

# PENERAPAN METODE *CLUSTERING K-MEANS* UNTUK MENENTUKAN PRIORITAS PENERIMA BANTUAN PROGRAM BERAS UNTUK RAKYAT MISKIN (RASKIN) STUDI KASUS : KECAMATAN SIULAK

Raudatul Kinanti<sup>1</sup>, Jasmir<sup>1</sup>, Fachruddin<sup>2,\*</sup>

Fakultas Ilmu Komputer, Magister Sistem Informasi, Universitas Dinamika Bangsa, Jambi, Indonesia

Email: <sup>1</sup>raudatulkinanti665@gmail.com, <sup>2</sup> jasmir@unama.ac.id, <sup>3</sup>fachruddin.stikom@gmail.com

Email Penulis Korespondensi: [raudatulkinanti665@gmail.com](mailto:raudatulkinanti665@gmail.com)

Artikel Info :

Artikel History :

Submitted : 01-09-2024

Accepted : 25-09-2024

Published : 30-09-2024

## Kata Kunci:

Kemiskinan;  
Bantuan Sosial;  
RASKIN; *K-Means*  
*Clustering*; *Data*  
*mining*; *SPSS*

## Keywords:

Poverty; Social  
Assistance; RASKIN;  
*K-Means*  
*CLUSTERING*; *Data*  
*mining*; *SPSS*

**Abstrak**– Kemiskinan di Indonesia merupakan masalah kompleks dan berkelanjutan. Pemerintah berupaya mengatasinya melalui program bantuan sosial seperti Bantuan Sosial Tunai (BST), Bantuan Langsung Tunai (BLT), RASKIN, dan Program Keluarga Harapan (PKH). Kabupaten Kerinci di Provinsi Jambi, dengan populasi 255.736 jiwa, masih memiliki banyak masyarakat yang membutuhkan bantuan, terutama di Kecamatan Siulak. Berdasarkan laporan BPS tahun 2021, distribusi bantuan RASKIN sering tidak sesuai kebutuhan karena pembagian yang tidak merata dan kurangnya validasi data. Penelitian ini menggunakan metode *K-Means CLUSTERING* untuk menentukan penerima bantuan RASKIN agar lebih tepat sasaran. Data penelitian berasal dari 349 penduduk Desa Siulak Panjang setelah proses pembersihan data dari total 355 data yang ada. Kriteria penerima bantuan meliputi jumlah anggota rumah tangga, luas lantai kurang dari 10 meter persegi, tingkat pendidikan rendah, kondisi ekonomi tidak stabil, dan kesehatan yang buruk. Delapan atribut digunakan dalam analisis. Hasil menunjukkan perbedaan jumlah penerima bantuan pada setiap kluster antara perhitungan manual dan menggunakan SPSS, yang disebabkan oleh standarisasi data pada SPSS. Dari perhitungan manual, *Cluster 1* memiliki 74 data (prioritas ketiga), *Cluster 2* memiliki 37 data (prioritas pertama), *Cluster 3* memiliki 73 data (prioritas kedua), *Cluster 4* memiliki 42 data (prioritas kelima), dan *Cluster 5* memiliki 123 data (prioritas keempat). Sebaliknya, hasil menggunakan SPSS menunjukkan *Cluster 1* dengan 32 data (prioritas pertama), *Cluster 2* dengan 68 data (prioritas kelima), *Cluster 3* dengan 124 data (prioritas kedua), *Cluster 4* dengan 73 data (prioritas keempat), dan *Cluster 5* dengan 52 data (prioritas ketiga). Berdasarkan hasil perbandingan, penggunaan SPSS lebih direkomendasikan karena memberikan hasil yang lebih akurat dan konsisten melalui standarisasi data, efisien dalam pengolahan data, dan menghasilkan kluster yang lebih sesuai dengan kondisi sebenarnya.

**Abstract**– Poverty in Indonesia is a complex and ongoing issue. The government attempts to address this through social assistance programs such as Cash Social Assistance (BST), Direct Cash Assistance (BLT), RASKIN, and the Family Hope Program (PKH). Kerinci Regency in Jambi Province, with a population of 255,736 people, still has many communities in need of assistance, especially in Siulak Subdistrict. According to a 2021 BPS report, the distribution of RASKIN assistance often does not meet needs due to uneven distribution and lack of data validation. This study uses the *K-Means CLUSTERING* method to determine RASKIN recipients more accurately. The research data consists of 349 residents of Siulak Panjang Village after cleaning data from the original 355 records. The criteria for receiving assistance include household size, floor area of less than 10 square meters, low education levels, unstable economic conditions, and poor health. Eight attributes were used in this analysis. Results show differences in the number of aid recipients in each *Cluster* between manual calculations and SPSS, due to SPSS data standardization. Manual calculations indicate *Cluster 1* with 74 data (third priority), *Cluster 2* with 37 data (first priority), *Cluster 3* with 73 data (second priority), *Cluster 4* with 42 data (fifth priority), and *Cluster 5* with 123 data (fourth priority). In contrast, SPSS results show *Cluster 1* with 32 data (first priority), *Cluster 2* with 68 data (fifth priority), *Cluster 3* with 124 data (second priority), *Cluster 4* with 73 data (fourth priority), and *Cluster 5* with 52 data (third priority). Based on this comparison, SPSS is more recommended as it provides more accurate and consistent results through data standardization, is efficient in data processing, and produces *Cluster s* that better match the actual conditions.

## 1. PENDAHULUAN

Kemiskinan merupakan permasalahan yang berkepanjangan dan kompleks di Indonesia[1], termasuk di Kabupaten Kerinci, Provinsi Jambi. Kecamatan Siulak di Kabupaten Kerinci menghadapi tantangan besar dengan banyak kelompok masyarakat yang memerlukan bantuan sosial, seperti keluarga miskin, lansia tanpa pendapatan tetap, anak yatim piatu, dan kelompok minoritas etnis. Untuk menangani masalah ini, pemerintah telah meluncurkan berbagai program bantuan sosial seperti Bantuan Sosial Tunai (BST), Bantuan Langsung Tunai (BLT), RASKIN (Rice for the Poor), dan Program Keluarga Harapan (PKH) yang bertujuan memberikan bantuan langsung, baik berupa uang tunai, beras, atau pendidikan[2].

Meskipun program-program tersebut dirancang untuk mengurangi kemiskinan, pelaksanaannya sering kali menghadapi berbagai kendala. Di Kecamatan Siulak, pelaksanaan Bantuan RASKIN sering kali tidak sesuai dengan kebutuhan

masyarakat[3]. Distribusi bantuan sering kali tidak merata dan tidak tepat sasaran, disebabkan oleh kurangnya koordinasi antara pemerintah pusat dan daerah serta ketidakakuratan data yang digunakan untuk menentukan penerima bantuan. Permasalahan ini menunjukkan perlunya evaluasi dan perbaikan dalam sistem distribusi bantuan sosial. Ketelitian dalam memvalidasi data sering diabaikan sehingga menimbulkan data yang tidak akurat[4][5]. Untuk memastikan bantuan RASKIN diberikan kepada yang benar-benar membutuhkan, diperlukan analisis dengan mengelompokkan data penduduk menggunakan algoritma *K-Means*. Pendekatan ini bertujuan untuk menjadikan program RASKIN lebih tepat sasaran, sehingga penerima bantuan adalah mereka yang benar-benar membutuhkan, dan membantu pemerintah dalam melaksanakan program RASKIN secara efektif.

Menurut [6] *Data mining* merupakan proses yang memanfaatkan teknik-teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengekstraksi serta mengidentifikasi informasi dan pengetahuan yang berguna dari berbagai *database* besar. Proses ini juga sering disebut sebagai penemuan pengetahuan, ekstraksi pengetahuan, analisis data atau pola, pengumpulan informasi, dan istilah lainnya. Tujuan dari *Data mining* adalah menggali, memambang suatu data besar untuk mendapatkan suatu informasi[7].

Metode *K-Means* dipilih karena lebih mudah diimplementasikan dibandingkan dengan metode *K-Medoids Clustering* atau *Fuzzy C-Means Clustering*. Selain itu, waktu yang dibutuhkan untuk proses pembelajaran relatif lebih cepat, metode ini sangat fleksibel, dan dapat dijelaskan dengan cara non-statistik. Oleh karena itu, metode *K-Means Clustering* dianggap lebih tepat untuk digunakan[8].

Penelitian terdahulu menunjukkan bahwa proses penentuan penerima bantuan sosial masih banyak bergantung pada pencatatan manual dan penggunaan MS Excel, yang seringkali tidak memadai dalam memastikan akurasi data. [4] menemukan bahwa proses manual ini menyebabkan ketidakakuratan dalam distribusi Bantuan Langsung Tunai (BLT) di Desa Cempaka Timur. Penelitian lain [5] menunjukkan bahwa klusterisasi data menggunakan algoritma *K-Means* dapat membantu dalam menentukan penerima bantuan dengan lebih tepat, mengatasi masalah kuota dan ketidaksesuaian kondisi penerima di Desa Pelangsan.

