

## Penerapan Data Mining untuk Mengcluster Data Penduduk Miskin Menggunakan Algoritma K-Means di Dusun Bagik Endep Sukamulia Timur

Suhartini<sup>1</sup>, Ria Yuliani<sup>2</sup>

<sup>1</sup>Program Studi Teknik Informatika, Universitas Hamzanwadi

<sup>2</sup>Program Studi Sistem Informasi Universitas Hamzanwadi

Suhartini311279@gmail.com

### Abstrak

Terjadinya kemiskinan pada masyarakat disebabkan oleh suatu kondisi ketidakmampuan kepala keluarga secara ekonomi untuk memenuhi kebutuhan primer/ pokok keluarganya yaitu kebutuhan sandang, pangan, papan serta pendidikan. Masyarakat miskin itu sendiri hampir terdapat pada setiap Negara, Kota maupun Daerah contohnya disalah satu Dusun Bagik Endep Desa Sukamulia Timur.[1] Berdasarkan kondisi tersebut perlu dilakukannya pengklasteran untuk membantu Pemerintah desa dalam pengelompokan keluarga miskin, sehingga bantuan dapat disalurkan dengan tepat. Dengan mengamati persoalan di atas maka diperlukan Data Mining untuk mengelompokkan penerima bantuan dengan menggunakan metode K-Means dalam pengklasteran penduduk miskin. Dimana metode Algoritme K-Means Clustering ini bertujuan untuk mengelompokkan data penduduk di wilayah Sukamulia Timur yang memang dikatakan tergolong penduduk miskin. Data yang digunakan adalah data Penduduk Sukamulia Timur pada tahun 2019 yang berjumlah 200 data dengan 9 atribut yaitu nama penduduk, pekerjaan, penghasilan/bulan, jumlah anak yang sekolah SD, jumlah anak yang sekolah SMP, jumlah anak yang sekolah SMA, jumlah anak yang kuliah, jumlah anak yang tidak sekolah serta jumlah anggota keluarga. Berdasarkan hasil pengujian yang dilakukan dengan menerapkan algoritma K-Means didapatkan hasil dengan Cluster 1 berjumlah 18 penduduk dengan kriteria Penduduk ekonomi tinggi, Cluster 2 berjumlah 72 Penduduk dengan kriteria Penduduk ekonomi sedang, dan Cluster 3 berjumlah 110 penduduk dengan kriteria Penduduk ekonomi rendah[2]. Metode K-Means ini diharapkan mampu membantu pemerintah Desa Sukamulia Timur dalam pengambilan keputusan dan menemukan informasi yang dibutuhkan untuk memecahkan masalah dalam mendata penduduk miskin dengan tepat.

**Kata Kunci :** Data Mining, K-Means Clustering, Kemiskinan

### Abstract

*The occurrence of poverty in the community is caused by a condition of the economic inability of the head of the family to meet the primary / basic needs of his family, namely the need for clothing, food, shelter and education. The poor community itself can be found in almost every country, city and region, for example in one of the Bagik Endep hamlets of East Sukamulia Village. Based on these conditions, it is necessary to carry out clustering to assist the village government in grouping poor families, so that assistance can be distributed appropriately. By observing the above problems, Data Mining is needed to classify aid recipients using the K-Means method in clustering the poor. Where the K-Means Clustering Algorithm method aims to classify population data in the East Sukamulia region who are said to be classified as poor. The data used is data on the population of East Sukamulia in 2019, amounting to 200 data with 9 attributes, namely the name of the population, occupation, income / month, the number of children attending elementary school, the number of children attending junior high school, the number of children attending high school, the number of children attending college , the number of children who are not in school and the number of family members. Based on the results of tests carried out by applying the K-Means algorithm, the results obtained are Cluster 1 totaling 18 residents with the criteria of high economic population, Cluster 2 totaling 72 residents with moderate economic population criteria, and Cluster 3 totaling 110 residents with low economic population criteria. The K-Means method is expected to be able to assist*

*the government of Sukamulia Timur Village in making decisions and finding the information needed to solve problems in recording the poor population accurately.*

**Keywords:** *Data Mining, K-Means Clustering, Proverty*

## 1. Pendahuluan

Bagi Pemerintah Indonesia masalah kemiskinan adalah salah satu masalah yang sudah lama dan belum bisa diselesaikan. Masalah kemiskinan ini terjadi hampir disetiap wilayah salah satunya wilayah Desa Sukamulia Timur. Kemiskinan dapat disebabkan oleh kelangkaan kebutuhan dasar terhadap pendidikan dan pekerjaan. Pemerintah sendiri telah melakukan beberapa upaya dalam mengatasi kemiskinan diantaranya melalui program bantuan sosial seperti tunjangan, Jaminan Kesehatan Masyarakat, Beras Miskin (Raskin), Program Keluarga Harapan (PKH), Bantuan Biaya Peningkatan Prestasi Akademik (BBP-PPA), Bantuan biaya pendidikan yang hanya ditujukan untuk calon Mahasiswa tidak mampu/miskin (BIDIKMISI) dan program-program lainnya, tetapi pada kenyataannya masalah kemiskinan belum dapat ditanggulangi.[3]

