

# Pengaplikasian Data Mining Dalam Mengelompokkan Data Penerima Bantuan Subsidi Rumah dengan Menggunakan Metode K-Means Clustering

Alvendo Wahyu Aranski<sup>1</sup>, Sarah Astiti<sup>2</sup>, Riko Andrian Putra<sup>3</sup>, Darmansah<sup>4,\*</sup>

<sup>1</sup> Fakultas Teknologi informasi, Prodi Sistem informasi, Institut Teknologi Batam, Batam, Indonesia

<sup>2</sup> Fakultas Informatika, Sistem Informasi, Institut Teknologi Telkom Purwokerto, Purwokerto, Indonesia

<sup>3</sup> Prodi Sistem informasi, Institut Teknologi Dan Ilmu Sosial Khatulistiwa, Pasaman, Indonesia

<sup>4\*</sup> Fakultas Teknik dan komputer, Program Studi Sistem Informasi, Universitas Putera Batam, Batam, Indonesia

Email: <sup>1</sup>alvendo@iteba.ac.id, <sup>2</sup>sarah@ittelkom-pwt.ac.id, <sup>3</sup>urangguma@gmail.com, <sup>4,\*</sup>darmansah071@gmail.com

Email Penulis Korespondensi: darmansah071@gmail.com

Submitted: 17/06/2024; Accepted: 30/06/2024; Published: 30/06/2024

**Abstrak**—Dari tahun 2015 hingga saat ini, pemerintah memberikan bantuan untuk renovasi rumah di daerah Tembesi kecamatan Sagulung Kota Batam. Namun, dalam penentuan pemberian bantuan, pemerintah kecamatan terkadang menghadapi masalah dalam menentukan masyarakat yang akan mendapatkan subsidi rumah dan belum adanya skema atau kategori dalam penentuan penerima bantuan. Oleh karena itu, penulis akan melakukan penelitian ini dengan mengelompokkan atau memclusterkan kelayakan penerima bantuan subsidi rumah dengan menggunakan Algoritma K-Means. algoritma pengelompokan K-Means dapat mengelompokkan setiap data ke dalam kumpulan, sehingga kumpulan data dengan karakteristik yang sama akan dikelompokkan dalam kumpulan yang sama, atau kumpulan data dengan karakteristik yang berbeda akan dikelompokkan dalam kumpulan yang berbeda. Tujuan pengelompokan adalah untuk menentukan bahwa kumpulan 0 dan kumpulan 1 layak menerima bantuan subsidi rumah, dan kumpulan 1 tidak. Penelitian ini menggunakan metrik seperti jumlah anggota keluarga, pekerjaan, kondisi rumah, dan pendapatan. Hasil dari penelitian ini di dapatkan data sebanyak 91 data dalam cluster 0 dan 79 data dalam cluster 1. Dengan demikian dari data yang berjumlah 170 data, 91 orang layak untuk mendapatkan bantuan subsidi rumah, dan 79 orang tidak layak.

**Kata Kunci:** Data Mining; Algoritma K-Means Clustering; Penerima Bantuan

**Abstract**—From 2015 until now, the government has provided assistance for house renovations in the Tembesi area, Sagulung sub-district, Batam City. However, in determining the provision of assistance, sub-district governments sometimes face problems in determining which people will receive housing subsidies and there is no scheme or category for determining recipients of assistance. Therefore, the author will conduct this research by grouping or clustering the eligibility of recipients of housing subsidy assistance using the K-Means algorithm. The K-Means clustering algorithm can group each data into sets, so that data sets with the same characteristics will be grouped in the same set, or data sets with different characteristics will be grouped in different sets. The purpose of grouping is to determine that group 0 and group 1 are eligible to receive housing subsidy assistance, and group 1 is not. This research uses metrics such as number of family members, employment, housing conditions, and income. The results of this research obtained 91 data in cluster 0 and 79 data in cluster 1. Thus, from the 170 data, 91 people were eligible for housing subsidy assistance, and 79 people were not eligible.

**Keywords:** Data Mining; K-Means Clustering Algorithm; Aid Recipients

## 1. PENDAHULUAN

Kemiskinan adalah masalah yang sering terjadi di negara berkembang dan salah satunya terjadi di Indonesia [1]. Banyak program pemerintah untuk mengurangi kemiskinan [2], seperti program bantuan renovasi rumah, bantuan sosial, bantuan langsung tunai (BLT), dan program keluarga harapan (PKH), antara lain [3][4][5][6]. Hasil survei menunjukkan bahwa warga pemerintah tidak menerima bantuan yang mereka butuhkan, yang mengakibatkan distribusi tunjangan yang tidak merata dan tidak konsisten [7],[8].

Kelurahan Tembesi terletak di kecamatan Sagulung kota Batam [9]. Banyak penduduk Tembesi bekerja di pasar, buruh, PNS, Honorer, atau sebagai pekerja keras. Beberapa dari mereka bekerja di luar kota, bahkan di luar negeri [10]. Akibatnya, Tembesi banyak mendapat bantuan dari program pemerintah seperti subsidi rumah bagi masyarakat yang kurang mampu. Namun, dengan bantuan tersebut, banyak penerima bantuan yang tidak tepat sasaran dengan program bantuan subsidi rumah, sehingga dampak yang ditimbulkan mengakibatkan bantuan yang tidak sesuai dengan masyarakat dikarenakan kurangnya penyelidikan yang lebih dalam [11], [3]. Oleh karena itu, penulis meneliti masalah ini lebih lanjut. Penelitian ini dilakukan selama sekitar dua bulan dan menemukan bahwa desa-desa yang terlibat dalam program subsidi rumah. Untuk menyelesaikan masalah ini, penulis menggunakan Algoritma K-Means Clustering untuk menentukan kelayakan penerima bantuan berdasarkan atribut seperti pendapatan, jumlah anggota keluarga, pekerjaan, dan kelakayaan rumah. Dalam pengelompokan penulis membagi menjadi dua kelompok yaitu cluster 0 layak dan cluster 1 tidak layak [12].