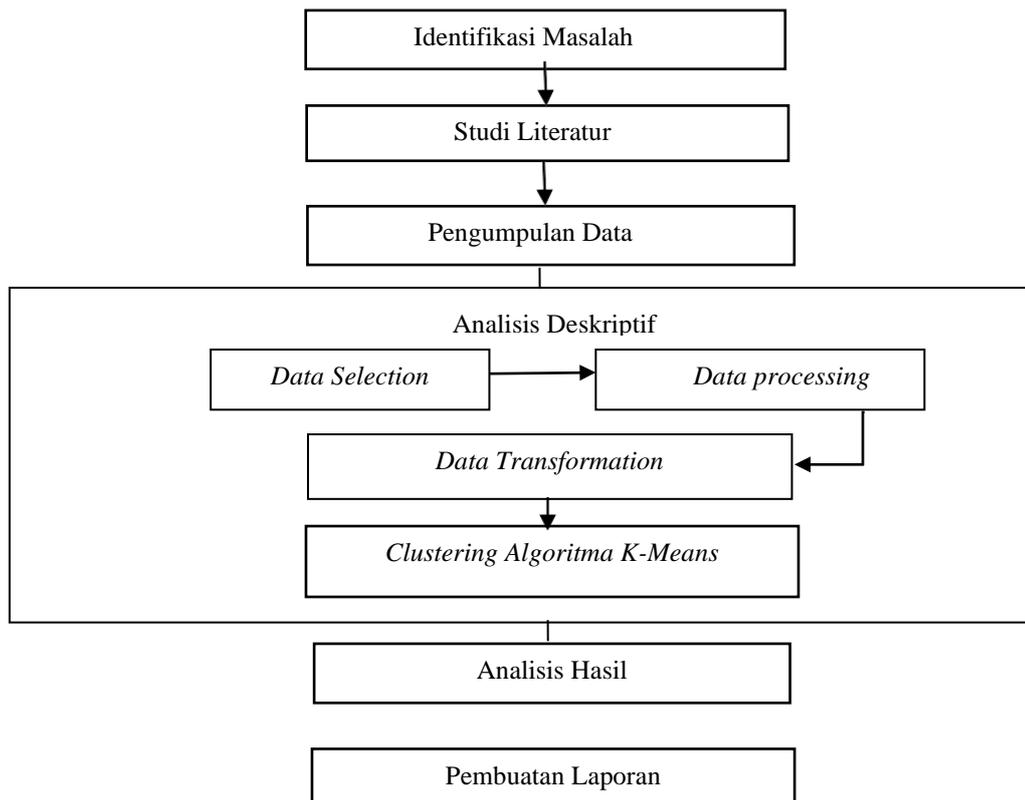
Lebih lanjut [9] mengembangkan aplikasi berbasis metode *K-Means* untuk menentukan penerima bantuan sosial beras di Kelurahan Landasan Ulin Timur, dengan hasil yang menunjukkan tingkat akurasi yang tinggi dalam identifikasi penerima bantuan. Metode ini terbukti efektif dalam memproses data besar dan menghasilkan keputusan yang lebih tepat. Berdasarkan temuan-temuan ini, terdapat kebutuhan untuk menerapkan metode *K-Means CLUSTERING* dalam konteks program RASKIN di Kecamatan Siulak.

Penelitian ini bertujuan untuk menerapkan metode *K-Means* dalam menentukan prioritas penerima bantuan RASKIN di Kecamatan Siulak, Kabupaten Kerinci. Dengan menggunakan teknik *CLUSTERING* ini, diharapkan dapat meningkatkan akurasi dalam distribusi bantuan dan memastikan bantuan yang diberikan benar-benar sampai kepada mereka yang membutuhkan. Penelitian ini diusulkan dalam bentuk tesis yang berjudul "Penerapan Metode *CLUSTERING K-Means* untuk Menentukan Prioritas Penerima Bantuan Program Beras untuk Keluarga Miskin (RASKIN): Studi Kasus di Kecamatan Siulak."

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Adapun kerangka kerja penelitian yang digunakan dapat dilihat pada gambar



## Gambar 2.1 Alur penelitian

1. **Identifikasi Masalah** Pada tahap ini, penulis mengidentifikasi dan merumuskan masalah dengan mengamati langsung situasi di Kecamatan Siulak, melakukan wawancara dengan warga dan perangkat desa, serta memantau berita terkait secara daring. Tujuan dari langkah ini adalah untuk memahami tantangan dalam penentuan prioritas bantuan RASKIN, sehingga penulis dapat merencanakan langkah-langkah selanjutnya dan mengidentifikasi data yang diperlukan untuk penelitian.
2. **Studi Literatur** Penulis melakukan kajian pustaka untuk memperoleh dasar teori yang relevan dengan penelitian. Ini mencakup studi terhadap buku, artikel, jurnal ilmiah, dan hasil penelitian serupa yang berkaitan dengan teknik *Data mining*, algoritma *K-Means*, dan penggunaan tools seperti SPSS. Tujuannya adalah agar penulis dapat memahami dan memformulasikan kerangka teori yang mendasari penelitian ini.
3. **Pengumpulan Data** Penulis mengumpulkan data sekunder berupa data SDGS penduduk dari Desa Siulak Panjang, serta data primer melalui wawancara dengan perangkat desa dan camat. Data sekunder diperoleh dari sumber-sumber yang sudah ada, sedangkan data primer dikumpulkan langsung dari pihak yang berwenang. Metode pengumpulan data meliputi observasi langsung, wawancara, dan studi literatur untuk memastikan informasi yang lengkap dan relevan.
4. **Data Selection** Pada tahap seleksi data, penulis memilih dan memilah data operasional yang relevan sebelum memulai proses *Data mining*. Data yang telah dipilih akan disimpan secara terpisah untuk memudahkan proses analisis lebih lanjut.
5. **Data Processing** Tahap ini melibatkan persiapan data sebelum analisis, termasuk ekstraksi, analisis, dan statistik data. Proses ini bertujuan untuk mengolah data mentah menjadi informasi yang bermanfaat untuk memecahkan masalah dan mendukung pengambilan keputusan.
6. **Data Transformation** Data Transformation adalah tahap di mana data mentah diubah secara sistematis untuk memenuhi kebutuhan analisis. Transformasi ini penting untuk meningkatkan kualitas data dan memastikan bahwa data yang digunakan akurat serta bebas dari bias.
7. **CLUSTERING Algoritma K-Means** Pada tahap ini, penulis menerapkan algoritma *K-Means CLUSTERING* menggunakan tools SPSS untuk mengelompokkan data dan menentukan prioritas penerima bantuan RASKIN. Proses ini bertujuan untuk menghasilkan penilaian yang akurat mengenai penerima bantuan yang paling membutuhkan.
8. **Analisis Hasil** Penulis menganalisis hasil dari penerapan algoritma *K-Means CLUSTERING* untuk menentukan penerima bantuan. Hasil analisis ini dibandingkan dengan data yang diperoleh untuk memastikan akurasi dan efektivitas penyaluran bantuan.
9. **Penulisan Laporan** Laporan akhir disusun berdasarkan kerangka penelitian yang mencakup pendahuluan, landasan teori, metodologi, analisis hasil, dan penutup. Laporan ini juga meliputi lampiran bukti hasil penelitian untuk memberikan dokumentasi lengkap mengenai proses dan temuan penelitian.

## 2.3 Data mining

*Data mining* merupakan proses eksplorasi data yang mendalam, yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk menganalisis dan menemukan informasi serta pengetahuan berharga dari berbagai *database* besar, yang kemudian digunakan untuk pengambilan keputusan bisnis yang relevan. [6][7][10].

## 2.4 Tahap-tahap Data mining

Adapun tahapan dalam *Data mining* dapat dijelaskan sebagai berikut [11] :

1. **Data Selection**  
Tahapan pertama untuk menggali informasi menggunakan KDD yaitu dengan memilih atau menyeleksi beberapa data yang dibutuhkan dari sekumpulan data yang banyak. Data yang telah dipilih akan digunakan pada tahapan *Data mining*. Untuk memudahkan penggunaan dan pencarian kembali data yang sudah diseleksi, data tersebut akan disimpan pada suatu berkas yang dibedakan dengan data lainnya.
2. **Pre-processing (Cleaning)**  
Untuk memastikan data yang disimpan merupakan data yang bermutu dan akurat, dilakukan tahapan ini yaitu dengan membuang data yang tidak lengkap atau tidak valid. Tahapan ini penting agar proses *Data mining* nanti dapat menghasilkan data yang bermutu dan valid sesuai dengan data yang dimasukkan.
3. **Transformation**  
Pada tahapan transformasi data, data akan disesuaikan dengan teknik *Data mining* yang akan dipakai. Diterapkan suatu format pada data sehingga nanti data yang tidak sesuai dengan format akan disingkirkan dan menyisakan data yang sesuai untuk digunakan pada tahapan selanjutnya dengan kualitas data yang tepat.
4. **Data mining**  
Pada proses *Data mining*, digunakan beberapa teknik atau metode tertentu untuk menganalisis data yang ada sehingga didapatkan hasilnya yaitu data yang berisi pengetahuan penting atau tersembunyi. Pemilihan teknik harus sesuai dengan tujuan KDD dilakukan sehingga hasilnya akan relevan.
5. **Evaluasi Pola (Pattern Evaluation)**  
Hasil data yang didapatkan dari proses *Data mining* akan diidentifikasi sehingga didapatkan suatu pola yang akan dimasukkan ke dalam knowledge based untuk kemudian dilakukan analisis.
6. **Presentasi Pengetahuan (Knowledge Presentation)**  
Pada tahapan ini, data yang ada di dalam knowledge based tersebut akan ditampilkan atau diperlihatkan terkait metode yang digunakan untuk memperolehnya kepada pengguna yang melakukan *Knowledge Discovery* tersebut.

