

Pengelompokan Data Pengeluaran Bulanan pada Aplikasi Pencatatan Keuangan Pribadi menggunakan Metode *K-Means Clustering*

¹Agung Riyanto, ²Mutaqin Akbar
^{1,2}Universitas Mercu Buana Yogyakarta

Alamat Surat

Email: ¹agungriyantor@gmail.com*, ²mutaqin@mercubuana-yogya.ac.id

Article History:

Diajukan: 14 September 2024; **Direvisi:** 2 Oktober 2024; **Accepted:** 25 Oktober 2024

ABSTRAK

Pengelolaan keuangan pribadi menjadi sangat penting dalam era modern yang kompleks saat ini, faktor gaya hidup yang berlebihan dan kurangnya pemahaman seseorang terhadap finansial membuat aplikasi Pencatatan Keuangan Pribadi menjadi salah satu peran penting dalam membantu mengelola pengeluaran pribadi, namun data pengeluaran pada aplikasi pencatatan keuangan pribadi masih terkesan rancu untuk membedakan pengeluaran prioritas dan bukan prioritas. Pada penelitian ini, peneliti mengimplementasi pengelompokan data pengeluaran bulanan menggunakan metode *K-Means Clustering*. Data yang digunakan adalah 30 data training dan 25 data uji yang diambil dari data aplikasi Pencatatan Keuangan Pribadi (CaKeP). Pada proses clustering menggunakan *K-Means Clustering* menghasilkan 3 kelompok yaitu pengeluaran rendah, sedang, dan tinggi dengan menghasilkan score 72% menggunakan *Silhouette Score*.

Kata kunci: pengelompokan pengeluaran bulanan, data mining, k-means clustering, pencatatan keuangan pribadi

ABSTRACT

Personal financial management is becoming crucial in today's complex modern era, excessive lifestyle factors and a lack of understanding of finance make the Personal Financial Records application one of the key roles in helping manage personal expenditure, but the expenditures data on the personal financial records applications are still impressed with ranks to distinguish priority and non-priority accounting. In this study, the researchers implemented the grouping of monthly expenditure data using the K-Means Clustering method. The data used is 30 training data and 25 test data taken from the Personal Financial Records application data. (CaKeP). In the process of clustering using K-Means Clustering produces three groups of low, medium, and high outputs with a 72% score using Silhouette Score.

Keywords: *monthly expenditure grouping, data mining, k-means clustering, personal financial records*

1. PENDAHULUAN

Pengelolaan keuangan pribadi yang baik merupakan suatu hal yang sangat penting untuk dicapai oleh setiap individu, terutama dalam era modern saat ini yang serba kompleks. Namun faktor gaya hidup berlebihan serta minimnya pemahaman mengenai finansial membuat seseorang sulit mengatur keuangannya. Aplikasi Pencatatan Keuangan Pribadi menjadi salah satu peran penting dalam membantu mengelola pengeluaran pribadi, namun data pengeluaran pada aplikasi pencatatan

keuangan pribadi masih terkesan rancu untuk membedakan pengeluaran prioritas dan bukan prioritas. Oleh karena itu, diperlukan suatu metode untuk mengelompokkan data pengeluaran bulanan agar mudah dipahami dan dikelola.

Terdapat beberapa metode yang dapat digunakan untuk mengelompokkan data khususnya pengeluaran keuangan pribadi yang memiliki kelebihan dan kekurangan masing-masing yaitu *K-Means Clustering* & *Algoritma Hierarchical Clustering*. Kelebihan dari *k-means clustering* yaitu sederhana dan mudah dipahami, cepat dan efisien dalam komputasi, dan cocok untuk data dengan dimensi yang tinggi, dan hasil clustering yang dihasilkan stabil ketika dijalankan beberapa kali dengan inisialisasi *centroid* yang berbeda (Alfina et al, 2012), sedangkan kelebihan dari *hierarchical clustering* yaitu hasil klaster dapat berbentuk *dendogram* sehingga dapat memvisualisasikan hubungan antar data, tidak perlu menentukan jumlah klaster sebelumnya dengan ukuran yang berbeda-beda (Aggarwal & Reddy, 2013).

