

## Implementasi Algoritma *K-Means* Dalam Pengelompokan Penerima Bantuan Bedah Rumah

Erwin Arry Kusuma<sup>1\*</sup>, Adani Dharmawati<sup>2</sup>, Nisrinah<sup>3</sup>

<sup>1,3</sup>Teknik Informatika, STMIK Banjarbaru, Banjarbaru, Indonesia

<sup>2</sup>Teknik Informatika, Universitas Islam Kalimantan, Banjarmasin, Indonesia

\*e-mail Corresponding Author: erwinarry@gmail.com

### Abstract

*Self-help housing, built through community initiatives, is a crucial solution to address substandard housing in Indonesia. The Self-Help Housing Stimulant Assistance Program (BSPS) supports the construction of these houses. This study aims to implement the K-Means method to prioritize recipients of house renovation assistance based on socioeconomic data and housing conditions. Using 207 recipient data from Hulu Sungai Tengah Regency, the attributes considered include the number of dependents, monthly income, and house size. The clustering process resulted in two priority recipient clusters. The analysis results show that the K-Means method can classify aid recipients with an accuracy of 61%. Results of this study can help the government allocate resources more efficiently and effectively.*

**Keywords:** Clustering; Data Mining; K-Means; House Renovation

### Abstrak

Rumah swadaya yang dibangun atas inisiatif masyarakat adalah solusi penting untuk mengatasi masalah rumah tidak layak huni (RTLH) di Indonesia. Program Bantuan Stimulan Perumahan Swadaya (BSPS) mendukung pembangunan rumah swadaya ini. Penelitian ini bertujuan mengimplementasikan metode *K-Means* untuk mengelompokkan prioritas penerima bantuan bedah rumah berdasarkan data sosial ekonomi dan kondisi rumah. Menggunakan 207 data penerima bantuan dari Kabupaten Hulu Sungai Tengah, atribut yang dipertimbangkan meliputi jumlah tanggungan, penghasilan per bulan, dan luas rumah. Proses pengelompokan menghasilkan dua klaster prioritas penerima bantuan. Hasil analisis menunjukkan bahwa metode *K-Means* dapat mengelompokkan penerima bantuan dengan akurasi 61%. Hasil penelitian ini dapat membantu pemerintah dalam mengalokasikan sumber daya secara lebih efisien dan tepat sasaran.

**Kata Kunci:** Clustering; Data Mining; K-Means; Bedah Rumah

### 1. Pendahuluan

Rumah swadaya atau rumah yang dibangun atas prakarsa dan upaya masyarakat merupakan solusi penting untuk mengatasi masalah rumah tidak layak huni (RTLH) di Indonesia. Program Bantuan Stimulan Perumahan Swadaya (BSPS) dari Kementerian Pekerjaan Umum dan Perumahan Rakyat (PUPR) adalah salah satu skema bantuan yang mendukung pembangunan rumah swadaya ini. Program BSPS atau Bantuan Stimulan Perumahan Swadaya adalah suatu tindakan dari pemerintah dimana program bantuan atau suntikan dana untuk masyarakat berpenghasilan rendah (MBR) dengan harapan untuk memperbaiki tempat tinggal yang masih tidak layak ditempati [1]. Program BSPS membantu masyarakat meningkatkan kualitas rumah mereka sehingga layak huni. Berdasarkan data terbaru dari Kementerian PUPR, jumlah RTLH di Indonesia mencapai sekitar 43 juta unit yang tersebar di 514 kabupaten/kota.

Pada era perkembangan teknologi informasi yang pesat, penggunaan metode data mining menjadi semakin penting dalam mendukung pengambilan keputusan di berbagai bidang, termasuk dalam peningkatan kualitas hidup masyarakat. Data mining adalah teknik yang digunakan untuk mengidentifikasi pola dalam data berjumlah besar. Teknik ini memiliki kelebihan dan kekurangan, serta mencakup berbagai metodologi dan tugas untuk berbagai aplikasi [2]. Salah satu tantangan utama yang dihadapi oleh pemerintah adalah masalah perumahan tidak layak huni (RTLH) yang masih tersebar luas di berbagai wilayah di Indonesia. Untuk mengatasi

masalah ini, pemerintah dan lembaga terkait telah menerapkan berbagai program bantuan, salah satunya adalah Program Bantuan Bedah Rumah.

