Active Power*, dan *Efficiency*. Hasil pengelompokan selanjutnya digunakan untuk mengidentifikasi tiga kondisi operasi utama, yaitu *high-efficiency regime*, *transition regime*, dan *low-efficiency regime*. Dengan demikian, pendekatan yang digunakan tidak hanya menghasilkan clustering secara matematis, tetapi juga memberikan interpretasi teknis terhadap karakteristik efisiensi turbin angin.

2.2. Pemodelan Efisiensi Menggunakan Metode K-Means

K-Means merupakan salah satu metode *unsupervised learning* yang digunakan untuk mengelompokkan data berdasarkan tingkat kemiripan karakteristik. Algoritma ini menentukan sejumlah cluster dan *Centroid* sebagai pusat setiap cluster. Setiap observasi ditempatkan pada cluster dengan *Centroid* terdekat berdasarkan jarak Euclidean, kemudian *Centroid* diperbarui berdasarkan rata-rata anggota cluster. Proses iteratif dilakukan sampai keanggotaan cluster dan posisi *Centroid* mencapai kondisi konvergen (Rodriguez dkk., 2023). Dalam penelitian ini, K-Means diterapkan pada tiga fitur yang telah dinormalisasi, yaitu *Wind Speed*, *Active Power*, dan *Efficiency*, dengan jumlah cluster kandidat $K=1$ hingga $K=5$.

Penerapan K-Means pada data turbin angin dapat digunakan untuk menemukan pola operasi yang tidak diketahui sebelumnya. Rodriguez dkk. (2023) menerapkan K-Means pada data SCADA turbin angin untuk analisis eksploratif kondisi operasi dan mendukung kebutuhan pemeliharaan prediktif. Pendekatan tersebut menunjukkan bahwa data operasional dapat dikelompokkan berdasarkan kemiripan karakteristik tanpa memerlukan label kondisi sebelumnya. Pada penelitian ini, prinsip tersebut dikembangkan lebih lanjut dengan memberikan interpretasi efisiensi terhadap setiap cluster.

Pada penelitian ini, K-Means menggunakan jarak Euclidean untuk menentukan kedekatan antara data dengan *Centroid*. Jarak $d(x_i, \mu_k)$ antara data x_i dan *Centroid* μ_k dihitung menggunakan Persamaan (1):

$$d(x_i, \mu_k) = \sqrt{\sum_{j=1}^p (x_{ij} - \mu_{kj})^2} \quad (1)$$

dengan x_{ij} merupakan nilai variabel ke- j pada data ke- i , μ_{kj} merupakan nilai variabel ke- j pada *Centroid* cluster ke- k , dan p merupakan jumlah variabel yang digunakan. Berdasarkan jarak tersebut, setiap data ditempatkan pada cluster dengan nilai jarak terkecil.

Penentuan jumlah cluster merupakan tahapan penting dalam K-Means karena jumlah cluster memengaruhi karakteristik hasil pengelompokan. Pada penelitian ini digunakan *Elbow Method* dengan

menguji $K=1$ sampai $K=5$ menggunakan nilai *Within-Cluster Sum of Squares* (WCSS). $K=3$ dipilih apabila penurunan WCSS mulai melandai pada tiga cluster. Pemilihan ini kemudian diperkuat menggunakan *Silhouette Score* untuk mengevaluasi kekompakan dan pemisahan cluster. Kombinasi kedua evaluasi tersebut digunakan agar pemilihan K tidak hanya didasarkan pada pertimbangan visual Elbow, tetapi juga pada kualitas hasil pengelompokan.

$$WCSS = \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - \mu_k\|^2 \quad (2)$$

WCSS (*Within – Cluster Sum of Squares*) adalah jumlah total kuadrat jarak setiap data terhadap *Centroid* cluster tempat data tersebut berada, K merupakan jumlah cluster, C_k merupakan cluster ke- k , x_i merupakan data ke- i yang termasuk dalam cluster C_k , μ_k merupakan *Centroid* cluster ke- k dan $\|x_i - \mu_k\|^2$ merupakan kuadrat jarak antara data x_i dan *Centroid* μ_k .

Nilai WCSS akan mengalami penurunan ketika jumlah cluster ditambah. Jumlah cluster optimal ditentukan berdasarkan titik *elbow*, yaitu ketika penurunan WCSS mulai melandai sehingga penambahan jumlah cluster berikutnya memberikan peningkatan yang relatif kecil. Pendekatan ini digunakan untuk memperoleh jumlah cluster yang mampu merepresentasikan struktur data secara efektif sebelum dilakukan karakterisasi terhadap masing-masing kelompok.

2.3. Desain Penelitian dan Jenis Data

Penelitian ini menggunakan desain kuantitatif dengan pendekatan analisis data berbasis *machine learning*. Data yang digunakan berupa data sekunder dari *Wind Turbine Power Generation Dataset*. Dataset terdiri atas 100 data pengamatan yang digunakan untuk menggambarkan kondisi operasi turbin angin. Tiga variabel utama yang dianalisis adalah *Wind Speed* (m/s), *Active Power* (kW), dan *Efficiency* (%). Penggunaan data sekunder memungkinkan analisis dilakukan berdasarkan data operasional yang telah tersedia tanpa pengukuran eksperimental tambahan.

Tiga variabel utama digunakan dalam penelitian ini, yaitu *Wind Speed* (m/s), *Active Power* (kW), dan *Efficiency* (%). *Wind Speed* digunakan untuk merepresentasikan kondisi energi angin yang tersedia, *Active Power* merepresentasikan daya listrik aktif yang dihasilkan, sedangkan *Efficiency* digunakan sebagai indikator kinerja konversi energi. Ketiga variabel dipilih karena secara simultan mewakili sisi masukan energi, keluaran daya, dan hasil konversi. Penggunaan ketiganya memungkinkan cluster tidak hanya dibedakan berdasarkan besarnya daya, tetapi juga berdasarkan kondisi angin dan tingkat efisiensi. Justifikasi ini penting karena daya keluaran yang tinggi tidak selalu identik dengan efisiensi yang tinggi.

Penggunaan beberapa parameter operasi dalam analisis turbin angin diperlukan karena perubahan daya tidak selalu menunjukkan perubahan efisiensi yang sebanding. Penelitian eksperimental menunjukkan bahwa perubahan daya turbin dapat menghasilkan variasi daya generator dan efisiensi pada kondisi operasi tertentu (Chalimah dkk., 2024). Selain itu, sistem monitoring PLTB dapat menyediakan beberapa parameter operasi secara bersamaan, seperti kecepatan angin, kecepatan putaran, tegangan, arus, dan daya (Fajri & Yuhendri, 2023). Kondisi tersebut mendukung penggunaan pendekatan multivariabel dalam karakterisasi performa turbin.

