

IMPLEMENTASI K-MEANS CLUSTERING PADA DATA PENDUDUK MISKIN LINGKUNGAN II DI KELURAHAN MANEMBO-NEMBO

Maxsi Bambang¹, Lestari Margatama²

¹Fakultas Teknologi Informasi, Sistem Informasi, Universitas Budi Luhur, Jakarta Selatan, Indonesia

Email: ¹1712500618@student.budiluhur.ac.id

²*lestari.margatama@budiluhur.ac.id

Abstrak-Penyaluran bantuan sosial di lingkungan II Kelurahan Manembo-Nembo masih dianggap kurang tepat sasaran. Bantuan sosial dari pemerintah seharusnya didapatkan oleh seluruh keluarga miskin yang ada. Kurangnya informasi mengenai jumlah dan daftar penduduk miskin adalah kendalanya. Oleh karena itu, perlu dilakukan pengelompokan penduduk miskin (keluarga miskin) berdasarkan dengan data-data yang dimiliki oleh Kelurahan Manembo-Nembo. Pada penelitian ini data penduduk dikategorikan menggunakan metode *Clustering* algoritma K-Means berdasarkan data penghasilan, Pendidikan terakhir, bantuan sosial dan data kondisi rumah yang diterima. Menggunakan software *RapidMiner Studio*, didapatkan hasil 3 cluster. Menggunakan metode elbow untuk mendapatkan jumlah cluster optimal dan menghitung *Sum of Square Error* dari setiap nilai cluster. Hasil penelitian dengan *RapidMiner* didapatkan hasil, untuk cluster 2 dengan kategori penduduk miskin berjumlah 119 kepala keluarga. Cluster 0 dengan kategori penduduk menengah berjumlah 78 kepala keluarga. Cluster 1 dengan kategori penduduk kaya berjumlah 75 kepala keluarga.

Kata Kunci: *Clustering, CRISP-DM, K-Means, Metode Elbow, Sum of Square Error.*

IMPLEMENTATION OF K-MEANS CLUSTERING ON DATA OF POOR RESIDENTS OF NEIGHBORHOOD II IN MANEMBO-NEMBO VILLAGE

Abstract- *The distribution of social assistance in neighborhood II of Manembo-Nembo Village is still considered less targeted. Social assistance from the government should be obtained by all poor families. The lack of information on the number and list of poor people is an obstacle. Therefore, it is necessary to categorize the poor (poor families) based on the data owned by Manembo-Nembo Village. In this research, population data is categorized using the K-Means algorithm clustering method based on income data, latest education, social assistance and data on housing conditions received. Using RapidMiner Studio software, the results obtained are 3 clusters. Using the elbow method to get the optimal number of clusters and calculate the Sum of Square Error of each cluster value. The results of research with RapidMiner obtained results, for cluster 2 with the category of poor population amounted to 119 family heads. Cluster 0 with the middle population category amounted to 78 family heads. Cluster 1 with the rich population category amounted to 75 family heads.*

Keywords: *Clustering, CRISP-DM, K-Means, Elbow-Method, Sum of Square Error*

1. PENDAHULUAN

Kelurahan Manembo – Nembo merupakan salah satu kampung dari hasil pembukaan hutan pada tahun 1894 oleh sekelompok orang yang datang dari “Negeri Tonsea” di bawah pimpinan Efraim Lengkong untuk dijadikan tempat tinggal dan berkebun, yang kemudian oleh mereka dinamakan “kampung Tulap” yang artinya “Pendengkean” atau sekali melompat sudah berada di sebelah. Manembo-Nembo adalah salah satu kelurahan di kecamatan Matuari, Kota Bitung, Sulawesi Utara, Indonesia. Kelurahan Manembo-Nembo yang melayani masyarakat secara langsung, setiap hari melakukan kegiatan pengelolaan data kependudukan seperti jumlah data penduduk, yaitu data penduduk berkecukupan, data penduduk kelas menengah dan data penduduk miskin. Data kependudukan yang lengkap dan terbaru (update) merupakan data wajib yang harus dimiliki dan diperhatikan oleh pihak kelurahan, yang nantinya akan di berikan ke kantor kecamatan agar bisa digunakan sebagai alat kontrol mengetahui pertumbuhan dan perkembangan penduduk.

Kemiskinan yang terjadi pada masyarakat merupakan masalah utama yang menjadi perhatian pemerintah di negara manapun di berbagai pemerintahan provinsi, kabupaten dan kota khususnya Kelurahan Manembo-Nembo Kecamatan Matuari Kota Bitung.

Dalam Penyaluran bantuan masih kurang tepat sasaran karena kurang informasi mengenai jumlah penduduk miskin yang perlu menerima bantuan di setiap RT/RW Kelurahan Manembo-nembo. Kondisi yang terjadi saat ini adalah bahwa di setiap RT/RW mendapatkan jumlah kuota bantuan yang sama, pada kenyataan jumlah yang berhak menerima bantuan di setiap RT/RW tidak sama. Kondisi tersebut salah satunya disebabkan karena data penduduk yang ada masih berupa buku catatan manual dan formulir yang diisi oleh setiap RT/RW.

