

Implementasi Data Mining Dalam Mengelompokkan Data penduduk Kurang Mampu Menggunakan Metode K-Means Clustering

Arisman Waruwu¹, Milfa Yetri², Feri Setiawan³

^{1,2,3} Sistem Informasi, STMIK Triguna Dharma

Email: ¹arismanwaruwu2001@gmail.com, ²milfa.anfa03@gmail.com, ³ferysetiawan13@gmail.com

Email Penulis Korespondensi: arismanwaruwu2001@email.com

Abstrak

Penduduk kurang mampu merupakan permasalahan bangsa yang sangat mendesak dan sering kali di tandai dengan tingginya tingkat pengangguran dan keterbelakangan. Salah satu aspek penting untuk mendukung langkah – langkah penanggulangan dalam penanganan penduduk kurang mampu, pemerintah dapat juga melakukan evaluasi kebijakannya terhadap penduduk serta membandingkan data penduduk kurang mampu antar waktu dan daerah serta menentukan targetnya agar dapat bisa menyelesaikan masalah yang di alami penduduk. Data mining merupakan istilah dari pattern recognition merupakan metode yang di gunakan untuk pengelolaan data yang akan di gunakan, untuk menemukan pola yang tersembunyi dari pengelolaan data. Algoritma K-Means merupakan salah satu metode yang menggunakan konsep descriptive dan dapat digunakan untuk menerapkan suatu algoritma dalam penentuan suatu objek kedalam cluster tertentu. Hasil penelitian ini bermanfaat bagi Desa Hiliwaele II dan dapat dikategorikan sebagai suatu proses pengelompokkan record, observasi, atau mengelompokkan data yang memiliki kesamaan objek, dengan klasifikasi tidak ada variabel target dalam melakukan pengelompokan pada proses clustering.

Kata Kunci: Penduduk, K-Means Clustering, Data Mining, Desa Hiliwaele II Kurang Mampu

Abstract

The underprivileged population is a very urgent national problem and is often marked by high levels of unemployment and underdevelopment. One of the important aspects to support countermeasures in dealing with underprivileged people, the government can also evaluate its policies on the population and compare data on the underprivileged population between time and regions and determine targets so that it can solve the problems experienced by the population. Data mining is a term for pattern recognition, which is a method used to manage data that will be used to find hidden patterns in data management. The K-Means algorithm is a method that uses a descriptive concept and can be used to apply an algorithm in determining an object into a particular cluster. The results of this study are useful for Hiliwaele II Village and can be categorized as a process of grouping records, observations, or grouping data that have the same object, with no target variable classification in grouping in the clustering process.

Keywords: Population, K-Means Clustering, Data Mining, Disadvantaged Hiliwaele II Village

1. PENDAHULUAN

Penduduk kurang mampu merupakan permasalahan bangsa yang sangat mendesak dan sering kali di tandai dengan tingginya tingkat pengangguran dan keterbelakangan. Salah satu aspek penting untuk mendukung langkah – langkah penanggulangan dalam penanganan penduduk kurang mampu, pemerintah dapat juga melakukan evaluasi kebijakannya terhadap penduduk serta membandingkan data penduduk kurang mampu antar waktu dan daerah serta menentukan targetnya agar dapat bisa menyelesaikan masalah yang di alami penduduk dengan tujuan memperbaiki kondisi ekonomi serta melakukan pendataan kepada penduduk kurang mampu. Pemerintahan Desa Hiliwaele-II memiliki permasalahan salah satunya, yaitu penduduk yang kurang mampu selama masa pandemi Covid 19 pada tahun 2021. Didesa Hiliwaele-II rata-rata mengalami ketidak stabilan pendapatan, berdasarkan dari data BPS dan hasil survei pemerintahan Desa setempat. Dengan Teknik data mining ini akan menghasilkan suatu pengetahuan baru yang bersumber dari hasil pengelolaan data tersebut. Data mining juga di artikan sebagai rangkaian untuk menemukan pola yang terbaik dari data jumlah besar kemudian data-data tersebut dapat di simpan di dalam database, data warehouse atau penyimpanan informasi lain nya[2]. Data mining sangat perlu untuk di lakukan di dalam pengelolaan data yang berjumlah besar untuk bisa memudahkan aktivitas recording suatu transaksi dan proses data agar dapat memudahkan memberikan informasi yang akurat bagi penggunaanya, alasan mengapa data mining sangat menarik perhatian industry informasi di beberapa tahun akhir ini, yaitu karena tersedianya data yang jumlahnya sangat besar karna dapat mengubah data tersebut menjadi informasi dan pengetahuan dari informasi yang berukuran besar yang nantinya sangat berguna untuk pengembangan informasi di masa depan.