Sebelum digunakan pada proses K-Means, data dilakukan preprocessing untuk memastikan data dapat diproses secara konsisten dan dapat direplikasi. Tahapan ini meliputi pemeriksaan data, penyesuaian format, pemilihan tiga variabel yang digunakan dalam clustering, serta normalisasi. Karena *Active Power* memiliki rentang numerik jauh lebih besar daripada *Wind Speed* dan *Efficiency*, normalisasi dilakukan menggunakan *Min-Max Scaling* agar ketiga variabel berada pada skala yang sebanding. Data hasil normalisasi kemudian digunakan sebagai input *Elbow Method* dan *K-Means clustering*.

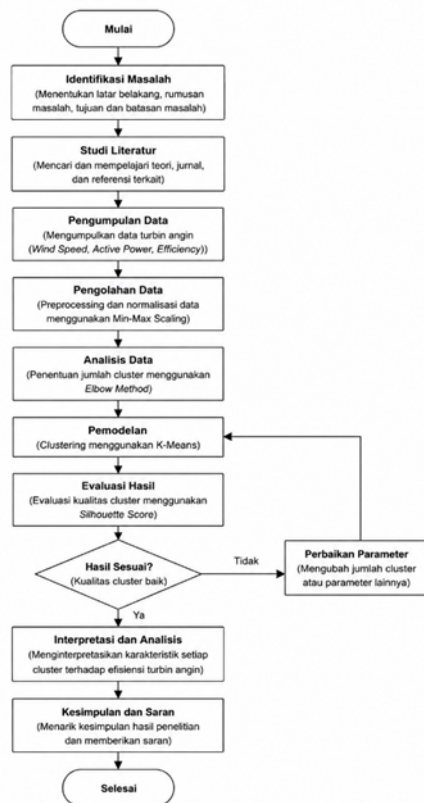
Proses *Min-Max Scaling* dilakukan menggunakan Persamaan (3):

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3)$$

dengan X' merupakan nilai hasil normalisasi, X merupakan nilai aktual, X_{min} merupakan nilai minimum, dan X_{max} merupakan nilai maksimum dari masing-masing variabel. Normalisasi diperlukan karena K-Means menggunakan perhitungan jarak dalam proses pembentukan cluster. Data hasil normalisasi kemudian digunakan sebagai input untuk menentukan jumlah cluster dan melakukan proses K-Means.

2.4. Tahapan Penelitian

Tahapan penelitian ditunjukkan pada Gambar 1, dimulai dari pengumpulan dataset yang terdiri atas 100 data pengamatan. Data kemudian diseleksi berdasarkan variabel yang digunakan dalam penelitian, yaitu *Wind Speed*, *Active Power*, dan *Efficiency*. Pemilihan variabel tersebut dilakukan untuk memperoleh representasi kondisi operasi yang mencakup parameter masukan, keluaran daya, dan indikator efisiensi turbin. Pendekatan berbasis parameter operasi juga digunakan pada sistem monitoring PLTB untuk memperoleh informasi kondisi pembangkit secara lebih terstruktur (Fajri & Yuhendri, 2023).



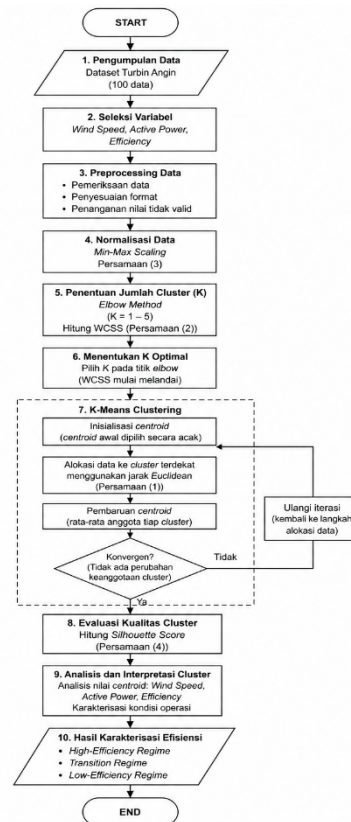
Gambar 1. Flowchart Tahapan Penelitian

Tahap berikutnya adalah *preprocessing* data dan normalisasi menggunakan Min-Max Scaling. Tahap ini bertujuan menyamakan skala ketiga variabel sehingga setiap parameter memiliki kontribusi yang seimbang dalam proses perhitungan jarak. Setelah normalisasi dilakukan, data digunakan untuk menentukan jumlah cluster menggunakan *Elbow Method*. Pengujian dilakukan dengan nilai ($K=1,2,3,4$) dan (5), kemudian nilai *WCSS* dari setiap jumlah cluster dibandingkan untuk menentukan titik *elbow*.

Setelah jumlah cluster optimal diperoleh, dilakukan proses *K-Means clustering*. Tahapan pada Metode K-Means ditunjukkan pada Gambar 2. Data yang telah dinormalisasi dikelompokkan berdasarkan jarak terdekat terhadap *Centroid*. *Centroid* diperbarui berdasarkan rata-rata anggota cluster dan proses iterasi dilakukan sampai diperoleh kondisi konvergen. Penggunaan K-Means pada data turbin angin menunjukkan bahwa metode ini dapat digunakan untuk mengidentifikasi kelompok kondisi operasi berdasarkan karakteristik data SCADA (Rodriguez dkk., 2023).

Tahap selanjutnya adalah evaluasi kualitas cluster menggunakan *Silhouette Score*. Evaluasi ini dilakukan untuk mengetahui tingkat kekompakan data dalam cluster dan tingkat pemisahan antarcluster. Nilai *Silhouette* untuk data ke- (i) dihitung menggunakan Persamaan (4):

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (4)$$



Gambar 2. Diagram Alir Tahapan pada Metode K-Means

Dengan $S(i)$ merupakan *Silhouette Score* untuk data ke- i , nilai $a(i)$ merupakan rata-rata jarak data terhadap anggota cluster yang sama, sedangkan $b(i)$ merupakan rata-rata jarak data terhadap cluster terdekat. $\max\{a(i), b(i)\}$ merupakan nilai terbesar antara nilai $a(i)$ dan $b(i)$. Nilai *Silhouette* yang semakin mendekati 1 menunjukkan bahwa data memiliki kedekatan yang lebih kuat dengan cluster tempatnya berada dibandingkan dengan cluster lainnya. Evaluasi menggunakan *Silhouette* digunakan untuk memastikan bahwa hasil pengelompokan memiliki kualitas pemisahan yang memadai.