Pada permasalahan di atas, peneliti mengimplementasikan Metode *K-Means Clustering* untuk mengelompokkan data pengeluaran bulanan menjadi beberapa kelompok. K-Means dipilih karena kelebihannya dalam menangani data dengan dimensi yang tinggi, mudah dipahami dan cepat dalam komputasi (Tan, Steinbach & Kumar, 2005).

1.1 Pengeluaran Bulanan

Pengeluaran bulanan adalah pengeluaran yang dilakukan dalam periode satu bulan. Pengeluaran bulanan dapat mencakup berbagai jenis pengeluaran, seperti pengeluaran untuk makanan, transportasi, tagihan listrik, tagihan internet, dan lain sebagainya (Adebayo, Adebayo, Oyebo, 2016).

1.2 Aplikasi Pencatatan Keuangan Pribadi

Aplikasi pencatatan keuangan pribadi adalah aplikasi yang digunakan untuk mencatat pengeluaran dan pemasukan pribadi. Aplikasi ini dapat membantu pengguna dalam mengelola keuangan pribadi dengan lebih efektif (Fadhilah & Nasution, 2020).

2. METODE

2.1 Data Mining

Data mining adalah proses penggalian dan analisis data untuk mengidentifikasi pola dan hubungan dalam data yang dapat membantu dalam pengambilan keputusan. Data mining dapat digunakan untuk memproses data besar dengan cepat dan efisien (Witten, Frank, & Hall, 2016).

2.2 Clustering

Clustering adalah suatu metode analisis data yang digunakan untuk mengelompokkan data menjadi beberapa kelompok berdasarkan kemiripan karakteristik antar data dalam satu kelompok dan perbedaan karakteristik antar kelompok. *Cluster* dapat membantu pengguna untuk memahami pola data, mengidentifikasi data yang anomali, dan membuat keputusan berdasarkan karakteristik data yang telah dikelompokkan. (Tan, Steinbach, & Kumar, 2019).

2.3 Metode K-Means Clustering

K-Means Clustering adalah salah satu algoritma clustering yang paling populer dan sering digunakan dalam pengolahan data. Algoritma ini bekerja dengan cara mengelompokkan data ke dalam kelompok (*k*) berdasarkan jarak antara data dengan pusat kelompok. Pusat kelompok dihitung dengan rata-rata data yang ada di dalam kelompok tersebut. Algoritma *K-Means Clustering* dapat menghasilkan kelompok-kelompok yang homogen dan memudahkan analisis data lebih lanjut (Hastie, Tibshirani, & Friedman, 2009).

2.4 Flowchart Sistem



Gambar 1 Flowchart Sistem

Dengan adanya flowchart diatas, maka tahapan secara umum untuk alur sistemnya adalah sebagai berikut:

1. Mulai
2. User masuk ke dalam aplikasi
3. User membuat periode pencatatan bulanan
4. User input data pemasukan dan pengeluaran
5. User melakukan simpan periode bulanan
6. Data tersimpan
7. User melihat detail periode pencatatan bulanan yang sudah dibuat

8. Lihat analisis hasil pengelompokan data pengeluaran bulanan

3. HASIL DAN PEMBAHASAN

3.1 Dataset

Data yang digunakan pada penelitian ini merupakan data pengeluaran bulanan milik salah satu user pada periode bulan tertentu dengan jumlah 30 untuk data training dan 25 data untuk data uji dari data yang diambil aplikasi CaKep.

Tabel 1. Dataset pelatihan

No	Nama Pengeluaran	Pengeluaran	Pemasukan
1	Turnamen	200000	13188000
2	Um 6	50000	13188000
3	Turnamen 5	200000	13188000
4	Senar	115000	13188000
5	Raket	300000	13188000
6	Uang Jjan Tep	500000	13188000
7	Kue Tep	96000	13188000
8	Om Leo	400000	13188000
9	Miniso	122500	13188000
10	Member	98000	13188000
11	Gojek	50000	13188000
12	Parfum	100000	13188000
13	Sepatu Badminton	600000	13188000
14	Kue	82000	13188000
15	Um 10/2	150000	13188000
16	Um 8	150000	13188000
17	Um 5	150000	13188000
18	Mtr Acik	200000	13188000
19	Turnamen	150000	13188000
20	Lawson	100000	13188000
21	Klinik	900000	13188000
22	Gopay Ojek	40000	13188000
23	Grab Car 5/2	65000	13188000
24	Barang	3521500	13188000
25	Hutang Santi	2000000	13188000
26	Um 7/2	150000	13188000
27	prive	300000	13188000
28	Tabungan Bln 2	500000	13188000
29	Ojek	25000	13188000
30	Ambil	1038500	13188000