Data Mining merupakan proses mencari informasi yang menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. *Data Mining* merupakan beberapa ilmu yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, *database* dan visualisasi untuk penanganan permasalahan pengambilan informasi dari *database* yang besar [1].

Clustering adalah teknik untuk mengkategorikan data penduduk calon penerima PKH (Program Keluarga Harapan), yang memiliki kesamaan karakteristik. Sampai saat ini pengelompokannya masih bersifat subyektif dari RT/RW sehingga berpotensi salah sasaran penerima bantuan. Penelitian ini bertujuan melakukan klasifikasi secara obyektif dengan menerapkan data mining pada data Penduduk Kelurahan Manembo-Nembo menggunakan algoritma K-Means Clustering.

Data pada penelitian ini di bagi jadi 3 kategori “masyarakat mampu”, “menengah” dan “tidak mampu”. Atribut yang digunakan pada penelitian ini adalah atribut Penghasilan, Pendidikan, Penerima bantuan, Jenis Lantai, Jenis Dinding, Fasilitas buang air, Sumber air, Bahan bakar memasak, Penerangan. Data tersebut akan di proses dengan Ms.Excel dan Rapid miner.

2. METODE PENELITIAN

2.1 Pengertian Data Mining

Data Mining merupakan disiplin ilmu yang mempelajari metode untuk mengekstrak pengetahuan atau menemukan pola dari suatu data. Jadi dapat disimpulkan bahwa *Data mining* adalah proses penggalian data secara mendalam untuk mengetahui hal yang berarti dan tidak diketahui keberadaanya. Clustering merupakan suatu metode untuk mencari dan mengelompokan data yang memiliki kemiripan karakteristik (*similarity*) antara satu data dengan data yang lain[2].

Knowledge Discovery in Database adalah keseluruhan proses *non-trivial* untuk mencari dan mengidentifikasi pola (*Pattern*) dalam data, dimana polayang ditemukan bersifat sah, baru dapat bermanfaat dan dapat dimengerti [3].

Knowledge discovery in Database (KKD) didefinisikan sebagai ekstraksi informasi potensial, implisit, dan tidak dikenal dari sekumpulan data. Proses *Knowledge discovery* melibatkan hasil dari proses *data mining* (proses mengekstrak kecenderungan pola suatu data), kemudian mengubah hasilnya secara akurat menjadi informasi yang mudah dipahami [4]

2.2 K-Means

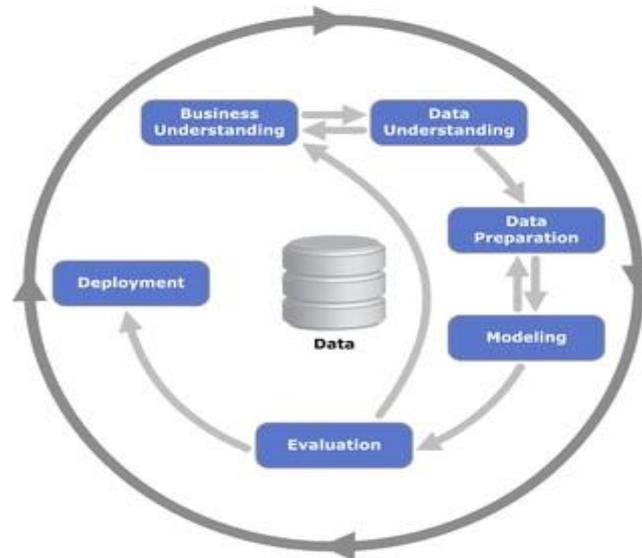
Metode K-means digunakan sebagai alternative metode kluster untuk data dengan ukuran yang lebih besar. K-Means merupakan metode pengelompokan yang paling terkenal dan banyak digunakan pada berbagai bidang karena sederhana dan mudah diimplementasikan. Langkah pada metode k-means yaitu dengan menentukan jumlah cluster secara acak, Mengalokasikan setiap data/objek ke cluster terdekat, mengulang Langkah tersebut hingga nilai cluster tidak berubah. [5]

2.3 CRISP-DM

CRISP-DM adalah model proses penambangan data (data mining) yang sering digunakan para ahli untuk memecahkan masalah. Jalannya penelitian ini mengacu pada enam tahap yaitu Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation dan Deployment[6].

2.4 RapidMiner

RapidMiner adalah aplikasi atau perangkat lunak yang berfungsi sebagai alat pembelajaran dalam ilmu data mining. Platfrom dikembangkan oleh perusahaan yang didedikasikan untuk semua Langkah yang melibatkan sejumlah besar data dalam bisnis komersial, penelitian, pendidikan, pelatihan, dan pembelajaran untuk pengelompokan, klasiikasi dan analisis regresi [7].



