Pengelompokan Rumah Tangga Berdasarkan Indikator Pembentuk Rumah Layak Huni Menggunakan Metode *K-Prototypes* di Provinsi Jawa Barat Tahun 2023

Rena Siti Nur Aliyah Hamzah *, Abdul Kudus

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

*renasitinuraliyah@gmail.com, abdul.kudus@unisba.ac.id

Abstract. The *K-Prototypes* method is an effective clustering technique for handling data with mixed types. The Susenas data involving several livable housing variables has 25,890 observations, consisting of 2 numerical variables and 9 categorical variables. Therefore, the *K-Prototypes* method can be used to cluster this data. The purpose of this study is to cluster households based on indicators of livable houses in West Java Province in 2023. *K-Prototypes* combines *K-Means* and *K-Modes*, using Euclidean distance for numeric variables and Simple Matching for categorical variables. This method works similar to *K-Means*, which is known to be efficient and easy to implement. Cluster evaluation is conducted using the Silhouette coefficient, with the optimal number of clusters determined by a Silhouette coefficient value close to 1. The optimal results of clustering households in West Java Province in 2023 using *K-Prototypes* show the existence of 2 clusters, with a Silhouette coefficient value of 0.219759, while other coefficient values are lower. Cluster 1 consists of 17,330 members, most of whom are from urban areas and have good livable houses, while cluster 2 consists of 8,560 members, most of whom are from rural areas and have less livable houses.

Keywords: *K-Prototypes, Pengelompokan, Liveable House, Euclidean Distance, Simple Matching Distance, Silhouette Coefficient.*

Abstrak. Metode *K-Prototypes* adalah teknik pengelompokan yang efektif untuk menangani data dengan tipe campuran. Data Susenas yang melibatkan variabel rumah layak huni memiliki 25.890 pengamatan, terdiri dari 2 variabel numerik dan 9 variabel kategorik. Oleh karena itu, metode *K-Prototypes* dapat digunakan untuk mengelompokkan data ini. Tujuan dari penelitian ini adalah untuk mengelompokkan rumah tangga berdasarkan indikator-indikator rumah layak huni di Provinsi Jawa Barat pada tahun 2023. *K-Prototypes* menggabungkan *K-Means* dan *K-Modes*, menggunakan jarak Euclidean untuk variabel numerik dan *Simple Matching* untuk variabel kategorik. Metode ini bekerja mirip dengan *K-Means*, yang dikenal efisien dan mudah diterapkan. Evaluasi klaster dilakukan menggunakan koefisien Silhouette, dengan optimalnya jumlah klaster ditentukan oleh nilai koefisien Silhouette yang mendekati 1. Hasil optimal pengelompokan rumah tangga di Provinsi Jawa Barat pada tahun 2023 menggunakan *K-Prototypes* menunjukkan adanya 2 klaster, dengan nilai koefisien Silhouette sebesar 0,219759, sementara nilai koefisien lainnya lebih rendah. Klaster 1 terdiri dari 17.330 anggota, yang mayoritas berasal dari perkotaan dan memiliki rumah layak huni yang baik, sedangkan klaster 2 terdiri dari 8.560 anggota, sebagian besar berasal dari pedesaan dan memiliki capaian rumah layak huni yang kurang.

Kata Kunci: *K-Prototypes, Pengelompokan, Rumah Layak Huni, Jarak Euclidean, Jarak Simple Matching, Koefisien Silhouette.*

A. Pendahuluan

Analisis kluster merupakan teknik pengelompokan n objek pengamatan ke dalam K kluster berdasarkan kesamaan karakteristiknya [1]. Metode yang paling sederhana dan efisien dalam analisis kluster adalah metode *K-Means*, dimana metode tersebut efektif jika digunakan pada data dengan tipe numerik [2]. *K-Means* dikembangkan menjadi suatu metode yang dapat menangani data dengan tipe kategorik yaitu *K-Modes*. *K-Means* dan *K-Modes* bekerja dengan menentukan terlebih dahulu jumlah kluster yang akan dibentuk, sehingga kedua metode tersebut masuk ke dalam kategori analisis kluster *non* hierarki. Analisis kluster *non* hierarki memiliki algoritma dan iterasi yang mudah sehingga akan lebih efisien apabila digunakan untuk data besar daripada analisis kluster hierarki [3].