Tabel 2. Dataset pengujian

No	Nama Pengeluaran	Pengeluaran	Pemasukan
1	Indomaret	99000	3185600
2	Lipstik Deodoran Popice	73000	3185600
3	Beny	650000	3185600
4	Point' Kids	225500	3185600
5	Belanja Pagi	38000	3185600
6	Teh, Sunligt Dll	7000	3185600
7	Makan Pagi & Siang	22000	3185600
8	Print	3500	3185600

No	Nama Pengeluaran	Pengeluaran	Pemasukan
9	Makan	40000	3185600
10	Arisan	200000	3185600
11	Cilok	10000	3185600
12	Pulsa	52000	3185600
13	Sangune Anak Tamu	20000	3185600
14	Sate Dan Beras	60000	3185600
15	Cireng	5000	3185600
16	Jajan Rokok	30000	3185600
17	Makan Siang Tyas	30000	3185600
18	Hiburan Pasar Malam	93000	3185600
19	Kado Untuk Reni	60000	3185600
20	Beny	150000	3185600
21	Rokok	25000	3185600
22	Tabungan Emas Galih	1000000	3185600
23	Belanja Beny	52500	3185600
24	Tempe	2000	3185600
25	Listrik	52000	3185600

3.2 Proses K-Means Clustering

Pada tahap ini, proses pengolahan data dilakukan menggunakan tools Google Collab dan menggunakan Python Language karena terdapat berbagai macam fungsi yang dapat membantu proses analisis pada library *Pandas*, *matplotlib.pyplot*, dan *sklearn*. Library *Pandas* berfungsi untuk analisis data dan memanipulasi tabel, library *matplotlib.pyplot* berfungsi untuk membuat visualisasi grafik dan plot pada data, dan library *sklearn* berfungsi untuk membantu prosesing data seperti mengubah skala data, mengisi missing value ataupun melakukan training data khususnya untuk menggunakan fungsi *K-Means Clustering*.

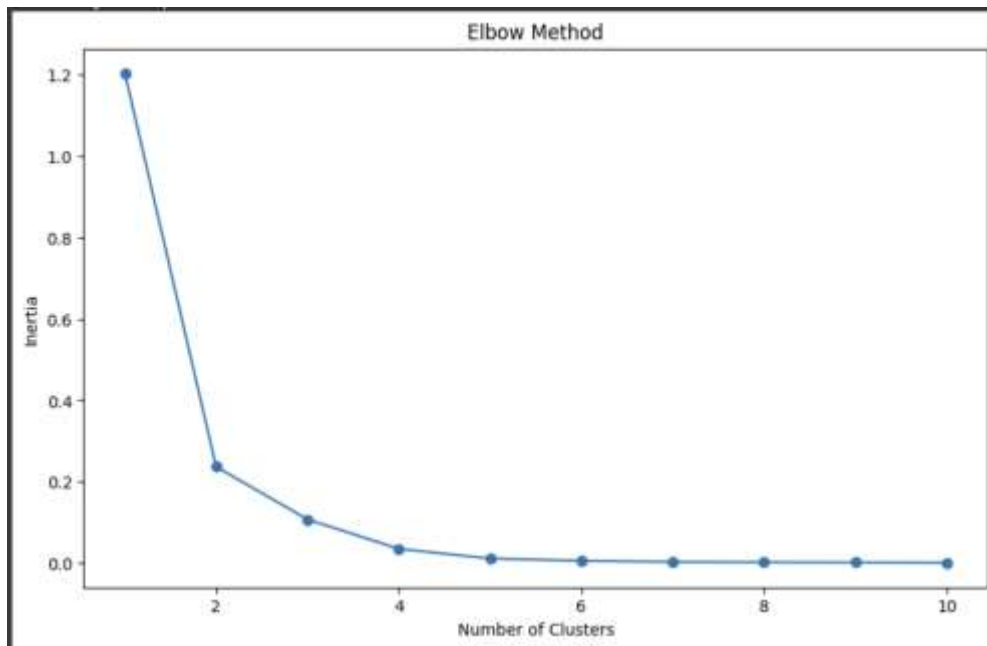
Tahap mengelompokkan data menggunakan metode K-Means Clustering:

