



Received 23rd May 2024
Accepted 23rd May 2024
Published 00th March 20xx

Open Access

DOI: 10.35472/x0xx0000

Implementasi *Clustering* dengan Algoritma *K-Means* Menggunakan *PySpark* untuk Analisis Data Meteran Listrik Rumah Tangga

Andrian Agustinus Lumban Gao¹, Aulia Wardani²,
Maria Gresia Hutabarat³, IDitta Winanda Putri⁴,
Ghozi Alvin Karim⁵

^a andrian.121450090@student.itera.ac.id

^b aulia.121450034@student.itera.ac.id

^c maria.121450057@student.itera.ac.id

^d ditta.121450006@student.itera.ac.id

^e ghozi.121450123@student.itera.ac.id

Abstract: *This research explores the use of electricity in households, which has increased with the development of technology and changes in modern lifestyles. Household electricity consumption is a crucial factor in the country's energy economy. With the increasing population and improving living standards, the demand of electricity continues to rise, necessitating an analysis of energy consumption patterns to ensure efficient use of electricity and cost reduction. In the era of big data, collecting energy consumption data can be done in real-time and processed using tools like PySpark. PySpark allows large-scale data processing and the implementation of machine learning, including the K-Means algorithm for clustering electricity consumption patterns. The results of this research are expected to provide detailed information on household electricity consumption patterns, helping the government and electricity providers design effective energy-saving strategies and support energy conservation and carbon emission reduction.*

Keywords: *Electricity, Households, PySpark, K-Means*

Abstrak: Penelitian ini menelusuri penggunaan listrik dalam rumah tangga yang mengalami peningkatan seiring dengan perkembangan teknologi dan perubahan gaya hidup modern. Konsumsi listrik rumah tangga menjadi faktor penting dalam perekonomian energi negara. Dengan bertambahnya populasi dan peningkatan taraf hidup, kebutuhan listrik terus meningkat, sehingga diperlukan analisis terhadap pola konsumsi energi untuk menjamin efisiensi penggunaan listrik dan pengurangan biaya. Di era *big data*, pengumpulan data konsumsi energi dapat dilakukan secara *real-time* dan diolah menggunakan alat seperti *PySpark*. *PySpark* memungkinkan pemrosesan data skala besar serta penerapan *machine learning*, termasuk algoritma *K-Means* untuk pengelompokan pola konsumsi daya listrik. Hasil penelitian ini diharapkan memberikan informasi rinci tentang pola konsumsi energi listrik rumah tangga, membantu pemerintah dan penyedia listrik merancang strategi penghematan energi yang efektif serta mendukung konservasi energi dan pengurangan emisi karbon.

Kata Kunci : *Listrik, Rumah Tangga, PySpark, Algoritma K-Means*





Pendahuluan

Latar Belakang

Penggunaan listrik dalam rumah tangga mengalami peningkatan yang signifikan seiring dengan berkembangnya teknologi dan perubahan gaya hidup modern [1]. Konsumsi listrik rumah tangga merupakan aspek penting dari perekonomian energi negara [2]. Seiring dengan bertambahnya jumlah penduduk dan peningkatan taraf hidup, kebutuhan akan listrik terus meningkat. Analisis mendalam terhadap pola konsumsi energi rumah tangga diperlukan untuk menjamin efisiensi energi dan efisiensi penggunaan listrik. Pemantauan dan analisis penggunaan listrik penting untuk mengoptimalkan efisiensi energi dan mengurangi biaya [3][4].

Ditengah era *big data* saat ini, pengumpulan data dari berbagai aktivitas dapat dilakukan secara *real-time*, termasuk dalam penggunaan energi listrik. Meski dengan volume yang besar, kini data dapat diolah dan dianalisis dengan cepat dan kompleks. Salah satu alat yang dapat mengolah *big data* adalah *PySpark*, sebuah kerangka kerja komputer yang memungkinkan pemrosesan data dalam skala besar dan dapat melibatkan penggunaan *machine learning* dalam pemrosesannya [5][6][7]. Sehingga analisis dapat dilakukan dengan lebih efisien.

Salah satu metode efektif untuk melakukan analisis ini adalah dengan menggunakan algoritma pengelompokan seperti *K-Means*. Algoritma ini memungkinkan kami mengelompokkan rumah tangga berdasarkan pola konsumsi listriknya, sehingga memudahkan merancang strategi penghematan energi yang lebih akurat dan efisien [8][9].