Tahap terakhir adalah analisis dan interpretasi karakteristik setiap cluster berdasarkan nilai *Centroid* *Wind Speed*, *Active Power*, dan *Efficiency*. Hasil *Centroid* digunakan untuk memberikan makna teknis terhadap kelompok yang terbentuk. Cluster dengan karakteristik efisiensi relatif tinggi diinterpretasikan sebagai *high-efficiency regime*, kelompok dengan karakteristik peralihan sebagai *transition regime*, dan kelompok dengan efisiensi relatif rendah sebagai *low-efficiency regime*. Dengan tahapan tersebut, penelitian menghasilkan karakterisasi kondisi operasi turbin angin berbasis data yang tidak hanya menunjukkan keanggotaan cluster, tetapi juga memberikan interpretasi terhadap pola efisiensi pada setiap kondisi operasi.

3. PEMBAHASAN

Penelitian menggunakan 100 data pengamatan dari *Wind Turbine Power Generation Dataset* dengan tiga variabel utama, yaitu *Wind Speed*, *Active Power*, dan *Efficiency*. Ketiga variabel tersebut digunakan secara simultan dalam proses K-Means untuk mengidentifikasi karakteristik kondisi operasi turbin angin. Data *Wind Speed* berada pada rentang 2,1–21,9 m/s, *Active Power* sebesar 18–9.372 kW, sedangkan *Efficiency* berada pada rentang 49,4–78,8%.

Nilai rata-rata *Wind Speed* sebesar 12,0 m/s dengan standar deviasi 5,80 m/s menunjukkan adanya variasi kondisi angin yang cukup besar. Sementara itu, *Active Power* memiliki rentang yang lebar, dari 18 kW hingga 9.372 kW. Variasi tersebut menunjukkan bahwa dataset mencakup kondisi operasi mulai dari kecepatan angin rendah dengan daya keluaran kecil hingga kecepatan angin tinggi dengan daya keluaran yang jauh lebih besar.

Variasi *Efficiency* sebesar 49,4–78,8% menunjukkan bahwa peningkatan daya keluaran tidak selalu menghasilkan efisiensi yang lebih tinggi. Kondisi tersebut menjadi dasar penting dalam penelitian karena karakterisasi hanya berdasarkan *Active Power* dapat memberikan interpretasi yang kurang tepat terhadap performa turbin. Oleh karena itu, *Wind Speed*, *Active Power*, dan *Efficiency* digunakan secara bersamaan untuk memperoleh kelompok operasi yang lebih representatif.

Sebelum proses clustering, data dinormalisasi menggunakan Min-Max Scaling. Tahapan ini diperlukan karena *Active Power* mempunyai rentang numerik yang jauh lebih besar dibandingkan *Wind Speed* dan *Efficiency*. Tanpa normalisasi, variabel *Active Power* berpotensi memberikan pengaruh dominan terhadap perhitungan jarak Euclidean. Hasil normalisasi membuat ketiga variabel memiliki skala yang sebanding sehingga pembentukan cluster dapat mempertimbangkan seluruh karakteristik yang digunakan.

Hasil karakteristik awal tersebut menunjukkan bahwa dataset memiliki variasi kondisi operasi yang sesuai untuk dianalisis menggunakan pendekatan *unsupervised learning*. Dengan demikian, K-Means tidak hanya digunakan untuk membagi data secara matematis, tetapi juga untuk menemukan pola hubungan antara kondisi angin, daya keluaran, dan efisiensi. Pendekatan tersebut sejalan dengan penelitian sebelumnya yang menunjukkan bahwa K-Means dapat digunakan untuk mengelompokkan data operasional turbin berdasarkan karakteristik *Wind Speed* dan *output power* (Rodriguez dkk., 2023).

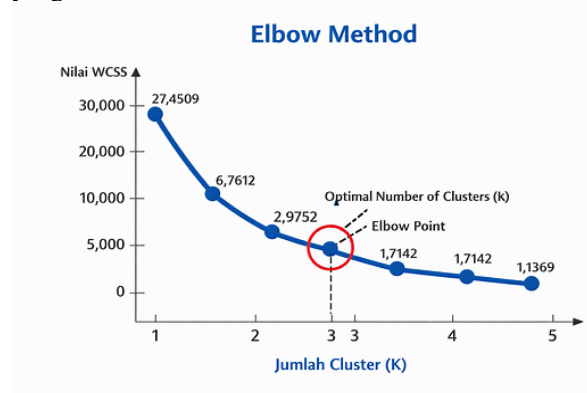
3.1 Penentuan Jumlah Cluster Menggunakan Elbow Method

Penentuan jumlah cluster dilakukan menggunakan *Elbow Method* dengan menghitung nilai *Within-Cluster Sum of Squares* (WCSS) pada (K=1) hingga (K=5). Hasil perhitungan menunjukkan nilai WCSS berturut-turut sebesar 27,4509; 6,7612; 2,9752; 1,7142; dan 1,1369.

Tabel 1. Hasil *Elbow Method*

Jumlah Cluster (K)	WCSS
1	27,4509
2	6,7612
3	2,9752
4	1,7142
5	1,1369

Nilai WCSS mengalami penurunan sangat besar dari (K=1) menjadi (K=2), kemudian kembali menurun hingga (K=3). Setelah (K=3), penurunan WCSS menjadi lebih kecil dibandingkan penurunan pada tahap sebelumnya. Kondisi tersebut menunjukkan bahwa penambahan cluster setelah tiga kelompok memberikan peningkatan homogenitas yang semakin kecil.



Gambar 3. Elbow Method untuk Penentuan K Optimal

Berdasarkan pola tersebut, titik *elbow* diperoleh pada K=3, sehingga tiga cluster dipilih sebagai jumlah cluster optimal. Pemilihan ini penting karena jumlah cluster yang terlalu sedikit berpotensi menggabungkan kondisi operasi yang berbeda, sedangkan jumlah cluster yang terlalu banyak dapat menghasilkan kelompok yang kurang representatif.

Hasil tersebut juga memperlihatkan bahwa tiga cluster mampu merepresentasikan perubahan kondisi operasi turbin dari kecepatan angin rendah hingga tinggi. Dengan demikian, penggunaan (K=3) tidak hanya didukung oleh perubahan WCSS, tetapi juga menghasilkan kelompok yang memiliki interpretasi teknis yang jelas terhadap tingkat efisiensi.