1. Menentukan K sebagai Jumlah Cluster: Pada tahap ini, peneliti melakukan pemilihan K terbaik menggunakan *Elbow Method* dengan percobaan cluster 1-11.
2. Inisialisasi Pusat Kelompok: Secara acak memilih 3 titik dari data sebagai pusat kelompok awal.
3. Menghitung Jarak: Menghitung jarak Euclidean antara setiap titik data dengan pusat kelompok yang ada.
4. Mengelompokkan data: Mengelompokkan setiap titik data ke dalam kelompok terdekat berdasarkan jarak yang telah dihitung.
5. Menghitung Pusat Kelompok Baru: Menghitung rata-rata jumlah untuk setiap kelompok dan menggunakan nilai ini sebagai pusat kelompok baru.
6. Konvergensi atau Iterasi: Mengulangi langkah-langkah 3-5 sampai pusat kelompok tidak mengalami perubahan signifikan.
7. Pengujian Hasil: Menguji validasi menggunakan *silhouette_score* dari library *python sklearn* untuk mendapatkan presentasi nilai yang optimal.

3.3 Pembahasan

3.3.1 Penentuan jumlah cluster menggunakan Elbow Method

Dalam proses menggunakan *Elbow Method*, peneliti melakukan percobaan dengan jumlah kluster yang bervariasi mulai dari 1 hingga 11 untuk mengetahui berapa jumlah kluster yang optimal. Pada perhitungan untuk mencari jumlah kluster yang optimal dibantu dengan bahasa pemrograman python menggunakan library *sklearn.cluster*. Detail kode dan hasil penentuan jumlah klaster dengan menggunakan *Elbow Method* dapat dilihat pada gambar dibawah:

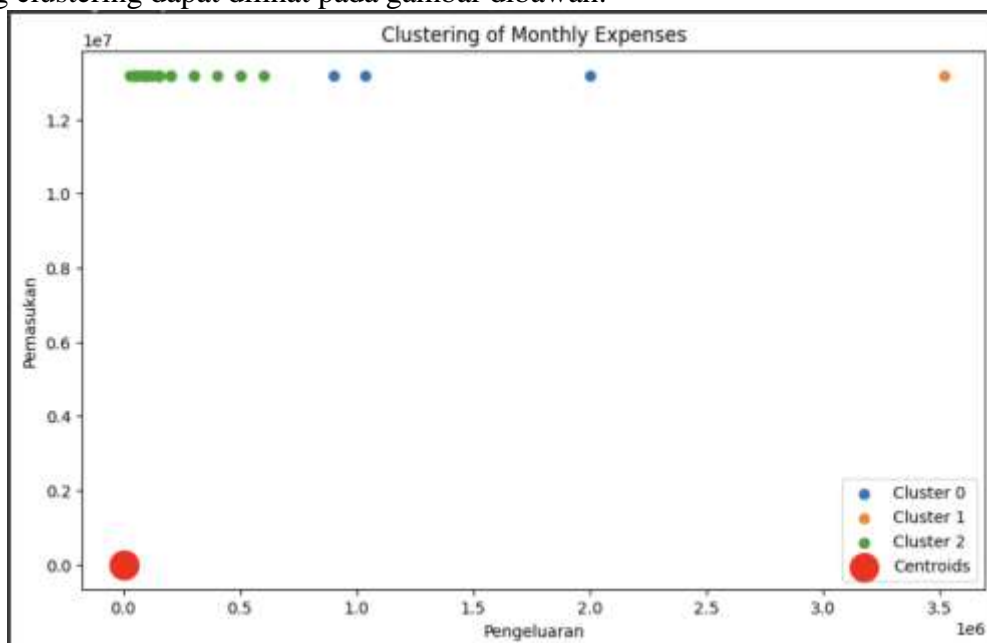


Gambar 2. Grafik elbow method

Pada Gambar 1 Grafik Elbow Method menunjukkan bahwa terdapat penurunan nilai inersia yang signifikan saat jumlah kluster meningkat dari 1 hingga 3. Oleh karena itu secara grafik kita dapat menyimpulkan bahwa jumlah cluster optimal adalah 3. Kluster 3 masih dianggap optimal karena nilai Inersia sudah tidak terlalu jauh jika kita menggunakan $k=4$ dan seterusnya.