## 2.5 CLUSTERING

Menurut [12] menyatakan bahwa, “*CLUSTERING* digunakan untuk menemukan kelompok data yang serupa. Fungsi dari *CLUSTERING* sebagai preprocessing, di mana *CLUSTERING* untuk mengidentifikasi anggota kelompok yang mirip”.

Menurut [13] menyatakan “*CLUSTERING* dilakukan dengan cara mengumpulkan data yang serumpun dari sebuah data set yang besar, untuk menyingkirkan suatu data yang tidak dibutuhkan”.

## 2.7 K-Means

Menurut [14] menyatakan bahwa bahwa Algoritma *K-Means* membentuk *cluster* data melalui proses iteratif. Jumlah *cluster* awal yang diinginkan diberikan sebagai input dan jumlah *cluster* akhir sebagai output. Jika algoritma dirancang untuk menghasilkan *K cluster*, maka jumlah *K* yang ditentukan di awal dan di akhir akan sama. Metode *K-Means* memilih *K* pola secara acak sebagai titik awal *centroid*. Jika posisi *centroid* baru tidak berubah, jumlah iterasi yang dibutuhkan untuk mencapai *cluster centroid* akan dipengaruhi oleh pemilihan *centroid* awal secara acak. Nilai *K*, yang digunakan sebagai pusat awal, dihitung menggunakan rumus jarak geometris, yang menentukan jarak terdekat antara titik *centroid* dan data atau objek. *Cluster* akan terdiri dari data yang dekat dengan *centroid* tersebut [15].

## 2.8 SPSS

Menurut *SPSS* merupakan program aplikasi yang unggul dalam analisis statistik dan manajemen data melalui antarmuka grafis yang mudah digunakan[17]. Awalnya dikembangkan untuk ilmu sosial, *SPSS* kini diterapkan dalam berbagai disiplin ilmu untuk analisis statistik, manajemen data, dan dokumentasi data. Program ini memudahkan pengguna dalam menjalankan berbagai aktivitas dengan menggunakan mouse dan menu deskriptif yang intuitif. Dengan fitur-fitur seperti pivot tables dan kemampuan untuk menyesuaikan *output*, *SPSS* menjadi alat yang efektif bagi peneliti dari berbagai bidang untuk mengolah dan menganalisis data secara komprehensif[18].

## 2.9 Raskin

Instruksi Presiden Nomor 8 Tahun 2008 tentang Kebijakan Perberasan mengarahkan Menteri, Kepala Lembaga Pemerintah Non-Departemen, serta Gubernur dan Bupati/Walikota di seluruh Indonesia untuk meningkatkan pendapatan petani, ketahanan pangan, ekonomi pedesaan, dan stabilitas ekonomi nasional. Salah satu Langkah konkret adalah Program Beras untuk Keluarga Miskin (*Raskin*)/Beras Sejahtera (*Rastra*), yang bertujuan mengurangi beban pengeluaran rumah tangga miskin dan mendukung ketahanan pangan serta kesejahteraan sosial. *Raskin*, yang merupakan bagian dari upaya penanggulangan kemiskinan dalam *Cluster I*, berfungsi sebagai perlindungan social berbasis keluarga untuk memenuhi kebutuhan pangan pokok dan telah beroperasi secara rutin sejak tahun 1998.

# 3. HASIL DAN PEMBAHASAN

## 3.1 Representasi Data

Berdasarkan wawancara dan observasi, peneliti mengumpulkan 355 data penduduk Desa Siulak Panjang, Kecamatan Siulak, Kabupaten Kerinci tahun 2022. Kriteria penerima bantuan *RASKIN* mencakup jumlah anggota keluarga, status lahan, status bangunan, jenis dan luas lantai, kondisi pekerjaan, jenis pekerjaan, serta pendidikan terakhir. Upaya ini bertujuan untuk meningkatkan ketahanan pangan dan memberikan perlindungan sosial kepada rumah tangga yang membutuhkan. Berdasarkan hasil wawancara dan observasi yang sudah dilakukan, peneliti memperoleh data-data penduduk Desa Siulak Panjang, Kecamatan Siulak, Kabupaten Kerinci tahun 2022.

**Tabel 3.1 Data Penduduk Desa Siulak Panjang**

Nama Data/Atribut	
<b>Atribut</b>	NAMA, Jumlah ART, RT, Desa, Kecamatan, Status bangunan, status lahan, Luas lantai, Jenis lantai, Jenis dinding, Jenis atap, penerang rumah, enrg masak, Fasilitas MCK, Sumber Air Mandi, Fasilitas BAB, Sumber Air Minum, Rumahh dibawah Sute/sutas, Rumah di lereng bukit/gunung, Rumahh di bantai sungai , Kondisi Rumah, Bantuan BLT, Bantuan PKH, Kondsi pekerjaan, Pekerjaan Utama, Jamsos Pekerja, Pendidikan terakhir.
<b>Data 1</b>	YONRIANTO , 4, 1, Siulak Panjang, Siulak, 1, 1, 70, 10, 1, 3, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 5, 1, 2, 3.
<b>Data 2</b>	PONZA MARTA, 4, 1, Siulak Panjang, Siulak, 4, 2, 80, 6, 2, 3, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 5, 10, 2, 6.
<b>Data 3</b>	DISMINA, 2, 1, Siulak Panjang, Siulak, 4, 2, 60, 2, 1, 3, 1, 1, 1, 2, 1, 1, 2, 2, 2, 2, 2, 3, 1, 2, 1.
<b>Data 4</b>	BUSTARAMIN, 3, 1, Siulak Panjang, Siulak, 1, 1, 16, 6, 1, 3, 1, 1, 1, 4, 4, 1, 2, 2, 2, 2, 2, 2, 1, 2, 2.
...	...
<b>Data 349</b>	MUNTAHAR, 4, 4, Siulak Panjang, Siulak, 2, 2, 78, 6, 1, 3, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 1, 2, 5, 1, 2, 2.

Dari tabel 3.1 dapat diketahui sebagian data dari keseluruhan data penduduk Desa Siulak Panjang memiliki 26 variabel/atribut. Selanjutnya dilakukan pemilihan data yang akan digunakan untuk perhitungan *CLUSTERING K-Means* dapat dilihat sebagai berikut :

**Tabel 3. 2 Data Penduduk Yang Digunakan Untuk Perhitungan K-Means**

No	Nama	Jumlah ART	Status lahan	Status bangunan	Jenis lantai	Luas lantai	Kondsi pekerjaan	Pekerjaan	Pendidikan terakhir
1	YONRIANTO	4	1	1	10	70	5	1	3
2	PONZA MARTA	4	2	4	6	80	5	10	6
3	DISMINA	2	2	4	2	60	3	1	1
4	BUSTARAMIN	3	1	1	6	16	2	1	2
5	NAZARUDIN	2	2	4	6	15	5	1	1
6	MUL ASRAL	2	1	1	6	48	5	1	1

7	NOPRIZAL	3	1	1	6	50	2	1	4
8	SAMSUL ISMAIL	2	1	1	2	70	5	1	4
9	RADISAH MUHAMMAD	5	2	4	6	64	5	1	2
10	ALIFGRA	3	1	1	6	20	5	14	2
11	RAWADI	4	1	1	2	80	5	1	4
12	DURADI	4	1	1	6	90	5	10	2
13	HASAN BASRI	2	2	4	6	10	5	1	2
14	ULIL AMRI	4	1	1	6	20	5	1	3
15	AFRIZAL	4	1	1	6	72	5	1	3
16	ADRIL PANI	3	1	1	6	64	5	1	4
17	LOZI ARENA	3	1	1	6	43	5	1	4
18	SIDARNI	3	2	4	6	100	2	1	2
19	IRNAYATI	3	1	1	6	100	5	1	4
..	.....	...	...	...	...	...	...	...	...
349	MUNTAHAR	4	2	2	6	78	5	1	2

Dari 3.2 peneliti mengambil beberapa atribut yang paling berpengaruh dalam penentuan penerima bantuan yaitu sebanyak 9 atribut yang terdiri dari Nama, Jumlah ART, Status lahan, Status bangunan, Jenis lantai, Luas lantai, Kondisi pekerjaan, Pekerjaan, Pendidikan terakhir. Sedangkan data penduduk yang digunakan sebanyak 349 penduduk di Desa Siulak Panjang, Kecamatan Siulak, Kabupaten Kerinci.