Dalam perkembangannya, *K-Means* dan *K-Modes* dikombinasikan menjadi sebuah metode yang dapat menangani data bertipe campuran yaitu metode *K-Prototypes* [3]. *K-Prototypes* bekerja dengan menggabungkan 2 teknik pengukuran jarak, sehingga metode ini dapat mengelompokkan data numerik dan kategorik secara bersama-sama. Adapun kelebihan lain dari *K-Prototypes* yaitu dapat secara efektif menangani data besar karena masuk ke dalam kategori analisis kluster *non* hierarki yang memiliki algoritma dan iterasi yang mudah serta efisien. Sehingga dalam penelitian ini akan dilakukan penerapan metode *K-Prototypes* pada studi kasus rumah layak huni yang memiliki data besar dan bertipe campuran.

Suatu rumah dapat dikatakan layak huni apabila memenuhi syarat keselamatan bangunan, kecukupan minimum luas bangunan, dan kesehatan bagi penghuninya [4]. Membangun rumah dan permukiman yang layak, aman, dan nyaman merupakan salah satu tujuan dalam mewujudkan Indonesia emas 2045 yang merupakan visi pemerintah dalam usaha menjadikan Indonesia yang lebih maju dan sejahtera di tahun 2045. Angka rumah layak huni di Indonesia 3 tahun terakhir ini masih belum stabil, tahun 2021 angka tersebut mencapai 60,90% dan mengalami penurunan di tahun 2022 menjadi 60,66%. Kemudian di tahun 2023 angka rumah layak huni kembali mengalami kenaikan menjadi 63,15% [5].

Berkaitan dengan hal ini, pemerintah dan beberapa lembaga terkait harus berupaya meningkatkan dan menstabilkan kembali angka rumah layak huni di Indonesia berdasarkan pada indikator pembentuk rumah layak huni. Beberapa faktor-faktor pembentuk rumah layak huni di antaranya: kecukupan luas tempat tinggal, ketahanan bangunan yang disusun berdasarkan jenis atap, dinding, dan lantai terluas tempat tinggal, akses terhadap sanitasi layak yang disusun berdasarkan fasilitas buang air besar, jenis kloset, dan tempat pembuangan tinja, serta akses terhadap air minum layak yang disusun berdasarkan sumber utama air minum dan sumber utama air mandi, cuci, dll [5].

Berdasarkan uraian tersebut maka akan dilakukan penelitian terkait pengelompokan rumah tangga berdasarkan indikator pembentuk rumah layak huni menggunakan metode *K-Prototypes* di Provinsi Jawa Barat tahun 2023 dengan tujuan yang ingin dicapai dari penelitian ini yaitu untuk mengetahui pengelompokan rumah tangga yang paling optimal di Provinsi Jawa Barat tahun 2023 dengan menggunakan metode *K-Prototypes* berdasarkan indikator pembentuk rumah layak huni.

B. Metodologi Penelitian

Penelitian ini menggunakan data Survei Sosial Ekonomi Nasional (Susenas) Maret tahun 2023 dengan variabel dan wilayah terpilih yang diperoleh dari Badan Pusat Statistik (BPS) Provinsi Jawa Barat. Unit penelitian berupa rumah tangga terpilih yang berjumlah sebanyak 25.890 rumah tangga.

Beberapa variabel numerik yang digunakan di antaranya: jumlah anggota rumah tangga dan luas lantai. Sedangkan variabel kategorik yang digunakan di antaranya: klasifikasi perkotaan/pedesaan, bahan atap terluas tempat tinggal, bahan dinding terluas tempat tinggal, bahan lantai terluas tempat tinggal, fasilitas buang air besar, jenis kloset, tempat pembuangan tinja, sumber utama air minum, dan sumber utama air mandi, cuci, dll. Analisis data dilakukan dengan teknik klustering menggunakan metode *K-Prototypes*.

Normalisasi Data

Salah satu persiapan data yang sangat diperlukan sebelum dilakukannya pengujian agar diperoleh hasil yang lebih optimal disebut sebagai normalisasi data [6]. Tujuan dari normalisasi data yaitu untuk membuat data yang sebelumnya memiliki skala satuan berbeda menjadi berada pada satuan rentang yang sama. Normalisasi data ini dilakukan hanya untuk data yang bertipe numerik.

Teknik normalisasi data yang digunakan adalah dengan *min-max scaling* pada nilai logaritma dari data asli karena ditemukannya data ekstrim atau data yang berpotensi sebagai pencilan [2]. Teknik ini akan mengubah skala data menjadi rentang [0, 1] dengan menggunakan persamaan berikut:

$$x_{new} = \frac{\log(x) - \min(\log(x))}{\max(\log(x)) - \min(\log(x))} \quad (1)$$

Dimana:

x_{new} : nilai objek baru.

x : nilai objek pada data numerik.