Hampir semua peralatan rumah tangga membutuhkan listrik, sehingga konsumsi saat ini tergantung pada kebutuhan masing-masing rumah tangga [10]. Seiring dengan meningkatnya konsumsi energi, pengelompokan data konsumsi listrik menjadi penting agar dapat dipantau penggunaannya. Hal ini membantu mengoptimalkan konsumsi daya. Banyak pengguna listrik yang tidak mengetahui berapa jumlah listrik yang dikonsumsi dalam jangka waktu tertentu. Oleh karena itu, diperlukan suatu sistem pendataan yang memberikan informasi penggunaan listrik secara jelas dan rinci.

Dalam studi ini akan dilakukan pengolahan data penggunaan listrik dengan menggunakan *PySpark* yang terintegrasi dengan *machine learning* untuk melakukan clustering menggunakan algoritma *K-Means* [11]. Fokus penelitian ini adalah menganalisis pola penggunaan listrik

rumah tangga dan mengelompokkannya berdasarkan kesamaan pola. Informasi ini dapat digunakan untuk merancang strategi penghematan energi dengan menyesuaikan pola konsumsi listrik pada masing-masing kelompok berdasarkan kebutuhannya [12].

Penelitian ini hanya berfokus pada analisis kelompok data daya listrik rumah tangga menggunakan algoritma *K-Means*. Data yang digunakan dalam penelitian ini adalah data meteran listrik rumah tangga yang telah disediakan, tanpa mempertimbangkan faktor-faktor lain seperti kondisi cuaca, jumlah penghuni atau jenis peralatan listrik yang digunakan. Penelitian ini tidak membahas implementasi strategi penghematan energi atau optimasi biaya secara spesifik, melainkan hanya memberikan wawasan tentang pola penggunaan energi. Analisis hanya menggunakan algoritma *K-Means* tanpa membandingkan dengan algoritma *clustering* lainnya.

Method / Metode

Data

Data yang kami gunakan pada clustering menggunakan algoritma *K-Means* ini adalah data meteran rumah tangga yang didapatkan dari pengukuran di sebuah rumah di Sceaux, Paris, Prancis yang tersedia di website UCI Machine Learning Repository. Data tersebut berisi informasi penggunaan listrik selama 47 bulan yang terhitung dari Desember 2006 sampai November 2010, dengan variabel atribut sebagai berikut:

NO	Atribut	Keterangan
1	date	Tanggal Pengukuran dalam format dd/mm/yyyy
2	time	Waktu pengukuran dalam format hh:mm:ss
3	global_active_power	Daya aktif total per menit rumah tangga (dalam kilowatt)
4	global_reactive_power	Daya reaktif total per menit rumah tangga (dalam kilovar)
5	voltage	Tegangan rata-rata per



		menit (dalam volt)
6	global_intensity	Intensitas arus total per menit rumah tangga (dalam ampere)
7	sub_metering_1	Energi aktif sub-meteran No. 1 (dalam watt-jam) untuk area dapur, meliputi mesin pencuci piring, oven, dan <i>microwave</i> .
8	sub_metering_2	Energi aktif sub-meteran No. 2 (dalam watt-jam) untuk area <i>laundry</i> , meliputi mesin cuci, pengering, kulkas, dan penerangan.
9	sub_metering_3	Energi aktif sub-meteran No. 3 (dalam watt-jam) untuk pemanas air listrik dan AC.

Tabel 1. Tabel Deskripsi Variabel Atribut Dataset

Data ini akan diolah menggunakan *PySpark* pada lingkungan *Apache Spark* untuk memungkinkan pengolahan data skala besar secara cepat dan efisien.

Preprocessing Data

Sebelum mengaplikasikan algoritma *K-Means*, dilakukan pemilihan variabel yang relevan, transformasi data, serta pembersihan data. Dilakukan pemilihan variabel yang relevan untuk menggambarkan pola pemakaian listrik pada area-area tertentu. Dari variabel tersebut dilakukan transformasi data untuk menyesuaikannya dengan kebutuhan model nanti. Selain itu dilakukan pembersihan data kosong pada setiap variabel [13].

Implementasi K-Means Clustering

Pada pengelompokan penggunaan daya listrik berdasarkan pola pemakaiannya, digunakan algoritma *K-Means clustering* dari *library machine learning* pada *PySpark*. *K-Means Clustering* sendiri adalah pengelompokan data didasari pemakaian konsumsi listrik di beberapa area, yaitu dapur, *laundry*, dan penggunaan pemanas air listrik dan ac.