3.3.2 Proses Clustering

Proses clustering dibantu dengan menggunakan bahasa python dengan init parameter $n_cluster=3$ sesuai dengan jumlah cluster optimal dan percobaan $max_iterasi$ 2-4. Pada proses *clustering* data awal yang digunakan adalah data training dengan jumlah 30 data. Detail proses training clustering dapat dilihat pada gambar dibawah:



Gambar 3. Hasil clustering

Pada Gambar 19 ditampilkan sebuah grafik hasil clustering yang mana terdapat 3 cluster yaitu cluster 0 = pengeluaran sedang, cluster 1 = pengeluaran tinggi, dan cluster 2 = pengeluaran

rendah. Kemudian proses selanjutnya adalah dengan melakukan clustering menggunakan model yang sudah dilatih pada data uji yang berjumlah 25 data.

	Nama	Pengeluaran	Pemasukan	Predicted_Cluster
0	Indomaret	99000	3185600	2
1	Lipstik Deodoran Popice	73000	3185600	2
2	Beny	650000	3185600	0
3	Point Kids	225500	3185600	0
4	Belanja Pagi	38000	3185600	2
5	Teh, Sunlight Di	7000	3185600	2
6	Makan Pagi & Siang	22000	3185600	2
7	Print	3500	3185600	2
8	Makan	40000	3185600	2
9	Arisan	200000	3185600	2
10	Cilok	10000	3185600	2
11	Pulsa	52000	3185600	2
12	Sangune Anak Tamu	20000	3185600	2
13	Sate Dan Beras	60000	3185600	2
14	Cireng	5000	3185600	2
15	Jajan Rokok	30000	3185600	2
16	Makan Slang Tyas	30000	3185600	2
17	Hiburan Pasar Malam	93000	3185600	2
18	Kado Untuk Rani	60000	3185600	2
19	Beny	150000	3185600	2
20	Rokok	25000	3185600	2
21	Tabungan Emas Galih	1000000	3185600	1
22	Belanja Beny	52500	3185600	2
23	Timpe	2000	3185600	2
24	Listrik	52000	3185600	2

Gambar 4. Hasil clustering data uji

3.3.3 Pengujian validasi clustering menggunakan silhoutte score

Dalam proses ini dilakukan uji validasi menggunakan *silhouette_score* dengan tujuan untuk mengetahui kualitas klasterisasi. Semakin tinggi nilai *silhouette_score*, maka semakin baik klasterisasi yang dihasilkan. Rentang nilai *silhouette_score* adalah berkisar antara -1 hingga 1. Berikut ini adalah interpretasi umum untuk rentang nilai *silhouette_score*:

- Nilai mendekati 1: Klasterisasi yang baik, dengan setiap sampel berada dalam klaster yang sesuai dengan jarak yang baik terhadap klaster lainnya.
- Nilai mendekati 0: Menunjukkan adanya tumpah tindih (*overlap*) antara klaster, atau ada kesulitan dalam mengklasifikasikan sampel ke klaster yang tepat.
- Nilai mendekati -1: Indikasi kuat bahwa sampel telah salah diklasifikasikan dan berada dalam klaster yang tidak sesuai.

Silhouette Score: 0.721052859092002

Gambar 5. Hasil uji validasi silhouette score

Dari hasil pengujian menggunakan *silhouette_score* diatas dapat disimpulkan bahwa hasil klasterisasi sudah cukup optimal karena nilai *silhouette_score* 0.72 mendekati 1 atau dalam presentase dapat disebutkan dengan nilai 72%.

3.3.4 Implementasi Sistem

Pada halaman Analisis Pengeluaran ditampilkan sebuah grafik yang menampilkan 3 kelompok yaitu kelompok pengeluaran rendah dengan total 86%, pengeluaran sedang 5%, dan pengeluaran tinggi 9% dari data pengeluaran bulanan. Serta ditampilkan juga sebuah list anggota data pengeluaran dari masing-masing kelompok.