Ukuran Jarak Kombinasi

Ukuran kesamaan merupakan suatu ukuran yang digunakan untuk melihat kemiripan suatu objek dengan objek lainnya. Dalam melakukan analisis kluster, ukuran kesamaan dapat dihitung dengan ukuran jarak. Semakin kecil ukuran jarak maka semakin mirip objek pengamatan dengan *centroid cluster*, begitupula sebaliknya semakin besar ukuran jarak maka semakin berbeda objek pengamatan dengan *centroid cluster* [3].

Dalam menangani data bertipe numerik dapat digunakan ukuran jarak Euclidean dan dalam menangani data bertipe kategorik dapat digunakan ukuran jarak *Simple Matching* [2]. Ukuran jarak Euclidean sering digunakan dalam analisis kluster *K-Means*, dimana ukuran jarak ini bekerja tanpa adanya hambatan dan sangat efektif untuk data numerik [7]. Ukuran jarak Euclidean umumnya dihitung melalui persamaan berikut ini [3]:

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (2)$$

Ukuran jarak *Simple Matching* merupakan total dari ketidakcocokan antara 2 objek pengamatan, dimana ketika nilainya semakin kecil berarti 2 objek pengamatan tersebut semakin mirip [8]. Secara umum, ukuran jarak *Simple Matching* dapat dihitung melalui persamaan berikut [8]:

$$d_{ij} = \sum_{k=1}^p \delta(x_{ik}, x_{jk}) \quad (3)$$

$$\delta(x_{ik}, x_{jk}) = \begin{cases} 0 & (x_{ik} = x_{jk}) \\ 1 & (x_{ik} \neq x_{jk}) \end{cases} \quad (4)$$

Dimana:

d_{ij} : Jarak antara objek ke- i dan objek ke- j .

p : Jumlah variabel.

x_{ik} : Nilai objek ke- i pada variabel ke- k ; $k = 1, 2, 3, \dots, p$.

x_{jk} : Nilai objek ke- j pada variabel ke- k ; $k = 1, 2, 3, \dots, p$.

$(x_{i1}, x_{i2}, \dots, x_{ip})$: Nilai variabel pada objek ke- i .

$(x_{j1}, x_{j2}, \dots, x_{jp})$: Nilai variabel pada objek ke- j .

Dalam penerapan *K-Prototypes*, ukuran jarak yang digunakan merupakan kombinasi dari ukuran jarak Euclidean untuk menangani data numerik dan ukuran jarak *Simple Matching* untuk

menangani data kategorik. Ukuran jarak kombinasi ini diseimbangkan oleh bobot penyeimbang jarak yang disebut koefisien gamma (γ) untuk menghindari ketidakseimbangan terhadap salah satu atributnya [8]. Ketika nilai koefisien gamma (γ) terlalu tinggi artinya variabel kategorik memiliki pengaruh yang dominan dan ketika nilai koefisien gamma (γ) terlalu rendah artinya variabel numerik yang memiliki pengaruh dominan. Oleh karena itu berdasarkan simulasi yang telah dilakukan pada penelitian sebelumnya, maka rentang nilai koefisien gamma (γ) ditetapkan sebagai berikut [9]:

$$\frac{1}{3}\bar{\sigma} \leq \gamma \leq \frac{2}{3}\bar{\sigma} \tag{5}$$

Dimana:

- $\bar{\sigma}$: Rata-rata simpangan baku data numerik.
- γ : Bobot penyeimbang jarak.

Sehingga ukuran jarak kombinasi yang digunakan untuk *K-Prototypes* dapat dihitung dengan menggunakan persamaan berikut:

$$d_{ik} = \sum_{j=1}^q (x_{ij} - x_{kj})^2 + \gamma \sum_{j=q+1}^p \delta(x_{ij}, x_{kj}) \tag{6}$$

Dimana:

- d_{ik} : Jarak antara objek ke-*i* dan *centroid cluster* ke-*k*.
- q : Jumlah variabel numerik.
- p : Jumlah variabel numerik dan kategorik.
- γ : Bobot penyeimbang jarak.
- x_{ij} : Nilai objek ke-*i* pada variabel ke-*j*.
- x_{jk} : Nilai *centroid cluster* ke-*k* pada variabel ke-*j*.