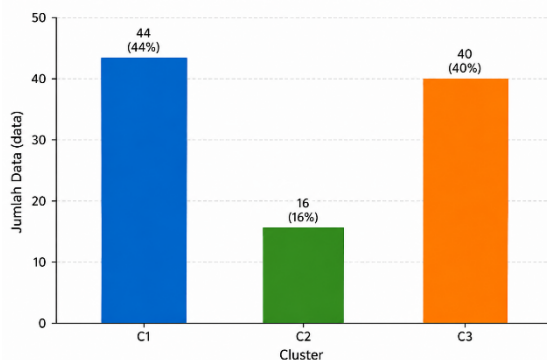
3.2 Hasil Pemodelan K-Means

Setelah jumlah cluster ditetapkan sebesar (K=3), proses K-Means dilakukan terhadap data yang telah dinormalisasi. Algoritma melakukan pengalokasian data berdasarkan jarak *Euclidean* terhadap *Centroid*, kemudian memperbarui *Centroid* berdasarkan rata-rata anggota masing-masing cluster. Hasil evolusi *Centroid* menunjukkan bahwa *Centroid* awal mengalami perubahan cukup besar pada iterasi pertama, kemudian tidak mengalami perubahan pada iterasi berikutnya.

Tabel 2. Karakteristik *Centroid* Hasil K-Means

Cluster	<i>Active Power</i> (kW)	<i>Wind Speed</i> (m/s)	<i>Efficiency</i> (%)	Karakteristik
C1	577,1	6,4	73,6	Efisiensi tinggi
C2	2.482	12,4	63,7	Efisiensi sedang/transisi
C3	6.122	18,0	54,7	Efisiensi rendah

Centroid akhir menunjukkan pola yang cukup jelas. C1 mempunyai *Wind Speed* rata-rata 6,4 m/s, *Active Power* 577,1 kW, dan *Efficiency* 73,6%. C2 memiliki *Wind Speed* 12,4 m/s, *Active Power* 2.482 kW, dan *Efficiency* 63,7%. Sementara itu, C3 memiliki *Wind Speed* 18,0 m/s, *Active Power* 6.122 kW, dan *Efficiency* 54,7%.



Gambar 4. Jumlah Data pada Setiap Cluster

Gambar 4 menunjukkan distribusi jumlah data pada tiga cluster yang dihasilkan melalui proses K-Means. Cluster 1 (C1) memiliki jumlah data terbesar, yaitu 44 data (44%), diikuti oleh Cluster 3 (C3) sebanyak 40 data (40%), sedangkan Cluster 2 (C2) memiliki jumlah data paling sedikit, yaitu 16 data (16%). Dari distribusi tersebut, terlihat bahwa C1 dan C3 mencakup sebagian besar dataset, yaitu 84% dari total 100 data, sementara C2 hanya mencakup 16%. Perbedaan jumlah anggota cluster ini menunjukkan bahwa karakteristik data turbin angin lebih banyak terkonsentrasi pada kondisi yang direpresentasikan oleh C1 dan C3, sedangkan C2 memiliki karakteristik yang lebih spesifik dan jumlah observasi yang relatif lebih sedikit.

Hasil tersebut menunjukkan adanya pola penting, yaitu peningkatan kecepatan angin diikuti peningkatan daya aktif, tetapi tidak diikuti peningkatan efisiensi. Dari C1 ke C3, *Wind Speed* meningkat dari 6,4 menjadi 18,0 m/s dan *Active Power* meningkat dari 577,1 menjadi 6.122 kW. Namun, *Efficiency* justru turun dari 73,6% menjadi 54,7%. Pola ini menjadi temuan utama dalam karakterisasi berbasis *machine learning* karena menunjukkan bahwa *Active Power* tidak dapat digunakan sebagai indikator tunggal keberhasilan operasi turbin. Temuan tersebut melengkapi studi clustering SCADA yang menekankan

pengelompokan berdasarkan karakteristik output dan *wind speed* (Rodriguez dkk., 2023), dengan menambahkan dimensi *Efficiency* sebagai dasar interpretasi regime.

Hasil tersebut memperkuat bahwa *Active Power* saja tidak dapat digunakan sebagai indikator tunggal keberhasilan operasi turbin. Daya yang semakin besar memang menunjukkan meningkatnya energi listrik yang dihasilkan, tetapi efisiensi menunjukkan seberapa efektif energi yang tersedia dikonversi menjadi keluaran. Temuan ini sejalan dengan penelitian eksperimental yang menunjukkan adanya hubungan antara variasi daya turbin dan efisiensi generator pada kondisi operasi tertentu (Chalimah dkk., 2024).

3.3 Karakteristik *High-Efficiency Regime*

Cluster 1 (C1) terdiri atas 44 data, dengan rentang *Wind Speed* 2,1–10,9 m/s, *Active Power* 18–1.700 kW, dan *Efficiency* 66,5–78,8%. Nilai rata-rata efisiensinya mencapai sekitar 73,6%, sehingga C1 dikategorikan sebagai *high-efficiency regime*.

Karakteristik C1 menunjukkan bahwa kondisi operasi dengan kecepatan angin rendah hingga menengah justru menghasilkan efisiensi yang relatif tinggi. Hal tersebut terlihat dari posisi *Centroid* C1 pada 6,4 m/s dengan efisiensi 73,6%. Artinya, turbin tidak harus menghasilkan daya maksimum untuk mencapai efisiensi maksimum.

Rentang efisiensi C1 yang relatif tinggi dan sebarannya yang lebih sempit juga menunjukkan kondisi operasi yang relatif stabil. Hasil boxplot dalam penelitian menunjukkan median efisiensi C1 berada di sekitar 74%, dengan rentang data yang relatif sempit.

Secara teknis, kondisi tersebut dapat diinterpretasikan sebagai daerah operasi yang lebih sesuai dengan karakteristik kerja turbin. Pada kondisi ini, energi angin yang tersedia cukup untuk menghasilkan daya, tetapi belum menyebabkan peningkatan beban dan kerugian operasi seperti yang terlihat pada kecepatan angin tinggi. Interpretasi ini perlu dipandang sebagai karakteristik yang diperoleh dari dataset, bukan sebagai batas universal untuk semua jenis turbin.