Gambar 2.1 Metode Cross Industry Proses for Data Mining

2.5 Metode Elbow

Menurut [8], Metode Elbow adalah untuk menghasilkan informasi dengan cara melihat perbandingan hasil antara jumlah cluster yang akan membentuk suatu titik terakhir, cluster nilai akan digunakan sebagai model data untuk cluster terbaik. Selain itu, hasil perhitungan akan digunakan untuk membandingkan jumlah cluster.

Metode *Elbow* merupakan salah satu metode untuk menentukan jumlah *cluster* terbaik dengan cara melihat presentase hasil perbandingan antara jumlah *cluster* yang membentuk siku pada suatu titik [9]. Metode ini memberikan gagasan dengan cara memilih nilai *cluster* dan menambahkan nilai *cluster* tersebut untuk dijadikan sebuah model data dalam penentuan *cluster* terbaik. Hasil presentase yang berbeda dari setiap nilai pada *cluster* dapat ditunjukkan dengan menggunakan grafik sebagai sumber informasi.

Untuk mendapatkan perbandingan adalah dengan menghitung dengan menghitung SSE (*Sum of Square Error*) dari masing-masing nilai *cluster*. Jika semakin besar jumlah *cluster* k maka nilai SSE akan semakin kecil [10]:

- Menginisialisasi awal nilai k
- Menaikan nilai k
- Menghitung hasil *sum of square error* dari tiap nilai k
- Melakukan analisis *sum of square error* dari nilai k yang mengalami penurunan secara drastic
- Menetapkan nilai k yang berbentuk siku

Rumus:

$$SSE = \sum_k \sum x_{is}^s k \| x_i - C_k \|^2 \quad (1)$$

2.6 Pengembangan Model

Model yang digunakan dalam penelitian ini adalah *Clustering* dengan Algoritma *K-means*. Berikut adalah tahapan yang akan dilakukan dalam mengolah data dengan algoritma *K-means clustering* sebagai berikut :

- Memilih jumlah *cluster* k yang terdiri dari 3 *cluster* yaitu warga mampu, menengah dan tidak mampu.
- Menentukan titik pusat awal dari setiap *cluster* secara acak dari dataset.
- Tempatkan setiap objek ke *cluster* terdekat. Menggunakan teori jarak *Euclidean distance* untuk menghitung jarak semua data ke setiap titik pusat *cluster* tersebut.

$$D(i,j) = \sqrt{(X_{i1} - X_{j1})^2 + (X_{i2} - X_{j2})^2 + \dots + (X_{ik} - X_{jk})^2} \quad (2)$$

Dimana:

- $D(i,j)$ = Jarak data ke i ke pusat cluster j
 X_{ki} = Data ke i atribut data ke k
 X_{kj} = Titik pusat ke j pada atribut ke k

- Hitung Kembali pusat *cluster* dengan keanggotaan *cluster* yang sekarang pusat *cluster* adalah rata-rata dari semua data dalam *cluster* tertentu. Jika dikehendaki bisa juga memakai median dari *cluster* tersebut.

$$R_k = \frac{1}{N_k} (X_{1k} + X_{2k} + \dots + X_{nk}) \quad (3)$$

Dimana:

R_k = Rata-rata baru

N_k = Jumlah training *pattern* pada cluster(k)

X_{nk} = Pola ke (n) yang menjadi bagian *cluster* (k)

- e. Tugaskan lagi setiap objek dengan memakai pusat *cluster* yang baru. Jika pusat *cluster* sudah tidak berubah lagi, maka proses peng*clusteran* selesai. Atau Kembali lagi ke Langkah ke tiga sampai pusat *cluster* tidak berubah lagi.

3. HASIL DAN PEMBAHASAN

Hasil dari penelitian ini didapatkan dengan mengolah data penduduk Lingkungan II Kelurahan Manembo-Nembo periode Mei-Juni 2023. Data diolah dengan menggunakan metode K-Means Clustering dan menghasilkan 3 cluster. Berikut merupakan tahapan penelitian yang dilakukan:

3.1 Pengumpulan Data

Dalam penelitian ini, data yang di gunakan adalah data primer yang diambil secara langsung melalui observasi dan wawancara dengan Kelurahan Manembo-Nembo dan Ketua RT di Lingkungan II. Data tersebut merupakan dataset Penduduk di Kelurahan Manembo-Nembo khususnya di Lingkungan II Periode Mei-Juni 2023.