K-Prototypes

K-Prototypes merupakan salah satu metode analisis kluster *non* hierarki, dimana metode ini lebih efisien untuk menangani data bertipe campuran dan data besar daripada metode lainnya [2]. Metode yang merupakan kombinasi dari *K-Means* dan *K-Modes* ini bekerja dengan mempertahankan efisiensi *K-Means* ketika berhadapan dengan data besar, serta dapat digunakan ketika berhadapan dengan data numerik dan kategorik secara bersama-sama [10].

Dalam *K-Prototypes*, *centroid cluster* akan terus diperbaharui dengan menggunakan nilai prototipenya yaitu nilai mean pada setiap kluster untuk data numerik dan nilai modus pada setiap kluster untuk data kategorik. Adapun nilai mean dan modus tersebut dapat dihitung dengan persamaan berikut:

$$\underline{\mu}_k = \frac{1}{N_k} \sum_{i=1}^{N_k} x_{ik} \tag{7}$$

$$\underline{Mod}_k = \underline{Mod}\{x_{ik} | k = 1, 2, \dots, p\} \tag{8}$$

Dimana:

- $\underline{\mu}_k$: Titik *centroid* pada kluster ke-*k*.
- N_k : Banyak objek pengamatan pada kluster ke-*k*.
- X_{ik} : Data ke-*i* pada kluster ke-*k* ; $i = 1, 2, \dots, N_k$.

Tahapan-tahapan dalam melakukan analisis data menggunakan metode *K-Prototypes* adalah sebagai berikut:

1. Menentukan banyaknya klaster yang akan dibentuk, yaitu sebanyak $K = 2$.
2. Menentukan nilai koefisien gamma dengan menggunakan Persamaan (5).
3. Menentukan data secara acak sebanyak K sebagai titik centroid cluster awal.
4. Melakukan perhitungan ukuran jarak pada setiap objek pengamatan dengan *centroid cluster* menggunakan Persamaan (6).
5. Alokasikan setiap objek pengamatan ke dalam klaster yang memiliki ukuran jarak paling dekat dan hitung total ukuran jarak setiap objek pengamatan dengan centroid cluster masing-masing.
6. Melakukan perhitungan *centroid cluster* yang baru dengan menggunakan Persamaan (7) untuk data numerik dan Persamaan (8) untuk data kategorik.
7. Melakukan perhitungan ukuran jarak pada setiap objek pengamatan dengan *centroid cluster* baru menggunakan Persamaan (6).
8. Mengidentifikasi apakah dengan memindahkan objek pengamatan ke dalam klaster lain akan mengurangi total ukuran jarak setiap objek pengamatan dengan *centroid cluster* masing-masing. Jika ya, maka pindahkan objek pengamatan tersebut.
9. Hitung selisih dari total ukuran jarak objek pengamatan dengan *centroid* lama dan total ukuran jarak objek pengamatan dengan *centroid* baru.
10. Melakukan kembali perhitungan pada langkah ke-6 dan lakukan iterasi secara berulang hingga diperoleh *centroid cluster* yang konvergen, yaitu ketika selisih total ukuran jarak objek pengamatan dengan centroid lama dan total ukuran jarak objek pengamatan dengan *centroid* baru sangat kecil atau dapat dikatakan bahwa ketika tidak ada lagi perpindahan objek pengamatan.
11. Naikkan K menjadi $K + 1$, kemudian lakukan kembali langkah 3 sampai 11 hingga mencapai nilai $K = 10$

Evaluasi Klaster

Untuk melihat apakah hasil dari pengelompokan itu baik atau tidak maka dilakukan proses evaluasi klaster [10]. Selain itu, evaluasi klaster juga digunakan untuk memilih proses pengelompokan yang paling optimal. Teknik evaluasi klaster yang digunakan dalam penelitian ini yaitu menggunakan nilai koefisien Silhouette karena merupakan teknik yang paling umum dan mudah digunakan.