Dengan demikian, C1 merupakan kelompok yang paling menguntungkan dari perspektif efisiensi. Temuan ini mendukung konsep bahwa evaluasi performa turbin perlu mempertimbangkan regime operasi, bukan hanya nilai daya maksimum. Penelitian tentang klasterisasi *wind speed* juga menunjukkan bahwa pengelompokan kondisi angin dapat membantu mengidentifikasi *wind regimes* yang berkaitan dengan diagnosis performa turbin (Martins dkk., 2024).

3.4 Karakteristik *Transition Regime*

Cluster 2 (C2) terdiri atas 16 data dengan *Wind Speed* 11,1–13,9 m/s, *Active Power* 1.866–3.252 kW, dan *Efficiency* 61,0–66,1%. Nilai rata-rata

efisiensinya sebesar 63,7%, sehingga cluster ini dikategorikan sebagai *transition regime*.

C2 memperlihatkan perubahan karakteristik yang berbeda dibandingkan C1. Ketika *Wind Speed* meningkat menjadi rata-rata 12,4 m/s, *Active Power* meningkat cukup signifikan hingga rata-rata 2.482 kW. Namun, efisiensi turun menjadi 63,7%. Hal ini menunjukkan bahwa peningkatan energi angin yang tersedia mulai menghasilkan tambahan daya yang besar, tetapi proporsi energi yang berhasil dikonversi menjadi keluaran tidak meningkat secara sebanding.

Kelompok ini dapat dipandang sebagai daerah transisi antara kondisi efisiensi tinggi dan rendah. Data pada C2 tidak lagi memiliki karakteristik efisiensi setinggi C1, tetapi belum menunjukkan penurunan efisiensi seperti C3. Karena itu, C2 penting dalam interpretasi hasil karena menunjukkan bahwa hubungan antara *Wind Speed* dan *Efficiency* bersifat tidak linier.

Secara operasional, penurunan efisiensi pada C2 dapat berkaitan dengan meningkatnya beban aerodinamis dan perubahan kondisi operasi sistem. Namun, karena dataset yang digunakan hanya mencakup *Wind Speed*, *Active Power*, dan *Efficiency*, faktor seperti temperatur, turbulensi, *tip speed ratio*, kondisi generator, dan rugi-rugi mekanis tidak dapat dibuktikan secara langsung dari penelitian ini.

Dengan demikian, C2 sebaiknya dipahami sebagai zona transisi performa, bukan sebagai kondisi kegagalan. Interpretasi tersebut penting agar hasil clustering tidak disalahartikan sebagai klasifikasi kerusakan. K-Means dalam penelitian ini digunakan untuk karakterisasi kondisi operasi berdasarkan kemiripan data, bukan untuk diagnosis kerusakan komponen.

Penurunan efisiensi pada kondisi tersebut dapat dikaitkan secara hipotetis dengan perubahan karakteristik operasi pada kecepatan angin tinggi. Penelitian terdahulu dapat memberikan penjelasan teknis mengenai kemungkinan pengaruh turbulensi, beban mekanis, atau keterbatasan pengendalian pada kondisi ekstrem, tetapi variabel penyebab tersebut tidak tersedia secara eksplisit dalam dataset penelitian ini. Oleh karena itu, interpretasi C3 sebagai *low-efficiency regime* merupakan hasil karakterisasi data dan tidak boleh dipahami sebagai bukti hubungan sebab-akibat antara kecepatan angin tinggi dan penurunan efisiensi.

3.5 Karakteristik Low-Efficiency Regime

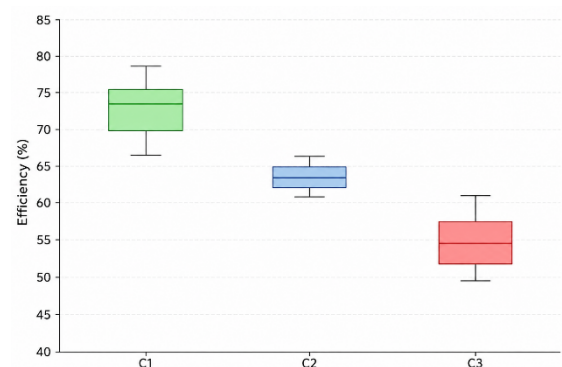
Cluster 3 (C3) terdiri atas 40 data, dengan *Wind Speed* 14,1–21,9 m/s, *Active Power* 3.366–9.372 kW, dan *Efficiency* 49,4–60,7%. Rata-rata efisiensi C3 sebesar 54,7%, sehingga cluster ini dikategorikan sebagai *low-efficiency regime*.

C3 memperlihatkan karakteristik yang berlawanan dengan C1. *Active Power* mencapai rata-rata 6.122 kW, jauh lebih besar dibandingkan C1 dan C2, tetapi efisiensi justru menjadi yang paling rendah. Hal ini menunjukkan bahwa daya keluaran tinggi tidak selalu identik dengan efisiensi tinggi.

Rentang *Wind Speed* C3 yang mencapai 14,1–21,9 m/s menunjukkan bahwa kelompok ini merepresentasikan kondisi angin tinggi. Visualisasi histogram juga menunjukkan bahwa C3 didominasi oleh *Active Power* 3.366–9.372 kW dan *Wind Speed* 14,1–21,9 m/s.

Penurunan efisiensi pada kondisi tersebut dalam penelitian sebelumnya dikaitkan dengan kemungkinan meningkatnya turbulensi, beban mekanis, dan keterbatasan sistem dalam mempertahankan kondisi operasi optimal. Namun, interpretasi tersebut harus ditempatkan sebagai penjelasan teknis yang mungkin, karena variabel penyebab seperti turbulensi dan beban mekanis tidak tersedia secara eksplisit dalam dataset penelitian.

Temuan C3 menjadi salah satu hasil penting penelitian karena memperlihatkan bahwa sistem dapat menghasilkan daya yang sangat tinggi ketika kecepatan angin tinggi, tetapi tingkat efisiensi justru menurun. Dengan demikian, karakterisasi berbasis clustering memberikan informasi yang lebih kaya dibandingkan analisis hubungan *Wind Speed*–*Power* saja.