NAMA	JENIS KELAMIN	SHDK	TEMPAT LAHIR	TANGGAL LAHIR	AGAMA	PENDIDIKAN	PEKERJAAN	PENGHASILAN	BANTUAN SOSIAL
Margaretha Palit Koagow	Perempuan	Kepala Keluarga	Lembean	03 Juli 1951	Katholik	SMA	Pensiunan	3.000.000-4.000.000	Tidak Ada
Herdian Badjuri	Laki-laki	Kepala Keluarga	Cilacap	07 September 1974	Katholik	SMA	Nelayan	1.000.000-3.000.000	Tidak Ada
Audi Ronny Kaseger	Laki-laki	Kepala Keluarga	Airmadidi	13 Oktober 1981	Katholik	SMA	Karyawan Swasta	1.000.000-3.000.000	Tidak Ada
Fransiskus Tiolong	Laki-laki	Kepala Keluarga	Mangarang	21 Maret 1952	Katholik	S1	Pensiunan	1.000.000-3.000.000	Tidak Ada
Barce Flooroh Tiolong	Laki-laki	Kepala Keluarga	Bitung	10 Maret 1983	Katholik	SMA	Karyawan Honorrer	1.000.000-3.000.000	Tidak Ada
Dessv Christu Pandaemanan	Perempuan	Kepala Keluarga	Manado	24 Desember 1978	Kristen	SMA	Tidak Bekerja	0	Tidak Ada

Gambar 3.1 Sebelum dilakukan Perubahan pada Data Penduduk

3.2 Data Understanding

Dalam penelitian ini, data yang di gunakan diambil berdasarkan data penduduk Lingkungan II Kelurahan Manembo-Nembo. Data tersebut diperoleh secara langsung dari Kelurahan dalam bentuk buku penduduk serta survey langsung bersama Ketua RT di Lingkungan II. Data yang di gunakan ini adalah data bulan Mei-Juni tahun 2023. Data disalin ke dalam Excel dengan 272 *record* dan 16 kolom. Atribut dalam data tersebut di antaranya:

- Nama, berisi nama penduduk Lingkungan 2 Kelurahan Manembo-Nembo.
- Jenis kelamin penduduk Lingkungan 2 Kelurahan Manembo-Nembo.
- SHDK (Status Hubungan dalam Keluarga) Lingkungan 2 Kelurahan Manembo-Nembo.
- Data tempat lahir Kelurahan Manembo-Nembo.
- Tanggal tahun lahir Kelurahan Manembo-Nembo.
- Agama penduduk Kelurahan Manembo-Nembo.
- Pendidikan, berisikan data Pendidikan terakhir penduduk Kelurahan Manembo-Nembo.
- Pekerjaan, berisikan data-data pekerjaan terbaru penduduk Kelurahan Manembo-Nembo
- Penghasilan, berisikan data-data penghasilan terbaru penduduk Kelurahan Manembo-Nembo.
- Bantuan Sosial, berisikan data-data penduduk penerima bantuan sosial di lingkungan II Kelurahan Manembo-Nembo.
- Jenis Lantai, berisikan data kondisi lantai dirumah penduduk lingkungan II Kelurahan Manembo-Nembo.
- Jenis Dinding, berisikan data kondisi dinding rumah penduduk lingkungan II Kelurahan Manembo-Nembo.
- Fasilitas Buang Air, berisikan data kondisi fasilitas buang air di rumah penduduk lingkungan II Kelurahan Manembo-Nembo.
- Sumber Air, berisikan data sumber air yang digunakan setiap rumah di lingkungan II Kelurahan Manembo-Nembo.
- Bahan Bakar Memasak, berisikan data bahan bakar yang digunakan untuk memasak dirumah penduduk lingkungan II Kelurahan Manembo-Nembo.
- Penerangan, berisikan data sumber penerangan yang digunakan di rumah setiap penduduk lingkungan II Kelurahan Manembo-Nembo.

3.3 Data Preprocessing

3.3.1 Transformasi Data

Pada tahap ini, dilakukan proses perubahan data menjadi 10 kategori (dengan menggunakan angka 1-5 untuk setiap atribut) agar data dapat diolah dan dilakukan proses Data Mining menggunakan algoritma K-Means Clustering.

NAMA	KATEGORI PENGASILAN(A)	KATEGORI PENDIDIKAN(B)	KATEGORI BANSOS(C)	KATEGORI LANTAI(D)	DINDING (E)	BUANG AIR(F)	AIR (G)	BAKAR (H)	PENERANGAN (J)
Margaretha Palit Koagow	2	3	1	3	2	2	3	2	4
Herdian Badjuri	2	3	1	3	2	2	3	2	4
Audi Ronny Kaseger	2	3	1	3	2	2	3	2	4
Fransiskus Tiolong	2	4	1	3	2	2	2	2	4
Barce Rooroh Tiolong	2	3	1	3	2	2	2	2	4

Gambar 3.2 Hasil Transformasi Data Penduduk

3.3.2 Data Selection

Data Selection merupakan seleksi atau pemilihan data yang akan digunakan dalam penelitian. Atribut terdiri dari nama, penghasilan, Pendidikan, bantuan sosial, jenis lantai, jenis dinding, fasilitas buang air, sumber air, bahan bakar memasak, dan penerangan. Atribut tersebut akan digunakan sebagai input dalam penelitian ini. Output yang dihasilkan dalam proses clustering adalah pengelompokan data penduduk untuk menentukan 3 kategori, “miskin”, “menengah” dan “kaya”. Untuk itu, digunakan 3 atribut utama yaitu penghasilan, Pendidikan dan bantuan sosial sebagai atribut yang paling berpengaruh dalam menentukan hasil. Kemudian didukung dengan 6 atribut pendukung yaitu data rumah penduduk yaitu, jenis dinding, jenis lantai, fasilitas buang air, sumber air, bahan bakar memasak dan juga sumber penerangan yang digunakan. Keseluruhan data tersebut berjumlah 272 data kepala keluarga lingkungan II Kelurahan Manembo-Nembo sepanjang periode Mei-Juni 2023.