Algoritma *K-Means* merupakan salah satu metode yang menggunakan konsep *descriptive* dan dapat digunakan untuk menerapkan suatu algoritma dalam penentuan suatu objek kedalam *cluster* tertentu. *K-Means* pertama kali diperkenalkan oleh MacQueen JB pada tahun 1976. Pada algoritma *K-Means* jumlah *cluster* telah ditentukan terlebih dahulu. *K-Means* mampu mengelompokkan objek besar dan pencilan objek dengan sangat cepat sehingga dapat mempercepat proses pengelolaan data[3]. *Clustering* merupakan suatu proses pengelompokkan record, observasi, atau mengelompokkan data yang memiliki kesamaan objek, dengan klasifikasi tidak ada variabel target dalam melakukan pengelompokan pada proses *clustering*.



Algoritma K-Means banyak digunakan dalam klasifikasi data karena algoritmanya yang sederhana dan klasifikasinya yang cepat, Algoritma K-Means ini memiliki tingkat ketelitian yang cukup tinggi terhadap klasifikasi objek, sehingga tidak mengherankan algoritma ini populer digunakan untuk pengklasifikasi objek dalam jumlah besar. Jika kita bandingkan dengan metode/Algoritma lain seperti Algoritma C4.5 yang dimana hanya mengembangkan dan membangun pohon keputusan sebagai atribut[4], Perbedaan lainnya, penelitian terdahulu ini tidak membangun sistem cerdas sebagaimana pada yang dilakukan pada penelitian di artikel ini.

Tujuan penelitian berdasarkan penjelasan dari batasan masalah yang telah di paparkan yaitu: Untuk menganalisa dalam mengelompokkan data penduduk kurang mampu berdasarkan data yang telah di peroleh, Untuk menerapkan algoritma *K-means* dalam mengelompokkan data penduduk kurang mampu di desa Hiliwaele II, Untuk merancang dan membangun sistem dengan menggunakan metode *K-Means* dalam mengelompokkan penduduk kurang mampu di desa Hiliwaele II.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini dapat dilakukan dengan beberapa metode penelitian yaitu sebagai berikut :

1. Data *Collecting* (Teknik Pengumpulan Data)
 - a. Pengamatan (*Observasi*)
 - b. Wawancara (*Interview*)
 - c. Studi Literatur

2.2 Penduduk Kurang Mampu

Secara garis besar kurang mampu memiliki banyak pengertian, konsep dan sudut pandang yang berbeda-beda yaitu di antaranya, bersifat alami, struktural, kultural, absolute ataupun relative. Berbagai macam pendekatan yang dapat digunakan untuk mengukur penduduk kurang mampu diantaranya Data Dasar Rumah Tangga (DDRT) dan Survey Rumah Tangga (SRT) [5]. Ketidakmampuan tidak hanya dihadapi oleh pemerintah pusat akan tetapi juga menjadi permasalahan yang serius yang di hadapi oleh pemerintah desa. Persoalan ini berkaitan erat dengan pendapatan penduduk yang rendah yang menyebabkan banyak penduduk kurang mampu sehingga banyak penduduk kesulitan dalam mencukupi kebutuhan kesehariannya. Melakukan sebuah upaya atau Tindakan dengan memberikan Program keluarga harapan (PKH), Bantuan langsung tunai (BLT), dan program-program lainnya [6].

2.3 Data Mining

Data adalah kumpulan informasi yang digunakan dalam proses pengambilan kesimpulan dengan menggunakan teknik statistik, matematika, dan kecerdasan buatan [7]. Data mining merupakan bidang dari beberapa keilmuan yang menyatukan Teknik statistika dan matematika sehingga menciptakan informasi yang penting dari data yang besar sehingga menghasilkan informasi dan pengetahuan sebagai dasar pengambilan keputusan[8]. Beberapa tahun terakhir, data semakin heterogen dan kompleks dengan volume yang meningkat cepat secara eksponensial. Oleh karena itu, saat ini di kenal istilah *big data*, yang menggambarkan volume data yang sangat besar, terstruktur maupun tidak terstruktur, yang membajiri dunia bisnis[9]. Data mining merupakan teknik memanfaatkan data dalam jumlah yang besar untuk memperoleh informasi berharga yang sebelumnya tidak diketahui dan dapat dimanfaatkan untuk pengambilan keputusan penting. Oleh karena itu dalam menggali informasi, *Data mining* menggunakan beberapa teknik antara lain[10].

- a. *Association Discovery*
- b. *Clustering*
- c. *Sequential Discovery*
- d. *Clasification*
- e. *Neural Network*