3.2 Perhitungan Manual K-Means CLUSTERING

Peneliti melakukan perhitungan dengan menggunakan Persamaan untuk menghitung jarak antar data pada K-Means menggunakan rumus Euclidian Distance (D).

Tahapan yang terlibat dalam pengelompokan menggunakan teknik K-Means adalah sebagai berikut:

Langkah 1: Pilih berapa banyak Cluster (Cluster = 5) yang akan digunakan.

Langkah 2: Centroid awalnya dipilih secara acak.

Saat ini, lima titik data acak (4,1,1,10,70,5,1,3) akan dipilih untuk berfungsi sebagai Centroid pertama dalam studi data populasi.

Langkah Ketiga: Tentukan jarak iterasi pertama ke centroid.

Pada titik ini, setiap bagian data akan ditautkan ke Centroid terdekat dan ditempatkan dalam grup yang sesuai dengan Centroid tersebut. Dengan menggunakan data populasi no. 1, tentukan jarak centroid setiap cluster menggunakan rumus di bawah ini:

Data : 4,1,1,10,70,5,1,3

Centroid m1 = (4,1,1,10,70,5,1,3)

Centroid m2 = (3,1,1,6,20,5,14,2)

Centroid m3 = (3,2,4,6,40,5,9,2)

Centroid m4 = (4,2,3,2,100,5,8,4)

Centroid m5 = (4,2,2,6,78,5,1,2)

$$DM1 = \sqrt{(4-4)^2 + (1-1)^2 + (1-1)^2 + (10-10)^2 + (70-70)^2 + (5-5)^2 + (1-1)^2 + (3-3)^2} = 0$$

$$DM2 = \sqrt{(4-3)^2 + (1-1)^2 + (1-1)^2 + (10-6)^2 + (70-20)^2 + (5-5)^2 + (1-14)^2 + (3-2)^2} = 51,83628073$$

$$DM3 = \sqrt{(4-3)^2 + (1-2)^2 + (1-4)^2 + (10-6)^2 + (70-40)^2 + (5-5)^2 + (1-9)^2 + (3-2)^2} = 31,4960315$$

$$DM4 = \sqrt{(4-4)^2 + (1-2)^2 + (1-3)^2 + (10-2)^2 + (70-100)^2 + (5-5)^2 + (1-8)^2 + (3-4)^2} = 31,9217794$$

$$DM5 = \sqrt{(4-4)^2 + (1-2)^2 + (1-2)^2 + (10-6)^2 + (70-78)^2 + (5-5)^2 + (1-1)^2 + (3-2)^2} = 9,110433579$$

Dari hasil perhitungan diatas, maka dimasukkan kedalam tabel Tabel 3.3

Tabel 3.3 Hasil Perhitungan Jarak Antara Masing-Masing Data Dengan Centroid (Iterasi 1)

No.	Nama	PENCARIAN CLUSTER PERTAMA					Jarak Terkecil	JARAK TERDEKAT					
		C1	C2	C3	C4	C5		C1	C2	C3	C4	C5	
1	YONRIANTO	0	51,83628073	31,4960315	31,9217794	9,110433579	0	v					
2	PONZA MARTA	14,69693846	60,35726965	40,22437072	20,61552813	10,24695077	10,24695077						v
3	DISMINA	13,6381817	42,43819035	22,04540769	40,82891133	18,78829423	13,6381817	v					
4	BUSTARAMIN	54,24942396	13,92838828	25,67099531	84,4985207	62,09669879	13,92838828		v				
5	NAZARUDIN	55,30822724	14,35270009	26,28687886	85,46344248	63,07138812	14,35270009		v				

6	MUL ASRAL	22,53885534	30,90307428	11,83215957	52,79204486	30,11644069	11,83215957	v	
7	NOPRIZAL	20,66397832	32,89376841	13,67479433	50,7937004	28,28427125	13,67479433	v	
8	SAMSUL ISMAIL	8,306623863	51,86520992	31,54362059	30,95157508	9,486832981	8,306623863	v	
9	RADISAH MUHAMMAD	8	46,03259715	25,37715508	36,97296309	14,17744688	8	v	
10	ALIFGRA	51,83628073	0	20,85665361	80,38656604	59,46427499	0	v	
11	RAWADI	12,84523258	61,56297589	41,1703777	21,30727575	5,099019514	5,099019514		v
12	DURADI	22,3159136	70,12132343	50,11985634	11,35781669	15,06651917	11,35781669		v
13	HASAN BASRI	60,25777958	16,73320053	31,06444913	90,41017642	68,05879811	16,73320053	v	
...	...	...	...	...	...	...	...	...	...
349	MUNTAHAR	9,110433579	59,46427499	38,89730068	23,53720459	0	0		v

Dari Tabel 3.3, terlihat hasil perhitungan jarak antara setiap data dengan *Centroid*. Kolom pencarian *Cluster* pertama menampilkan hasil jarak *Centroid* awal yang dipilih secara acak untuk data penduduk. Kolom C1 hingga C5 menunjukkan jarak antara *Centroid* m1 hingga m5 dengan setiap data penduduk. Kolom jarak terdekat mencatat pengelompokan data penduduk ke dalam *Cluster* berdasarkan nilai jarak terkecil.

Dari Tabel 3.3 didapatkan hasil keanggotaan penduduk Desa Siulak Panjang, Kecamatan Siulak, Kabupaten Kerinci (Iterasi 1)

*Cluster* 1 = (1, 3, 8, 9, 15, 16, 21, 25, 35, 48, 60, 66, 70, 75, 78, 79, 85, 87, 94, 97, 103, 106, 118, 120, 123, 124, 134, 136, 145, 151, 152, 153, 164, 170, 173, 177, 179, 180, 182, 184, 185, 188, 189, 194, 200, 212, 215, 216, 220, 225, 228, 231, 233, 234, 235, 238, 243, 248, 265, 267, 274, 321, 322, 323, 329, 332, 333, 339)

*Cluster* 2 = (4, 5, 10, 13, 14, 20, 24, 44, 63, 67, 68, 69, 74, 81, 82, 84, 91, 115, 117, 122, 130, 131, 132, 135, 146, 148, 155, 157, 165, 166, 168, 171, 186, 192, 193, 211, 217, 224)

*Cluster* 3 = (6, 7, 17, 23, 26, 32, 33, 34, 37, 39, 41, 45, 46, 47, 50, 52, 54, 55, 56, 57, 61, 62, 72, 73, 77, 80, 86, 88, 89, 90, 93, 95, 98, 99, 100, 101, 102, 104, 107, 111, 112, 119, 121, 125, 129, 137, 141, 143, 150, 156, 158, 159, 161, 163, 169, 175, 176, 181, 183, 187, 190, 195, 198, 199, 204, 205, 210, 213, 218, 223, 227, 229, 241, 245.)

*Cluster* 4 = (12, 18, 19, 22, 27, 28, 29, 36, 40, 42, 43, 51, 53, 65, 76, 83, 92, 105, 110, 113, 127, 133, 138, 139, 140, 147, 149, 160, 167, 172, 178, 202, 206, 208, 209, 214, 222, 226, 230, 232, 236, 237, 240, 244, 246, 247, 251, 252, 253, 254, 256, 257, 258, 260, 262, 263, 264, 271, 277, 278, 279, 280, 281, 282, 286, 288, 289, 291, 298, 301, 302, 307, 308, 310, 313, 314, 319, 327, 335, 336, 345.)

*Cluster* 5 = (2, 11, 30, 31, 38, 49, 58, 59, 64, 71, 96, 108, 109, 114, 116, 126, 128, 142, 144, 154, 162, 174, 191, 196, 197, 201, 203, 207, 219, 221, 239, 242, 249, 250, 255, 259, 261, 266, 268, 269, 270, 272, 273, 275, 276, 283, 284, 285, 287, 290, 292, 293, 294, 295, 296, 297, 299, 300, 303, 304, 305, 306, 309, 311, 312, 315, 316, 317, 318, 320, 324, 325, 326, 328, 330, 331, 334, 337, 338, 340, 341, 342, 343, 344, 346, 347, 348, 349.)