3.4 Modeling

Pada tahapan ini dilakukan pemodelan dengan menggunakan algoritma K-Means Clustering dengan menggunakan software RapidMiner Studio. Berikut tahapannya:

3.4.1.1 Perhitungan Algoritma K-Means

Berikut langkah perhitungan secara manual dalam cluster menggunakan algoritma K-Means:

- Menentukan Jumlah cluster
Jumlah cluster yang dipakai pada data penduduk miskin sebanyak 3 cluster yaitu kaya, bekecukupan, miskin.
- Menentukan centroid
Menentukan pusat cluster awal (Centroid) ditentukan secara acak atau random yang diambil dari data yang sudah ada. Nilai cluster 0 diambil dari barisan ke-267, cluster 1 diambil dari barisan ke-250, cluster 2 diambil dari barisan ke-272.
- Menghitung jarak dari centroid
Menghitung jarak titik centroid dengan titik tiap objek dengan menggunakan Euclidean Distance. Adapun perhitungan manual untuk mendapatkan centroid awal sebagai berikut:

$$D(ij) = \sqrt{(X_{ii} - X_{ij})^2 + (X_{2i} - X_{2j}) + \dots + (X_{ki} - X_{kj})} \quad (4)$$

Jarak dihitung dari data kemiskinan pertama ke pusat cluster pertama dengan persamaan, sebagai berikut:

$$\begin{aligned} D(1,0) &= \sqrt{((2)-4)^2 + (3-3)^2 + (1-1)^2 + (3-3)^2 + (2-2)^2 + (2-3)^2 + (3-2)^2 + (2-2)^2 + (4-3)^2} \\ &= 2,645751 \end{aligned}$$

Hasil dari perhitungan jarak, didapatkan hasil jarak antara data pertama dengan cluster pertama yaitu 2,645751. Setelah perhitungan jarak data pertama dengan pusat cluster kedua sebagai berikut:

$$\begin{aligned} D(1,1) &= \sqrt{((2)-3)^2 + (3-4)^2 + (1-1)^2 + (3-3)^2 + (2-2)^2 + (2-3)^2 + (3-2)^2 + (2+2)^2 + (4-4)^2} \\ &= 2 \end{aligned}$$

Hasil dari perhitungan jarak, hasil yang di dapatkan jarak antara data pertama dengan pusat cluster kedua 2. Selanjutnya perhitungan jarak data pertama dengan cluster ketiga sebagai berikut:

$$\begin{aligned} D(1,2) &= \sqrt{((2)-3)^2 + (3-4)^2 + (1-1)^2 + (3-2)^2 + (2-2)^2 + (2-3)^2 + (3-2)^2 + (2-2)^2 + (4-3)^2} \\ &= 3 \end{aligned}$$

Hasil perhitungan tersebut didapatkan jarak antara data pertama dengan pusat cluster ketiga 3. Hasil dari ketiga perhitungan tersebut, dapat disimpulkan bahwa jarak data ke cluster 0 merupakan jarak terdekat. Hasil

perhitungan manual setiap masing-masing data dengan 3 atribut pada iterasi pertama dapat dilihat pada lampiran (lihat pada table penerapan rumus Eucludien Distance).

Disaat semua data dimasukan kedalam kelas yang sama berdasarkan cluster terdekat, lalu hitung Kembali dipusat cluster untuk menentukan iterasi kedua berdasarkan rata-rata setiap anggota yang ada dicluster tersebut.

Tabel 3.1 Nilai Pusat Awal Cluster

Data Ke	Centroid	A	B	C	D	E	F	G	H	I
267	1	4	3	1	3	2	3	2	2	3
250	2	3	4	1	3	2	3	2	2	4
272	3	1	1	1	2	2	3	2	2	3

1. Perhitungan centroid baru

Perhitungan centroid baru pada cluster 1

$$\begin{aligned} \text{Atr 1} &= (4+4+4+4)/4 \\ &= 4 \end{aligned}$$

$$\begin{aligned} \text{Atr 2} &= (3+3+3+3)/4 \\ &= 3 \end{aligned}$$

$$\begin{aligned} \text{Atr 3} &= (1+2+1+1)/4 \\ &= 1,25 \end{aligned}$$

Perhitungan centroid baru pada cluster 2

$$\begin{aligned} \text{Atr 1} &= (2+2+2+2+2+\dots+3)/108 \\ &= 2,12963 \end{aligned}$$

$$\begin{aligned} \text{Atr 2} &= (3+3+3+4+3+\dots+3)/108 \\ &= 3,287037 \end{aligned}$$

$$\begin{aligned} \text{Atr 3} &= (1+1+1+1+1+\dots+2)/108 \\ &= 1,12037 \end{aligned}$$