Tingkat kemiskinan ini juga dipengaruhi oleh salah satunya yaitu lapangan kerja yang kurang memadai atau pekerjaan yang dimiliki mempunyai penghasilan yang tidak sesuai dengan tanggungan hidup seperti tanggungan dalam keluarga, anak dan lain sebagainya. Selain lapangan pekerjaan yang kurang memadai, ternyata salah satu penyebab

masalah kemiskinan belum bisa diselesaikan adalah dari hasil survey yang dilakukan bantuan yang diberikan pemerintah kepada penduduk tidak sesuai dengan apa yang mereka butuhkan. Salah satu kesulitan yang terkadang dihadapi Pemerintah dalam proses penanganan kemiskinan adalah proses pembagian bantuan sosial yang tidak merata dan tidak tepat sasaran. Ini disebabkan karena validasi data sering diabaikan sehingga sering menimbulkan data yang tidak akurat. Dengan adanya permasalahan tersebut, penulis akan mencoba menggunakan metode yaitu dengan pendekatan metode data mining. Dengan pendekatan data mining, penulis juga menerapkan metode Algoritma *K-Means Clustering*. Dimana metode Algoritme *K-Means Clustering* ini bertujuan mengelompokkan data penduduk di wilayah Sukamulia Timur yang memang dikatakan tergolong penduduk miskin. Data yang digunakan adalah data Penduduk Sukamulia Timur pada tahun 2019. Dengan metode cluster ini, dapat membantu Kantor Desa Sukamulia Timur dalam pengambilan keputusan dan menemukan informasi yang dibutuhkan untuk memecahkan masalah dalam mendata penduduk miskin dengan tepat. maka peneliti tertarik untuk melakukan pengkajian

tentang data penduduk miskin dengan judul "Penerapan Data Mining untuk Clustering Data Penduduk Miskin Menggunakan Algoritma K-Means di Dusun Bagik Endep Sukamulia Timur".[4]

## 2. Tinjauan Pustaka

### 2.1. Penelitian Terkait

- Penelitian yang dilakukan oleh Riyani Wulan Sari, Anjar Wanto, Agus Perdana Windarto dalam jurnal yang berjudul "Implementasi Rapidminer Dengan Metode K-Means (Study Kasus: Imunisasi Campak Pada Balita Berdasarkan Provinsi)", Data yang digunakan dalam penelitian ini yaitu data dari tahun 2004 2017 yang terdiri dari 34 provinsi .[5]
- Selanjutnya penelitian dilakukan oleh Devi Sartika, Juju Jumadi, dalam jurnal yang berjudul "Clustering Penilaian Kinerja Dosen Menggunakan Algoritma K-Means (Studi Kasus: Universitas Dehasen Bengkulu), Permasalahan yang dihadapi dalam penelitian ini adalah kesulitannya pembuat kebijakan yang dalam hal ini adalah pihak akademik kampus Universitas Dehasen Bengkulu untuk mengelompokkan dosen yang melakukan kegiatan mengajar dengan kualitas sangat baik, baik, cukup baik dan kurang baik sehingga dibutuhkan suatu teknik klasterisasi data.[6]
- Penelitian berikutnya dilakukan oleh Windania Purba, Willy Siawin, Hardih, dalam jurnal yang berjudul "Implementasi Data Mining Untuk Pengelompokkan Dan Prediksi Karyawan Yang Berpotensi Phk Dengan Algoritma K-Means Clustering" .[3]
- Penelitian yang dilakukan oleh Asroni, Ronald Adrian, dalam jurnal yang berjudul "Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang" Penelitian ini menguji data yang telah ada di data warehouse Universitas Muhammadiyah Magelang untuk memudahkan untuk mencari 5 orang mahasiswa pada jurusan Teknik Informatika dalam melakukan penyeleksian untuk mengikuti lomba.[7]
- Penelitian yang dilakukan oleh Donni prima dalam " judul Pengelompokan Kecamatan-Kecamatan Di Jakarta Barat Berdasarkan Jumlah Penyandang Masalah Kesejahteraan Sosial Tahun 2013 Dengan Menggunakan Metode Algoritma *K-Means*" Peneliti melakukan penelitian dengan melihat gambaran secara umum masalah kesejahteraan di Jakarta Barat Masalah kesejahteraan yang tertinggi di kota Jakarta Barat adalah fakir miskin, lansia terlantar dan anak terlantar. Kecamatan yang

memiliki masalah kesejahteraan sosial terkomples selaras dengan jumlah penyandang terbanyak adalah kecamatan Kembangan. [8]