Pada langkah ini dihitung pula rasio antara besaran BCV (*Between Cluster Variation*) dengan WCV (*within Cluster Variation*): Karena *Centroid* m1, m2, m3, m4, dan m5 maka:

*Centroid* m1 = (4,1,1,10,70,5,1,3)

*Centroid* m2 = (3,1,1,6,20,5,14,2)

*Centroid* m3 = (3,2,4,6,40,5,9,2)

*Centroid* m4 = (4,2,3,2,100,5,8,4)

*Centroid* m5 = (4,2,2,6,78,5,1,2)

Cara menghitung BCV (*Between Cluster Variation*)

$$D(m1,m2) = \sqrt{\frac{(4-3)^2 + (1-1)^2 + (1-1)^2 + (10-6)^2 + (70-20)^2 + (5-5)^2 + (1-14)^2 + (3-2)^2}{}} = 51,83628073$$

$$D(m1,m3) = \sqrt{\frac{(4-3)^2 + (1-2)^2 + (1-4)^2 + (10-6)^2 + (70-40)^2 + (5-5)^2 + (1-9)^2 + (3-2)^2}{}} = 31,4960315$$

$$D(m1,m4) = \sqrt{\frac{(4-4)^2 + (1-2)^2 + (1-3)^2 + (10-2)^2 + (70-100)^2 + (5-5)^2 + (1-8)^2 + (3-4)^2}{}} = 31,9217794$$

$$D(m1,m5) = \sqrt{\frac{(4-4)^2 + (1-2)^2 + (1-2)^2 + (10-6)^2 + (70-78)^2 + (5-5)^2 + (1-1)^2 + (3-2)^2}{}} = 9,110433579$$

$$D(m2,m3) = \sqrt{\frac{(3-2)^2 + (1-2)^2 + (1-4)^2 + (6-6)^2 + (20-40)^2 + (5-5)^2 + (14-9)^2 + (2-2)^2}{}} = 20,85665361$$

$$D(m2,m4) = \sqrt{\frac{(3-4)^2 + (1-2)^2 + (1-3)^2 + (6-2)^2 + (20-100)^2 + (5-5)^2 + (14-8)^2 + (2-4)^2}{}} = 80,38656604$$

$$D(m2,m5) = \sqrt{\frac{(3-4)^2 + (1-2)^2 + (1-2)^2 + (6-6)^2 + (20-78)^2 + (5-5)^2 + (14-1)^2 + (2-2)^2}{}} = 59,46427499$$

$$D(m3,m4) = \sqrt{\frac{(3-4)^2 + (2-2)^2 + (4-3)^2 + (6-2)^2 + (40-100)^2 + (5-5)^2 + (9-8)^2 + (2-4)^2}{}} = 60,19136151$$

$$D(m3,m5) = \sqrt{(3-4)^2 + (2-2)^2 + (4-2)^2 + (6-6)^2 + (40-78)^2 + (5-5)^2 + (9-1)^2 + (2-2)^2} = 38,89730068$$

Hasil perhitungan WCV (Within Cluster Variation) dengan memilih jarak terkecil antara data dengan *Centroid* pada masing-masing *Cluster* .

$$WCV = 02 + 10,246950772 + 13,63818172 + 13,928388282 + 14,352700092 + 11,832159572 + 13,674794332 + 8,3066238632 + 82 + 02 + \dots + 02 = 53177$$

Sehingga besaran rasio =  $BCV/WCV = 407,6978866/53177 = 0,007666809$

Karena langkah ini merupakan iterasi 1 maka lanjutkan ke langkah berikutnya.

Langkah 4: Pembaruan *Centroid* dengan menghitung nilai rata-rata masing-masing *Cluster* dari pengelompokan *Cluster* tersebut. Hasil dari pemisahan *Cluster* terbagi menjadi 5 *Cluster* , *Cluster* pertama terdiri dari 68 data penduduk, *Cluster* kedua terdiri dari 38 data penduduk, *Cluster* ketiga terdiri dari 74 data penduduk, *Cluster* keempat terdiri dari 81 data penduduk dan *Cluster* kelima terdiri dari 88 data penduduk.

Sehingga di dapatkan *Centroid* baru yaitu :

$$Centroid\ m1 = (3.117647059, 1.264705882, 1.602941176, 4.632352941, 65.72058824, 4.735294118, 2.220588235, 3.073529412)$$

$$Centroid\ m2 = (3,342105263, 1,631578947, 2,842105263, 4,552631579, 16,10526316, 4,421052632, 2,973684211, 2,815789474)$$

$$Centroid\ m3 = (3,094594595, 1,189189189, 1,72972973, 3,945945946, 42,40540541, 4,378378378, 4,148648649, 3,067567568)$$

$$Centroid\ m4 = (3,24691358, 1,185185185, 1,456790123, 3,419753086, 102,8518519, 4,802469136, 5,37037037, 3,679012346)$$

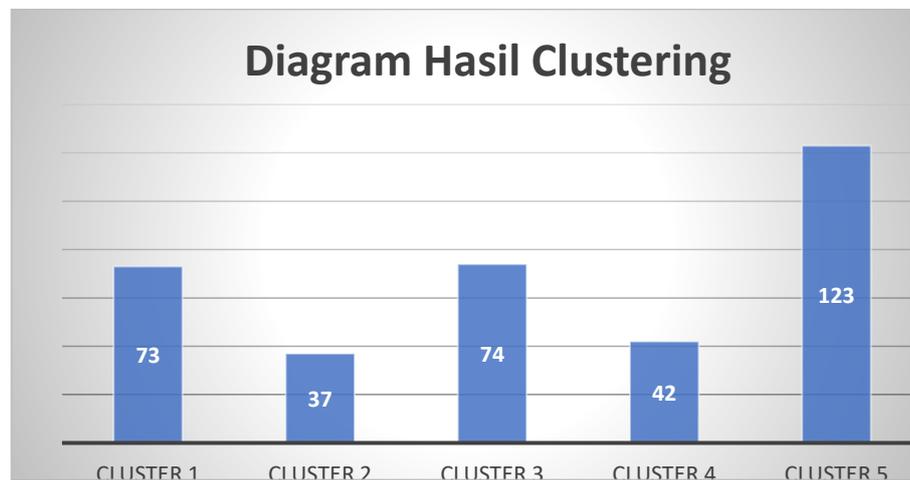
$$Centroid\ m5 = (3,022727273, 1,340909091, 1,727272727, 4,727272727, 82,20454545, 4,920454545, 2,795454545, 2,943181818)$$

Langkah 5: (iterasi 2) kembali kelangkah 3, Jika data *Cluster* masih berpindah, nilai *Centroid* masih di atas ambang, atau nilai fungsi objektif yang akan digunakan masih di atas ambang, langkah ini dilanjutkan dengan penempatan data kembali ke *Centroid* terdekat, seperti yang dilakukan pada langkah 3.

Proses perhitungan dilakukan sebanyak 5 iterasi. Jika dibandingkan, rasio saat ini (0,015826513) lebih kecil dibandingkan rasio sebelumnya (0,015839612), sehingga iterasi dihentikan.

### 3.3 Kesimpulan Perhitungan Manual

Kesimpulan perhitungan penentuan penerima bantuan RASKIN penduduk Desa Siulak Panjang, Kecamatan Siulak, Kabupaten Kerinci, dengan cara manual dan hasil rekomendasi bantuan dikelompokkan menjadi 5 *Cluster* seperti terlihat pada gambar 3.1 :



Gambar 3.1 Diagram Hasil CLUSTERING

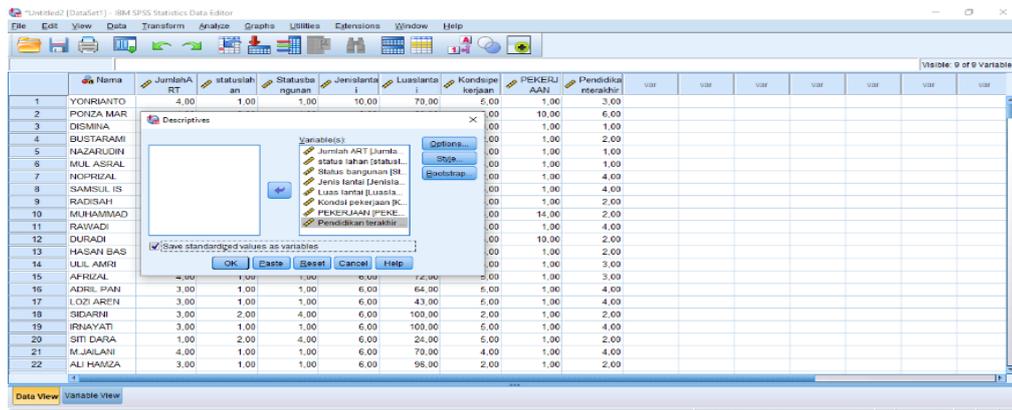
Kesimpulan perhitungan penentuan penerima bantuan RASKIN dengan cara manual dan hasil rekomendasi dikelompokkan menjadi 5 *Cluster* yaitu:

1. *Cluster* 1 terdiri dari 73 data penduduk yang akan menjadi prioritas ketiga dalam menerima bantuan RASKIN. Berdasarkan analisis variabel Luas Lantai, Jumlah Anggota Keluarga, Pendidikan Terakhir, Pekerjaan, dan Status Lahan, *Cluster* ini menunjukkan bahwa terdapat 0 penduduk yang "Sangat Layak", 0 penduduk "Layak", 73 penduduk "Cukup Layak", 0 penduduk "Kurang Layak", dan 0 penduduk "Tidak Layak" untuk menerima bantuan RASKIN
2. *Cluster* 2 terdiri dari 37 data penduduk yang akan menjadi prioritas pertama dalam menerima bantuan RASKIN. Berdasarkan analisis variabel Luas Lantai, Jumlah Anggota Keluarga, Pendidikan Terakhir, Pekerjaan, dan Status Lahan, *Cluster* ini menunjukkan bahwa terdapat 37 penduduk yang "Sangat Layak", 0 penduduk "Layak", 0 penduduk "Cukup Layak", 0 penduduk "Kurang Layak", dan 0 penduduk "Tidak Layak" untuk menerima bantuan RASKIN
3. *Cluster* 3 terdiri dari 74 data penduduk yang akan menjadi prioritas kedua dalam menerima bantuan RASKIN. Berdasarkan analisis variabel Luas Lantai, Jumlah Anggota Keluarga, Pendidikan Terakhir, Pekerjaan, dan Status Lahan,

- Cluster* ini menunjukkan bahwa terdapat 0 penduduk yang "Sangat Layak", 74 penduduk "Layak", 0 penduduk "Cukup Layak", 0 penduduk "Kurang Layak", dan 0 penduduk "Tidak Layak" untuk menerima bantuan RASKIN
- Cluster* 4 terdiri dari 42 data penduduk yang akan menjadi prioritas kelima dalam menerima bantuan RASKIN. Berdasarkan analisis variabel Luas Lantai, Jumlah Anggota Keluarga, Pendidikan Terakhir, Pekerjaan, dan Status Lahan, *Cluster* ini menunjukkan bahwa terdapat 0 penduduk yang "Sangat Layak", 0 penduduk "Layak", 0 penduduk "Cukup Layak", 0 penduduk "Kurang Layak", dan 42 penduduk "Tidak Layak" untuk menerima bantuan RASKIN
  - Cluster* 5 terdiri dari 123 data penduduk yang akan menjadi prioritas keempat dalam menerima bantuan RASKIN. Berdasarkan analisis variabel Luas Lantai, Jumlah Anggota Keluarga, Pendidikan Terakhir, Pekerjaan, dan Status Lahan, *Cluster* ini menunjukkan bahwa terdapat 0 penduduk yang "Sangat Layak", 0 penduduk "Layak", 0 penduduk "Cukup Layak", 123 penduduk "Kurang Layak", dan 0 penduduk "Tidak Layak" untuk menerima bantuan RASKIN.

### 3.4 Proses perhitungan data menggunakan SPSS

#### 3.4.1 Standardisasi/Transformasi Variabel



Gambar 3.2 Standardisasi/Transformasi Variabel

Gambar 3.2 menunjukkan proses transformasi variabel menjadi bentuk Z-Score. Langkah pertama adalah memasukkan seluruh variabel penilaian ke dalam kotak VARIABLES, yaitu Jumlah Keluarga, Status Bangunan, Penerang Rumah, Kondisi Pekerjaan, Pekerjaan Utama, Jamsos Pekerjaan, dan Pendapatan Pertahun. Selanjutnya, aktifkan opsi "Save standardized values as variables". Abaikan bagian lainnya, lalu tekan OK untuk menampilkan *output* dari aplikasi program SPSS.

#### 3.4.2 Output Deskripsi Dari Keseluruhan Variabel

*Output* yang didapat dari proses standardisasi/transformasi variabel yang dapat dilihat sebagai berikut :

Tabel 3.4 Output Deskripsi Dari Keseluruhan Variabel

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
Jumlah ART	349	1,00	6,00	3,1433	1,16070
status lahan	349	1,00	4,00	1,4011	,54096
Status bangunan	349	1,00	6,00	1,8911	1,25929
Jenis lantai	349	1,00	10,00	4,5931	2,01019
Luas lantai	349	5,00	150,00	68,1490	29,19712
Kondisi pekerjaan	349	1,00	5,00	4,6877	,85258
PEKERJAAN	349	1,00	14,00	3,5874	4,03668
Pendidikan terakhir	349	1,00	8,00	3,1519	1,61792
Valid N (listwise)	349				

Tabel 3.4 menunjukkan *output* dari proses standardisasi/transformasi variabel, yang mencakup penjelasan tentang semua variabel, termasuk nilai maksimum, nilai minimum, rata-rata, dan standar deviasi. Penjelasan ini digunakan sebagai dasar untuk menghitung Z-Score yang dihasilkan.

### 3.5 Hasil Z-Score Variabel

Selanjutnya buka tampilan "data view" dari tabel data. Kemudian di dapatkan hasil Z-Score dapat dilihat sebagai berikut :

ika	ZJumlahART	Zstatuslahan	ZStatusbangunan	Zjenislantai	ZLuaslantai	ZKondisipekerjaan	ZPEKERJAAN	ZPendidikanterakhir	YBR
1 .00	,73812	-,74154	-,70763	2,69973	,06340	-,36633	-,64097	-,09396	1,76037
2 .00	,73812	1,10701	1,67466	,69987	,40590	-,36633	1,58858	-,133002	1,76037
3 .00	-,98498	1,10701	1,67466	-,128999	-,27910	-,197950	-,64097	-,133002	1,76037
4 .00	-,12343	-,74154	-,70763	,69987	-,178610	-,3,15242	-,64097	-,71194	1,76037
5 .00	-,98498	1,10701	1,67466	,69987	-,1,82035	-,36633	-,64097	-,133002	1,76037
6 .00	-,98498	-,74154	-,70763	,69987	-,69010	-,36633	-,64097	-,133002	1,76037
7 .00	-,12343	-,74154	-,70763	,69987	-,62160	-,3,15242	-,64097	-,52421	1,76037
8 .00	-,98498	-,74154	-,70763	-,1,28999	,06340	-,36633	-,64097	-,52421	1,76037
9 .00	1,59966	1,10701	1,67466	,69987	-,14210	-,36633	-,64097	-,71194	1,76037
10 .00	-,12343	-,74154	-,70763	,69987	-,1,64910	-,36633	2,57950	-,71194	1,76037
11 .00	,73812	-,74154	-,70763	-,1,28999	,40590	-,36633	-,64097	-,52421	1,76037
12 .00	,73812	-,74154	-,70763	,69987	,74840	-,36633	1,58858	-,71194	1,76037
13 .00	-,98498	1,10701	1,67466	,69987	-,1,99160	-,36633	-,64097	-,71194	1,76037
14 .00	,73812	-,74154	-,70763	,69987	-,1,64910	-,36633	-,64097	-,09396	1,76037
15 .00	,73812	-,74154	-,70763	,69987	,13190	-,36633	-,64097	-,09396	1,76037
16 .00	-,12343	-,74154	-,70763	,69987	-,14210	-,36633	-,64097	-,52421	1,76037
17 .00	-,12343	-,74154	-,70763	,69987	-,86135	-,36633	-,64097	-,52421	1,76037
18 .00	-,12343	1,10701	1,67466	,69987	1,99090	-,3,15242	-,64097	-,71194	1,76037
19 .00	-,12343	-,74154	-,70763	,69987	1,99090	-,36633	-,64097	-,52421	1,76037
20 .00	-,184652	1,10701	1,67466	,69987	-,1,91910	-,36633	-,64097	-,71194	1,76037
21 .00	,73812	-,74154	-,70763	,69987	,06340	-,80659	-,64097	-,52421	1,76037
22 .00	-,12343	-,74154	-,70763	,69987	,95390	-,3,15242	-,64097	-,71194	1,76037

Gambar 3.2 Z-Score Variabel

Gambar 3.3 menunjukkan hasil Z-Score yang akan digunakan sebagai dasar untuk analisis Cluster. Namun, jika data yang dikumpulkan tidak memiliki variabilitas satuan, analisis Cluster dapat dilakukan secara instan tanpa melakukan transformasi atau standarisasi sebelumnya.

### 3.6 HASIL ANALISIS CLUSTER MENGGUNAKAN SPSS

Pada langkah selanjutnya adalah menginterpretasikan hasil Cluster menggunakan SPSS dapat dilihat sebagai berikut:

#### 3.7 Initial Cluster

Tabel 3. 5 Initial Cluster

	Initial Cluster Centers				
	Cluster				
	1	2	3	4	5
Zscore: Jumlah ART	2,46121	,73812	-,12343	,73812	-,184652
Zscore: status lahan	1,10701	1,10701	-,74154	-,74154	1,10701
Zscore: Status bangunan	,08646	1,67466	-,70763	-,70763	1,67466
Zscore: Jenis lantai	,69987	,69987	,69987	-,1,28999	-,1,28999
Zscore: Luas lantai	-,1,23810	-,1,92310	-,1,64910	2,80339	,74840
Zscore: Kondsi pekerjaan	-,4,32533	,36633	,36633	,36633	,36633
Zscore: Pekerjaan	-,64097	-,64097	2,57950	-,64097	-,64097
Zscore: Pendidikan terakhir	-,71194	2,99652	-,71194	2,99652	-,1,33002

#### 3.8 Iterasi Analisis Cluster

Agar kita dapat mengetahui berapa kali dilakukan proses iterasi yang dilakukan pada objek sebanyak 349 sampel, dapat anda lihat di bawah ini:

Tabel 3.6 Iterasi Analisis Cluster Iteration History<sup>a</sup>

Iteration	Change in Cluster Centers				
	1	2	3	4	5
1	2,812	2,512	2,706	2,804	2,422
2	,408	,543	,458	,482	,417
3	,238	,193	,228	,098	,164
4	,000	,116	,110	,086	,163
5	,000	,070	,163	,093	,364
6	,082	,000	,164	,109	,493
7	,000	,089	,049	,085	,147
8	,000	,122	,055	,091	,170
9	,000	,048	,064	,109	,063
10	,000	,032	,104	,158	,084

a. Iterations stopped because the maximum number of iterations was performed. Iterations failed to converge. The maximum absolute coordinate change for any center is ,093. The current iteration is 10. The minimum distance between initial centers is 5,703.