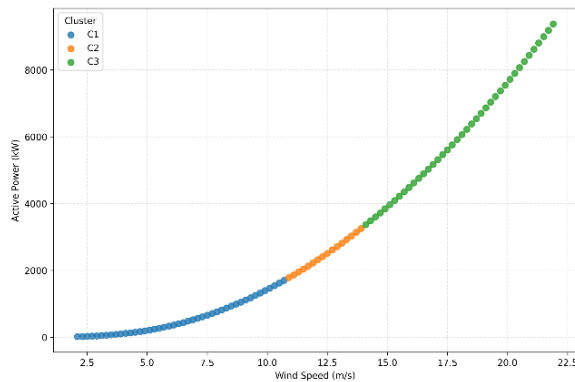


Gambar 5. Distribusi Efficiency pada Setiap Cluster

Sebagaimana ditunjukkan pada Gambar 5, *distribusi Efficiency* C1 berada pada tingkat yang lebih tinggi dibandingkan C2 dan C3

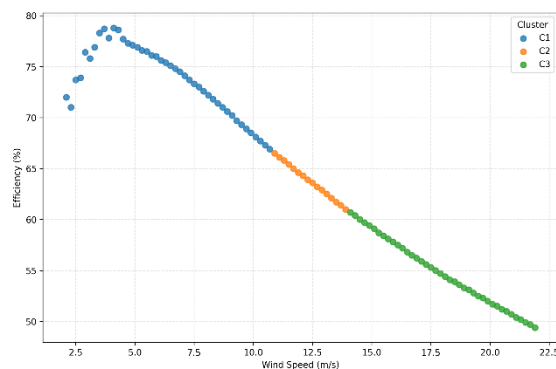
3.6 Analisis Pola Wind Speed, Active Power, dan Efficiency

Hasil clustering secara keseluruhan menunjukkan adanya pola yang konsisten dari C1 hingga C3. *Wind Speed* meningkat dari 6,4 m/s → 12,4 m/s → 18,0 m/s, sedangkan *Active Power* meningkat dari 577,1 kW → 2.482 kW → 6.122 kW. Sebaliknya, *Efficiency* menurun dari 73,6% → 63,7% → 54,7%. Hal ini ditunjukkan pada Gambar 6 dimana peningkatan *Wind Speed* menghasilkan peningkatan *Active Power*. Peningkatan *Wind Speed* tidak selalu menghasilkan peningkatan *Efficiency*.



Gambar 6. Wind Speed terhadap Active Power Berdasarkan Cluster

Pola tersebut (Gambar 6) menunjukkan hubungan yang menarik antara energi angin yang tersedia dan efisiensi konversi. Ketika kecepatan angin meningkat, daya keluaran meningkat secara nyata, tetapi peningkatan tersebut tidak berlangsung secara proporsional terhadap efisiensi. Dengan kata lain, turbin menghasilkan lebih banyak daya pada kondisi angin tinggi, tetapi persentase pemanfaatan energi relatif menurun.



Gambar 7. Wind Speed terhadap Efficiency Berdasarkan Cluster

Hasil *scatter plot* pada Gambar 7 juga memperlihatkan pola penurunan efisiensi dari kondisi angin rendah menuju tinggi. C1 berada pada kisaran kecepatan angin rendah dengan efisiensi sekitar 70–78%, C2 berada pada kecepatan menengah dengan efisiensi sekitar 60–67%, sedangkan C3 berada pada kecepatan tinggi dengan efisiensi yang lebih rendah.

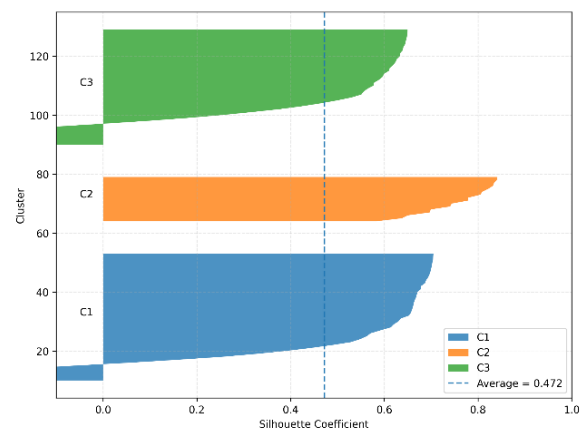
Temuan ini menunjukkan kelebihan pendekatan *Machine Learning-Based Efficiency Characterization*. Algoritma tidak hanya mengelompokkan data berdasarkan satu parameter, tetapi secara simultan mempertimbangkan tiga karakteristik operasi. Hasil akhirnya menghasilkan tiga *regime* yang memiliki makna teknis, yaitu *high-efficiency*, *transition*, dan *low-efficiency regime*.

Hasil ini juga memiliki kesesuaian dengan penelitian terdahulu yang menggunakan pendekatan *data-driven* untuk menganalisis operasi turbin. Rodriguez dkk. (2023) menunjukkan bahwa K-Means efektif dalam mengelompokkan turbin

berdasarkan karakteristik *output* dan *wind speed*, sedangkan Martins dkk. (2024) menunjukkan bahwa klusterisasi dapat digunakan untuk mengidentifikasi *wind regimes* yang relevan dengan diagnosis performa.

3.7 Validasi Kualitas Cluster

Kualitas hasil clustering dievaluasi menggunakan *Silhouette Score*. Hasil pengujian menghasilkan nilai 0,5863. Berdasarkan kriteria evaluasi yang digunakan dalam penelitian, nilai 0,50–0,70 termasuk kategori baik, sehingga nilai 0,5863 menunjukkan bahwa data memiliki tingkat kekompakan dan pemisahan cluster yang cukup baik.



Gambar 8. Silhoutte Analysis

Nilai tersebut menunjukkan bahwa tiga cluster yang terbentuk bukan sekadar hasil pembagian data secara arbitrer. Sebagian besar data memiliki karakteristik yang relatif lebih dekat dengan cluster tempat data tersebut berada dibandingkan dengan cluster lainnya. Hal ini memperkuat pemilihan $K=3$ yang sebelumnya diperoleh melalui *Elbow Method*.

Kesesuaian antara hasil *Elbow Method* dan *Silhouette Score* menjadi penting dalam penelitian ini. *Elbow Method* menunjukkan bahwa penurunan WCSS mulai melandai pada $K=3$, sedangkan *Silhouette Score* sebesar 0,5863 menunjukkan kualitas clustering dalam kategori baik. Kombinasi kedua evaluasi tersebut memberikan dasar yang lebih kuat untuk menggunakan tiga cluster sebagai representasi kondisi operasi turbin. Meskipun demikian, nilai *Silhouette* 0,5863 belum menunjukkan pemisahan yang sangat kuat; karena itu, hasil ini lebih tepat dipandang sebagai karakterisasi awal yang masih dapat ditingkatkan melalui dataset yang lebih besar, penambahan variabel operasi, dan perbandingan dengan algoritma clustering lainnya.