Sebelum pengelompokan dilakukan penentuan jumlah *cluster* dengan menggunakan *Elbow Method*, yaitu pemantauan nilai *inertia* yang divisualisasikan dalam sebuah plot [14]. Jumlah cluster terbaik ditentukan dari titik siku atau "*elbow point*" pada plot yang menggambarkan penurunan nilai *inertia* yang mulai melambat secara signifikan. Penurunan ini menunjukkan penambahan nilai cluster tidak lagi memberikan pemisahan yang berarti [15][16]. *Inertia* sendiri adalah ukuran yang menggambarkan seberapa tersebar titik-titik data dalam sebuah *cluster* [17]. Semakin kecil *inertia*, maka semakin baik pemisahan data dalam *cluster* [18]. *Inertia* dalam *K-Means* dirumuskan sebagai berikut

$$I = \sum_{i=1}^K \sum_{x \in C_i} ||x - \mu_i||^2 \quad (1)$$

Keterangan:

I = Inertia

K = Jumlah Klaster

C_i = Klaster ke- i

x = Titik dalam Klaster C_i

μ_i = Pusat Klaster C_i

Dari hasil *inertia* tersebut, dibentuk plot garis yang menggambarkan penurunan *inertia* pada setiap jumlah *cluster*. Penurunan yang signifikan menandakan *cluster* membantu mengelompokkan data dengan baik. Sebaliknya, penurunan yang tidak signifikan menunjukkan tidak adanya perbedaan yang berarti dalam pengelompokan data.

Setelah mendapatkan jumlah *cluster* yang optimal, dilakukan penerapan algoritma *K-Means*, dimana algoritma akan menentukan titik *centroid* (pusat klaster) dari masing-masing kelompok secara acak [19]. Secara sistematis pusat *cluster* dapat dihitung dengan rumus sebagai berikut

$$\mu_1 = \frac{1}{n_i} \sum_{x \in C_i} x \quad (2)$$

Keterangan:

μ_i = Pusat *cluster*

n_i = Jumlah titik dalam *cluster* ke- i

x = Titik data dalam *cluster* C_i

Dari titik *centroid* tersebut dilakukan pengelompokan data-data terdekat dari pusat *cluster*. Jarak antar data ke titik pusat *cluster* dapat menggunakan persamaan Euclidean dengan rumus sebagai berikut

$$d = \sqrt{\sum_{i=1}^n (x - \mu_i)^2}$$

(3)

Keterangan:

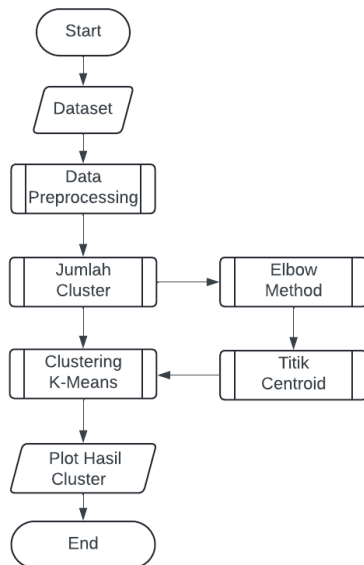
d = Jarak data ke pusat *cluster*

x_i = Data ke- i

μ_i = Pusat *cluster* C_i

Dari pusat serta persebaran data dalam sebuah *cluster* tersebut, dibentuk sebuah plot yang memberikan gambaran tentang persebaran data pada setiap *cluster* menurut variabelnya.

Flowchart



Gambar 1. Alur Penelitian

Hasil Dan Pembahasan

Untuk pemrosesan data, semua dilakukan dengan Apache Spark, yang diakses melalui Python dengan library PySpark. Penghubungan Spark dengan lingkungan Python dilakukan dengan menginisialisasi sesi Spark dengan menggunakan objek *builder* dari PySpark

Data

Data yang digunakan dalam percobaan ini adalah data meteran rumah tangga yang didapatkan dari pengukuran di sebuah rumah di Sceaux, Paris, Prancis yang tersedia di website UCI Machine Learning Repository. Data tersebut berisi informasi penggunaan listrik selama 47 bulan yang terhitung dari Desember 2006 sampai November 2010, dengan total jumlah data secara keseluruhan adalah 1.048.575 data.