Dari penelitian sebelumnya menyebutkan bahwa Algoritma K-Means adalah teknik clustering yang membagi kumpulan data menjadi kluster K yang tidak tumpang tindih [13]. Setiap kluster memiliki centroid terdekat. Meminimalkan variansi kluster total atau kesalahan kuadrat total adalah tujuan algoritma ini [14]. Selain itu algoritma ini membagi dataset ke dalam kelompok K dengan mengurangi jumlah total kesalahan kuadrat, atau total kesalahan kuadrat, dari setiap titik data ke centroid terdekatnya [15]. Kemudian Para peneliti telah menggunakan algoritma K-Means Clustering untuk mengelompokkan data. Salah satu peneliti sebelumnya dari penelitian ini adalah untuk

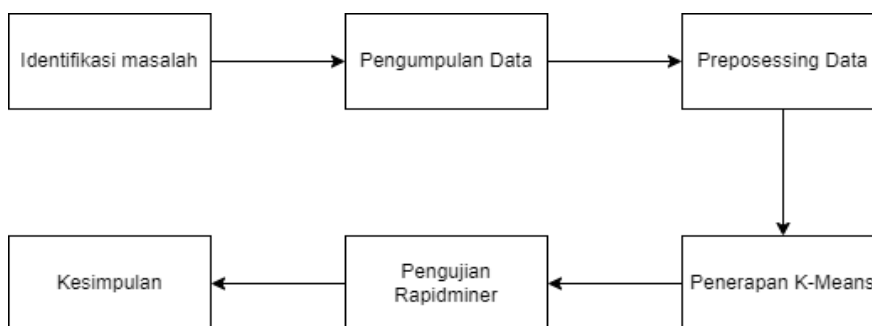
menemukan pengelompokan data saat ini dari siswa yang layak mendapatkan beasiswa [1]. Penelitian ini dilakukan dengan menggunakan pemrosesan data cepat yang terdiri dari tiga kelompok: kelompok tingkat tinggi, kelompok tingkat sedang, dan kelompok tingkat rendah[16].

Penelitian berikutnya bertujuan untuk mengelompokkan penduduk miskin ke dalam tiga kelompok. Penelitian ini menguji aplikasi RapidMiner untuk memastikan bahwa hasil dari proses mengelompokkan data untuk mengidentifikasi penerima PKH yang menggunakan aplikasi tersebut sama dengan hasil yang dihasilkan di sistem yang sama [17]. Penelitian ini diharapkan bisa menjadi acuan oleh pemerintah setempat dalam memberikan subsidi rumah kepada masyarakat di kawasan kelurahan tembesi [18]. Selain itu metode K-Means clustering dinilai mampu untuk memberikan hasil yang baik dalam berbagai penelitian terdahulu [19], [20]. Oleh karena itu diharapkan penelitian ini bisa memberikan dampak yang positif dan bisa memberikan rujukan oleh pemerintah daerah atau kecamatan sagulung dalam memberikan bantuan subsidi rumah .

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan penelitian

Pada penelitian ini agar terarah dan sesuai dengan apa yang diharapkan maka penulis membuat beberapa tahapan yang dilakukan dalam penelitian ini yaitu seperti gambar 1 di bawah ini:



**Gambar 1.** Tahapan Penelitian

Pada gambar diatas dapat dijelaskan tahapan penelitian yang dilakukan seperti berikut ini:

a. Mengidentifikasi Masalah

Langkah pertama adalah identifikasi masalah, pada tahapan ini penulis berusaha untuk melihat dan memahami masalah saat bantuan subsidi rumah tidak tepat sasaran.

b. Study Literatur

Pada bagian ini, penulis melakukan sejumlah tugas yang berkaitan dengan pengumpulan daftar pustaka, membaca, dan mencatat. Selain itu, menyelesaikan masalah dengan melihat sumber tulisan sebelumnya, mengolah bahan penelitian, atau mencari teori yang relevan dengan masalah atau kasus penelitian ini.

c. Pengumpulan Data

Peneliti menggunakan dua metode untuk mendapatkan data: observasi dan wawancara. Dengan observasi, mereka mengamati lokasi penelitian penulis untuk mendapatkan hasil yang akurat. Dengan wawancara, secara langsung mewawancarai Sekretaris kelurahan tembesi.

d. Preprocessing Data

Data preprocessing menjelaskan jenis proses yang menggunakan data mentah untuk mempersiapkan prosedur lain. Preprocessing dalam data mining adalah untuk membuat transformasi data menjadi lebih mudah dan efisien untuk digunakan. Setelah data primer (mentah) dipreposing, tahap pengolahan data dimulai.

e. Pemodelan Menggunakan K-Means

Pada tahap ini, penerapan algoritma k-means dilakukan secara manual.

f. Pengujian Menggunakan Rapid Miner

Setelah pemodelan selesai, pengujian dilakukan di aplikasi rapidminer untuk mendapatkan hasil yang sesuai dengan pemodelan.

g. Kemudian Kesimpulan

Pada tahapan ini penulis menarik Kesimpulan dari hasil analisis yang telah dilakukan.

### 2.2 Algoritma K-Means Clustering

K-Means adalah algoritma clustering non-hierarchical yang bekerja pada atribut numerik dan membagi data ke dalam cluster. Ini juga termasuk dalam partitioning clustering, yang memisahkan data ke k daerah bagian yang



terpisah. Algoritma K-Means sangat terkenal karena kemudahan dan kemampuan untuk mengkluster data besar dan outlier dengan sangat cepat.