2.4 K-Means Clustering

K-Means Clustering merupakan salah satu metode yang menjalankan satu atau lebih kelompok objek, untuk menganalisis dan mengekstraksi pengetahuan secara otomatis. Prediksi, perbandingan, klasifikasi, clustering, dan perkiraan adalah jenis pendekatan dalam data mining. Pendekatan algoritma k-means merupakan metode yang mengelompokkan objek-objek berdasarkan kesamaan karakteristik yang sama dengan jumlah kluster ditetapkan terlebih dahulu[11]. Pada dasarnya *clustering* merupakan suatu metode yang digunakan untuk mengelompokkan rangkaian data menjadi sebuah informasi yang memiliki karakteristik (*similarity*) antara satu dengan yang lain. Pengelompokan data dimulai dengan menentukan objek-objek yang memiliki kesamaan yang mendekati, kemudian tahapan selanjutnya melakukan pemrosesan terhadap objek yang lainnya yang memiliki kedekatan kedua. Sehingga *cluster* akan membentuk semacam pohon dimana ada hierarki (tingkatan) dari yang memiliki kemiripan hingga paling tidak mirip. Dendogram biasanya digunakan untuk membantu dalam melakukan proses hierarki tersebut[12]. Metode ini adalah metode pengelompokan data yang sederhana dan banyak digunakan dalam aplikasi data mining. Setiap *cluster* direpresentasikan oleh sebuah titik pusat (*centroid*).

Setiap data akan di kelompokkan pada *cluster* dengan titik pusat yang terdekat dari data tersebut [13].

Adapun Langkah-langkah untuk melakukan *clustering* dengan metode K-Means adalah sebagai berikut :

1. Pilih jumlah *cluster* (k) dari dataset yang ada
2. Menentukan k sebagai *Centroid*, biasanya di lakukan secara acak (*random*)
3. Hitung jarak setiap data ke masing-masing centroid menggunakan persamaan Euclidean Distance, sebagai berikut :

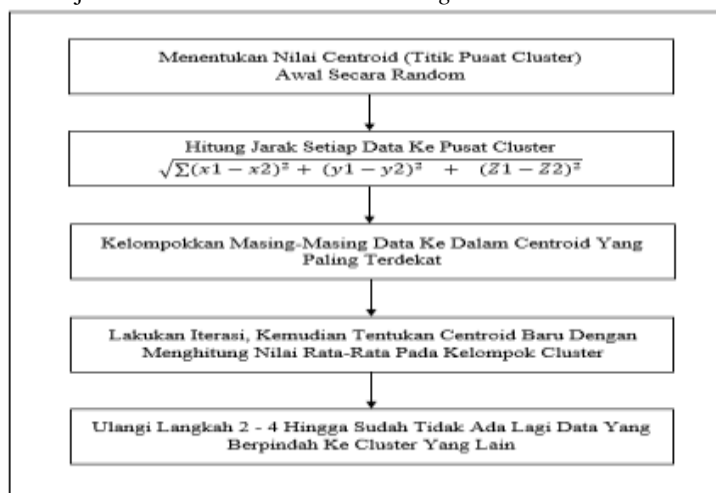
$$d(P + Q) = \sum_{j=1}^p (X_j(P) - x_j(Q))^2$$

4. Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan centroidnya.
5. Tentukan posisi centroid baru (k)
6. kembali ke langkah 3 jika posisi centroid baru dengan centroid lama tidak sama.

3. HASIL DAN PEMBAHASAN

3.1 Penerapan Metode K-Means Clustering

Berikut ini adalah Langkah kerja dari metode *K-Means Clustering* :



Gambar 1 Kerangka Kerja K-Means Clustering

Berikut adalah data yang digunakan sebagai sampel dalam penelitian ini

Tabel 1. Data Penduduk Kurang Mampu

NO	NAMA PENDUDUK	JENIS KELAMIN	RATA-RATA PENDAPATAN SETIAP BULAN	STATUS KEPEMILIKKAN RUMAH	JUMLAH TANGGUNGAN
1	YA'ARO WARUWU	LAKI-LAKI	700000	10	7
2	OTONI HALAWA	LAKI-LAKI	700000	10	5
3	YANUARI HALAWA	LAKI-LAKI	600000	6	5
4	MARLINUS WARUWU	LAKI-LAKI	658333	8	4
5	FONASOKHI HALAWA	LAKI-LAKI	641667	10	5
6	TEMASOKHI WARUWU	LAKI-LAKI	691667	8	4
7	NOVERMAN WARUWU	LAKI-LAKI	691667	6	2
8	AMIRUDI WARUWU	LAKI-LAKI	666667	6	4
9	NURMASA WARUWU	PEREMPUAN	700000	10	3
10	ERINA ZAMASI	PEREMPUAN	641667	6	6
...
34	TOLOSOKHI WARUWU	LAKI-LAKI	733333	10	3

Tabel 2. Tabel Penilaian

Alt	NAMA	V1	V2	V3
P1	YA'ARO WARUWU	700000	10	7
P2	OTONI HALAWA	700000	10	5
P3	YANUARI HALAWA	600000	6	5
P4	MARLINUS WARUWU	658333	8	4
P5	FONASOKHI HALAWA	641667	10	5
P6	TEMASOKHI WARUWU	691667	8	4
P7	NOVERMAN WARUWU	691667	6	2
P8	AMIRUDI WARUWU	666667	6	4
P9	NURMASA WARUWU	700000	10	3
P10	ERINA ZAMASI	641667	6	6
...
P34	TOLOSOKHI WARUWU	733333	10	3