Date	Time	Global_active_power	Global_reactive_power	Voltage
16/12/2006	2024-05-23 17:24:00	4.216	0.418	234.840
16/12/2006	2024-05-23 17:25:00	5.360	0.436	233.630
16/12/2006	2024-05-23 17:26:00	5.374	0.498	233.290
16/12/2006	2024-05-23 17:27:00	5.388	0.502	233.740
16/12/2006	2024-05-23 17:28:00	3.666	0.528	235.680

Voltage	Global_intensity	Sub_metering_1	Sub_metering_2	Sub_metering_3
234.840	18.400	0.000	1.000	17.0
233.630	23.000	0.000	1.000	16.0
233.290	23.000	0.000	2.000	17.0
233.740	23.000	0.000	1.000	17.0
235.680	15.800	0.000	1.000	17.0

Gambar 2. Dataframe Konsumsi Tenaga Listrik Rumah Tangga Perorangan

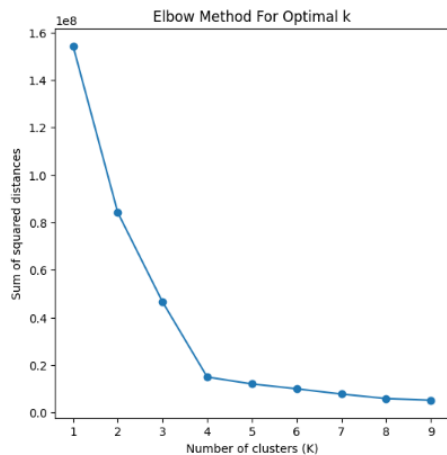
Kemudian dilakukan pemilihan variabel, dimana variable yang digunakan berfokus pada energi aktif sub-meteran 1, 2 dan 3 yang merepresentasikan konsumsi listrik di beberapa area, yaitu dapur, *laundry*, dan penggunaan pemanas air listrik dan ac. Variabel tersebut kemudian ditransformasi ke bentuk desimal agar model dapat membaca nilai tersebut. Kemudian dilakukan pemeriksaan nilai kosong, dan terdapat 4069 data kosong yang kemudian diisi dengan nilai rata-rata dari variabel tersebut. Setelah itu dilakukan pembentukan variabel baru yang berisikan vektor dari gabungan tiga variabel. Pembentukan vektor ini dibutuhkan dalam menggunakan *machine learning* pada PySpark agar membantu efisiensi komputasi.

Sub_metering_1	Sub_metering_2	Sub_metering_3	features
0.0	1.0	17.0	[0.0,1.0,17.0]
0.0	1.0	16.0	[0.0,1.0,16.0]
0.0	2.0	17.0	[0.0,2.0,17.0]
0.0	1.0	17.0	[0.0,1.0,17.0]
0.0	1.0	17.0	[0.0,1.0,17.0]
0.0	2.0	17.0	[0.0,2.0,17.0]
0.0	1.0	17.0	[0.0,1.0,17.0]
0.0	1.0	17.0	[0.0,1.0,17.0]
0.0	2.0	16.0	[0.0,2.0,16.0]
0.0	1.0	17.0	[0.0,1.0,17.0]
0.0	1.0	17.0	[0.0,1.0,17.0]
0.0	1.0	16.0	[0.0,1.0,16.0]
0.0	2.0	17.0	[0.0,2.0,17.0]
0.0	1.0	17.0	[0.0,1.0,17.0]
0.0	0.0	17.0	[0.0,0.0,17.0]
0.0	0.0	17.0	[0.0,0.0,17.0]
0.0	0.0	17.0	[0.0,0.0,17.0]
0.0	0.0	18.0	[0.0,0.0,18.0]
0.0	0.0	17.0	[0.0,0.0,17.0]

Gambar 3. DataFrame Cluster Konsumsi Tenaga Listrik Rumah Tangga Perorangan

Implementasi K-Means Clustering

Sebelum pemrosesan data pada model *K-Means*, dilakukan penentuan jumlah *cluster* (*k*) yang optimal dengan menggunakan Metode Elbow [20][21]. Nilai *K* yang diuji berada pada rentang 1 sampai 10. Model *K-Means* kemudian dilatih dengan nilai *K* yang berbeda beda dan maksimal 100 kali perulangan untuk mencapai nilai *inertia* yang cukup stabil atau konvergen (perubahan yang tidak signifikan). Dari pelatihan tersebut didapatkan plot yang menunjukkan nilai *inertia* di setiap percobaan nilai *K*.