Untuk membentuk pusat cluster awal, algoritma K-Means pertama-tama mengambil sebagian dari banyaknya komponen populasi. Pusat cluster ini dipilih secara acak dari sekumpulan populasi data. Selanjutnya, algoritma K-Means menguji setiap komponen di dalam populasi data dan menandai masing-masing komponen ke salah satu pusat cluster yang telah ditetapkan tergantung pada jarak minimum antara masing-masing komponen dan pusat cluster. Posisi pusat cluster akan dihitung. Berikut ini adalah proses Algoritma K-Means :

- a. Tentukan jumlah kluster (K)
- b. Tentukan titik centroid awal secara acak
- c. Menghitung jarak setiap objek.
- d. Kelompokkan nilai centroid yang sama ke dalam kelompok yang sama.
- e. Lakukan iterasi, kemudian tentukan posisi centroid baru dengan menggunakan persamaan 1
- f. Ulangi langkah 3 jika posisi centroid baru tidak sama.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Identifikasi Masalah

Pada tahap awal penelitian, penulis menemukan masalah untuk diteliti dan dirumuskan sebagai tujuan utama penelitian. Berdasarkan analisis mereka, penulis menemukan masalah yang akan digunakan sebagai rumusan tujuan penelitian. Fokus utama dari rumusan masalah ini adalah menentukan layak dan tidak layak penerima bantuan, yang akan dibagi menjadi dua kelompok dengan menggunakan algoritma k-means untuk menentukan layak dan tidak layak.

#### 3.2 Processing Data

Tahap persiapan adalah proses mengubah data mentah menjadi format baru yang lebih efisien untuk memungkinkan melanjutkan proses.

##### a. Pembersihan data

Pembersihan data adalah proses menemukan dan memperbaiki kesalahan dan ketidakakuratan yang ditemukan selama pengumpulan data. Berikut adalah data awal seperti pada tabel 1 di bawah ini:

**Tabel 1.** Data Awal

| No | Keluarga   | Kepala Keluarga | Pekerjaan    | KualitasRumah     | Pendapatan |
|----|--|-----------------|--------------|-------------------|------------|
| 1  | Sukadi<br>Ernita<br>Ani sertina<br>Aldi<br>Tomi<br>wena                    | Sukadi          | Buruh        | Rumah Tidak Layak | 1700000    |
| 2  | Andi Ritonga<br>Weni sitorus<br>mariana<br>morulu<br>Alex<br>Gana<br>Sindi | Andi Ritonga    | Buruh        | Rumah Tidak Layak | 1900000    |
| 3  | Simbolon<br>Renata<br>Lusmiar<br>Saut                                      | Simbolon        | Petani/Kebun | Rumah Layak       | 3600000    |
| 4  | Koslen<br>Renata<br>Iusmider<br>Sindi<br>wini                              | Koslen          | Pegawai      | Rumah Layak       | 2700000    |
| 5  | David<br>Lenni<br>Della<br>Surya<br>Erni                                   | David           | Buruh        | Rumah Tidak Layak | 1700000    |



| No  | Keluarga  | Kepala Keluarga | Pekerjaan      | Kualitas Rumah    | Pendapatan |
|-----|---|-----------------|----------------|-------------------|------------|
| 6   | Layla<br>Adlo<br>Indi<br>Indra<br>Juliana<br>Partiwi<br>Sekar<br>Susi | Indra           | Buruh          | Rumah Layak       | 1600000    |
| 7   | Erianto<br>Santi<br>Kayla   | Erianto         | Buruh          | Rumah Layak       | 1600000    |
| 8   | Amin<br>Sibarani<br>Rama<br>Yanti                                     | Amin            | Buruh          | Rumah Layak       | 1600000    |
| 9   | Anto<br>Meta  | Anto            | Pedagang       | Rumah Layak       | 3400000    |
| 10  | Sandi<br>Sri  | Sandi           | Buruh          | Rumah Layak       | 1900000    |
| ... | ...   | ...             | ...            | ...               | ...        |
| 170 | Salman<br>Siti  | Salman          | Petani/Pekebun | Rumah Tidak Layak | 2500000    |

Tabel 1 di atas menunjukkan bahwa fitur jumlah anggota keluarga belum ditentukan; bagaimanapun, cara menghitung jumlah anggota keluarga didasarkan pada jumlah keluarga, seperti yang ditunjukkan pada baris pertama dari keluarga Sukadi, di mana ada enam anggota. Untuk menghitung jumlah anggota keluarga dari kepala keluarga, hasilnya dapat dilihat di bawah ini. Pada tabel 2 merupakan tabel hasil pembersihan data

**Tabel 2.** Hasil Pembersihan Data

| No  | Nama Kepala Keluarga | Pekerjaan      | Jumlah Anggota Keluarga | Kualitas Rumah    | Pendapatan |
|-----|----------------------|----------------|-------------------------|-------------------|------------|
| 1   | Sukadi               | Buruh          | 6                       | Rumah Tidak Layak | 1700000    |
| 2   | Andi Ritonga         | Buruh          | 7                       | Rumah Tidak Layak | 1900000    |
| 3   | Simbolon             | Petani/Pekebun | 4                       | Rumah Layak       | 3600000    |
| 4   | Koslen               | Pegawai        | 5                       | Rumah Layak       | 2700000    |
| 5   | David                | Buruh          | 8                       | Rumah Tidak Layak | 1700000    |
| 6   | Indra                | Buruh          | 5                       | Rumah Layak       | 1600000    |
| 7   | Erianto              | Buruh          | 3                       | Rumah Layak       | 1600000    |
| 8   | Amin                 | Buruh          | 4                       | Rumah Layak       | 1600000    |
| 9   | Anto                 | Pedagang       | 2                       | Rumah Layak       | 3400000    |
| 10  | Sandi                | Buruh          | 2                       | Rumah Layak       | 1900000    |
| 11  | ...                  | ...            | ...                     | ...               | ...        |
| 170 | Salman               | Buruh          | 2                       | Rumah Layak       | 1800000    |

b. Pemberian bobot

Tahap pemberian bobot adalah transformasi data yang berupa teks ke dalam data label untuk meningkatkan keakuratan clustering menggunakan k-means. Pada tahap ini, pemberian bobot dapat membantu algoritma k-means menyesuaikan dengan lebih baik terhadap karakteristik data dan menghasilkan hasil yang lebih relevan dengan atribut yang digunakan, seperti pekerjaan dan kualitas rumah. Berikut ini adalah pembobot seperti yang terlihat pada tabel 3 dan 4 dibawah ini:

**Tabel 3.** Bobot Nama Pekerjaan

| Atribut        | Kriteria       | Bobot |
|----------------|----------------|-------|
| Nama Pekerjaan | Buruh          | 4     |
|                | Petani/Pekebun | 3     |
|                | Pedagang       | 2     |
|                | Pegawai        | 1     |



**Tabel 4.** Bobot Kualitas Rumah

| Atribut        | Kriteria          | Bobot |
|----------------|-------------------|-------|
| Kualitas Rumah | Rumah Tidak Layak | 2     |
|                | Rumah Layak       | 1     |

Pada Tabel 4 diatas adalah kualitas rumah. Berikut ini adalah tabel hasil dari pembobotan rumah seperti tabel 5 dibawah ini:

**Tabel 5.** Hasil Pembobotan

| No  | Nama Kepala Keluarga | Bobot Pekerjaan | Jumlah Anggota Keluarga | Bobot Kualitas Rumah | Pendapatan |
|-----|----------------------|-----------------|-------------------------|----------------------|------------|
| 1   | Sukadi               | 4               | 6                       | 2                    | 1700000    |
| 2   | Andi Ritonga         | 4               | 7                       | 2                    | 1900000    |
| 3   | Simbolon             | 3               | 4                       | 1                    | 3600000    |
| 4   | Koslen               | 1               | 5                       | 1                    | 2700000    |
| 5   | David                | 4               | 8                       | 2                    | 1700000    |
| 6   | Indra                | 4               | 5                       | 1                    | 1600000    |
| 7   | Erianto              | 4               | 3                       | 1                    | 1600000    |
| 8   | Amin                 | 4               | 2                       | 1                    | 1600000    |
| 9   | Anto                 | 2               | 2                       | 1                    | 3400000    |
| 10  | Sandi                | 4               | 2                       | 1                    | 1900000    |
| 11  | ...                  | ...             | ...                     | ...                  | ...        |
| 170 | Salman               | 4               | 2                       | 1                    | 1800000    |

Untuk menghitung nilai z-score, nilai average, nilai standart deviasi, dan nilai z-score dihitung dalam tabel berikut: Tabel 6 di bawah ini menunjukkan hasil nilai average untuk setiap atribut yang digunakan.

**Tabel 6.** Hasil Nilai Avarage

| No | Bobot Pekerjaan | Jumlah Anggota Keluarga | Bobot Kualitas Rumah | Pendapatan |
|----|-----------------|-------------------------|----------------------|------------|
| 1  | 4               | 6                       | 2                    | 1700000    |
| 2  | 4               | 7                       | 2                    | 1900000    |
| 3  | 3               | 4                       | 1                    | 3600000    |
| 4  | 1               | 5                       | 1                    | 2700000    |
| 5  | 4               | 8                       | 2                    | 1700000    |
| 6  | 4               | 5                       | 1                    | 1600000    |
| 7  | 4               | 3                       | 1                    | 1600000    |
| 8  | 4               | 2                       | 1                    | 1600000    |
| 9  | 2               | 2                       | 1                    | 3400000    |
| 10 | 4               | 2                       | 1                    | 1900000    |
| 11 | ---             | ---                     | ---                  | ---        |
| 12 | 4               | 2                       | 1                    | 1800000    |

Adapun proses menghitung rata – rata dapat dilihat dibawah ini:

$$\bar{x} = \frac{4+4+3+1+4+4+4+4+2+4+3+4+4+3+3+4+2+4+3+4+\dots+3}{170}$$

$$\bar{x} = \frac{537}{170} = 3,158823529$$

Menghitung nilai rata-rata untuk atribut lain juga sama dengan menjumlahkan semua data kemudian membagi total data untuk mendapatkan nilai rata-rata. Setelah mengetahui nilai rata-rata untuk masing-masing atribut, nilai setiap objek kemudian dihitung. Pada Tabel 7 dibawah ini merupakan nilai bobot pekerjaan deviasi  $x_i$

**Tabel 7.** Nilai Bobot Pekerjaan Deviasi  $x_i$

| Bobot Pekerjaan | $(x_i - \bar{x})$ | $(x_i - \bar{x})^2$ |
|-----------------|-------------------|---------------------|
| 4               | 0,841176471       | 0,707577855         |
| 4               | 0,841176471       | 0,707577855         |
| 3               | -0,158823529      | 0,025224913         |
| 1               | -2,158823529      | 4,660519031         |
| 4               | 0,841176471       | 0,707577855         |
| 4               | 0,841176471       | 0,707577855         |



| Bobot Pekerjaan | $(x_i - \bar{x})$ | $(x_i - \bar{x})^2$ |
|-----------------|-------------------|---------------------|
| 4               | 0,841176471       | 0,707577855         |
| 4               | 0,841176471       | 0,707577855         |
| 2               | -1,158823529      | 1,342871972         |
| 4               | 0,841176471       | 0,707577855         |
| 3               | 0,841176471       | 0,707577855         |
| ...             | ...               | ...                 |
| 3               | -0,158823529      | 0,025224913         |

Untuk menghitung proses nilai objek deviasi dapat dilihat dibawah ini:

$$(x_1 - \bar{x}) = (4 - 3,158823529) = 0,841176471 \quad (x_1 - \bar{x})^2 = (0,841176471)^2 = 0,707577855$$

Di bawah ini adalah contoh hasil tabel lainnya. Tabel 8 berikut menunjukkan nilai total dari semua anggota keluarga deviasi xi.