Selama ini penentuan penerima Program Bantuan Bedah Rumah dilakukan secara konvensional. Pengambilan Keputusan yang tidak tepat dapat menyebabkan permasalahan pada alokasi bantuan. Misalnya, bantuan mungkin tidak diberikan kepada mereka yang paling membutuhkannya atau sebaliknya, diberikan kepada yang kurang membutuhkan. Selain itu, ketika jumlah data penerima bantuan bertambah, pengelompokan manual menjadi semakin tidak praktis dan sulit dikelola. Dalam konteks ini, penggunaan metode pengelompokan seperti *K-Means clustering* dapat menjadi alat yang efektif untuk mengidentifikasi dan mengelompokkan prioritas penerima bantuan bedah rumah. Clustering adalah teknik pengelompokan objek berdasarkan kesamaan antar objek [3]. *K-Means* merupakan salah satu metode clustering yang umum digunakan dalam analisis data, yang bertujuan untuk mengelompokkan data ke dalam beberapa kelompok berdasarkan kemiripan karakteristik tertentu. Algoritma *K-Means* dalam teknologi data mining secara efektif mengelompokkan dan mengelola data, mendukung aktivitas inovasi, analisis optimasi, dan pengambilan keputusan di perusahaan [4].

Penelitian ini bertujuan untuk mengimplementasikan metode *K-Means* dalam pengelompokan prioritas penerima bantuan bedah rumah, dengan memanfaatkan data sosial ekonomi dan kondisi rumah sebagai variabel input. Melalui analisis *clustering*, diharapkan dapat dihasilkan kelompok-kelompok penerima bantuan yang memiliki tingkat prioritas yang berbeda, sehingga memungkinkan pemerintah dan lembaga terkait untuk mengalokasikan sumber daya secara efisien dan merencanakan intervensi yang sesuai untuk setiap kelompok. Analisis cluster merupakan suatu analisis multivariat yang digunakan untuk mengelompokkan objek-objek berdasarkan kemiripan karakteristik yang dimiliki [5].

## 2. Tinjauan Pustaka

Penelitian berjudul "*Clustering Menggunakan Metode K-Means Untuk Menentukan Prioritas Penerima Bantuan Bedah Rumah (Studi Kasus: Desa Ciomas Bogor)*" mengembangkan metode penilaian prioritas bagi penduduk miskin dengan menggunakan data mining dan pengklasteran *K-Means*. Metode ini dapat membantu dalam pengambilan keputusan mengenai kelompok penduduk yang layak menerima bantuan bedah rumah. Selain itu, hasil dari penelitian ini bisa dijadikan acuan bagi kepala desa untuk merumuskan kebijakan lanjutan berdasarkan informasi yang diperoleh dari metode pengklasteran *K-Means*. [6].

Hasil dari penelitian yang berjudul "*Algoritma K-Means Clustering dalam Memprediksi Penerima Bantuan Langsung Tunai (BLT) Dana Desa*" bertujuan untuk meramalkan penerima BLT-DD yang tepat sasaran. Penelitian ini menggunakan data usulan penerima BLT-DD Nagari Taluk pada tahun 2022. Setelah mengolah data dari 25 sampel, ditemukan bahwa 11 data termasuk dalam cluster 1 dengan status penerima bantuan yang layak, 5 data termasuk dalam cluster 2 dengan status penerima yang dipertimbangkan, dan 9 data termasuk dalam cluster 3 dengan status tidak layak. Dari hasil pengujian, tingkat akurasi sebesar 83,33% berhasil diperoleh, sehingga rekomendasi dapat diberikan kepada pihak pemerintah Nagari untuk membantu dalam pengambilan kebijakan yang tepat [7].

Penelitian dengan judul "*Algoritma K-Means untuk Klasterisasi Penerima Program Keluarga Harapan di Nanaenoe Nusa Tenggara Timur*" bertujuan untuk memberikan prioritas kepada keluarga berpenghasilan rendah dalam menerima bantuan Program Keluarga Harapan. Metode yang digunakan adalah *K-Means clustering*, yang melibatkan langkah-langkah pengumpulan data, data latih, dan data uji dengan mempertimbangkan kriteria seperti nama lengkap, tempat lahir, status perkawinan, pekerjaan, dusun, dan pendidikan terakhir. Melalui pengujian dengan  $k=10$  menggunakan *K-Means clustering* dan Davies Bouldin index, ditemukan urutan  $K_2=0.883$ ,  $K_3=1.225$ ,  $K_4=1.059$ ,  $K_5=0.934$ ,  $K_6=0.838$ ,  $K_7=0.764$ ,  $K_8=0.817$ ,  $K_9=0.835$ , dan  $K_{10}=0.866$ . Kesimpulan dari proses klasterisasi menunjukkan bahwa nilai Davies Bouldin Index terbaik adalah 0.764 pada urutan  $K_7$ . Penggunaan metode *K-Means clustering* memungkinkan identifikasi kelompok penerima manfaat dari Program Keluarga Harapan di Desa Nanaenoe [8].