Gambar 4. Grafik Nilai *Inertia* terhadap nilai *K* dengan *Elbow Method*

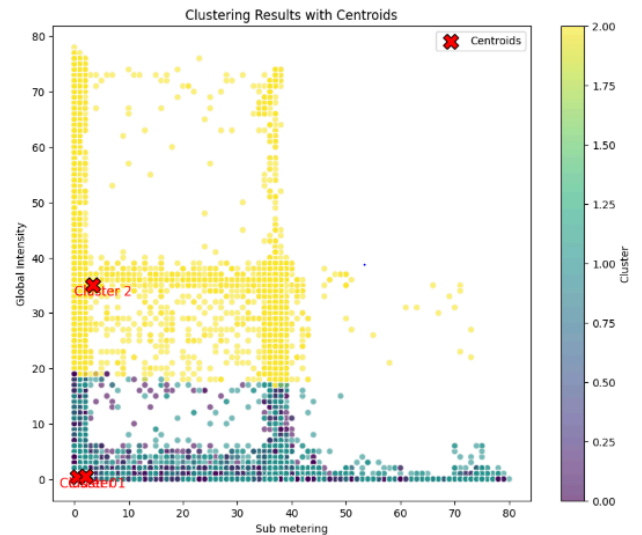
Berdasarkan plot *inertia* terhadap nilai *K*, dapat dilihat saat *K* berada pada nilai 4 sampai 9 terjadi perlambatan penurunan nilai *inertia*. Sementara itu, penurunan nilai *inertia* yang signifikan terakhir terjadi saat *K* dengan nilai 3. Titik ini adalah *Elbow Point*, yang menandakan nilai *K* tersebut adalah nilai *K* yang paling optimal untuk digunakan pada algoritma *K-Means* dengan data ini.

Setelah mendapatkan nilai *K* yang optimal, dilakukan pengaplikasian *K-Means* pada data, untuk menentukan *cluster* pada setiap data [22][23][24]. *K-Means* pada percobaan ini akan menentukan 3 pusat *cluster* pada 3 variabel, yang berarti membentuk *cluster* 3 dimensi. Berdasarkan penerapan didapatkan pusat *cluster* sebagai berikut

Cluster	sub_metering_1	sub_metering_2	sub_metering_3
1	0.60644932	0.40417119	0.26722586
2	2.18255066	0.49945946	17.55722148
3	3.36820389	35.06604881	10.68403847

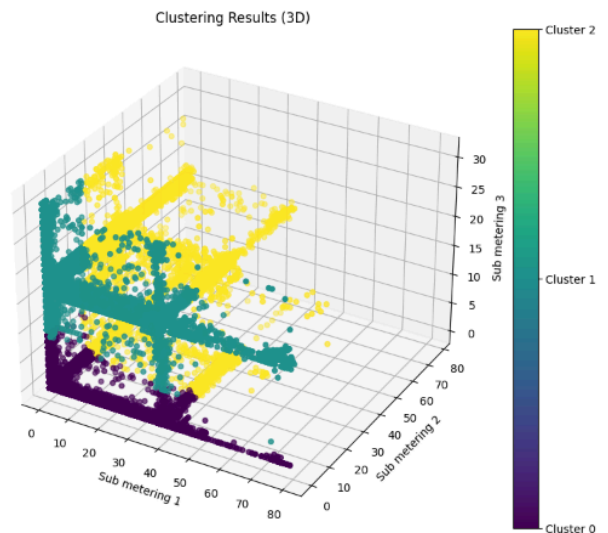
Tabel 2. Tabel *Cluster Centers* Sub-metering

Setelah menemukan ketiga pusat *cluster*, dilakukan pengelompokan data berdasarkan ketiga pusat *cluster* tersebut, dan didapatkan persebaran data yang dapat dilihat dari grafik titik berikut.



Gambar 5 Grafik Hasil Pengelompokan dengan *Centroids*

Dari grafik tersebut dapat dilihat persebaran data sesuai pusat *cluster* yang sudah ditemukan. Pengelompokan data dapat dilihat berdasarkan ketiga warna yang dibentuk, dengan pusat *cluster* yang bertandakan "x". Untuk gambaran persebaran data yang lebih jelas dapat dilihat pada grafik tiga dimensi berikut



Gambar 6 Visualisasi Cluster 3D

Visualisasi hasil *Cluster* pada **Gambar 6**. data konsumsi listrik rumah tangga individu, menunjukkan pola pengelompokan berdasarkan tiga sub-metering variabel, yaitu sub-metering 1, sub-metering 2, dan sub-metering 3. Data tersebut dikelompokkan dengan menggunakan algoritma *K-Means* menjadi tiga *cluster* atau kelompok yang ditampilkan pada **Gambar 6**. dengan grafik 3D. *Cluster* ini ditandai dengan warna berbeda yaitu, ungu (*cluster* 0), hijau (*cluster* 1), dan kuning (*cluster* 2).