**Tabel 8.** Nilai Jumlah Anggota Keluarga Deviasi  $x_i$

| Jumlah Anggota Keluarga | $(x_i - \bar{x})$ | $(x_i - \bar{x})^2$ |
|-------------------------|-------------------|---------------------|
| 6                       | 1,558823529       | 2,429930796         |
| 7                       | 2,558823529       | 6,547577855         |
| 4                       | -0,441176471      | 0,194636678         |
| 5                       | 0,558823529       | 0,312283737         |
| 8                       | 3,558823529       | 12,66522491         |
| 5                       | 0,558823529       | 0,312283737         |
| 3                       | -1,441176471      | 2,076989619         |
| 2                       | -2,441176471      | 5,959342561         |
| 2                       | -2,441176471      | 5,959342561         |
| 2                       | -2,441176471      | 5,959342561         |
| 2                       | -2,441176471      | 5,959342561         |
| ...                     | ...               | ...                 |
| 6                       | 1,558823529       | 2,429930796         |

Pada Tabel 9 dibawah ini merupakan nilai bobot kualitas rumah pendapatan deviasi  $x_i$

**Tabel 9.** Nilai Bobot Kualitas Rumah Pendapatan Deviasi  $x_i$

| Bobot Kualitas Rumah | $(x_i - \bar{x})$ | $(x_i - \bar{x})^2$ |
|----------------------|-------------------|---------------------|
| 2                    | 0,694117647       | 0,481799308         |
| 2                    | 0,694117647       | 0,481799308         |
| 1                    | -0,305882353      | 0,093564014         |
| 1                    | -0,305882353      | 0,093564014         |
| 2                    | 0,694117647       | 0,481799308         |
| 1                    | -0,305882353      | 0,093564014         |
| 1                    | -0,305882353      | 0,093564014         |
| 1                    | -0,305882353      | 0,093564014         |
| 1                    | -0,305882353      | 0,093564014         |
| 1                    | -0,305882353      | 0,093564014         |
| 1                    | -0,305882353      | 0,093564014         |
| 1                    | -0,305882353      | 0,093564014         |
| ...                  | ...               | ...                 |
| 2                    | 0,694117647       | 0,481799308         |

Pada Tabel 10 dibawah ini merupakan nilai pendapatan deviasi  $x_i$  seperti dibawah ini:

**Tabel 10.** Nilai Pendapatan Deviasi  $x_i$

| Pendapatan | $(x_i - \bar{x})$ | $(x_i - \bar{x})^2$ |
|------------|-------------------|---------------------|
| 1700000    | -677647           | 4,59205             |
| 1900000    | -477647           | 2,28147             |
| 3600000    | 1222353           | 1,49415             |
| 2700000    | 322353            | 1,03911             |
| 1700000    | -677647           | 4,59205             |
| 1600000    | -777647           | 6,04735             |
| 1600000    | -777647           | 6,04735             |
| 1600000    | -777647           | 6,04735             |
| 3400000    | 1022353           | 1,04521             |



|         |             |             |
|---------|-------------|-------------|
| 1900000 | -477647     | 2,28147     |
| 3800000 | 1422353     | 2,02309     |
| ...     | ...         | ...         |
| 2500000 | 122352,9412 | 14970242215 |

Setelah semua nilai objek deviasi didapatkan kemudian menghitung standart deviasi dengan rumus pada persamaan 4 yang dijelaskan sebelumnya

Standart deviasi bobot pekerjaan

$$S = \frac{\sqrt{1700000}}{170 - 1} = \sqrt{1,010128785} = 1,005051633$$

Standart deviasi jumlah anggota keluarga

$$S = \frac{\sqrt{1900000}}{170 - 1} = \sqrt{4,20066133} = 2,049551495$$

Standart deviasi bobot kualitas rumah

$$S = \frac{\sqrt{3600000}}{170 - 1} = \sqrt{0,213574661} = 0,462141386$$

Standart deviasi pendapatan

$$S = \frac{\sqrt{2700000}}{170 - 1} = \sqrt{6,37249} = 798,278$$

Setelah standart deviasi dan average didapatkan kemudian dapat melanjutkan menghitung z-score, berikut salah satu proses dalam menghitung z-score. Berikut tabel 11 normalisasi setelah dilakukan perhitungan z-score seperti dibawah ini:

**Tabel 11.** Hasil Normalisasi

| No  | Bobot Pekerjaan | Jumlah Anggota Keluarga | Bobot Kualitas Rumah | Pendapatan   |
|-----|-----------------|-------------------------|----------------------|--------------|
| 1   | 0,836948514     | 0,760568121             | 1,501959504          | -0,848885316 |
| 2   | 0,836948514     | 1,248479746             | 1,501959504          | -0,598346247 |
| 3   | -0,158025244    | -0,215255129            | -0,66188046          | 1,531235839  |
| 4   | -2,147972759    | 0,272656496             | -0,66188046          | 0,403810029  |
| 5   | 0,836948514     | 1,736391371             | 1,501959504          | -0,848885316 |
| 6   | 0,836948514     | 0,272656496             | -0,66188046          | -0,97415485  |
| 7   | 0,836948514     | -0,703166753            | -0,66188046          | -0,97415485  |
| 8   | 0,836948514     | -1,191078378            | -0,66188046          | -0,97415485  |
| 9   | -1,152999001    | -1,191078378            | -0,66188046          | 1,28069677   |
| 10  | 0,836948514     | -1,191078378            | -0,66188046          | -0,598346247 |
| 11  | -0,158025244    | -1,191078378            | -0,66188046          | 1,781774908  |
| ... | ...             | ...                     | ...                  | ...          |
| 170 | -0,158025244    | 0,760568121             | 1,501959504          | 0,15327096   |

### 3.3 Algoritma K-Means Clustering

Setelah normalisasi selesai, algoritma k-means dapat digunakan untuk memclustering data. Langkah pertama pada tahap k-means adalah menemukan titik pusat centroid, yang ditemukan secara kebetulan dengan data yang tidak memiliki kesamaan antaranya. Selanjutnya, titik pusat centroid ditemukan. Berikut ini adalah tabel 12 letak titik centroid:

**Tabel 12.** Titik Pusat Centroid

| Cluster | Bobot Pekerjaan | Jumlah Anggota Keluarga | Bobot Kualitas Rumah | Pendapatan   |
|---------|-----------------|-------------------------|----------------------|--------------|
| C0      | 0,836948514     | 1,248479746             | 1,501959504          | -0,598346247 |
| C1      | -1,152999001    | -1,191078378            | -0,66188046          | 1,28069677   |