Gambar 6. Halaman lihat analisis

4. SIMPULAN DAN SARAN

Pengelompokan data pengeluaran bulanan pada aplikasi pencatatan keuangan pribadi menggunakan metode *K-means clustering* telah tersajikan pada artikel ini. Berdasarkan hasil pengujian dan pembahasan yang telah dilakukan maka dapat disimpulkan bahwa pengelompokan data pengeluaran dengan *k-means* telah berhasil disajikan, dengan penentuan jumlah kluster dengan menggunakan *Elbow Method* menghasilkan 3 sebagai jumlah kluster optimal, yaitu untuk kelompok pengeluaran rendah (cluster 2), pengeluaran sedang (cluster 0), dan pengeluaran tinggi (cluster 1). Hasil pengujian klusterisasi menggunakan silhouette menghasilkan skor 72%, artinya klusterisasi cukup baik, dengan setiap sampel berada dalam kluster yang sesuai dengan jarak yang baik terhadap kluster lainnya.

5. DAFTAR PUSTAKA

- Adebayo, A. R, Adebayo, A. O, & Oyeboode, O. (2016). Monthly Expenditure of Nigerian Households on Food: Evidence from Ibadan. *Journal of Environmental and Social Sciences*, 3(11), 1–14.
- Adiya, M. H., & Desnelita, Y. (2019). *Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru*. 05(01).
- Aggarwal, C. C, & Reddy, C. K. (2013). *Data clustering: Algorithms and applications*. CRC Press.
- Alfina, T., Santosa, B., & Hakim, J. A. R. (2012). *Analisa Perbandingan Metode Hierarchical Clustering, K-means dan Gabungan Keduanya dalam Cluster Data (Studi kasus: Problem Kerja Praktek Jurusan Teknik Industri ITS)*. 1.
- Almayda, A., & Saepudin, S. (2021). *PENERAPAN DATA MINING K-MEANS CLUSTERING UNTUK MENGELOMPOKKAN BERBAGAI JENIS MERK SMARTPHONE*.

- Chandra, S., & Singh, V. (2019). *AN EFFICIENT K-MEANS CLUSTERING ALGORITHM FOR DATA MINING*. *International Journal of Computer Applications*, 182(16), 23–27.
- Fadhilah, A. N., & Nasution, M. N. (2020). Analisis Pengaruh Aplikasi Pencatatan Keuangan Pribadi Terhadap Pengelolaan Keuangan Pribadi. *Jurnal Online Mahasiswa Fakultas Ekonomi Universitas Riau*, 7(1), 1–12.
- Hidayatullah, D., & Widayanti, R. (2017). Penerapan Metode K-Means Clustering pada Data Transaksi Penjualan Online di Tokopedia. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi (JNTETI)*, 6(1), 1–8.
- Lee, C., Lee, J., & Lee, S. (2019). A Personal Financial Management System Using a Machine Learning Algorithm. *Sustainability*, 11(13), 3679.
- Rahman, M. M., & Rabbani, M. G. (2018). Klasifikasi Data Berbasis Cluster Menggunakan Algoritma K-Means. *International Journal of Computer Science and Information Security (IJCSIS)*, 16(1), 76–83.
- Risdianto, R., Wardoyo, R., & Wardana, I. N. (2016). Clustering data pengeluaran untuk penerapan cash management pada usaha kecil dan menengah. *Jurnal Teknik Informatika Dan Sistem Informasi*, 2(1), 1–6.
- Septioko, T., Parhusip, H. A., & Mahatma, T. (n.d.). *Aplikasi K-means Untuk Pengelompokan Rumah Tangga Di Salatiga Berdasarkan Data SUSENAS 2011*.
- Suhartono, T., & Winarno, W. (2020). Analisis Perilaku Konsumen dalam Pengelolaan Keuangan Pribadi. *Jurnal Manajemen Dan Kewirausahaan*, 22(2), 97–108.
- Tan, P.-N., Steinbach, M., & Kumar, V. (n.d.). *Introduction to Data Mining*.
- Tibshirani, S., & Friedman, H. (n.d.). *Valerie and Patrick Hastie*.
- Witten, I. H., Frank, E., & Hall, M. A. (2016). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann.
- Zainudin, E., & Wijaya, E. Y. (n.d.). Analisis Pengelompokan Data Penggunaan Energi Listrik Rumah Tangga Menggunakan Metode K-Means Clustering. *Jurnal Elektronika Dan Telekomunikasi (JET)*.