## 2.2. Landasan Teori

### 1. Penduduk Miskin

Kemiskinan adalah apabila pendapatan suatu komunitas berada dibawah garis kemiskinan tertentu. Kemiskinan juga berarti kekurangan kebutuhan social, termasuk keterkucilan social, ketergantungan dan ketidakmampuan berpartisipasi dalam kehidupan masyarakat yang layak. Secara arti sempitnya kemiskinan dipahami sebagai keadaan kekurangan uang dan barang untuk menjamin kelangsungan hidup yang seharusnya. Menurut Hasya Yanto: "kemiskinan adalah keadaan dimana seseorang yang tidak bisa memenuhi kebutuhan hidupnya". Kemiskinan adalah keadaan dimana terjadi ketidakmampuan untuk memenuhi kebutuhan dasar seperti makanan, pakaian, tempat berlindung dan pendidikan. Kemiskinan secara konseptual dibedakan menurut kemiskinan relative dan kemiskinan absolute. Dimana perbedaannya terletak pada standar penilaiannya. Standar penilaian kemiskinan relative merupakan standar kehidupan yang ditentukan dan ditetapkan secara subyektif oleh masyarakat setempat dan bersifat local serta mereka yang berada dibawah

standar penilaian tersebut dikategorikan sebagai miskin secara relatif. Sedangkan standar penilaian secara absolute merupakan standar kehidupan minimum yang dibutuhkan untuk memenuhi kebutuhan dasar yang diperlukan, baik makan maupun non makanan. Standar kehidupan minimum untuk memenuhi kebutuhan dasar ini disebut sebagai garis kemiskinan. Jadi dapat disimpulkan bahwa kemiskinan dimana seseorang tidak dapat memenuhi kebutuhan hidupnya seperti kebutuhan makanan, pakaian, tempat tinggal, dan pendidikan. [9]

Menyambung hal diatas, kemiskinan di Indonesia yang banyak dibicarakan ini tentunya mempunyai dasar. Oleh karena itu, kemiskinan yang dimaksud ini telah di dasari Undang-Undang RI nomor 13 tahun 2011 tentang penanganan Penduduk Miskin. Pasal 1(1) UU tersebut menyebutkan bahwa "Penduduk Miskin adalah orang yang sama sekali tidak mempunyai mata pencaharian dan/atau mempunyai sumber penghasilan tetapi tidak mempunyai kemampuan untuk memenuhi kebutuhan dasar yang layak bagi kehidupan dirinya dan keluarganya.

### 2. Data Mining

Data mining merupakan salah satu solusi untuk menjelaskan proses penambangan informasi dalam suatu basis data yang berskala besar. [10] Saat suatu organisasi baik itu

perusahaan maupun suatu institusi yang mempunyai data yang kompleks, tidak menutup kemungkinan banyak sekali informasi yang dapat diperoleh, serta bagaimana solusi data mining bisa diterapkan dengan berbagai teknik diantaranya yaitu classification, association dan clustering.[11]

a. Deskripsi

Para peneliti biasanya mencoba menemukan cara untuk mendeskripsi pola dan trend yang tersembunyi dalam data.

b. Estimasi

Estimasi mirip dengan klasifikasi, kecuali variabel tujuan yang lebih kearah numerik dari pada kategori.

c. Prediksi

Prediksi memiliki kemiripan dengan estimasi dan klasifikasi. Hanya saja, prediksi hasilnya menunjukkan sesuatu yang belum terjadi (mungkin terjadi dimasa depan).

d. Klasifikasi

Dalam klasifikasi variabel, tujuan bersifat kategorik. Misalnya, kita akan mengklasifikasikan pendapatan dalam tiga kelas, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

e. Clustering

Clustering lebih ke arah pengelompokan record, pengamatan, atau kasus dalam kelas yang memiliki kemiripan.

f. Asosiasi

Mengidentifikasi hubungan antara berbagai peristiwa yang terjadi pada satu waktu.