Koefisien Silhouette bekerja dengan mengukur kesamaan pada setiap objek pengamatan. Nilai koefisien Silhouette berada pada rentang -1 hingga 1, ketika nilai tersebut semakin besar dan mendekati 1, artinya klaster yang dibentuk akan semakin optimal [11]. Ketika nilai Silhouette < 0 artinya suatu objek pengamatan tidak berada dalam klaster yang tepat dan ketika nilai Silhouette = 0 artinya suatu objek pengamatan berada di antara 2 klaster [12]. Nilai koefisien Silhouette dapat dihitung dengan menggunakan persamaan berikut:

$$S_{ik} = \frac{b_{ik} - a_{ik}}{\max(a_{ik}, b_{ik})} \quad (9)$$

$$S_k = \frac{1}{N_k} \sum_{i=1}^{N_k} S_{ik} \quad (10)$$

$$S = \frac{1}{K} \sum_{k=1}^K S_k \quad (11)$$

Dimana:

a_{ik} : Rata-rata jarak objek ke- i pada klaster ke- k dengan seluruh objek yang berada di dalam klaster yang sama.

b_{ik} : Rata-rata jarak objek ke- i pada klaster ke- k dengan seluruh objek yang berada pada

klaster terdekat.

S_{ik} : Nilai Silhouette pada objek ke- i dan klaster ke- k .

S_k : Nilai Silhouette pada klaster ke- k .

S : Nilai Silhouette total

C. Hasil Penelitian dan Pembahasan

Normalisasi Data

Dalam penelitian ini digunakan 2 variabel numerik yang memiliki skala satuan berbeda yaitu variabel X1 yang merupakan jumlah anggota rumah tangga dengan satuan jumlah orang dan X2 yang merupakan luas lantai dengan satuan m². Maka dari itu, dilakukan normalisasi data dengan teknik *min-max scaling* pada nilai logaritma dari nilai data asli. Berikut merupakan hasil dari normalisasi data numerik:

Tabel 1. Data Numerik Hasil Normalisasi Data

No	X1	X2
1	0,540	0,627
2	0,270	0,428
3	0,428	0,604
⋮	⋮	⋮
25.889	0,540	0,611
25.890	0,270	0,681

Terlihat pada Tabel 1, setelah dilakukan normalisasi data pada variabel numerik maka data pada variabel X1 dan X2 berada dalam rentang 0 hingga 1. Selanjutnya data tersebut dapat digunakan untuk analisis data lanjutan yaitu dengan menggunakan metode *K-Prototypes*.

Hasil Pengelompokan dengan K-Prototypes

Dalam melakukan analisis klaster *K-Prototypes*, langkah pertama yang harus dilakukan adalah menentukan banyaknya klaster yang akan dibentuk. Dalam penelitian ini, dilakukan uji coba terhadap pembentukan klaster sebanyak $K = 2, 3, 4, 5, 6, 7, 8, 9, 10$.

Selanjutnya dilakukan juga penentuan nilai koefisien gamma (γ), dimana rentang nilai tersebut berada di antara $\frac{1}{3}\bar{\sigma} \leq \gamma \leq \frac{2}{3}\bar{\sigma}$. Koefisien gamma (γ) yang digunakan dalam penelitian ini juga dengan uji coba terhadap 3 nilai, berikut merupakan nilai koefisien gamma (γ) yang digunakan:

Tabel 2. Nilai Koefisien Gamma

σ_{X1}	σ_{X2}	$\overline{\sigma_{X1,X2}}$	γ
0,188	0,103	0,146	$\frac{1}{3}\bar{\sigma} = 0,049$
			$\frac{1}{2}\bar{\sigma} = 0,073$
			$\frac{2}{3}\bar{\sigma} = 0,097$

Tabel 2 menunjukkan 3 nilai koefisien gamma (γ) yang digunakan dalam penelitian, dimana nilai tersebut merupakan nilai paling kecil, nilai tengah, dan nilai paling tinggi dalam rentang nilai koefisien gamma (γ). 3 nilai koefisien gamma tersebut akan digunakan dan dipilih nilai yang paling optimal berdasarkan pada evaluasi klaster.

Langkah selanjutnya adalah penentuan *centroid cluster* secara acak sebanyak K . Sebagai contoh akan dilakukan pembentukan 2 buah klaster, maka tentukan 2 objek pengamatan secara acak yang akan menjadi *centroid* dari klaster 1 dan klaster 2. Kemudian jarak antara

objek data dengan *centroid* awal akan dihitung dan dilakukan pengalokasian setiap objek pengamatan ke seluruh kluster.