Perhitungan centroid baru pada cluster 3

$$\begin{aligned} \text{Atr 1} &= (1+3+3+4+3+\dots+1)/160 \\ &= 1,875 \end{aligned}$$

$$\begin{aligned} \text{Atr 2} &= (3+3+3+4+3+\dots+1)/160 \\ &= 1,725 \end{aligned}$$

$$\begin{aligned} \text{Atr 3} &= (1+1+1+1+1+\dots+1)/160 \\ &= 1,25625 \end{aligned}$$

Tabel 3.2 Nilai Centroid Iterasi Kelima

Centroid	A	B	C	D	E	F	G	H	I
1	1,961039	3,012987	1,194805	2,25974	1,38961	2,662338	2,181818	1,948052	2,285714
2	2,432432	3,324324	1,013514	3,202703	2	2,527027	2,243243	2	3,72973
3	1,77686	1,363636	1,322314	1,942149	1,123967	2,561983	2,198347	1,942149	2,347107

Pada iterasi kelima masih terjadi perubahan, sehingga masih dilakukan perhitungan.

1. Perhitungan *centroid* baru pada *cluster* 1

$$\begin{aligned} \text{Atr 1} &= \frac{3+4+3+2+\dots+4}{21} \\ &= 2,761905 \end{aligned}$$

$$\begin{aligned} \text{Atr 2} &= \frac{3+4+3+3+3+\dots+4}{21} \\ &= 3,047619 \end{aligned}$$

$$\begin{aligned} \text{Atr 3} &= \frac{1+2+1+1+\dots+1}{21} \\ &= 1,142857 \end{aligned}$$

$$\begin{aligned} \text{Atr 4} &= \frac{3+2+3+3+\dots+2}{21} \end{aligned}$$

$$\begin{aligned} &= 2,809524 \\ \text{Atr 5} &= \frac{2+2+2+1+\dots+2}{21} \end{aligned}$$

$$\begin{aligned} &= 1,714286 \\ \text{Atr 6} &= \frac{2+2+3+2+\dots+2}{21} \end{aligned}$$

$$\begin{aligned} &= 2,666667 \\ \text{Atr 7} &= \frac{2+2+2+3+\dots+2}{21} \end{aligned}$$

$$\begin{aligned} &= 2,142857 \\ \text{Atr 8} &= \frac{2+2+2+2+\dots+2}{21} \end{aligned}$$

$$\begin{aligned} &= 2 \\ \text{Atr 9} &= \frac{3+3+2+2+\dots+4}{21} \end{aligned}$$

$$= 2,333333$$

2. Perhitungan *centroid* baru pada *cluster 2*

$$\text{Atr 1} = \frac{2+2+2+2+2+\dots+3}{86}$$

$$= 2,313953$$

$$\text{Atr 2} = \frac{3+3+3+4+3+\dots+3}{86}$$

$$= 3,383721$$

$$\text{Atr 3} = \frac{1+1+1+1+1+\dots+2}{86}$$

$$= 1,081395$$

$$\text{Atr 4} = \frac{3+3+3+4+3+\dots+3}{86}$$

$$= 2,94186$$

$$\text{Atr 5} = \frac{2+2+2+2+2+\dots+2}{86}$$

$$= 1,848837$$

$$\text{Atr 6} = \frac{2+2+3+2+2+\dots+3}{86}$$

$$= 2,55814$$

$$\text{Atr 7} = \frac{2+2+2+2+2+\dots+3}{86}$$

$$= 2,27907$$

$$\text{Atr 8} = \frac{3+2+2+2+3+\dots+3}{86}$$

$$= 1,988372$$

$$\text{Atr 9} = \frac{4+4+3+4+3+\dots+3}{86}$$

$$= 3,546512$$

3. Perhitungan *centroid* baru pada *cluster 3*

$$\text{Atr 1} = \frac{1+2+2+1+2+\dots+1}{165}$$

$$= 1,751515$$

$$\text{Atr 2} = \frac{1+2+1+2+2+\dots+1}{165}$$

$$= 1,745455$$

$$\text{Atr 3} = \frac{1+1+1+1+2+\dots+1}{165}$$

$$= 1,272727$$

$$\text{Atr 4} = \frac{2+2+3+2+2+\dots+1}{165}$$

$$= 2,024242$$

$$\text{Atr 5} = \frac{2+1+1+1+1+\dots+2}{165}$$

$$= 1,187879$$

$$\text{Atr 6} = \frac{2+2+2+2+2+\dots+3}{165}$$

$$= 2,581818$$

$$\text{Atr 7} = \frac{2+2+2+3+2+\dots+2}{165}$$

$$= 2,175758$$

$$\text{Atr 8} = \frac{2+2+2+2+2+\dots+2}{165}$$

$$= 1,939394$$

$$\text{Atr } 9 = \frac{3+3+3+3+3+\dots+3}{165}$$

$$= 2,315152$$

Tabel 3.3 Nilai Centroid Iterasi Keenam

Centroid	A	B	C	D	E	F	G	H	I
1	1,897436	3	1,192308	2,25641	1,384615	2,628205	2,179487	1,935897	2,269231
2	2,466667	3,32	1,026667	3,186667	2	2,546667	2,24	2	3,72
3	1,789916	1,344538	1,319328	1,941176	1,117647	2,571429	2,201681	1,94958	2,352941

Setelah hasil yang didapatkan sama, selanjutnya menampilkan kelompok data berdasarkan penempatannya pada setiap cluster. Dari sebanyak 272 data, berikut hasil akhir dari proses clustering.