Tabel 3. Atribut

Atribut	Variabel
Rata-Rata Pendapatan Setiap Bulan	V1
Status Kepemilikan Rumah	V2
Jumlah Tanggungan	V3

Tabel 4. Nilai Bobot

Atribut	Nilai
Sangat Baik	10
Baik	8
Cukup	6

Tabel 5. Cluster

Cluster	Keterangan
1	Normal
2	Rendah

Penetapan jumlah *cluster* yaitu 2 *cluster*. Setelah menetapkan jumlah *cluster*, tentukan titik pusat awal *cluster* (*centroid*). Berikut ini titik *centroid* yang telah dipilih :

Tabel 6. Centroid Awal Cluster

No	Centroid	Alt	M1	M2	M3
1	M1	P1	700000	10	7
2	M2	P17	625000	8	2

Menghitung jarak setiap variabel pada sampel data terhadap pusat *Cluster* yang sudah di tentukan.

a. Dengan *Centroid* M1 (700000,10,7)

1. Jarak antara P1 dengan titik M1

$$\begin{aligned}
 &= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (Z_1 - Z_2)^2} \\
 &= \sqrt{(700000 - 700000)^2 + (10 - 10)^2 + (7 - 7)^2} \\
 &= 0
 \end{aligned}$$

2. Jarak antara P2 dengan titik M1

$$\begin{aligned}
 &= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (Z_1 - Z_2)^2} \\
 &= \sqrt{(700000 - 700000)^2 + (10 - 10)^2 + (5 - 7)^2} \\
 &= 2
 \end{aligned}$$

3. Jarak antara P3 dengan titik M1

$$\begin{aligned}
 &= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (Z_1 - Z_2)^2} \\
 &= \sqrt{(600000 - 700000)^2 + (6 - 10)^2 + (5 - 7)^2} \\
 &= 100000
 \end{aligned}$$

4. Jarak antara P4 dengan titik M1

$$= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (Z_1 - Z_2)^2}$$

$$= \sqrt{(658333 - 700000)^2 + (8 - 10)^2 + (4 - 7)^2}$$

$$= 41667$$
5. Jarak antara P5 dengan titik M1

$$= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (Z_1 - Z_2)^2}$$

$$= \sqrt{(641667 - 700000)^2 + (10 - 10)^2 + (5 - 7)^2}$$

$$= 58333$$

Lanjutkan perhitungan sampai variabel P34

- b. Dengan *Centroid* M2 (625000, 8, 2)
 1. Jarak P1 dengan M2

$$= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (Z_1 - Z_2)^2}$$

$$= \sqrt{(700000 - 625000)^2 + (10 - 8)^2 + (7 - 2)^2}$$

$$= 75000$$
 2. Jarak antara P2 dengan titik M2

$$= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (Z_1 - Z_2)^2}$$

$$= \sqrt{(700000 - 625000)^2 + (10 - 8)^2 + (5 - 2)^2}$$

$$= 75000$$
 3. Jarak antara P3 dengan titik M2

$$= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (Z_1 - Z_2)^2}$$

$$= \sqrt{(600000 - 625000)^2 + (6 - 8)^2 + (5 - 2)^2}$$

$$= 25000$$
 4. Jarak antara P4 dengan titik M2

$$= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (Z_1 - Z_2)^2}$$

$$= \sqrt{(658333 - 625000)^2 + (8 - 8)^2 + (4 - 2)^2}$$

$$= 33333$$
 5. Jarak antara P5 dengan titik M2

$$= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (Z_1 - Z_2)^2}$$

$$= \sqrt{(641667 - 625000)^2 + (10 - 8)^2 + (5 - 2)^2}$$

$$= 16667$$

Lanjutkan perhitungan sampai variabel P34

Berdasarkan proses perhitungan di atas maka di peroleh hasil jarak setiap variabel dengan *centroid* awal M1 dan M2 sebagai berikut.

Tabel 7. Hasil Perhitungan Iterasi Ke 1

No	Alt	Jarak ke titik M1	Jarak ke titik M2	Kedekatan	Keterangan
1	P1	0	75000	0	Cluster 1
2	P2	2	75000	2	Cluster 1
3	P3	100000	25000	25000	Cluster 2
4	P4	41667	33333	33333	Cluster 2
5	P5	58333	16667	16667	Cluster 2
6	P6	8333	66667	8333	Cluster 1
7	P7	8333	66667	8333	Cluster 1
8	P8	33333	41667	33333	Cluster 1
9	P9	4	75000	4	Cluster 1
10	P10	58333	16667	16667	Cluster 2
..
34	P34	33333	108333	33333	Cluster 1

Setelah mengetahui jarak terdekat setiap *cluster*. Urutkan sesuai *cluster* nya masing-masing.