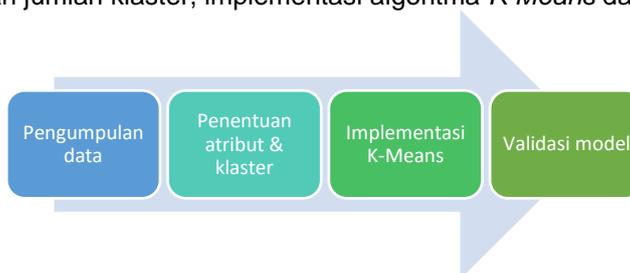
Penelitian terakhir yang dilakukan oleh [9], membahas Penentuan Penerima Bantuan Sosial Beras untuk Masyarakat Miskin Menggunakan Metode *K-Means*. Penelitian ini berhasil mengembangkan aplikasi untuk melakukan pengoperasian metode klaster *K-Means* dan berdasarkan pada pengujian metode dengan menggunakan *Confussion Matrix* didapatkan

tingkat akurasi sebesar 88%, tingkat presisi sebesar 94%, angka recall sebesar 76% dan angka F-1 Score sebesar 84% dari total data sebanyak 50 data.

Dalam penelitian ini akan dibuat sebuah aplikasi yang di dalamnya akan diterapkan algoritma *K-Means* untuk mengelompokkan penerima bantuan bedah rumah. Kriteria yang digunakan adalah jumlah tanggungan, penghasilan perbulan, dan luas rumah. Perbedaan mendasar dengan penelitian sebelumnya, pada penelitian ini menggunakan jumlah dataset lebih banyak yaitu sebanyak 207 data.

### 3. Metodologi

Pada penelitian ini terdapat tahapan-tahapan yang dilakukan, yaitu pengumpulan data, penentuan atribut dan jumlah kluster, implementasi algoritma *K-Means* dan validasi model.



Gambar 1. Tahapan penelitian

#### 3.1. Pengumpulan data

Metode pengumpulan data yang digunakan dalam penelitian ini adalah wawancara, observasi, dan kajian literatur. Data dikumpulkan melalui wawancara langsung dengan Dinas Perumahan dan Permukiman Kabupaten Hulu Sungai Tengah. Selain itu, data juga diperoleh melalui pengamatan langsung dan pencatatan hasil observasi dari Perumahan dan Permukiman Kabupaten Hulu Sungai Tengah. Untuk melengkapi data, dilakukan kajian literatur dengan mengkaji beberapa paper, artikel, berita, buku, dan sumber lain yang relevan dengan penelitian ini. Pengumpulan data adalah alat mendasar untuk penelitian, dengan metode seperti observasi, wawancara, kuesioner, dan akses database menjadi hal yang penting untuk mencapai tujuan penelitian [10].

Penelitian ini menggunakan metode populasi. Data diperoleh dari Dinas Perumahan dan Permukiman Kabupaten Hulu Sungai Tengah tentang calon penerima bantuan bedah rumah di desa Murung, Kindingan, dan Tilahaan, Kecamatan Hantakan, pada tahun 2018 dengan total 207 data. Data yang diperoleh mencakup informasi NIK, nama lengkap, jenis kelamin, alamat, jenis bantuan, jumlah tanggungan, pekerjaan, penghasilan per bulan, luas rumah, kondisi lantai, kondisi dinding, dan kondisi atap. Memahami metode pengambilan sampel dalam desain penelitian sangat penting untuk menganalisis dan menarik kesimpulan dari data [11].

Tabel 1. Sampel Dataset

No	NIK	Nama Lengkap	Jenis Kelamin	Alamat	Jenis Bantuan	Jumlah Tanggungan	Pekerjaan	Penghasilan Per Bulan	Luas Rumah	Kondisi Lantai	Kondisi Dinding	Kondisi Atap
1	63070907056003457	Hamsan	L	Murung B Rt 01. Rw. 01	PK	-	Petani	300.000	45	Tidak Layak	Tidak Layak	Tidak Layak
2	63070947057007689	Abdullah	L	Murung B Rt 01. Rw. 01	PK	-	Petani	400.000	15	Tidak Layak	Tidak Layak	Layak
3	63070907056000001	Sarjono	L	Murung B Rt 01. Rw. 01	PK	2	Petani	400.000	50	Layak	Layak	Layak
4	63070947057000002	Rasidah	P	Murung B Rt 01. Rw. 01	PK	1	Petani	300.000	24	Layak	Layak	Layak
5	6307094107570063	Salamiah	P	Murung B Rt 01. Rw. 01	PK	-	Petani	500.000	24	Layak	Layak	Layak
6	6307094207830255	Bariah	P	Murung B Rt 01. Rw. 01	PK	2	Petani	500.000	18	Layak	Layak	Layak
7	6307094107722114	Sahliah	P	Murung B Rt 02. Rw. 01	PK	5	Petani	500.000	36	Tidak Layak	Tidak Layak	Layak
8	6307094107620090	Asmiah	P	Murung B Rt 02. Rw. 01	PK	2	Petani	500.000	64	Layak	Layak	Layak
9	6307094107630192	Saiyah	P	Murung B Rt 01. Rw. 01	PK	1	Petani	300.000	18	Layak	Layak	Layak
10	6307094107560024	Jumaiyah	P	Murung B Rt 03. Rw 02	PK	2	Petani	300.000	28	Tidak Layak	Tidak Layak	Tidak Layak
11	63070904033400001	Andrus	L	Murung B Rt 03. Rw 02	PK	1	Petani	500.000	28	Layak	Tidak Layak	Layak
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....