Setelah menemukan titik pusat centroid, langkah berikutnya adalah menghitung jarak antara masing-masing data. Ini dilakukan dengan data yang telah dinormalisasi, yang dapat ditemukan dalam tabel 11 diatas hasil normalisasi. Cara menghitung jarak antara masing-masing data adalah sebagai berikut:

$$D(1 - 0) = \sqrt{(0,836 - 0,836)^2 + (0,760 - 1,248)^2 + (1,501 - 1,501)^2 + ((-0,848) - 0,598)^2} = 0,54847751$$



$$D(1 - 1) = \sqrt{(0,836 - 1,152)^2 + (0,760 - 1,191)^2 + (1,501 - 0,661)^2 + ((-0,848) - 1,280)^2} = 4,121424319$$

Proses di atas diambil dari salah satu data pertama, sehingga jarak dari masing-masing data dapat dilihat pada tabel di bawah ini. Tabel 13 di bawah ini menunjukkan hasil dari Iterasi ke 1.

**Tabel 13.** Iterasi ke 1

| Nama Kepala Keluarga | Bobot Pekerjaan | Jumlah Anggota |         | Bobot Kualitas |       |        | Cluster |
|----------------------|-----------------|----------------|---------|----------------|-------|--------|---------|
|                      |                 | Keluarga       | Rumah   | Pendapatan     | C0    | C1     |         |
| Sukadi               | 0,8369          | 0,7606         | 1,50196 | -0,8489        | 0,548 | 4,1214 | 0       |
| Andi Ritonga         | 0,8369          | 1,2485         | 1,50196 | -0,5983        | 0     | 4,2573 | 0       |
| Simbolon             | -0,158          | -0,215         | -0,6619 | 1,5312         | 3,514 | 1,416  | 1       |
| Koslen               | -2,148          | 0,2727         | -0,6619 | 0,4038         | 3,943 | 1,9752 | 1       |
| David                | 0,8369          | 1,7364         | 1,50196 | -0,8489        | 0,548 | 4,6634 | 0       |
| Indra                | 0,8369          | 0,2727         | -0,6619 | -0,9742        | 2,403 | 3,3447 | 0       |
| Erianto              | 0,8369          | -0,703         | -0,6619 | -0,9742        | 2,938 | 3,0467 | 0       |
| Amin                 | 0,8369          | -1,191         | -0,6619 | -0,9742        | 3,283 | 3,0074 | 1       |
| Anto                 | -1,153          | -1,191         | -0,6619 | 1,2807         | 4,257 | 0      | 1       |
| Sandi                | 0,8369          | -1,191         | -0,6619 | -0,5983        | 3,261 | 2,7369 | 1       |
| ...                  | ...             | ...            | ...     | ...            | ...   | ...    | ...     |
| Salman               | -0,158          | 0,7606         | 1,50196 | 0,1533         | 1,339 | 3,2791 | 0       |

Setelah menghitung jarak masing-masing data, iterasi ke 1 ditemukan, seperti yang ditunjukkan pada tabel 13 di atas. Selanjutnya, hasil cluster ditemukan pada tabel 14 di bawah ini, yang menunjukkan iterasi ke 1 hasil cluster.

**Tabel 14.** Hasil Cluster Iterasi Ke 1

| Cluster | Total |
|---------|-------|
| 0       | 83    |
| 1       | 87    |

Sampai hasil cluster tidak berubah, iterasi akan berlanjut. Sebelum melanjutkan iterasi berikutnya, perlu menentukan titik centroid baru. Proses menentukan titik centroid baru digambarkan di bawah ini:

$$\begin{aligned} \text{bobot pekerjaan}(0) &= 0,48930708 \\ &= \frac{(-0,848) + (-0,598) + (-0,848) + (-0,974) + (-0,974) + (-1,09) + 0,153 + \dots + 0,153}{83} \end{aligned}$$

$$\begin{aligned} \text{pendapatan}(0) &= -0,468548898 \\ &= \frac{1,531 + 0,403 + (-0,974) + 1,280 + (-0,598) + 1,781 + (-0,723) + 1,030 + \dots + 0,153}{87} \end{aligned}$$

$$\text{pendapatan}_{(1)} = 0,44700642$$

Tabel centroid baru dari iterasi sebelumnya ditunjukkan di bawah ini, di mana titik pusat centroid baru iterasi ke 1 ditunjukkan pada tabel 15 berikut ini:

**Tabel 15.** Titik Pusat Centroid Baru Iterasi Ke 1

| Cluster | Bobot Pekerjaan | Jumlah Anggota Keluarga | Bobot Kualitas Rumah | Pendapatan   |
|---------|-----------------|-------------------------|----------------------|--------------|
| C0      | 0,48930708      | 0,560700708             | 0,511285786          | -0,468548898 |
| C1      | -0,466810203    | -0,534921366            | -0,487778394         | 0,44700642   |

Untuk menghitung Iterasi selanjutnya sama seperti sebelumnya untuk mendapatkan Iterasi ke 13, berikut ini adalah hasil iterasi ke 13 kalinya seperti tabel 16 berikut ini:

**Tabel 16.** Iterasi Ke 13

| Nama Kepala Keluarga | Bobot Pekerjaan | Jumlah Anggota Keluarga | Bobot Kualitas Rumah | Pendapatan | C0    | C1     | Cluster |
|----------------------|-----------------|-------------------------|----------------------|------------|-------|--------|---------|
| Sukadi               | 0,8369          | 0,7606                  | 1,50196              | -0,8489    | 1,685 | 2,9966 | 0       |
| Andi Ritonga         | 0,8369          | 1,2485                  | 1,50196              | -0,5983    | 1,966 | 3,0207 | 0       |
| Simbolon             | -0,158          | -0,215                  | -0,6619              | 1,5312     | 2,612 | 1,1732 | 1       |
| Koslen               | -2,148          | 0,2727                  | -0,6619              | 0,4038     | 3,242 | 1,5382 | 1       |
| David                | 0,8369          | 1,7364                  | 1,50196              | -0,8489    | 2,299 | 3,3768 | 0       |
| Indra                | 0,8369          | 0,2727                  | -0,6619              | -0,9742    | 0,744 | 2,6658 | 0       |