Tabel 8. Hasil Perhitungan Sesuai Cluster

No	Alt	Jarak ke titik M1	Jarak ke titik M2	Kedekatan	Keterangan
1	P1	0	75000	0	Cluster 1
2	P2	2	75000	2	Cluster 1
3	P6	8333	66667	8333	Cluster 1
4	P7	8333	66667	8333	Cluster 1
5	P8	33333	41667	33333	Cluster 1
6	P9	4	75000	4	Cluster 1
7	P13	16667	58333	16667	Cluster 1
8	P14	33333	41667	33333	Cluster 1
9	P19	25000	100000	25000	Cluster 1
10	P20	16667	91667	16667	Cluster 1
...
34	P34	75000	4	4	Cluster 2

Setelah mengetahui hasil perhitungan sesuai cluster. Maka data di urutkan sesuai cluster nya masing-masing.

Tabel 9. Data Diurutkan Sesuai Cluster

No	Alt	Rata-Rata Pendapatan Selama Satu Tahun	Status Kepemilikan Rumah	Jumlah Tanggungan	Cluster
1	P1	700.000,00	10	7	Cluster 1
2	P2	700.000,00	10	5	Cluster 1
3	P6	691.667,00	8	4	Cluster 1
4	P7	691.667,00	6	2	Cluster 1
5	P8	666.667,00	6	4	Cluster 1
6	P9	700.000,00	10	3	Cluster 1
7	P13	683.333,00	6	1	Cluster 1
8	P14	666.667,00	8	4	Cluster 1
9	P19	725.000,00	10	5	Cluster 1
10	P20	716.667,00	10	3	Cluster 1
...
34	P31	625.000,00	8	6	Cluster 2

Berdasarkan hasil perhitungan iterasi ke 1 jarak variabel pada *centroid* awal maka masing-masing cluster memiliki anggota sebagai berikut:

1. C1 memiliki jumlah anggota sebanyak 17 orang
2. C2 memiliki jumlah anggota sebanyak 17 orang

Lanjutkan iterasi sampai posisi anggota cluster sudah tidak ada yang berubah lagi.

Berdasarkan hasil perhitungan iterasi ke 2 jarak variabel pada *centroid* awal maka masing-masing cluster memiliki anggota sebagai berikut,

1. C1 memiliki jumlah anggota sebanyak 15 orang
2. C2 memiliki jumlah anggota sebanyak 19 orang

Terjadi perpindahan beberapa anggota C1 dan C2

Lanjutkan iterasi sampai posisi anggota cluster sudah tidak ada yang berubah lagi.

Setelah mengetahui anggota tiap-tiap cluster kemudian pusat cluster baru di hitung untuk melakukan iterasi -2 berdasarkan data anggota tiap-tiap cluster sesuai dengan rumus pusat anggota cluster yaitu mencari nilai rata-rata dari setiap cluster. Sehingga di dapat perhitungan sebagai berikut:

- a. Titik Pusat awal (M1) yaitu :
= rata-rata (M1)
= (700000, 9, 3)
- b. Titik Pusat awal (M2) yaitu :
= rata-rata (M2)
= (640790, 8, 4)

Tabel 10. *Centroid* Iterasi Ke 3

No	<i>Centroid</i>	K1	K2	K3
1	M1	700000	8	3
2	M2	640790	7	4

Pada tahap ini dilakukan perhitungan jarak pada setiap variabel pada data terhadap pusat *cluster* yang baru.

a. Dengan *centroid* M1 (700000, 9, 3)

1. Jarak antara P1 dengan M1

$$\begin{aligned}
 &= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (Z_1 - Z_2)^2} \\
 &= \sqrt{(700000 - 700000)^2 + (10 - 8)^2 + (7 - 3)^2} \\
 &= 4
 \end{aligned}$$

2. Jarak antara P2 dengan M1

$$\begin{aligned}
 &= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (Z_1 - Z_2)^2} \\
 &= \sqrt{(700000 - 700000)^2 + (10 - 8)^2 + (5 - 3)^2} \\
 &= 2
 \end{aligned}$$

3. Jarak antara P3 dengan M1

$$\begin{aligned}
 &= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (Z_1 - Z_2)^2} \\
 &= \sqrt{(691667 - 700000)^2 + (8 - 8)^2 + (4 - 3)^2} \\
 &= 8333
 \end{aligned}$$

4. Jarak antara P4 dengan M1

$$\begin{aligned}
 &= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (Z_1 - Z_2)^2} \\
 &= \sqrt{(691667 - 700000)^2 + (6 - 8)^2 + (2 - 3)^2} \\
 &= 8333
 \end{aligned}$$

5. Jarak antara P5 dengan M1

$$\begin{aligned}
 &= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (Z_1 - Z_2)^2} \\
 &= \sqrt{(700000 - 700000)^2 + (10 - 8)^2 + (3 - 3)^2} \\
 &= 1
 \end{aligned}$$

Lanjutkan perhitungan sampai variabel P34

b. Dengan *centroid* M2 (640790, 8, 4)

1. Jarak antara P1 dengan M2

$$\begin{aligned}
 &= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (Z_1 - Z_2)^2} \\
 &= \sqrt{(700000 - 640790)^2 + (10 - 7)^2 + (7 - 4)^2} \\
 &= 59210
 \end{aligned}$$