Cluster 0 (ungu) mengelompokkan rumah tangga dengan konsumsi energi yang relatif rendah untuk semua sub-metering. Rumah tangga dalam kelompok ini memiliki kebiasaan konsumsi listrik yang lebih baik dengan asumsi penggunaan peralatan elektronik dapur, laundry, pemanas air dan pendingin udara yang cukup minim.

Cluster 1 (hijau) mengelompokkan rumah tangga dengan konsumsi energi yang lebih bervariasi pada ketiga sub-metering. Namun terdapat penggunaan energi yang cukup besar pada sub-metering 3, yaitu peralatan pemanas air dan pendingin udara. Berdasarkan hal tersebut dapat diasumsikan penggunaan perangkat elektronik pada dapur dan peralatan *laundry* rumah tangga dalam kelompok ini relatif lebih minim dan lebih banyak menggunakan pemanas air dan pendingin udara.

Cluster 2 (kuning) mengelompokkan rumah tangga dengan konsumsi energi tinggi untuk semua sub-metering variabel. Namun pada sub-metering 1, penggunaan daya masih relatif rendah. Berdasarkan hal tersebut dapat diasumsikan penggunaan perangkat elektronik khususnya pada peralatan *laundry*, pemanas air, dan pendingin udara cukup intensif. Namun, penggunaan peralatan elektronik dapur relatif rendah namun cukup signifikan pada kelompok rumah tangga ini.

Penggunaan algoritma *K-Means* efektif untuk mengelompokkan rumah tangga berdasarkan pola konsumsi energinya. Hasil pengelompokan ini dapat digunakan untuk menyusun strategi penghematan energi yang lebih akurat dan efisien [25]. Contohnya, rumah tangga di kelompok 0 mungkin menerima saran penghematan energi sederhana, sementara rumah tangga di kelompok 2 mungkin memerlukan tindakan yang lebih efektif untuk mengurangi konsumsi listrik. Secara keseluruhan, kelompok ini membantu memahami pola konsumsi energi dan mendukung konservasi energi dan pengurangan karbon.

Conclusions / Kesimpulan

Penelitian ini membahas implementasi *clustering* dengan algoritma *K-Means* menggunakan PySpark untuk analisis data meteran listrik rumah tangga. Data yang digunakan dalam penelitian kami merupakan data meteran listrik rumah tangga yang didapatkan dari pengukuran di sebuah rumah di Sceaux, Paris, Prancis yang tersedia di website UCI *Machine Learning Repository*. Data tersebut berisi informasi penggunaan listrik selama 47 bulan yang terhitung dari Desember 2006 sampai November 2010, dengan total jumlah data secara keseluruhan adalah 1.048.575 data.

Dalam penelitian ini, dilakukan *preprocessing data* untuk memilih variabel yang relevan, transformasi data, serta pembersihan data. Variabel yang digunakan berfokus pada energi aktif sub-meteran 1, sub-metering 2, hingga sub-metering 3 yang merepresentasikan konsumsi listrik di beberapa area, yaitu dapur, *laundry*, dan penggunaan pemanas air listrik dan ac.

Dari hasil analisis, diperoleh 3 (tiga) *cluster* yang optimal untuk mengelompokkan data konsumsi listrik rumah tangga. Hasil analisis menunjukkan bahwa *cluster* pertama memiliki pola konsumsi listrik yang relatif stabil dan rendah, *cluster* kedua memiliki pola konsumsi listrik yang relatif tinggi dan tidak stabil, dan *cluster* ketiga memiliki pola konsumsi listrik yang relatif tinggi dan stabil.

Dengan demikian, penelitian ini menunjukkan bahwa algoritma *K-Means* dapat digunakan untuk mengelompokkan data konsumsi listrik rumah tangga berdasarkan pola konsumsi listriknya. Hasil analisis dapat digunakan untuk merancang strategi penghematan energi yang lebih akurat dan efisien berdasarkan kebutuhan masing-masing kelompok.