Tabel 3.4 Hasil Dari Clustering

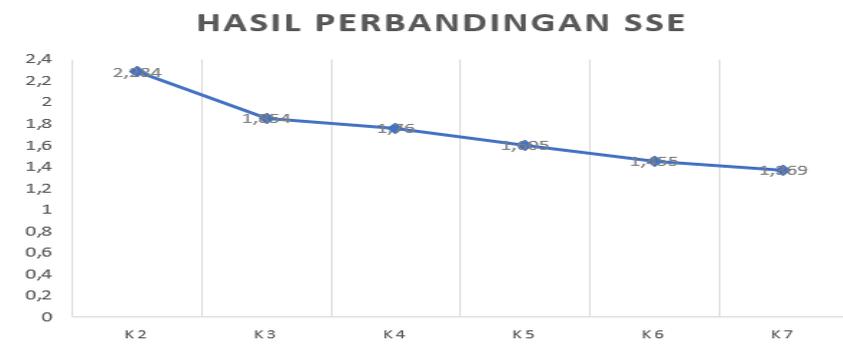
Centroid ke 5	Jumlah Data
Cluster 1	78
Cluster 2	75
Cluster 3	119
Total	272

Dari hasil Perhitungan menggunakan RapidMiner dan hasil perhitungan manual didapatkan hasil yang sama

3.5 Pengujian

Tabel 3.5 Perbandingan Nilai SSE

Cluster 1	Hasil SSE	Selisih
1	2,284	2,248
2	1,854	0,43
3	1,76	0,094
4	1,605	0,155


Gambar 3.3 Grafik SSE tiap Cluster

3.6 Evaluation

Dari hasil penerapan metode k-means clustering pada data penduduk lingkungan II Kelurahan Manembo-Nembo diperoleh hasil seperti berikut:

1. Cluster pertama dengan Centroid (1,897436, 3, 1,192308, 2,25641,1,384615, 2,628205, 2,179487, 1,935897, 2,269231) dengan jumlah 78 kepala keluarga.
2. Cluster kedua dengan Centroidb (2,466667, 3,23, 1,026667, 3,186667, 2, 2,546667, 2,24, 2, 3,72) dengan jumlah 75 Kepala Keluarga.

3. Cluster Ketiga dengan Centroid (1,789916, 1,344538, 1,319328, 1,941176, 1,117647, 2,571429, 2,201681, 1,94958, 2,352941) dengan jumlah 119 Kepala Keluarga

3.7 Hasil Penelitian

Hasil penelitian ini menghasilkan model perhitungan manual dengan menggunakan algoritma K-Means Clustering dengan perhitungan dilakukan menggunakan Microsoft Excel. Dengan data yang tidak berubah pada iterasi keenam dengan hasil centroid terbaik. Kemudian, perhitungan dengan menggunakan software RapidMiner Studio.

Tabel 3.6 Hasil Centroid Terbaik

Centroid	A	B	C	D	E	F	G	H	I
1	1,897436	3	1,192308	2,25641	1,384615	2,628205	2,179487	1,935897	2,269231
2	2,466667	3,32	1,026667	3,186667	2	2,546667	2,24	2	3,72
3	1,789916	1,344538	1,319328	1,941176	1,117647	2,571429	2,201681	1,94958	2,352941

Cluster	KATEGORI ...	KATEGORI ...	KATEGORI ...	KATEGORI ...	DINDING (E)	BUANG AIR(F)	AIR (G)	BAKAR (H)	PENERANG...
Cluster 0	1.867	3.027	1.200	2.227	1.360	2.613	2.187	1.933	2.253
Cluster 1	2.474	3.282	1.026	3.179	2	2.564	2.231	2	3.679
Cluster 2	1.790	1.345	1.319	1.941	1.118	2.571	2.202	1.950	2.353

Gambar 3.4 Hasil Centroid Pada Tapid Miner

Tabel 3.7 Hasil Cluster Model Terbaik

Cluster 0	75 kepala keluarga
Cluster 1	78 kepala keluarga
Cluster 2	119 kepala keluarga
Total Number of Items	272 kepala keluarga

Bisa disimpulkan pada gambar 4.21 bahwa *centroid* data yang digunakan di *RapidMiner Studio* adalah sebagai berikut:

1. *Cluster 0*, merupakan *cluster* paling tinggi dengan memiliki penghasilan tinggi, Pendidikan tinggi, bansos rendah, dan data rumah yang tergolong baik yang diberikan kepada penduduk dengan kategori "kaya".
2. *Cluster 1*, merupakan *cluster* menengah dengan memiliki penghasilan rendah Pendidikan tinggi, bansos rendah, dan data rumah yang tergolong baik yang diberikan kepada penduduk dengan kategori "menengah".
3. *Cluster 2*, merupakan *cluster* paling rendah dengan memiliki penghasilan rendah, pendidikan rendah, bansos rendah, dan data rumah yang tergolong masih rendah yang diberikan kepada penduduk dengan kategori "miskin".