2. Jarak antara P2 dengan M2

$$\begin{aligned}
 &= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (Z_1 - Z_2)^2} \\
 &= \sqrt{(700000 - 640790)^2 + (10 - 7)^2 + (5 - 4)^2} \\
 &= 59210
 \end{aligned}$$

3. Jarak antara P3 dengan M2

$$\begin{aligned}
 &= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (Z_1 - Z_2)^2} \\
 &= \sqrt{(691667 - 640790)^2 + (8 - 7)^2 + (4 - 4)^2} \\
 &= 50877
 \end{aligned}$$

4. Jarak antara P4 dengan M2

$$\begin{aligned}
 &= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (Z_1 - Z_2)^2} \\
 &= \sqrt{(691667 - 640790)^2 + (6 - 7)^2 + (2 - 4)^2} \\
 &= 50877
 \end{aligned}$$

5. Jarak antara P5 dengan M2

$$\begin{aligned}
 &= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (Z_1 - Z_2)^2} \\
 &= \sqrt{(700000 - 640790)^2 + (10 - 7)^2 + (3 - 4)^2} \\
 &= 59210
 \end{aligned}$$

Lanjutkan perhitungan sampai variabel P34

Berdasarkan perhitungan di atas maka di peroleh hasil jarak setiap variabel dengan *centroid* baru M1 dan M2 sebagai berikut.

Tabel 11. Perhitungan Hasil Iterasi Ke 3

No	Alt	Jarak ke titik M1	Jarak ke titik M2	Kedekatan	Keterangan
1	P1	4	59,21	4	Cluster 1
2	P2	2	59,21	2	Cluster 1
3	P6	8,333	50,877	8,333	Cluster 1
4	P7	8,333	50,877	8,333	Cluster 1
5	P9	1	59,21	1	Cluster 1
6	P13	16,667	42,543	16,667	Cluster 1
7	P19	25	84,21	25	Cluster 1
8	P20	16,667	75,877	16,667	Cluster 1
9	P21	8,333	67,543	8,333	Cluster 1
10	P22	16,667	42,543	16,667	Cluster 1
...
34	P31	75	15,79	15,79	Cluster 2

Setelah mengetahui jarak terdekat setiap *cluster*. Urutkan sesuai *cluster* nya masing-masing.

Tabel 12. Hasil Perhitungan Sesuai *Cluster*

No	Alt	Jarak ke titik M1	Jarak ke titik M2	Kedekatan	Keterangan
1	P1	4	59,21	4	Cluster 1
2	P2	2	59,21	2	Cluster 1
3	P6	8,333	50,877	8,333	Cluster 1
4	P7	8,333	50,877	8,333	Cluster 1
5	P9	1	59,21	1	Cluster 1
6	P13	16,667	42,543	16,667	Cluster 1
7	P19	25	84,21	25	Cluster 1
8	P20	16,667	75,877	16,667	Cluster 1
9	P21	8,333	67,543	8,333	Cluster 1
10	P22	16,667	42,543	16,667	Cluster 1
...
34	P31	75	15,79	15,79	Cluster 2

Setelah mengetahui hasil perhitungan sesuai *cluster*. Maka data di urutkan sesuai *cluster* nya masing-masing.

Tabel 13. Data Diurutkan Sesuai *Cluster*

No	Alt	Rata-Rata Pendapatan Selama Satu Tahun	Status Kepemilikan Rumah	Jumlah Tanggungan	Cluster
1	P1	700.000,00	10	7	Cluster 1
2	P2	700.000,00	10	5	Cluster 1
3	P6	691.667,00	8	4	Cluster 1
4	P7	691.667,00	6	2	Cluster 1
5	P9	700.000,00	10	3	Cluster 1
6	P13	683.333,00	6	1	Cluster 1
7	P19	725.000,00	10	5	Cluster 1
8	P20	716.667,00	10	3	Cluster 1
9	P21	708.333,00	10	5	Cluster 1
10	P22	683.333,00	8	4	Cluster 1
...
34	P31	625.000,00	8	6	Cluster 2

Berdasarkan hasil perhitungan iterasi ke 3 jarak variabel pada *centroid* awal maka masing-masing *cluster* memiliki anggota sebagai berikut,

1. C1 memiliki jumlah anggota sebanyak 15 orang

2. C2 memiliki jumlah anggota sebanyak 19 orang

Berdasarkan pengelompokan di atas, tidak ada terjadi perpindahan anggota antar *cluster*. Maka iterasi di hentikan.

Dari hasil *clusterisasi* pada data di hasilkan pengelompokan data penduduk kurang mampu dengan dua *cluster* yaitu *cluster* normal dan rendah, dengan algoritma K-Means *Clustering* sebanyak 3 iterasi. Maka di dapat hasil pada *cluster* pertama sebanyak 15 orang berpenghasilan normal dan pada *cluster* ke dua sebanyak 19 orang berpenghasilan rendah.