|         |        |        |         |         |       |        |     |
|---------|--------|--------|---------|---------|-------|--------|-----|
| Erianto | 0,8369 | -0,703 | -0,6619 | -0,9742 | 0,98  | 2,746  | 0   |
| Amin    | 0,8369 | -1,191 | -0,6619 | -0,9742 | 1,371 | 2,9106 | 0   |
| Anto    | -1,153 | -1,191 | -0,6619 | 1,2807  | 3,148 | 1,4354 | 1   |
| Sandi   | 0,8369 | -1,191 | -0,6619 | -0,5983 | 1,377 | 2,6798 | 0   |
| ...     | ...    | ...    | ...     | ...     | ...   | ...    | ... |
| Salman  | -0,158 | 0,7606 | 1,50196 | 0,1533  | 2,149 | 1,9936 | 1   |

Berikut hasil cluster dari Iterasi ke 13. Pada tabel 17 dibawah ini merupakan hasil cluster iterasi ke 13

Tabel 17. Hasil Cluster Iterasi Ke 13

| Cluster | Total |
|---------|-------|
| 0       | 91    |
| 1       | 79    |

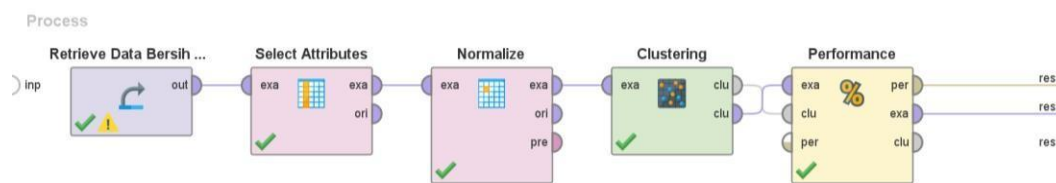
Selain itu, hasil centroid terakhir dapat dilihat di bawah ini. Hasil titik centroid terakhir ditunjukkan dalam tabel 18 di bawah ini.

Tabel 18. Titik Centroid Terakhir

| Cluster | Bobot Pekerjaan | Jumlah Anggota Keluarga | Bobot Kualitas Rumah | Pendapatan   |
|---------|-----------------|-------------------------|----------------------|--------------|
| C0      | 0,771345849     | -0,006150146            | 0,003916452          | -0,807587667 |
| C1      | -0,888512306    | 0,007084346             | -0,004511357         | 0,930259211  |

### 3.4 Pengujian Rapid Miner

Langkah terakhir adalah pengujian system di rapid miner, Berikut ini adalah hasil dari pengujian dengan aplikasi rapidminer seperti yang terlihat pada gambar 2 di bawah ini:



Gambar 2. Proses Rapid Miner

Proses pengolahan yang digambarkan pada gambar 2 di atas. Proses ini dimulai dengan pengolahan data yang telah dijelaskan sebelumnya, memilih fitur yang akan digunakan, masuk ke tahap normalisasi, dan kemudian menggunakan k-means untuk memcluster. Hasil cluster yang dihasilkan dapat dilihat pada gambar 3 dibawah ini:

## Cluster Model

```
Cluster 0: 91 items
Cluster 1: 79 items
Total number of items: 170
```

Gambar 3. Hasil Cluster Rapid Miner

Cluster menunjukkan hasil C0 sebanyak 91 dan C1 sebanyak 79, yang sama dengan hasil pemodelan sebelumnya, yang dapat dilihat pada tabel 17. Selanjutnya, hasil akhir centroid pada penambang cepat ditunjukkan pada gambar 3. Berikut ini adalah gambar 4 yang menunjukkan hasil centroid dengan aplikasi rapiminer:

| Attribute               | cluster_0 | cluster_1 |
|-------------------------|-----------|-----------|
| BOBOT PEKERJAAN         | 0.771     | -0.889    |
| JUMLAH ANGGOTA KELUARGA | -0.006    | 0.007     |
| BOBOT KUALITAS RUMAH    | 0.004     | -0.005    |
| PENDAPATAN              | -0.808    | 0.930     |

Gambar 4. Hasil Centroid Pada Rapid Miner

## 4. KESIMPULAN

Dari Analisa yang dilakukan menggunakan algoritma K-means clustering dari jumlah data sebanyak 170 dengan dibagi kedalam 2 cluster yaitu layak dan tidak layak mendapatkan bantuan subsidi rumah, maka dapat disimpulkan bahwa

cluster 0 atau layak mendapatkan bantuan subsidi adalah sebanyak 91 kepala keluarga, dan cluster 1 atau tidak layak ada sebanyak 79 kepala keluarga tidak layak mendapatkan bantuan. Hasil ini dilakukan berdasarkan beberapa atribut dan pembobotan atribut yang sudah dijelaskan pada tahapan analisis data.