Conflicts of Interest / Konflik Kepentingan

Dalam analisis data meteran rumah tangga ini, tidak ditemukannya *conflicts of interest* yang berpotensi mempengaruhi hasil analisis. sebab apabila ditinjau dari segi pola konsumsi listrik rumah tangga, menggunakan algoritma *K-Means*, tidak ditemukannya keterkaitan dengan kepentingan lain yang dapat mempengaruhi hasilnya.

Demikian juga pada analisis tidak terikat suatu organisasi atau individu yang mempengaruhi hasil analisis. Sehingga dapat diambil kesimpulan bahwa tidak terjadi konflik kepentingan dalam hasil penelitian tersebut.

Acknowledgements

Pada akhirnya, kami mengucapkan terima kasih atas dukungan dari Bapak dan Ibu Dosen Analisis Big Data, dan rekan rekan, yang telah memberikan bimbingan, dukungan, selama pengerjaan tugas besar ini berlangsung. Terakhir

kami berharap penelitian mengenai *"Implementasi Clustering Dengan Algoritma K-Means Menggunakan Pyspark Untuk Analisis Data Meteran Listrik Rumah Tangga"* dapat memberikan manfaat bagi pembacanya.

References

- [1] R. Rizaldy, T. Nurlambang, M. H. Dewi Susilowati, and H. Anggrahita, "Differences in electricity consumption based on electricity user sectors in Pramuka Island, Panggang Island and Tidung Island of Kepulauan Seribu Regency, DKI Jakarta," IOP Conference Series: Earth and Environmental Science, vol. 649, no. 1, p. 012011, Feb. 2021, doi: <https://doi.org/10.1088/1755-1315/649/1/012011>.
- [2] Batubara, I.F., & Lubis, F.Z. (2023). CLUSTERING DATA PELANGGAN PLN HELVETIA MENGGUNAKAN K-MEANS CLUSTER. Kohesi: Jurnal Sains dan Teknologi, 2(1), 81-90.
- [3] F. Karanfil and Y. Li, "Electricity consumption and economic growth: Exploring panel-specific differences," Energy Policy, vol. 82, pp. 264–277, Jul. 2015, doi: <https://doi.org/10.1016/j.enpol.2014.12.001>.
- [4] X. Xu and C. Chen, "Energy efficiency and energy justice for U.S. low-income households: An analysis of multifaceted challenges and potential," Energy Policy, vol. 128, pp. 763–774, May 2019, doi: <https://doi.org/10.1016/j.enpol.2019.01.020>.
- [5] V. Riabchuk, L. Hagel, F. Germaine, and A. Zharova, "Utility-Based Context-Aware Multi-Agent Recommendation System for Energy Efficiency in Residential Buildings," arXiv preprint arXiv:2205.02704, 2022. [Online]. Available: <https://doi.org/10.48550/arXiv.2205.02704>.
- [6] H. Kazmi and P. Pinson, "On the contribution of pre-trained models to accuracy and utility in modeling distributed energy resources," arXiv preprint arXiv:2302.11679, 2023. [Online]. Available: <https://doi.org/10.48550/arXiv.2302.11679>.
- [7] S. Moayed et al., "Energy Optimization Technologies in Smart Homes," IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society, Singapore, 2020, pp. 1974–1979, doi: 10.1109/IECON43393.2020.9254345. keywords: {Monitoring;Wireless sensor networks;Energy consumption;Smart homes;Wireless communication;Optimization;Internet of Things;Energy management;Smart buildings;IoT;STDR;ILM;NILM;Machine learning;Energy-storage},
- [8] M. Zaharia et al., "Apache Spark," Communications of the ACM, vol. 59, no. 11, pp. 56–65, Oct. 2016, doi: <https://doi.org/10.1145/2934664>.
- [9] J. Li and R. E. Just, "Modeling household energy consumption and adoption of energy efficient technology," Energy Economics, vol. 72, pp. 1–11, 2018. [Online]. Available: <https://doi.org/10.1016/j.eneco.2018.04.019>.
- [10] D. Oliveira, E. M. G. Rodrigues, T. D. P. Mendes, J. P. S. Catalão, and E. Pouresmaeil, "Model predictive control technique for energy optimization in residential appliances," in 2015 IEEE International Conference on Smart Energy Grid Engineering (SEGE), Oshawa, ON, Canada, 2015, pp. 1–6. doi: 10.1109/SEGE.2015.7324578.