Tabel 14. Hasil Perhitungan

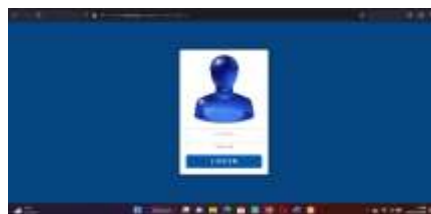
NO	NAMA PENDUDUK	Hasil			Cluster
		M1	M2	Kedekatan	
1	YAARO WARUWU	4	59,21	4	Cluster 1
2	OTONI HALAWA	2	59,21	2	Cluster 1
3	TEMASOKHI WARUWU	8,333	50,877	8,333	Cluster 1
4	NOVERMAN WARUWU	8,333	50,877	8,333	Cluster 1
5	NURMASA WARUWU	1	59,21	1	Cluster 1
6	TEHELI WARUWU	16,667	42,543	16,667	Cluster 1
7	SEMI WARUWU	25	84,21	25	Cluster 1
8	HEZATULO WARUWU	16,667	75,877	16,667	Cluster 1
9	RAMILIA WARUWU	8,333	67,543	8,333	Cluster 1
10	YAREDI WARUWU	16,667	42,543	16,667	Cluster 1
...
34	FATOUOSA WARUWU	75	15,79	15,79	Cluster 2

3.2 Implementasi Sistem

Berikut ini adalah tahapan dari pengaplikasian Data Mining dengan menggunakan metode *K-Means* dalam mengelompokkan data penduduk kurang mampu di Desa Hiliwaele II :

1. *Form* login

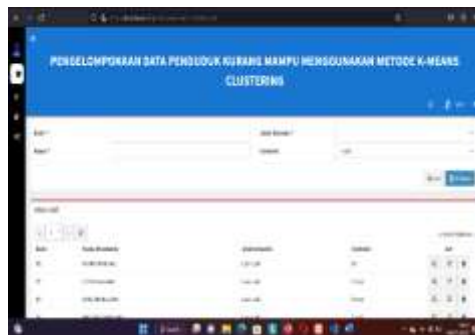
Form login merupakan tampilan yang pertama muncul Ketika sistem diakses. Sebelum masuk dan mengakses aplikasi, *user* harus melakukan *login* dengan cara meng-*input* *username* dan *password* dengan benar sesuai dengan sistem *database*. Jika *username* dan *password* benar, *user* akan masuk ke halaman menu utama. Berikut ini adalah tampilan *form* login :



Gambar 2. Tampilan *Form* Login

2. *Form* Alternatif

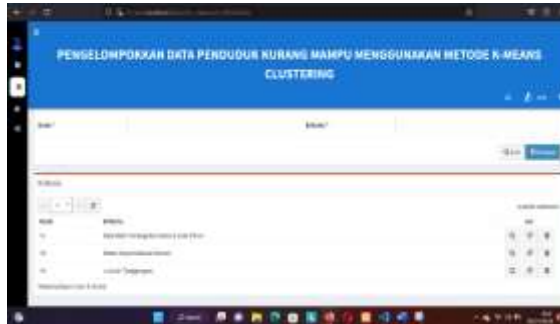
Menu alternatif merupakan tampilan awal setelah berhasil login ke sistem. Pada menu alternatif menu yang digunakan untuk mengelola pengimputan data penduduk kurang mampu, Data yang dimasukkan ke dalam sistem meliputi kode, nama, jenis kelamin, dan centroid. Berikut ini adalah tampilan *form* menu utama :



Gambar 3. Tampilan *Form* Menu Alternatif

3. *Form* menu kriteria

Form menu kriteria ini merupakan *form* yang digunakan untuk mengelola dan mengimput kriteria-kriteria yang akan di gunakan, Data yang dimasukkan ke dalam sistem meliputi kode dan kriteria. Berikut merupakan tampilan *form* menu kriteria :



Gambar 4. Tampilan *Form* Menu Kriteria

3. *Form* Menu Data Nilai

Form menu data nilai merupakan hasil Penilaian Pada semua variabel yang digunakan. Data yang di masukkan ke dalam sistem meliputi, nama, kriteria dan nilai. Berikut merupakan tampilan *form* menu data nilai :



Gambar 5. Tampilan *Form* Menu Data Nilai

4. *Form* Menu Proses Analisa

Form proses analisa merupakan dimulainya proses perhitungan dan muncul iterasi. Berikut tampilan *form* menu proses analisa.