Dari tabel 3.6, terlihat bahwa iterasi dilakukan sebanyak 10 kali untuk menentukan Cluster yang optimal. Jarak minimum antara pusat Cluster yang dihasilkan dari proses iterasi ini tercatat sebesar 5,703.

#### 3.9 Output Analisis Cluster Dengan SPSS

Hasil akhir dari proses CLUSTERING dapat anda lihat seperti di bawah ini:

Tabel 3. 7 Output Analisis Cluster Dengan SPSS

Final Cluster Centers	
	Cluster

	1	2	3	4	5
Zscore: Jumlah ART	,06503	,49739	-,14427	,39586	-,90214
Zscore: status lahan	-,27940	1,18857	-,72663	-,58960	1,17811
Zscore: Status bangunan	-,11206	1,11412	-,70763	-,62061	1,17071
Zscore: Jenis lantai	,32677	,27556	,09810	-,62216	,07804
Zscore: Luas lantai	-,59806	-,42013	,17885	,51756	-,23563
Zscore: Kondsi pekerjaan	-2,82253	,28008	,35687	,30206	,09565
Zscore: PEKERJAAN	-,25389	,24065	-,55906	1,29335	-,64097
Zscore: Pendidikan terakhir	-,30633	,40605	-,40789	,88829	-,61685

Dengan ketentuan berikut, data di atas masih terkait dengan proses standarisasi yang mengacu pada *Z-Score*, seperti yang ditunjukkan dalam *output* tabel 3.7:

- a. Nilai negatif (-) menunjukkan bahwa data berada di bawah rata-rata total.
- b. Nilai positif (+) menunjukkan bahwa data berada di atas rata-rata total.

Dari tabel 3.7 *output* analisis *Cluster* dengan *SPSS*, dengan ketentuan yang telah dijabarkan diatas pula, dapat didefinisikan sebagai berikut :

a. *Cluster 1*

Dalam *Cluster 1*, penduduk memiliki nilai di bawah rata-rata untuk Status Lahan, Status Bangunan, Luas Lantai, Kondisi Pekerjaan, Pekerjaan, dan Pendidikan Terakhir. Sebaliknya, Jumlah ART dan Jenis Lantai berada di atas rata-rata. Oleh karena itu, *Cluster 1* menjadi prioritas pertama untuk penerima bantuan, karena memiliki nilai variabel yang paling banyak di bawah rata-rata.

b. *Cluster 2*

Penduduk di *Cluster 2* menunjukkan nilai di atas rata-rata untuk Jumlah ART, Status Lahan, Status Bangunan, Jenis Lantai, Kondisi Pekerjaan, Pekerjaan, dan Pendidikan Terakhir. Luas Lantai adalah satu-satunya variabel di bawah rata-rata. *Cluster 2* termasuk prioritas kelima penerima bantuan karena sebagian besar variabelnya berada di atas rata-rata.

c. *Cluster 3*

Di *Cluster 3*, penduduk memiliki nilai di bawah rata-rata untuk Jumlah ART, Status Lahan, Status Bangunan, Pekerjaan, dan Pendidikan Terakhir. Sementara itu, Jenis Lantai, Luas Lantai, dan Kondisi Pekerjaan berada di atas rata-rata. Dengan demikian, *Cluster 3* termasuk prioritas kedua penerima bantuan, karena memiliki nilai variabel yang cukup banyak di bawah rata-rata.

d. *Cluster 4*

Penduduk dalam *Cluster 4* memiliki nilai di bawah rata-rata untuk Status Lahan, Status Bangunan, dan Jenis Lantai. Sebaliknya, Jumlah ART, Luas Lantai, Kondisi Pekerjaan, Pekerjaan, dan Pendidikan Terakhir berada di atas rata-rata. *Cluster 4* menjadi prioritas keempat untuk penerima bantuan karena banyak variabelnya berada di atas rata-rata.

e. *Cluster 5*

Di *Cluster 5*, penduduk menunjukkan nilai di bawah rata-rata untuk Jumlah ART, Luas Lantai, Pendidikan Terakhir, dan Pekerjaan. Sedangkan Status Lahan, Status Bangunan, Jenis Lantai, dan Kondisi Pekerjaan berada di atas rata-rata. Oleh karena itu, *Cluster 5* masuk dalam prioritas ketiga untuk penerima bantuan, karena memiliki nilai variabel yang cukup banyak di bawah rata-rata.

**3.10 Perbedaan Variabel Pada Cluster Yang Terbentuk**

Tahapan berikutnya adalah menganalisis perbedaan variabel pada *Cluster* yang terbentuk, seperti yang dapat dilihat pada gambar berikut

**Tabel 3.8 Perbedaan Variabel Pada Cluster yang Terbentuk**

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Zscore: Jumlah ART	18,325	4	,799	344	22,948	,000
Zscore: status lahan	65,396	4	,251	344	260,318	,000
Zscore: Status bangunan	61,572	4	,296	344	208,238	,000
Zscore: Jenis lantai	9,587	4	,900	344	10,650	,000
Zscore: Luas lantai	12,464	4	,867	344	14,381	,000
Zscore: Kondsi pekerjaan	70,799	4	,188	344	375,829	,000
Zscore: PEKERJAAN	47,058	4	,464	344	101,321	,000
Zscore: Pendidikan terakhir	28,058	4	,685	344	40,938	,000

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

Dari Tabel 3.8, nilai F dan nilai probabilitas (Sig) masing-masing variabel digunakan untuk menilai perbedaan variabel pada klaster yang terbentuk. "*MS Between*" ditunjukkan oleh "*Means Square*" pada kolom *Cluster*, sedangkan "*MS Within*" ditunjukkan oleh "*Means Square*" pada kolom *Error*. Rumus uji F Anova adalah sebagai berikut:

$$F = \frac{MS\ Between}{MS\ Within}$$

Semakin besar nilai F dan ( $\text{sig} < 0,05$ ), maka semakin signifikan perbedaan variabel pada *Cluster* yang terbentuk. Hasil penelitian menunjukkan bahwa instrumen "Kondisi Pekerjaan" adalah variabel yang paling menonjol dalam membedakan penduduk di lima *Cluster*, dengan nilai  $F = 375,829$  dan  $\text{sig} = 0,000$ .

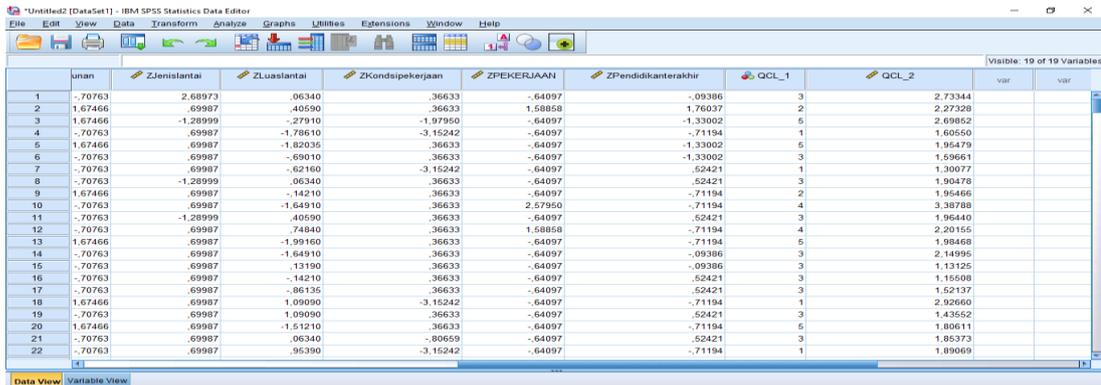
### 3.11 Jumlah Anggota Masing-Masing Cluster

Selanjutnya untuk mengetahui jumlah anggota masing-masing *Cluster* yang terbentuk dapat dilihat pada tabel *output* berikut ini :

Tabel 3. 9 Jumlah Anggota Masing-Masing Cluster

Number of Cases in each Cluster		
Cluster	1	32,000
	2	68,000
	3	124,000
	4	73,000
	5	52,000
Valid		349,000
Missing		,000