No	NIK	Nama Lengkap	Jenis Kelamin	Alamat	Jenis Bantuan	Jumlah Tanggungan	Pekerjaan	Penghasilan Per Bulan	Luas Rumah	Kondisi Lantai	Kondisi Dinding	Kondisi Atap
197	6307091001870002	RIDUANSYAH	L	KINDINGAN RT.03 RW.02	PK	2	Petani	150.000	16	Tidak Layak	Tidak Layak	Layak
198	6307011112470001	AS'ARI	L	KINDINGAN RT.03 RW.02	PK	-	Petani	50.000	6	Tidak Layak	Tidak Layak	Tidak Layak
199	6307090107710113	PUDIN	L	KINDINGAN RT.03 RW.02	PK	2	Petani	150.000	10	Layak	Layak	Layak
200	6307094208730001	PADILLAH	P	KINDINGAN RT.04 RW.03	PK	-	Petani	50.000	10	Tidak Layak	Tidak Layak	Tidak Layak
201	6307094107430101	PURNAMA	P	KINDINGAN RT.04 RW.03	PK	3	Petani	150.000	16	Layak	Layak	Layak
202	6307090106830004	WARI	L	KINDINGAN RT.04 RW.03	PK	3	Petani	50.000	10	Tidak Layak	Tidak Layak	Layak
203	6307090107730231	JIDI	L	KINDINGAN RT.04 RW.04	PK	4	Petani	150.000	24	Layak	Layak	Layak
204	6307090107680221	MURSIDI	L	KINDINGAN RT.04 RW.04	PK	3	Petani	150.000	16	Layak	Layak	Layak
205	6307090107670075	TAUPIK RAHMAN	L	KINDINGAN RT.04 RW.04	PK	4	Petani	150.000	16	Tidak Layak	Tidak Layak	Tidak Layak
206	6306030107820045	M.ARIANI	L	KINDINGAN RT.04 RW.04	PK	3	Petani	150.000	18	Tidak Layak	Tidak Layak	Tidak Layak
207	6307090107730274	JUHAN	L	KINDINGAN RT.04 RW.04	PK	3	Petani	150.000	18	Tidak Layak	Tidak Layak	Tidak Layak

### 3.2. Penentuan Atribut dan Klaster

Sampel data yang digunakan adalah data penerima bantuan bedah rumah pada tahun 2018 pada desa Murung, Kindingan, dan Tilahaan Kecamatan Hantakan. Pengaturan atribut awal proses perhitungan adalah sebagai berikut:

Tabel 2. Penentuan Atribut & Klaster

Jumlah Klaster	Jumlah data	Jumlah atribut
2	207	3

### 3.3. Implementasi K-Means

Langkah-langkah pengelompokan menggunakan *K-Means*:

- 1) Tentukan nilai K sebagai jumlah *cluster* yang ingin dibentuk, K = 2
- 2) Bangkitkan k centroid (titik pusat cluster awal) awal secara acak:
  - a. Diambil data ke-7 sebagai pusat *cluster*  
K1 = (Jumlah tanggungan = 3 orang Jumlah penghasilan = 175000, luas rumah = 14 M2)
  - b. Diambil data ke-11 sebagai pusat *cluster*  
K2 = (Jumlah tanggungan = 2 orang Jumlah penghasilan = 150000, luas rumah = 12 M2)

Tabel 3. Pusat Cluster

Data ke	No.KTP	Jumlah Tanggungan	Penghasilan Per Bulan	Luas Rumah
5	6307090811850002	0	600000	35
72	6307090107570021	7	400000	40