Dalam *K-Prototypes*, *centroid cluster* akan diperbaharui dengan menggunakan nilai mean untuk data bertipe numerik dan nilai modus untuk data bertipe kategorik. Selanjutnya, ukuran jarak akan dihitung kembali antara objek pengamatan dengan *centroid* terbaru lalu dilakukan identifikasi apakah ketika objek pengamatan dipindahkan ke dalam kluster lain akan mengurangi total dari ukuran jarak setiap objek dengan *centroid* masing-masing. Ketika hal tersebut terjadi, maka objek pengamatan akan mengalami perpindahan kluster. Proses ini akan dilakukan secara berulang hingga tercapai *centroid* yang konvergen.

Adapun hasil dari pengelompokan *K-Prototypes* dengan $K = 2, 3, 4, 5, 6, 7, 8, 9, 10$ dan menggunakan nilai $\gamma = 0,049; 0,073; \text{ dan } 0,097$ adalah sebagai berikut:

Tabel 3. Jumlah Anggota Kluster dengan $\gamma = 0,049$

K	k									
	1	2	3	4	5	6	7	8	9	10
2	17.319	8.571	-	-	-	-	-	-	-	-
3	6.004	7.368	12.518	-	-	-	-	-	-	-
4	5.025	5.803	3.155	11.907	-	-	-	-	-	-
5	7.360	3.444	6.885	2.693	5.508	-	-	-	-	-
6	3.569	3.908	4.429	2.396	7.275	4.313	-	-	-	-
7	4.420	5.540	2.314	2.448	3.331	3.690	4.147	-	-	-
8	4.757	2.866	3.251	3.609	2.708	2.704	3.941	2.054	-	-
9	2.941	3.089	2.774	2.564	1.876	3.338	2.346	5.271	1.691	-
10	2.324	1.987	2.562	2.842	4.655	2.370	1.836	1.967	2.395	2.952

Tabel 4. Jumlah Anggota Kluster dengan $\gamma = 0,073$

K	k									
	1	2	3	4	5	6	7	8	9	10
2	17.330	8.560	-	-	-	-	-	-	-	-
3	7.439	5.306	13.145	-	-	-	-	-	-	-
4	5.474	5.312	11.909	3.195	-	-	-	-	-	-
5	8.441	4.830	4.359	5.201	3.059	-	-	-	-	-
6	3.900	2.456	7.024	4.132	4.090	4.288	-	-	-	-
7	3.655	2.506	7.261	3.195	4.160	2.485	2.628	-	-	-
8	3.088	3.134	3.876	2.270	5.791	1.998	3.133	2.600	-	-
9	3.046	2.612	3.834	2.383	5.273	609	2.218	3.906	2.009	-
10	1.348	3.296	2.651	5.640	2.200	989	2.991	1.921	3.686	1.168

Tabel 5. Jumlah Anggota Kluster dengan $\gamma = 0,097$

K	k									
	1	2	3	4	5	6	7	8	9	10
2	8.559	17.331	-	-	-	-	-	-	-	-
3	14.915	4.433	6.542	-	-	-	-	-	-	-
4	3.220	5.488	11.894	5.288	-	-	-	-	-	-
5	4.539	10.440	3.480	4.609	2.822	-	-	-	-	-
6	3.555	4.781	2.142	8.359	4.092	2.961	-	-	-	-

K	k									
	1	2	3	4	5	6	7	8	9	10
7	5.399	2.648	2.936	3.157	4.152	4.821	2.777	-	-	-
8	3.372	1.312	6.211	2.405	4.906	2.466	3.496	1.722	-	-
9	1.094	3.056	5.530	2.417	3.639	2.775	3.023	2.206	2.150	-
10	4.105	3.532	1.980	2.711	2.688	837	1.027	2.883	3.302	2.825

Tabel 3, Tabel 4, dan Tabel 5 menunjukkan jumlah anggota pada setiap kluster yang terbentuk. Selanjutnya untuk menentukan pembentukan kluster mana yang paling optimal maka akan dilakukan evaluasi pada kluster.

Hasil Evaluasi Kluster

Evaluasi kluster dilakukan untuk melihat pembentukan kluster yang paling optimal, dimana evaluasi yang digunakan menggunakan nilai koefisien Silhouette. Berikut nilai koefisien Silhouette untuk setiap pembentukan kluster dan setiap nilai koefisien gamma:

Tabel 6. Nilai Koefisien Silhouette

K	γ		
	0,049	0,073	0,097
2	0,117382	0,219759	0,219732
3	0,041077	0,063504	0,123931
4	0,048804	0,041437	0,041054
5	-0,053532	0,034073	0,042517
6	-0,008244	-0,028823	0,03241
7	-0,008891	0,016838	0,013818
8	-0,026341	-0,00231	0,020454
9	-0,007469	-0,010455	0,019381
10	-0,026802	0,013896	0,008131

Dari Tabel 6 terlihat untuk nilai koefisien Silhouette yang paling besar dan mendekati 1 adalah nilai pada pembentukan K = 2 dengan $\gamma = 0,073$, dimana nilai koefisien Silhouette yang diperoleh sebesar 0,219759. Artinya pembentukan 2 kluster dengan nilai koefisien gamma (γ) sebesar 0,073 merupakan pembentukan kluster yang paling optimal.