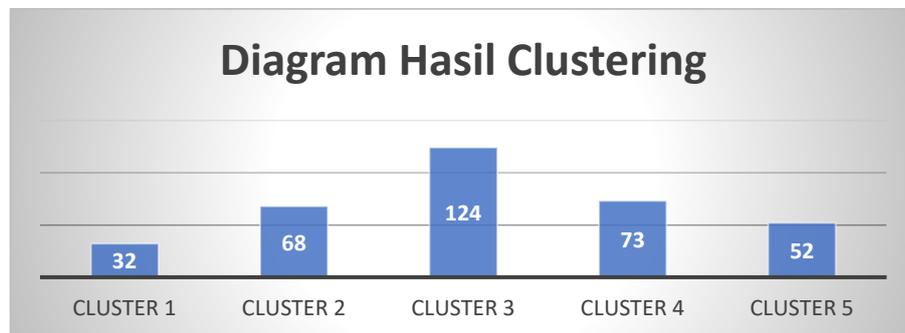
Dari Tabel 3.9 terlihat bahwa *Cluster-1* memiliki 32 penduduk, *Cluster-2* memiliki 68 penduduk, *Cluster -3* memiliki 124 penduduk, *Cluster -4* memiliki 73 penduduk, dan *Cluster -5* memiliki 52 penduduk, masing-masing. Jumlah penduduk masing-masing *Cluster* dapat dilihat pada kolom terakhir dari tampilan "data view" sebagai berikut:



Gambar 3. 3 Hasil Pengelompokkan Cluster

Dua kolom terakhir pada gambar 3.4 menunjukkan informasi berikut: "QCL\_1" menunjukkan nomor hasil pengelompokan penduduk *Cluster*, dan "QCL\_2" menunjukkan jarak antara objek dan pusat pengelompokan penduduk.

### 3.12 Kesimpulan Perhitungan tools SPSS



Gambar 3.5 Hasil CLUSTERING Menggunakan SPSS

Sebanyak 349 data penduduk di Desa Siulak Panjang, Kecamatan Siulak, Kabupaten Kerinci terbagi menjadi 5 *Cluster*, menurut hasil *Cluster* yang dilakukan menggunakan alat SPSS.

- Cluster 1*  
Karena banyak variabelnya di bawah rata-rata, data dari *Cluster -1* yang terdiri dari 32 penduduk adalah yang paling penting bagi penerima bantuan.
- Cluster 2*  
*Cluster-2* memiliki 68 data penduduk, yang menjadi prioritas kelima penerima bantuan karena banyak variabelnya di atas rata-rata.
- Cluster 3*

Data dari 124 orang di *Cluster -3* menjadi prioritas kedua penerima bantuan karena banyak variabelnya di bawah rata-rata.

d. *Cluster 4*

*Cluster-4* mengandung 73 data penduduk, yang menjadi prioritas keempat penerima bantuan karena memiliki sejumlah variabel yang jauh di atas rata-rata.

e. *Cluster 5*

*Cluster-5* memiliki 52 data penduduk, yang menjadi prioritas ketiga penerima bantuan karena banyak variabelnya di bawah rata-rata.

## 4. KESIMPULAN

Penelitian ini menggunakan data penduduk Desa Siulak Panjang, Kecamatan Siulak Mukai, Kabupaten Kerinci. Setelah pembersihan data, dari total 355 data penduduk, hanya 349 yang memenuhi kriteria untuk dianalisis. Kriteria penerima bantuan RASKIN meliputi jumlah anggota rumah tangga (ART), luas lantai kurang dari 10 meter persegi, tingkat pendidikan rendah, kondisi ekonomi tidak stabil, dan kondisi kesehatan yang kurang baik. Penulis menggunakan delapan atribut dalam perhitungan: jumlah ART, status lahan, status bangunan, jenis lantai, luas lantai, kondisi pekerjaan, pekerjaan, dan pendidikan terakhir. Penelitian ini menganalisis penentuan penerima bantuan RASKIN menggunakan metode *K-Means CLUSTERING*. Penulis merekomendasikan lima klaster, yang dihitung secara manual dan menggunakan alat SPSS. Hasil perbandingan menunjukkan perbedaan antara perhitungan manual dan SPSS. Perhitungan manual tidak melakukan standarisasi data, sedangkan SPSS melakukan standarisasi data sebelum pengolahan. Hasil klasterisasi menunjukkan perbedaan jumlah penduduk yang masuk dalam prioritas penerima bantuan di setiap klaster antara perhitungan manual dan SPSS. Berdasarkan hasil perbandingan perhitungan manual dan SPSS dalam menentukan penerima bantuan RASKIN menggunakan metode *K-Means CLUSTERING*, dapat disimpulkan bahwa penggunaan SPSS lebih direkomendasikan. SPSS memberikan hasil yang lebih akurat dan konsisten karena melakukan standarisasi data sebelum pengolahan. Selain itu, SPSS lebih efisien dalam mengolah data, menghasilkan klaster yang lebih representatif, dan lebih tepat sasaran. Oleh karena itu, SPSS adalah alat yang lebih efektif untuk menentukan prioritas penerima bantuan RASKIN di Kecamatan Siulak.

## REFERENCES

- [1] N. KARIM, "KEBIJAKAN BERORIENTASI KESEJAHTERAAN (KASUS DI PROVINSI JAMBI)," vol. 5, no. 2, pp. 82–92, 2016.
- [2] Program-program Kemensos Antisipasi Kemiskinan Akibat Pandemi, "No Title," *Kementerian Sosial RI*, 2020. <https://kemensos.go.id/program-program-kemensos-antisipasi-kemiskinan-akibat-pandemi>
- [3] penerima B. K. B. O. Mampu, "No Title," *BPK RI*, 2021. <https://jambi.bpk.go.id/penerima-bansos-kerinci-banyak-orang-mampu/>
- [4] H. Hasanah, P. Riswanto, and ..., "Penerapan *Data mining* Menggunakan *K-Means* Untuk Penentuan Penerima Bantuan Langsung Tunai (BLT) Dana Desa Pada Desa Cempaka Timur," *JTKSI (Jurnal Teknol. ...)*, pp. 236–241, 2023, [Online]. Available: <https://jurnal.ftikomibn.ac.id/index.php/jtkksi/article/view/1508>
- [5] N. Nurahman and J. Susanto, "Klasterisasi Data Penerima Bantuan Langsung Tunai Menggunakan Algoritma *K-Means*," *JURIKOM (Jurnal Ris. Komputer)*, vol. 10, no. 2, p. 461, 2023, doi: 10.30865/jurikom.v10i2.5807.
- [6] M. Arhami, M. Kom, and S. T. Muhammad Nasir, *Data mining-Algoritma dan Implementasi*. Penerbit Andi, 2020.
- [7] and L. W. S. Amna, Wahyuddin S, I. Gede Iwan Sudipa, Tri Andi E. Putra, Ahmad Jurnaidi Wahidin, Wara Alfa Syukrilla, Anindya Khrisna Wardhani, Nono Heryana, Tutuk Indriyani, *Data mining Data mining. Vol. 2*. 2023.
- [8] M. K. Hidayat and R. Fitriana, "Implementasi *K-Means* dan *K-Medoids* dalam pengelompokan wilayah potensial produksi daging ayam," *J. Teknol. Ind. Pertan.*, vol. 32, no. 3, pp. 239–247, 2022.
- [9] B. Susarianto and T. Nizami, "Penentuan Penerima Bantuan Sosial Beras untuk Masyarakat Miskin Menggunakan Metode *K-Means*," *Jutisi J. Ilm. Tek. Inform. dan Sist. Inf.*, vol. 12, no. 2, pp. 772–783, 2023.
- [10] P. W. Rahayu *et al.*, *Buku Ajar Data mining*. PT. Sonpedia Publishing Indonesia, 2024.
- [11] D. Nofriansyah, S. Kom, and M. Kom, *Konsep Data mining vs sistem pendukung keputusan*. Deepublish, 2015.
- [12] I. Werdiningsih, M. Kom, B. Nuqoba, M. Kom, and S. S. Muhammadun, *Data mining Menggunakan Android, Weka, dan SPSS*. Airlangga University Press, 2020.
- [13] D. Jollyta, W. Ramdhan, and M. Zarlis, *Konsep Data mining dan penerapan*. Deepublish, 2020.
- [14] I. G. I. Sudipa *et al.*, *Metode Penelitian Bidang Ilmu Informatika (Teori & Referensi Berbasis Studi Kasus)*. PT. Sonpedia Publishing Indonesia, 2023.
- [15] P. T. I. U. P. Y. Nurirwan Saputra, "Pengenalan *Data mining*," 2023.
- [16] D. T. Larose and C. D. Larose, *Discovering knowledge in data: an introduction to Data mining*, vol. 4. John Wiley & Sons, 2014.
- [17] C. Gunawan, *Mahir menguasai SPSS panduan praktis mengolah data penelitian new edition buku untuk orang yang (merasa) tidak bisa dan tidak suka statistika*. Deepublish, 2020.
- [18] S. Santoso, *Panduan Lengkap SPSS 26*. Elex Media Komputindo, 2020.