- [11] H. W. Choi, N. M. F. Qureshi, and D. R. Shin, "Analysis of Electricity Consumption at Home Using K-means Clustering Algorithm," in 2019 21st International Conference on Advanced Communication Technology (ICACT), PyeongChang, Korea (South), 2019, pp. 1–5. doi: 10.23919/ICACT.2019.8701981.
- [12] M. N. Morgül Tumbaz and H. Tekiner Moğulkoç, "Profiling energy efficiency tendency: A case for Turkish households," Energy Policy, vol. 118, pp. 35–46, 2018. doi: 10.1016/j.enpol.2018.04.064.
- [13] Y. Amri, A. L. Fadhilah, Fatmawati, N. Setiani, and S. Rani, "Analysis clustering of electricity usage profile using K-means algorithm," IOP Conference Series: Materials Science and Engineering, vol. 105, no. 1, p. 012020, 2016. doi: 10.1088/1757-899X/105/1/012020.
- [14] S. Hendrawan, F. K. S. Dewi, dan Pranowo, "Clustering evaluasi dosen Universitas Atma Jaya Yogyakarta menggunakan metode K-means," Jurnal Ilmiah Ahli Informatika JIAJ, vol. 4, no. 1, 2020. doi: 10.24002/jiaj.v4i1.7436.
- [15] S. Shin et al., "Clustering and prediction of long-term functional recovery patterns in first-time stroke patients," Frontiers in Neurology, vol. 14, p. 1130236, Mar. 2023. doi: 10.3389/fneur.2023.1130236.
- [16] E. Schubert, "Stop using the elbow criterion for k-means and how to choose the number of clusters instead," ACM SIGKDD Explorations Newsletter, vol. 25, no. 1, pp. 36–42, 2023. doi: 10.1145/3606274.3606278.
- [17] F. Afrianto, M. S. Roychansyah, dan Y. Herwangi, "Bagaimana kompleksitas metropolitan di Indonesia? Sebuah pengenalan awal dimensi fraktal," Pranatacara Bhumandala: Jurnal Riset Planologi, vol. 4, no. 2, 2023. doi: 10.32795/pranatacara_bhumandala.v4i2.4956.
- [18] R. Ishak dan Amiruddin, "Clustering prestasi akademik lulusan menggunakan metode K-Means," Jambura Journal of Electrical and Electronics Engineering, vol. 6, no. 1, Januari 2024, e-ISSN: 2715-0887, p-ISSN: 2654-7813.
- [19] N. Wakhidah, "CLUSTERING MENGGUNAKAN K-MEANS ALGORITHM," Jurnal Transformatika, vol. 8, no. 1, p. 33, Jul. 2010, doi: <https://doi.org/10.26623/transformatika.v8i1.45>.
- [20] M. Selimi, L. Cerda-Alabern, L. Wang, A. Sathiaselalan, L. Veiga, dan F. Freitag, "Bandwidth-aware service placement in community network clouds," arXiv:1603.08119 [cs.NI], 2016. doi: 10.48550/arXiv.1603.08119.
- [21] A. Yulistira dan R. Andika, "Pengelompokan data nilai siswa menggunakan metode K-Means Clustering," Jurnal Aplikasi Ilmu Teknologi Informasi, vol. 1, no. 1, 2020. doi: 10.58602/jaiti.v1i1.22.
- [22] A. P. Riani, A. Voutama, dan T. Ridwan, "Penerapan K-Means Clustering dalam pengelompokan hasil belajar peserta didik dengan metode elbow," Jurnal Singaperbangsa Karawang, vol. 6, no. 1, 2020. doi: 10.53513/jsk.v6i1.7351.
- [23] Y. R. Nasution dan M. Eka, "Penerapan algoritma K-Means Clustering pada aplikasi menentukan berat badan ideal," Algoritma, vol. 2, no. 1, 2020. doi: 10.30829/algoritma.v2i1.1620.
- [24] N. Sharma, P. Sharma, dan K. Tiwari, "A review on clustering method based on unsupervised learning approach," International Journal of Computer Applications, vol. 181, no. 19, September 2018.
- [25] Li, K., Wang, F., Zhen, Z., Mi, Z., Sun, H., Liu, C., Wang, B., & Lu, J. (2016). Analysis on residential electricity consumption behavior using improved k-means based on simulated annealing algorithm. In 2016 IEEE Power and Energy Conference at Illinois (PECI) (pp. 1–6). IEEE. DOI: 10.1109/PECI.2016.7459209.

Lampiran

https://colab.research.google.com/drive/12xJfVWQpVkJbY_vBioSeyMdXOfAUyQF?usp=sharing#scrollTo=OKCn66tnVTfs