Interpretasi Hasil Pengelompokan

Karena pembentukan 2 kluster dengan nilai koefisien gamma (γ) = 0,073 merupakan proses pengelompokan yang paling optimal, sehingga diperoleh anggota untuk setiap kluster adalah sebagai berikut:

Tabel 7. Anggota Setiap Kluster

	Anggota Kluster	
	Kluster 1	Kluster 2
Urut Rumah Tangga	1, 2, 4, 5, 6, 7, 11, 12, 13, 15, 16, 18, 19,... , 880, 881, 882, 883, 885, 886, 887, 888, 889, 890	3, 8, 9, 10, 14, 17, 20, 22, 33, 35, 38, 44, 45, 46,... , 821, 822, 828, 829, 838, 845, 859, 870, 871, 884
Jumlah	17.330	8.560

Adapun karakteristik dari setiap kluster dapat di interpretasi berdasarkan persentase capaian kelayakan untuk setiap indikator rumah layak huni. Berikut merupakan persentase dari capaian kelayakan indikator rumah layak huni untuk setiap kluster:

Tabel 8. Persentase Capaian Kelayakan Indikator Rumah Layak Huni Setiap Klaster

k	Kecukupan Luas Tempat Tinggal		Ketahanan Bangunan		Akses Terhadap Sanitasi Layak		Akses Terhadap Air Minum Layak	
	Layak	Tidak Layak	Layak	Tidak Layak	Layak	Tidak Layak	Layak	Tidak Layak
1	91,177%	8,823%	79,908%	20,092 %	77,398%	22,602%	95,897%	4,103%
2	94,544%	5,456%	75,292%	24,708 %	71,963%	28,037%	87,769%	12,231%

Persentase capaian kelayakan indikator rumah layak huni dihitung berdasarkan pada kriteria rumah layak huni yang telah dibentuk sebelumnya [5]. Adapun kriteria suatu rumah dapat dikatakan layak huni apabila [5]:

1. Memiliki luas lantai perkapita minimal 7,2 m². Dimana angka tersebut dihitung dari hasil bagi luas tempat tinggal dengan jumlah anggota rumah tangga.
2. Memiliki jenis atap terluas tempat tinggal berupa beton, genteng, seng, dan kayu/sirap.
3. Memiliki jenis dinding terluas tempat tinggal berupa tembok, plesteran anyaman bambu/kawat, kayu/papan, dan batang kayu.
4. Memiliki jenis lantai terluas tempat tinggal berupa marmer/granit, keramik, parket/vinil/karpet, ubin/tegel/teraso, kayu/papan, dan semen/bata merah.
5. Memiliki fasilitas tempat buang air besar yang digunakan sendiri, atau digunakan bersama dengan ART dari rumah tangga tertentu, atau menggunakan fasilitas di MCK umum/siapapun menggunakan.
6. Menggunakan jenis kloset berupa leher angsa.
7. Menggunakan tangki septik, atau IPAL, atau lubang tanah (khusus untuk rumah di pedesaan) sebagai tempat pembuangan akhir tinja.
8. Menggunakan air minum yang bersumber dari leding, sumur bor/pompa, sumur terlindung, mata air terlindung, dan air hujan.
9. Menggunakan air minum yang bersumber dari air bermerk dan air isi ulang apabila air untuk mandi/cuci/dll bersumber dari leding, sumur bor/pompa, sumur terlindung, mata air terlindung, dan air hujan.

Berdasarkan kriteria yang telah diuraikan, diperoleh persentase untuk melihat capaian kelayakan terhadap indikator pembentuk rumah layak huni sebagaimana Tabel 8. Terlihat untuk Tabel 8 yang diberi warna latar hijau merupakan nilai dengan persentase capaian tidak layak tertinggi pada setiap indikator dan angka yang bercetak tebal merupakan nilai dengan persentase capaian tidak layak tertinggi pada setiap klaster.