4. KESIMPULAN

Dari penelitian yang telah dilakukan dan berdasarkan perhitungan yang telah dijelaskan diatas, dapat ditarik kesimpulan sebagai berikut: Dari pengolahan data dengan menggunakan metode Clustering K-Means dan telah dioptimalisasi dengan menggunakan Metode Elbow didapatkan hasil posisi kurva optimal cluster terbaik adalah 1,854 dengan 3 cluster. Cluster 2, merupakan cluster paling rendah dengan memiliki penghasilan rendah, pendidikan rendah, bansos rendah, dan data rumah kondisi rendah yang diberikan kepada penduduk dengan kategori "miskin" dengan jumlah 119 Kepala Keluarga. Cluster 0, merupakan cluster paling tinggi dengan memiliki penghasilan tinggi, Pendidikan tinggi, dan bansos rendah, serta data rumah yang diberikan kepada penduduk dengan kategori "kaya" dengan jumlah 75 Kepala Keluarga. Cluster 1, merupakan cluster menengah dengan memiliki penghasilan rendah, Pendidikan tinggi, dan bansos rendah serta data rumah yang diberikan kepada penduduk dengan kategori "menengah" dengan jumlah 78 Kepala Keluarga. Hasil pengolahan metode K-Means dengan menggunakan RapidMinerStudio sama dengan perhitungan manual yang dilakukan. Sehingga didapatkan hasil jumlah penduduk miskin yang tepat agar bantuan sosial dapat disalurkan dengan baik dan tepat sasaran.

DAFTAR PUSTAKA

- [1] Y. Mardi, “Jurnal Edik Informatika Data Mining : Klasifikasi Menggunakan Algoritma C4.5,” *JurnalEdikInformatika*, vol. 2, no. 2, pp. 213–219, 2017.
- [2] D. Astuti, A. R. Iskandar, and A. Febrianti, “Penentuan Strategi Promosi Usaha Mikro Kecil Dan Menengah (UMKM) Menggunakan Metode CRISP-DM dengan Algoritma K-Means Clustering,” *Journal of Informatics, Information System, Software Engineering and Applications*, vol. 1, no. 2, pp. 60–72, 2019, doi: 10.20895/INISTA.V1I2.
- [3] Mabur, A. Ginanjar, and R. Lubis, “Penerapan Data Mining untuk Memprediksi Kriteria Nasabah Kredit,” *Jurnal Komputer dan Informatika (KOMPUTA)*, 2012.
- [4] Asroni and R. Adrian, “Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang,” *JURNAL ILMIAH SEMESTA TEKNIKA*, vol. 18, no. 1, pp. 76–82, 2015.
- [5] I. Nasution, A. P. Windarto, and M. Fauzan, “Penerapan Algoritma K-Means Dalam Pengelompokan Data Penduduk Miskin Menurut Provinsi,” *Technology and Science (BITS)*, vol. 2, no. 2, pp. 76–83, 2020, [Online]. Available: <https://www.bps.go.id>.
- [6] B. E. Adiana, I. Soesanti, and A. E. Permasari, “Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model dan Teknik Clustering,” *Jurnal Terapan Teknologi Informasi*, vol. 2, no. 1, pp. 23–32, 2018, doi: 10.21460/jutei.2017.21.76.
- [7] S. Angra and S. Ahuja, “Analysis of Student’s Data using Rapid Miner,” *Journal on Today’s Ideas - Tomorrow’s Technologies*, vol. 4, no. 2, pp. 109–117, Dec. 2016, doi: 10.15415/jotitt.2016.42007.
- [8] N. P. E. Merliana, Ernawati, and A. J. Santoso, “ANALISA PENENTUAN JUMLAH CLUSTER TERBAIK PADA METODE K-MEANS CLUSTERING,” *PROSIDING SEMINAR NASIONALMULTI DISIPLIN ILMU & CALL FOR PAPERSUNISBANK*, pp. 978–979, 2015.
- [9] T. S. Madhulatha, “An Overview on Clustering Methods,” *IOSR Journal of Engineering*, vol. 2, pp. 719–725, 2012.
- [10] Y. Irwanto, Purwananto, and R. Soelaiman, “Optimasi Kinerja Algoritma Klasterisasi K-Means untuk Kuantisasi Warna Citra,” vol. 1, no. 1, pp. 197–202, 2012.