## REFERENCES

- [1] U. S, “Penerapan Data Mining Dengan Mengimplementasikan Algoritma K-Means Dalam Proses Clustering Untuk Pengelompokan Mahasiswa Calon Penerima Beasiswa KIP,” *Build. Informatics, Technol. Sci.*, vol. 5, no. 1, 2023, doi: 10.47065/bits.v5i1.3411.
- [2] D. N. Alfiansyah, V. R. S. Nastiti, and N. Hayatin, “Penerapan Metode K-Means pada Data Penduduk Miskin Per Kecamatan Kabupaten Blitar,” *J. Repos.*, vol. 4, no. 1, pp. 49–58, 2022, doi: 10.22219/repositor.v4i1.1416.
- [3] S. Saragi, M. U. Batoebara, and N. A. Arma, “Analisis Pelaksanaan Program Keluarga Harapan (Pkh) Di Desa Kota Rantang Kecamatan Hamparan Perak,” *Publik J. Manaj. Sumber Daya Manusia, Adm. dan Pelayanan Publik*, vol. 8, no. 1, pp. 1–10, 2021, doi: 10.37606/publik.v8i1.150.
- [4] N. Fitra Tsania, B. Setiawati, and S. R. Arfah, “Implementasi Program Bantuan Langsung Tunai (Blt) Bagi Masyarakat Miskin Di Desa Laringgi Kabupaten Soppeng,” *KIMAP (Kajian Ilmu Mhs. Adm. Publik)*, vol. 4, no. 4, pp. 2245–2256, 2023, [Online]. Available: <https://journal.unismuh.ac.id/index.php/kimap/index>.
- [5] N. Ngabito, D. A. Razak, and B. Raf, “Implementasi Program Bantuan Rumah Hunian Bagi Masyarakat Miskin Di Provinsi Gorontalo,” *Provid. J. Ilmu Pemerintah.*, vol. 2, no. 1, pp. 76–93, 2023, doi: 10.59713/projip.v2i1.515.
- [6] K. N. Azizah, A. K. Nuzuli, and W. Oktaviana, “Implementasi Program Bantuan Langsung Tunai (BLT) Bagi Masyarakat Miskin di Nagari Batang Arak Tapan,” *J. Pengabd. Masy. dan Ris. Pendidik.*, vol. 2, no. 1, pp. 241–245, 2023, doi: 10.31004/jerkin.v2i1.154.
- [7] A. Ahyar, “Optimalisasi Pelayanan Bantuan Hukum Bagi Masyarakat Miskin,” *J. Penelit. Huk. Jure*, vol. 20, no. 3, p. 409, 2020, doi: 10.30641/dejure.2020.v20.409-434.
- [8] A. Kurniawan, N. Farkhatin, and M. Hidayah, “Sistem Pendukung Keputusan Penerimaan Bantuan Program Keluarga Harapan Menggunakan Metode Ahp,” *Semnas Ristek (Seminar Nas. Ris. dan Inov. Teknol.)*, vol. 8, no. 01, pp. 171–175, 2024, doi: 10.30998/semnasristek.v8i01.7152.
- [9] R. Daswito and N. A. Cahyadi, “PH , suhu air , dan perilaku pemberantasan sarang nyamuk terhadap keberadaan jentik nyamuk Aedes sp di Tembesi Lama , Kota Batam PH , water temperature , and eradication of mosquito den behavior on the presence of Aedes sp mosquito larvae in Tembesi Lama ,” vol. 04, no. 01, pp. 1–9, 2024.
- [10] R. Antonio, “Kajian Faktor-Faktor Yang Mempengaruhi Perkembangan Morfologi Kota Batam, Studi Pada Kawasan Jodoh,” *Sigma Tek.*, vol. 6, no. 2, pp. 511–513, 2023, doi: 10.33373/sigmateknika.v6i2.5542.
- [11] S. Amaliyah, Jasmir, and S. Rianti, “Penerapan Data Mining Untuk Menentukan Kelompok Prioritas Penerima Bantuan PKH Menggunakan Metode Clustering K-Means Pada Desa Kuala Dendang,” *J. Inform. Dan Rekayasa Komputer(JAKAKOM)*, vol. 3, no. 1, pp. 453–458, 2023, doi: 10.33998/jakakom.2023.3.1.802.
- [12] D. Darmansah, “Analisa Penyebab Kerusakan Tanaman Cabai Menggunakan Metode K-Means,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 7, no. 2, pp. 126–134, 2020, doi: 10.35957/jatisi.v7i2.309.
- [13] I. Iin, R. Fadila, A. Rizki Rinaldi, and F. Fathurrohman, “Penerapan Data Mining Dalam Mengelompokkan Jumlah Umkm Berdasarkan Kabupaten Kota Menggunakan K-Means Clustering,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 2, pp. 1446–1450, 2024, doi: 10.36040/jati.v8i2.8427.
- [14] E. F. L. Awalina and W. I. Rahayu, “Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail,” *J. Teknol. dan Inf.*, vol. 13, no. 2, pp. 122–137, 2023, doi: 10.34010/jati.v13i2.10090.
- [15] F. Febriansyah and S. Muntari, “Penerapan Algoritma K-Means untuk Klasterisasi Penduduk Miskin pada Kota Pagar Alam,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 8, no. 1, pp. 66–77, 2023, doi: 10.14421/jiska.2023.8.1.66-77.
- [16] G. Sonia and R. A. Putri, “Penerapan Metode K-Means Clustering Untuk Mengelompokkan Data Kelayakan Penerima Bantuan Renovasi Rumah,” *Build. Informatics, Technol. Sci.*, vol. 5, no. 2, pp. 442–455, 2023, doi: 10.47065/bits.v5i2.4298.
- [17] P. Dwi Lestari and M. Mulyawan, “Datamining Pada Penjualan Air Bersih Di Spam Akidah Menggunakan Algoritma K-Means Clustering Menggunakan Rapidminer,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 412–416, 2023, doi: 10.36040/jati.v7i1.6315.
- [18] W. Winayah, R. Kurniawan, and Y. Arie Wijaya, “Penerapan Data Mining Clustering Menggunakan Algoritma X-Means Pada Data Penerima Bantuan Program Keluarga Harapan Di Desa Gebang Kulon Kabupaten Cirebon,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 1, pp. 982–988, 2024, doi: 10.36040/jati.v8i1.8827.
- [19] Y. I. Kurniawan, “Pengelompokan Prioritas Negara Yang Membutuhkan Bantuan Menggunakan Clustering K-Means dengan Elbow dan Silhouette,” *J. Pendidik. dan Teknol. Indones.*, vol. 3, no. 10, pp. 455–463, 2023, [Online]. Available: <https://doi.org/10.52436/1.jpti.343>.
- [20] E. Dwiguna and A. Bahtiar, “Penerapan Data Mining Untuk Menentukan Penerima Bantuan Blt Menggunakan Metode Clustering K-Means Pada Desa Pamulihan,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 2, pp. 1382–1388, 2024, doi: 10.36040/jati.v8i2.9029.