Maka dari itu, dapat disimpulkan karakteristik dari klaster 1 dan 2 dibedakan sebagai berikut:

1. Klaster 1
 - Memiliki 17.330 anggota yang sebagian besar berasal dari perkotaan.
 - Memiliki capaian rumah layak huni yang baik karena hanya 1 indikator yang memiliki angka persentase tidak layak tertinggi.
2. Klaster 2
 - Memiliki 8.560 anggota yang sebagian besar berasal dari pedesaan.
 - Memiliki capaian rumah layak huni yang kurang karena terdapat 3 indikator yang memiliki angka persentase tidak layak tertinggi.

D. Kesimpulan

Berdasarkan hasil dan pembahasan dari penelitian ini, maka diperoleh beberapa kesimpulan sebagai berikut:

1. Hasil dari pengelompokan rumah tangga yang paling optimal di Provinsi Jawa Barat tahun 2023 dengan menggunakan metode *K-Prototypes* berdasarkan indikator

pembentuk rumah layak huni diperoleh sebanyak 2 klaster. Klaster 1 memiliki anggota sebanyak 17.330 rumah tangga dan klaster 2 memiliki anggota sebanyak 8.560 rumah tangga.

2. Klaster 1 menjadi kelompok rumah tangga yang sebagian besar berasal dari perkotaan dan memiliki capaian rumah layak huni yang baik.
3. Klaster 2 menjadi kelompok rumah tangga yang sebagian besar berasal dari pedesaan dan memiliki capaian rumah layak huni yang kurang.

Acknowledge

Puji dan Syukur penulis panjatkan kehadirat Allah SWT., yang telah memberi rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan penelitian ini. Penulis juga ucapkan terima kasih kepada Bapak Abdul Kudus, S.Si., M.Si., Ph.D., selaku dosen pembimbing yang telah memberi saran dan bimbingannya dalam pelaksanaan penelitian ini.

Daftar Pustaka

- [1] Annas S, Irwan, S RH, et al. K-Prototypes Algorithm for Clustering The Tectonic Earthquake in Sulawesi Island. *Jurnal Varian* 2022; 5: 191–198.
- [2] Yin S, Gan G, Valdez EA, et al. Applications of Clustering with Mixed Type Data in Life Insurance. *Risks* 2021; 9: 1–19.
- [3] Ganmanah M. *Penerapan Algoritme K-Prototypes untuk Pengelompokan Desa-Desa di Provinsi Jawa Barat Berdasarkan Indikator Indeks Desa Membangun Tahun 2020*. Skripsi Tidak Dipublikasikan, Universitas Islam Bandung, 2021.
- [4] Septianingsih A. Analisis K-Means Clustering Pada Pemetaan Provinsi Indonesia Berdasarkan Indikator Rumah Layak Huni. *Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika* 2022; 3: 224–241.
- [5] Direktorat Statistik Kesejahteraan Rakyat. *Indikator Perumahan dan Kesehatan Lingkungan 2023*. Jakarta, December 2023.
- [6] Harmain A, Paiman, Kurniawan H, et al. Normalisasi Data Untuk Efisiensi K-Means Pada Pengelompokan Wilayah Berpotensi Kakaran Hutan dan Lahan Berdasarkan Sebaran Titik Panas. *TEKNIMEDIA* 2021; 2: 83–89.
- [7] Harmouch M. 17 Types of Similarity and Dissimilarity Measures Used in Data Science. *Medium.com*.
- [8] Huang Z. Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values. *Data Mining and Knowledge Discovery* 2 1998; 283–304.
- [9] Huang Z. *Clustering Large Data Sets with Mixed Numeric And Categorical Values*. 1997.
- [10] Nooraeni R. Metode Cluster Menggunakan Kombinasi Algoritma Cluster K-Prototype dan Algoritma Genetika Untuk Data Bertipe Campuran. *Jurnal Aplikasi Statistika & Komputasi Statistik* 2015; 7: 81–98.
- [11] Novidianto R, Fithriasari K. Algoritma ClusterMix K-Prototypes Untuk Menangkap Karakteristik Pasien Berdasarkan Variabel Penciri Mortalitas Pasien Dengan Gagal Jantung. *Inferensi* 2021; 4: 37–46.
- [12] Orisa M. Optimasi Cluster pada Algoritma K-Means. *Metaverse: Peluang dan Tantangan Pendidikan Tinggi di Era Industri 50* 2022; 430–437.