- 3) Hitung jarak setia data ke masing-masing *centroid* menggunakan rumus

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \dots \dots \dots (1)$$

Ket:

dij: Jarak objek antara objek x dan y

n: Jumlah Atribut

Xi: Objek Data

Yi: Data Cluster

- Cluster Data ke-1 (0,300000 dan 45)

$$C1 = \sqrt{(0 - 0)^2 + (300000 - 600000)^2 + (45 - 35)^2} = 300000$$

$$C2 = \sqrt{(0 - 7)^2 + (300000 - 400000)^2 + (45 - 40)^2} = 100000$$

- Cluster Data ke-2 (0,400000 dan 15)

$$C1 = \sqrt{(0 - 0)^2 + (400000 - 600000)^2 + (15 - 35)^2} = 200000$$

- $C2 = \sqrt{(0 - 7)^2 + (400000 - 400000)^2 + (15 - 40)^2} = 26$
- Cluster Data ke-3 (2,400000 dan 50)
- $C1 = \sqrt{(2 - 0)^2 + (400000 - 600000)^2 + (50 - 35)^2} = 200000$
- $C2 = \sqrt{(2 - 7)^2 + (400000 - 400000)^2 + (50 - 40)^2} = 11$
- 

Berdasarkan sampey perhitungan diatas maka lebih lengkapnya data tersebut didalam tabel perhitungan jarak *cluster*.

Tabel 4. Perhitungan jarak pada cluster

No	NIK	Jumlah Tanggungan	Penghasilan Per Bulan	Luas Rumah	C1	C2	Jarak Terpendek
1	63070907056003457	-	300.000	45	300000	100000	100000
2	63070947057007689	-	400.000	15	200000	26	26
3	6307090705600001	2	400.000	50	200000	11	11
4	6307094705700002	1	300.000	24	300000	100000	100000
5	6307094107570063	-	500.000	24	100000	100000	100000
6	6307094207830255	2	500.000	18	100000	100000	100000
7	6307094107722114	5	500.000	36	100000	100000	100000
8	6307094107620090	2	500.000	64	100000	100000	100000
9	6307094107630192	1	300.000	18	300000	100000	100000
10	6307094107560024	2	300.000	28	300000	100000	100000
11	63070904033400001	1	500.000	28	100000	100000	100000
.....	.....	.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....	.....	.....
197	6307091001870002	2	150.000	16	450000	250000	250000
198	6307011112470001	-	50.000	6	550000	350000	350000
199	6307090107710113	2	150.000	10	450000	250000	250000
200	6307094208730001	-	50.000	10	550000	350000	350000
201	6307094107430101	3	150.000	16	450000	250000	250000
202	6307090106830004	3	50.000	10	550000	350000	350000
203	6307090107730231	4	150.000	24	450000	250000	250000
204	6307090107680221	3	150.000	16	450000	250000	250000
205	6307090107670075	4	150.000	16	450000	250000	250000
206	6306030107820045	3	150.000	18	450000	250000	250000
207	6307090107730274	3	150.000	18	450000	250000	250000

Keterangan:

- Data ke- : Urutan *Cluster* data ke-n
- No.KTP : Nomor Kartu Tanda Penduduk
- Jumlah Tanggungan : Jumlah tanggungan dalam satu keluarga
- Penghasilan perbulan : Penghasilan perbulan yang diperoleh
- Luas Rumah : Luas rumah dalam satu keluarga
- C1 : Hasil perhitungan jarak dari pusat *cluster*
- C2 : Hasil perhitungan jarak dari pusat *cluster*
- Jarak Terpendek : Jarak terpendek antara nilai C1 dan C2

4) Pengelompokkan data jarak terdekat antara data dengan *centroid* (Iterasi 1)

Tabel 5. Tabel Hasil Pengelompokan Berdasarkan centroid

Data ke	NIK	Tanggungan	Penghasilan Perbulan	Luas Rumah	C1	C2	Jarak Terpendek	C1	C2
2	6307094705 7007689	-	400.00 0	15	200000	26	26		OK
3	6307090705 600001	2	400.00 0	50	200000	11	11		OK
4	6307094705 700002	1	300.00 0	24	300000	1000 00	100000		OK
5	6307094107 570063	-	500.00 0	24	100000	1000 00	100000	OK	
6	6307094207 830255	2	500.00 0	18	100000	1000 00	100000	OK	
7	6307094107 722114	5	500.00 0	36	100000	1000 00	100000		OK
8	6307094107 620090	2	500.00 0	64	100000	1000 00	100000		OK
9	6307094107 630192	1	300.00 0	18	300000	1000 00	100000		OK
10	6307094107 560024	2	300.00 0	28	300000	1000 00	100000		OK
11	63070904033 400001	1	500.00 0	28	100000	1000 00	100000	OK	
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
197	63070910018 70002	2	150.00 0	16	450000	2500 00	250000		OK
198	63070111124 70001	-	50.000	6	550000	3500 00	350000		OK
199	63070901077 10113	2	150.00 0	10	450000	2500 00	250000		OK
200	63070942087 30001	-	50.000	10	550000	3500 00	350000		OK
201	63070941074 30101	3	150.00 0	16	450000	2500 00	250000		OK
202	63070901068 30004	3	50.000	10	550000	3500 00	350000		OK
203	63070901077 30231	4	150.00 0	24	450000	2500 00	250000		OK
204	63070901076 80221	3	150.00 0	16	450000	2500 00	250000		OK
205	63070901076 70075	4	150.00 0	16	450000	2500 00	250000		OK
206	63060301078 20045	3	150.00 0	18	450000	2500 00	250000		OK
207	63070901077 30274	3	150.00 0	18	450000	2500 00	250000		OK