Dengan demikian, hasil validasi menunjukkan bahwa K-Means mampu membentuk kelompok yang cukup jelas berdasarkan kombinasi *Wind Speed*, *Active Power*, dan *Efficiency*. Pendekatan ini juga sejalan dengan penelitian berbasis K-Means pada data operasional turbin yang menunjukkan bahwa clustering dapat digunakan untuk mengidentifikasi

kelompok performa yang berbeda (Rodriguez dkk., 2023).

Secara keseluruhan, nilai *Silhouette Score* 0,5863 memperkuat bahwa hasil karakterisasi *high-efficiency*, *transition*, dan *low-efficiency regime* memiliki dasar statistik yang cukup baik. Namun, nilai tersebut belum menunjukkan pemisahan yang sangat kuat, sehingga penelitian lanjutan masih dapat meningkatkan kualitas model melalui penambahan jumlah data, variabel operasi, atau perbandingan dengan algoritma clustering lainnya.

3.8 Implikasi Hasil Karakterisasi Efisiensi

Hasil penelitian memberikan implikasi bahwa evaluasi performa turbin angin sebaiknya tidak hanya berorientasi pada besarnya daya listrik yang dihasilkan. C3 menghasilkan daya aktif paling tinggi, yaitu rata-rata 6.122 kW, tetapi memiliki efisiensi paling rendah, yaitu 54,7%. Sebaliknya, C1 menghasilkan daya lebih kecil, yaitu 577,1 kW, tetapi memiliki efisiensi tertinggi sebesar 73,6%.

Perbedaan tersebut menunjukkan bahwa operasi dengan daya maksimum belum tentu merupakan operasi paling efisien. Informasi ini dapat digunakan sebagai dasar evaluasi kondisi operasi turbin karena memungkinkan pengelola membedakan antara kondisi yang menghasilkan daya besar dan kondisi yang menghasilkan konversi energi secara lebih efektif.

Dari perspektif *machine learning*, hasil tersebut memperlihatkan bahwa K-Means dapat berfungsi sebagai alat karakterisasi awal kondisi operasi. Model menghasilkan kelompok berdasarkan pola data aktual, kemudian *Centroid* memberikan representasi numerik dari masing-masing kondisi. Pendekatan ini dapat dikembangkan lebih lanjut untuk sistem monitoring berbasis data dengan menambahkan data real-time dan variabel operasi lainnya.

Meskipun demikian, penelitian ini belum dapat menyatakan bahwa kecepatan angin tinggi secara langsung menyebabkan penurunan efisiensi. Dataset yang digunakan tidak mencakup seluruh faktor yang dapat memengaruhi efisiensi, seperti turbulensi, temperatur, arah angin, *tip speed ratio*, kondisi generator, dan rugi-rugi mekanis. Oleh karena itu, hubungan yang ditemukan harus dipahami sebagai pola karakteristik pada dataset penelitian, bukan hubungan sebab-akibat yang bersifat universal.

Kontribusi utama penelitian terletak pada pengembangan pendekatan *K-Means-based efficiency characterization* yang mengubah data operasi turbin menjadi tiga *efficiency regimes*. Hasil tersebut memperluas penggunaan K-Means dari sekadar teknik pengelompokan menjadi pendekatan interpretatif untuk memahami perbedaan kondisi efisiensi. Berbeda dengan penelitian Rodriguez dkk. (2023) yang menekankan eksplorasi SCADA untuk kelompok performa dan pemeliharaan prediktif, penelitian ini menggunakan *Centroid* tiga variabel sebagai dasar untuk menafsirkan *high-efficiency*, *transition*, dan *low-efficiency regime*. Dengan

demikian, novelty penelitian terletak pada integrasi *clustering* multivariat dan interpretasi *regime* efisiensi, bukan pada pengembangan algoritma K-Means baru.

Tabel 3. Ringkasan Hasil Utama

Parameter	C1 – <i>High Efficiency</i>	C2 – <i>Transition</i>	C3 – <i>Low Efficiency</i>
Jumlah data	44	16	40
<i>Wind Speed Centroid</i>	6,4 m/s	12,4 m/s	18,0 m/s
<i>Active Power Centroid</i>	577,1 kW	2.482 kW	6.122 kW
<i>Efficiency Centroid</i>	73,6%	63,7%	54,7%
Interpretasi	Kondisi paling efisien	Kondisi transisi	Kondisi efisiensi rendah

Hasil tersebut memperlihatkan tren utama penelitian: semakin tinggi *Wind Speed*, *Active Power* meningkat, tetapi *Efficiency* cenderung menurun. Pola inilah yang menjadi dasar utama bahwa pendekatan K-Means berhasil mengidentifikasi tiga *regime* operasi turbin angin yang berbeda.

4. KESIMPULAN

Berdasarkan hasil penelitian, penerapan *K-Means-based Efficiency Characterization* berhasil mengidentifikasi karakteristik kondisi operasi turbin angin berdasarkan *Wind Speed*, *Active Power*, dan *Efficiency*. Dari 100 data pengamatan, diperoleh tiga cluster optimal ($K=3$) berdasarkan *Elbow Method*, dengan WCSS sebesar 27,4509; 6,7612; dan 2,9752 untuk $K=1$, $K=2$, dan $K=3$. Tiga cluster tersebut dapat diinterpretasikan sebagai *high-efficiency*, *transition*, dan *low-efficiency regime* dengan *Centroid Efficiency* masing-masing sebesar 73,6%, 63,7%, dan 54,7%. Nilai *Silhouette Score* sebesar 0,5863 memperkuat bahwa pengelompokan berada pada kategori baik.

Temuan utama menunjukkan bahwa peningkatan *Wind Speed* berkaitan dengan peningkatan *Active Power*, tetapi tidak selalu menghasilkan *Efficiency* yang lebih tinggi. C1 pada *Wind Speed* 6,4 m/s memiliki efisiensi tertinggi sebesar 73,6%, sedangkan C3 pada 18,0 m/s menghasilkan *Active Power* tertinggi sebesar 6.122 kW tetapi efisiensi terendah sebesar 54,7%. Temuan ini menunjukkan kontribusi pendekatan *clustering* sebagai alat karakterisasi kondisi operasi, karena performa tidak hanya dinilai dari besarnya daya keluaran tetapi juga dari tingkat efisiensi.

Keterbatasan penelitian terutama terletak pada jumlah data yang masih terbatas, yaitu 100 pengamatan, serta variabel yang hanya mencakup *Wind Speed*, *Active Power*, dan *Efficiency*. Oleh karena itu, hasil *clustering* belum dapat menjelaskan faktor penyebab penurunan efisiensi secara spesifik dan tidak dapat digunakan sebagai diagnosis

kerusakan turbin. Penelitian selanjutnya disarankan menggunakan dataset yang lebih besar, periode pengamatan yang lebih panjang, parameter operasi tambahan seperti arah angin, turbulensi, kecepatan rotor, temperatur, dan *tip speed ratio*, serta membandingkan K-Means dengan metode *clustering* lainnya.