Gambar 6. Tampilan *Form* Menu Proses Analisa

5. *Form* Hasil Perhitungan

Form ini berfungsi untuk menampilkan hasil dari pengelompokan data penduduk kurang mampu yang telah di hitung melalui *form* proses perhitungan. Adapun tampilan dari *form* hasil perhitungan sebagai berikut



Gambar 7. Tampilan *Form* Hasil Perhitungan

3. KESIMPULAN

Kesimpulan merupakan jawaban dari rumusan masalah yang menggambarkan hasil dari penelitian yang dilakukan. Maka dibutuhkan sebuah sistem yang mampu mengelompokkan data penduduk kurang mampu, maka dari itu di rancanglah sebuah sistem yang mampu menerapkan pemodelan algoritma *K-Means Clustering* sehingga dapat membantu pihak pemerintahan desa dalam mengelompokkan data penduduk kurang mampu di desanya. Aplikasi yang di bangun berbasis Website telah mampu mengimplementasikan Data Mining dalam melakukan pengelompokkan data penduduk kurang mampu di Desa Hiliwaele II.

UCAPAN TERIMAKASIH

Terimakasih di ucapkan kepada Tuhan yang MahaEsa yang telah memberikan Rahmat dan karunia sehingga mampu menyelesaikan jurnal ini. Kemudian kepada Ibu Milfa Yetri, dan Bapak Feri Setiawan atas arahan dan bimbingannya selama proses pengerjaan Skripsi hingga sampai ke penyusunan jurnal ini dan seluruh jajaran Manajemen, Dosen serta pegawai kampus STMIK Triguna Dharma yang telah banyak membantu baik dari segi informasi ataupun dukungan lainnya.

DAFTAR PUSTAKA

- [1] F. Andrianti and R. Firmansyah, "Penerapan Clustering Data Kurang Mampu Di Desa Situmekar Menggunakan Algoritma K-Means," vol. 1, no. 1, p. 88, 2020, [Online]. Available: <http://eprosiding.ars.ac.id/index.php/pti>
- [2] I. Nasution, A. P. Windarto, and M. Fauzan, "Penerapan Algoritma K-Means Dalam Pengelompokan Data Penduduk Miskin Menurut Provinsi," *Build. Informatics, Technol. Sci.*, vol. 2, no. 2, pp. 76–83, 2020, doi: 10.47065/bits.v2i2.492.
- [3] J. Arianto, "Penerapan Data Mining Untuk Pengelompokan Penduduk Kurang Mampu Desa Sambirejo Timur Dengan Algoritma K-Medoids (Studi Kasus Kantor Kepala Desa Sambirejo Timur)," vol. 3, pp. 569–573, 2019, doi: 10.30865/komik.v3i1.1660.
- [4] R. J. HOSTING, "Data Mining: Pengertian, Fungsi, Metode & Penerapannya," *jagoan hosting*, 2022. <https://www.jagoanhosting.com/blog/>
- [5] C. J. Silalahi, A. Situmorang, and J. F. Naibaho, "Implementasi Metode K-Means Clustering Untuk Memetakan Daerah Potensial Penghasil Padi di Provinsi Sumatera Utara," vol. 2, pp. 49–57, 2022.
- [6] W. Widayani and H. Harliana, "Perbandingan Algoritma K-Means dan SFCM Pada Pengelompokkan Rumah Tangga Miskin," *J. Sains dan Inform.*, vol. 6, no. 1, pp. 1–9, 2020, doi: 10.34128/jsi.v6i1.200.
- [7] J. Hutagalung and F. Sonata, "Penerapan Metode K-Means Untuk Menganalisis Minat Nasabah Asuransi," *J. MEDIA Inform. BUDIDARMA*, vol. 5, no. 3, pp. 1187–1194, 2021, doi: 10.30865/mib.v5i3.3113.
- [8] B. Fibriyanti, "Penerapan Algoritma K-Means Untuk Menentukan Prioritas Penerimaan Bantuan PKH," vol. 2, no. 12, pp. 1–19, 2022.
- [9] N. Sepriyanti, R. S. Nahampun, and M. H. Zikri, "Implementation of K-Means Clustering to Group Poverty Levels in Riau Province Penerapan K-Means Clustering Untuk Mengelompokkan Tingkat Kemiskinan di Provinsi Riau," pp. 59–65, 2022.
- [10] S. N. Br Sembiring, H. Winata, and S. Kusnasari, "Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means," *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 1, no. 1, p. 31, 2022, doi: 10.53513/jursi.v1i1.4784.
- [11] C. Satria and A. Anggrawan, "Aplikasi K-Means berbasis Web untuk Klasifikasi Kelas Unggulan," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 1, pp. 111–124, 2021, doi: 10.30812/matrik.v21i1.1473.
- [12] T. Noviana, J. Jasmir, and Y. Novianto, "Penerapan Data Mining Menentukan Kelompok Prioritas Penerima Bantuan Beras Rastra Dengan Clustering K-Means," *Progr. Stud. Tek. Inform. Stikom Din. Bangsa*, pp. 159–174, 2019.
- [13] J. Hutagalung, N. L. W. S. R. Ginantra, G. W. Bhawika, W. G. S. Parwita, A. Wanto, and P. D. Panjaitan, "COVID-19 Cases and Deaths in Southeast Asia Clustering using K-Means Algorithm," *J. Phys. Conf. Ser.*, vol. 1783, no. 1, pp. 1–6, 2021, doi: 10.1088/1742-6596/1783/1/012027.