Keterangan:

Data ke-	: Urutan <i>Cluster</i> data ke-n
No.KTP	: Nomor Kartu Tanda Penduduk
Jumlah Tanggungan	: Jumlah tanggungan dalam satu keluarga
Penghasilan perbulan	: Penghasilan perbulan yang diperoleh
Luas Rumah	: Luas rumah dalam satu keluarga
C1	: Hasil perhitungan jarak dari pusat <i>cluster</i>
C2	: Hasil perhitungan jarak dari pusat <i>cluster</i>
Jarak Terpendek	: Jarak terpendek antara nilai C1 dan C2
C0	: Pengelompokan berdasarkan <i>Centroid</i> (OK/-)
C1	: Pengelompokan berdasarkan <i>Centroid</i> (OK/-)

Setelah iterasi 1 selesai, lanjutkan ke iterasi 2, 3, 4 dan seterusnya. Iterasi adalah pengulangan proses, berhenti ketika hasil iterasi telah konvergen [12] atau dengan kata lain iterasi bisa selesai jika pada suatu itersi ke-n tidak ada pertukaran anggota sama sekali berdasarkan centroid (anggota masing-masing cluster sama dengan iterrasi sebelumnya).

### 3.4. Evaluasi Model

Evaluasi model *K-Means* pada penelitian ini menggunakan *Confusion Matrix*. Confusion matrix adalah tabel yang digunakan untuk menilai kinerja model klasifikasi biner dengan membandingkan sampel aktual dan prediksi, menyatakan jumlah data uji yang benar dan salah diklasifikasikan [13] [14].

*Confusion matrix* menggunakan empat parameter: *precision*, *recall*, *accuracy* dan *Fscore*. Parameter evaluasi tersebut diperoleh dengan menggunakan Persamaan berikut [15]:

$$\begin{aligned}
 precision &= \frac{TP}{TP+FP} \times 100 \\
 recall &= \frac{TP}{TP+FN} \times 100 \\
 recall &= \frac{TP+TN}{TP+TN+FP+FN} \times 100 \dots\dots\dots (2)
 \end{aligned}$$

**4. Hasil dan Pembahasan**

**4.1. Implementasi Program**

Berikut komponen-komponen sistem yang dapat ditunjukkan di bawah ini:

- 1) Halaman Masuk



Gambar 2. Tampilan Masuk

Halaman masuk adalah halaman pertama akan muncul pada saat aplikasi digunakan, berfungsi untuk mengakses manu bagi admin.

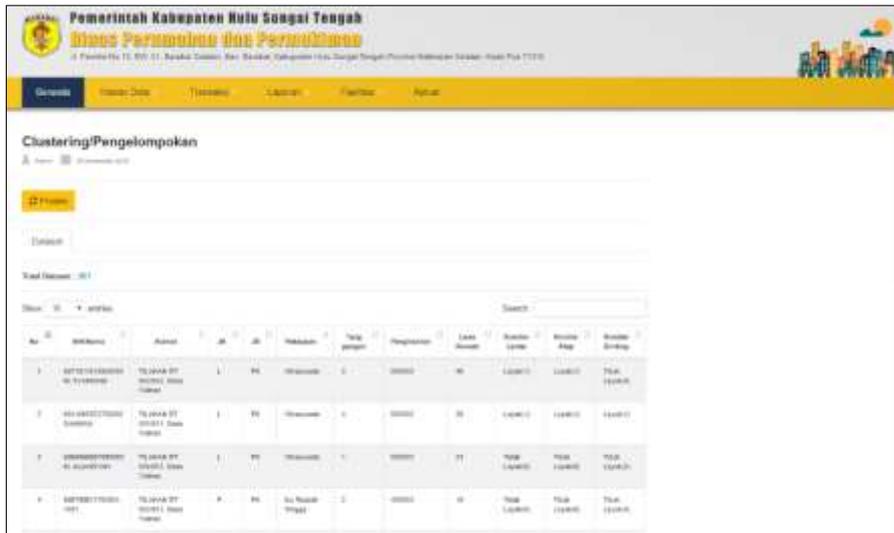
- 2) Menu Utama



Gambar 3. Tampilan Menu Utama

Menu utama adalah halaman yang tampil pertama kali muncul setelah pengguna berhasil login, pada tampilan ini terdapat seluruh sistem informasi dari sistem input hingga sistem output. Menu utama dapat mengontrol semua halaman yang ada yaitu Data Penduduk, Transaksi, Laporan dan Keluar Aplikasi.