PUSTAKA

- Allal, Z., Noura, H. N., Vernier, F., Salman, O., & Chahine, K. (2024). Wind turbine fault detection and identification using a two-tier machine learning framework. *Intelligent Systems with Applications*, 22, 200372. <https://doi.org/10.1016/j.iswa.2024.200372>
- Astolfi, D., Pandit, R., Celesti, L., Lombardi, A., & Terzi, L. (2022). SCADA data analysis for long-term wind turbine performance assessment: A case study. *Sustainable Energy Technologies and Assessments*, 52, 102357. <https://doi.org/10.1016/j.seta.2022.102357>
- Baihaqi, M. D. (2023). *Rancang bangun sistem pemantauan efisiensi daya DC pada turbin angin (Studi kasus turbin angin sumbu vertikal Darrieus tipe H)*. Doctoral dissertation, Universitas Gadjah Mada.
- Chalimah, S., Setyobudi, R., Sadrina, A., Satria, M. H. B., & Sari, T. P. (2024). Pengaruh variasi daya turbin pada turbin angin poros vertikal terhadap efisiensi daya generator. *TESME*, 1(1), 78–85.
- Fajri, S., & Yuhendri, M. (2023). Monitoring pembangkit listrik tenaga angin menggunakan Human Machine Interface. *JTEIN: Jurnal Teknik Elektro Indonesia*, 4(1), 434–444. <https://doi.org/10.24036/jtein.v4i1.421>
- Kurniawati, I., Reynanda Bagus W. A., Akbar, R., & Rais, M. A. (2026). Evaluasi Performansi Propagasi LoRa E220-900T30D di Lingkungan Rural untuk Monitoring Turbin Angin Berbasis IoT. *Jurnal Teknik*, 18(1), 69–78. <https://doi.org/10.30736/jt.v18i1.1614>
- Martins, A. A., Vaz, D. C., Silva, T. A. N., Cardoso, M., & Carvalho, A. (2024). Clustering of wind speed time series as a tool for wind farm diagnosis. *Mathematical and Computational Applications*, 29(3), 35. <https://doi.org/10.3390/mca29030035>
- Marti-Puig, P., & Núñez-Vilaplana, C. (2024). Dynamic clustering of wind turbines using SCADA signal analysis. *Energies*, 17(11), 2514. <https://doi.org/10.3390/en17112514>
- Paik, C., Chung, Y., & Kim, Y. J. (2023). Power curve modeling of wind turbines through clustering-based outlier elimination. *Applied System Innovation*, 6(2), 41. <https://doi.org/10.3390/asi6020041>
- Peco Chacón, A. M., Segovia Ramirez, I., & García Márquez, F. P. (2023). False alarm detection in wind turbine by classification models. *Advances in Engineering Software*, 177, 103409. <https://doi.org/10.1016/j.advensoft.2023.103409>
- Rodriguez, P. C., Marti-Puig, P., Caiafa, C. F., Serra-Serra, M., Cusidó, J., & Solé-Casals, J. (2023). Exploratory analysis of SCADA data from wind turbines using the K-Means clustering algorithm for predictive maintenance purposes. *Machines*, 11(2), 270. <https://doi.org/10.3390/machines11020270>
- Sugeng, S., Nizar, T. N., Jatmiko, D. A., Hartono, R., & Kerlooza, Y. Y. (2024). Kalibrasi sensor monitoring cuaca pada area lokal untuk meningkatkan akurasi pada sensor biaya rendah. *Komputika: Jurnal Sistem Komputer*, 13(2), 277–287. <https://doi.org/10.34010/komputika.v13i2.13949>
- Sun, S., Hu, W., Liu, Y., Wang, T., & Chu, F. (2023). Matching contrastive learning: An effective and intelligent method for wind turbine fault diagnosis with imbalanced SCADA data. *Expert Systems with Applications*, 223, 119891. <https://doi.org/10.1016/j.eswa.2023.119891>
- Tian, X., Jiang, Y., Liang, C., Liu, C., Ying, Y., Wang, H., Zhang, D., & Qian, P. (2022). A novel condition monitoring method of wind turbines based on GMDH neural network. *Energies*, 15(18), 6717. <https://doi.org/10.3390/en15186717>
- Tyass, I., Khalili, T., Rafik, M., Abdelouahed, B., Raihani, A., & Mansouri, K. (2023). Wind speed prediction based on statistical and deep learning models. *International Journal of Renewable Energy Development*, 12(2), 288–299. <https://doi.org/10.14710/ijred.2023.48672>
- Wei, L., Wang, L., Liu, F., & Qian, Z. (2023). Clustering analysis of wind turbine alarm sequences based on domain knowledge-fused Word2vec. *Applied Sciences*, 13(18), 10114. <https://doi.org/10.3390/app131810114>
- Wiranata, D. T., Prahasto, T., & Widodo, A. (2021). Analisis prognostik terhadap kerusakan bantalan pada poros kecepatan tinggi turbin angin menggunakan machine learning dengan pendekatan Support Vector Regression (SVR). *Jurnal Teknik Mesin*, 9(2), 181–190.
- Wu, Z., Li, Y., & Wang, P. (2024). A hierarchical modeling strategy for condition monitoring and fault diagnosis of wind turbine using SCADA data. *Measurement*, 238, 114325. <https://doi.org/10.1016/j.measurement.2024.114325>
- Yandri, Y., Khwee, K. H., & Kurnianto, R. (2024). Studi komparatif turbin angin berdaya sama dengan kurva daya yang berbeda berbasis distribusi Weibull. *Jurnal Teknologi Elektro*, 15(1), 36–46. <https://doi.org/10.22441/jte.2024.v15i1.007>

- Yang, L., Wang, K., & Zhou, J. (2024). Multimode high-dimensional time series clustering and monitoring for wind turbine SCADA data. *Quality and Reliability Engineering International*, 40(8), 4285–4306. <https://doi.org/10.1002/qre.3626>
- Zhang, G., Li, Y., & Zhao, Y. (2023). A novel fault diagnosis method for wind turbine based on adaptive multivariate time-series convolutional network using SCADA data. *Advanced Engineering Informatics*, 57, 102031. <https://doi.org/10.1016/j.aei.2023.102031>