3) Proses



Gambar 4. Tampilan Proses

Proses adalah halaman yang digunakan untuk memproses data penduduk dengan memuat data lalu dikelompokkan dan menghasilkan output berupa hasil pengelompokan.

4) Laporan



Gambar 5. Tampilan Laporan

4.2. Hasil Pengujian Sistem

Untuk pengujian akurasi *Clustering* yang akan dilakukan pada penelitian ini yaitu menggunakan precision, recall dan f-score, menggunakan tabel *Confusion matrix* adalah sebuah tabel yang menyatakan jumlah data uji yang benar diklasifikasikan dan jumlah data uji yang salah diklasifikasikan. Contoh confusion matrix untuk pengelompokan ditunjukkan pada

Tabel 6. *Confusion Matrix*

	Data	Aktual	
		Kelas Positive	Kelas Negative
Hasil Prediksi	Kelas Positive	True Positive (TP)	False Positive (FP)
	Kelas Negative	False Negative (FN)	True Negative (TN)
Total		P	N

Pada Penelitian ini diberikan nilai berdasarkan hasil cluster pada analisa data  
 TP = 90  
 FP = 117  
 FN = 0  
 TN = 0

1) *Precision*

*Precision* (p) = jumlah sampel berkategori positif diklasifikasi benar dibagi dengan total sampel yang diklasifikasi sebagai sample positif

$$\begin{aligned} \text{Precision} &= TP/(TP+FP) \\ &= 90/(90+117) \\ &= 90/207 = 0,43 \end{aligned}$$

Dari hasil di atas, precision adalah 0.43. Artinya, dari semua prediksi positif yang dibuat oleh model, hanya 43% yang benar-benar positif. Ini menunjukkan bahwa ada banyak false positives (FP = 117), yang mengindikasikan bahwa model ini sering salah mengklasifikasikan sampel negatif sebagai positif.

2) *Recall*

*Recall* (r) = jumlah sampel diklasifikasi positif dibagi total sampel dalam testing set berkategori positif

$$\begin{aligned} \text{Recall} &= TP/(TP + FN) \\ &= 90/(90 + 0) = 1 \end{aligned}$$

Dari hasil di atas, recall adalah 1. Artinya, model berhasil menemukan semua sampel positif yang ada. Tidak ada false negatives (FN = 0), yang menunjukkan bahwa model tidak melewatkan satu pun sampel positif.

3) *F-score*

*F-Measure* (F1) adalah *harmonic mean* dari *precision* dan *recall*, range dari F1 adalah 0-1.

$$\begin{aligned} F_1 &= 2 \cdot (\text{precision} \cdot \text{recall}) / (\text{precision} + \text{recall}) \\ &= 2 \cdot (0,43 \cdot 1) / (0,43 + 1) \\ &= 2 \cdot 0,43 / 1,43 = 0,61 \end{aligned}$$

Dari hasil di atas, F1 Score adalah 0.60. F1 Score ini menunjukkan keseimbangan antara *precision* dan *recall*. Dengan *precision* yang rendah (0.43) namun *recall* yang tinggi (1), F1 Score berada di tengah-tengah pada 0.61 atau 61%.

*Precision* yang rendah menunjukkan bahwa model memiliki banyak false positives, yang berarti model sering salah mengklasifikasikan sampel negatif sebagai positif. *Recall* yang tinggi menunjukkan bahwa model sangat baik dalam menangkap semua sampel positif tanpa melewatkan satu pun. F1 Score menunjukkan keseimbangan antara *precision* dan *recall*. Dalam kasus ini, meskipun *recall* sangat baik, rendahnya *precision* menurunkan F1 Score.

Metode *K-Means* dapat membantu dalam pengambilan keputusan mengenai penentuan penerima bantuan, sebagaimana temuan penelitian [6] [7] [8], namun hasil menunjukkan akurasi yang menurun dibanding penelitian [9] dengan menggunakan jumlah data yang lebih sedikit.

## 5. Simpulan

Penerapan Metode *K-Means* pada penerima bantuan bedah rumah berdasarkan kriteria yang ada di Dinas Perumahan dan Permukiman Hulu Sungai Tengah berhasil diterapkan kedalam aplikasi agar meringankan para pegawai dalam pendataan dan pengelompokan sehingga penerima bantuan bedah rumah lebih tepat sasaran. Penerapan Metode *K-Means* pada penerima bantuan bedah rumah berdasarkan kriteria yang ada di Dinas Perumahan dan Permukiman Hulu Sungai Tengah dengan 207 data menghasilkan 0,61 atau sama dengan 61% atau mengalami penurunan dengan menggunakan *F-score*. Dan untuk menghasilkan hasil yang lebih akurat lagi, dapat dilakukn penelitan lebih lanjut tentang hasil yang sudah didapat bisa ditambah metode perengkingan agar penerima bantuan bisa disesuaikan dengan jumlah penerima maupun anggaran yang sudah ditetapkan dari hasil pendataan.

## Daftar Referensi

- [1] M. A. Aziz, A. Bachtiar dan K. Asmara, "Implementasi Program Bantuan Stimulan Perumahan Swadaya (BSPS) Di Desa Gedangan Oleh Satker Penyediaan Perumahan Jawa IV," *Jurnal Ilmiah Wahana Pendidikan*, vol. 9, no. 22, pp. 645-652, 2023.
- [2] J. O. Ogunleye, *The Concept of Data Mining*, Rijeka: IntechOpen, 2021.

- [3] A. Pangestu dan T. Ridwan, "Penerapan Data Mining Menggunakan Algoritma Kmeans Pengelompokan Pelanggan Berdasarkan Kubikasi Air Terjual Menggunakan WEKA," *Just IT : Jurnal Sistem Informasi, Teknologi Informasi dan Komputer*, vol. 11, no. 3, pp. 67-71, 2021.
- [4] M. Cheng and D. Han, "A K-means Algorithm for Construction of Enterprise Innovation System Based on Data Mining Technology," in *2023 International Conference on Networking, Informatics and Computing (ICNETIC)*, Palermo, 2023.
- [5] I. Ayuningtias, N. N. Debatara dan N. Imro'ah, "Analisis Cluster Non-Hirarki Dengan Metode K-Modes," *Bimaster : Buletin Ilmiah Matematika, Statistika dan Terapannya*, vol. 8, no. 4, pp. 909-916, 2019.
- [6] Y. Kusnadi dan M. S. Putri, "Clustering Menggunakan Metode K-Means Untuk Menentukan Prioritas Penerima Bantuan Bedah Rumah (Studi Kasus : Desa Ciomas Bogor)," *Jurnal Teknologi Informatika dan Komputer*, vol. 7, no. 1, pp. 17-24, 2021.
- [7] Y. Filki, "Algoritma K-Means Clustering dalam Memprediksi Penerima Bantuan Langsung Tunai (BLT) Dana Desa," *Jurnal Informatika Ekonomi Bisnis*, vol. 4, no. 4, pp. 166-171, 2022.
- [8] E. Moruk, M. Martanto dan U. Hayati, "Algoritma K-Means Untuk Clustering Penerima Program Keluarga Harapan Di Nanaenoe Nusa Tenggara Timur," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 2, pp. 1681-1687, 2024.
- [9] B. Susarianto dan T. Nizami, "Penentuan Penerima Bantuan Sosial Beras untuk Masyarakat Miskin Menggunakan Metode K-Means," *Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, vol. 12, no. 2, pp. 772-783, 2023.
- [10] S. A. Mazhar, R. Anjum, A. I. Anwar dan A. A. Khan, "Methods of Data Collection: A Fundamental Tool of Research," *Journal of Integrated Community Health*, vol. 10, no. 1, pp. 6-10, 2021.
- [11] D. P. Turner, "Sampling Methods in Research Design," *Headache: The Journal of Head and Face Pain*, vol. 60, no. 1, pp. 8-12, 2020.
- [12] R. Muliono dan Z. Sembiring, "Data Mining Clustering Menggunakan Algoritma K-Means Untuk Klasterisasi Tingkat Tridarma Pengajaran Dosen," *CESS (Journal of Computer Engineering System and Science)*, vol. 4, no. 2, pp. 272-279, 2019.
- [13] M. F. Amin, "Confusion Matrix in Binary Classification Problems: A Step-by-Step Tutorial," *Journal of Engineering Research*, vol. 6, no. 5, pp. T1-T12, 2022.
- [14] M. F. Rahman, M. I. Darmawidjadja dan D. Alamsah, "Klasifikasi Untuk Diagnosa Diabetes Menggunakan Metode Bayesian Regularization Neural Network (RBNN)," *JURNAL INFORMATIKA*, vol. 11, no. 1, pp. 36-45, 2017.
- [15] N. Puspitasari, A. Septiarini dan A. R. Aliudin, "Metode K-Nearest Neighbor Dan Fitur Warna Untuk Klasifikasi Daun Sirih Berdasarkan Citra Digital," *PROSISKO: Jurnal Pengembangan Riset dan Observasi Sistem Komputer*, vol. 10, no. 2, pp. 165-172, 2023.