### 3. Algoritma K-Means

K-Means merupakan sebuah algoritma clustering pada data mining untuk dapat menghasilkan kelompok dari data yang jumlahnya banyak dengan metode partisi yang berbasis titik dengan waktu komputasi yang cepat dan efisien. Selain itu definisi K-Means merupakan metode clustering berbasis jarak yang membagi data ke dalam sejumlah cluster dan algoritma ini hanya bekerja pada atribut numerik. Dapat disimpulkan bahwa algoritma K-Means akan membuat data-data yang ingin dilakukan cluster nantinya akan menuju beberapa pusat/titik dimana titik tersebutlah yang digunakan sebagai acuan dalam pengelompokan dan data-data yang dilakukan klasterisasi harus numerik. Pada dasarnya Algoritma K-Means hanya mengambil sebagian dari banyaknya dari komponen yang didapatkan untuk kemudian dijadikan pusat *cluster awal*, pada penentuan pusat *cluster* ini dipilih secara acak dari populasi data.

### 4. K-Means Clustering

Data Clustering merupakan salah satu metode Data Mining yang bersifat tanpa arahan (unsupervised). K-Means merupakan salah satu metode data clustering non hirarki yang

berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster/ kelompok. Metode ini mempartisi data ke dalam cluster/ kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain.[8] Manfaat Clustering adalah sebagai Identifikasi Object (Recognition) misalnya dalam bidang Image Processing, Computer Vision atau robot vision. Selain itu adalah sebagai Sistem Pendukung Keputusan dan Data Mining seperti Segmentasi pasar, pemetaan wilayah, Manajemen marketing dan lain-lain.

Karakteristik K-means :

- K-means sangat cepat dalam proses clustering.
- K-means sangat sensitive pada pembangkitan centroid awal secara random.
- Memungkinkan suatu cluster tidak mempunyai anggota.
- Hasil clustering dengan K-means bersifat unik (selalu berubah-ubah,
- Terkadang baik, terkadang jelek).

### 2.3. Tahapan Penelitian

Adapun tahapan dari penelitian ini meliputi beberapa tahapan diantaranya:

1. Tahap pengumpulan informasi mengenai data-data penduduk Dusun Bagik endep Sukamulia Timur.
2. Tahap Analisis  
Pada tahap analisis ini sistem meliputi: pengumpulan fakta sebagai dasar untuk mempelajari suatu sistem yang ada seperti : kondisi prekonomian penduduk serta masalah yang dihadapi.
3. Tahap Pemahaman Data  
Data penduduk yang penulis gunakan ini diambil dari data Penduduk Dusun Bagik Endep Sukamulia Timur. Didalam data ini terdapat 200 penduduk, dan dalam data ini ada beberapa data penduduk yang dijadikan sebagai atribut yaitu : Pekerjaan, Penghasilan, jumlah anak yang masih dalam jenjang pendidikan dan jumlah anggota keluarga.
4. Tahap Persiapan Data  
Pada tahap ini data sebanyak 200 dan terdiri dari 9 atribut yang akan dilakukan beberapa penyelesaian untuk menghasilkan data yang dibutuhkan. Tahapannya yaitu :
  - a. Pembersihan Data : untuk membersihkan nilai yang kosong
  - b. Pengurangan Data : jumlah atribut data yang digunakan tidak semua yang terdapat dalam database, oleh karena itu atribut yang tidak diperlukan atau tidak digunakan akan dihapus[12]

### 3. Metode Penelitian

Untuk mendapatkan data dan informasi yang dibutuhkan ada beberapa cara yang penulis lakukan untuk mendapatkan data dan informasi yang benar dan jelas yaitu sebagai berikut:

- a. Observasi : Dalam metode ini peneliti mengumpulkan data dengan melakukan observasi / pengamatan ke Kantor Desa Sukamulia Timur..
- b. Wawancara : Wawancara atau tanya jawab dilakukan dengan staf Kantor Desa Sukamulia Timur untuk mengetahui gambaran tentang kondisi prekonomian Penduduk.
- c. Studi Pustaka : Metode ini dilakukan untuk menunjang metode wawancara dan observasi yang telah dilakukan. Dalam pengumpulan informasi yang dibutuhkan perlu mencari referensi-referensi yang berhubungan dengan penelitian referensi yang dapat diperoleh dari buku maupun internet.

### 4. Hasil dan Pembahasan

#### 4.1 Tahap Clustering dengan Metode K-Means

Tahapan melakukan clustering atau pengelompokan dengan metode K-Means adalah sebagai berikut [13]

- a. Tentukan jumlah cluster k.
- b. Pilih pusat cluster secara acak. Pemilihan pusat cluster ini bisa dilakukan dengan berbagai cara, yang paling sering dilakukan adalah dengan cara random (acak). Namun pada penelitian ini pusat cluster ditentukan dari yang tertinggi ke terendah.
- c. Jarak antara data dan pusat *cluster* dihitung menggunakan Euclidean Distance. Untuk menghitung semua jarak data ke setiap titik pusat *cluster* dapat menggunakan teori jarak Euclidean yang dirumuskan sebagai berikut :

$$d(x_j, c_j) = \sqrt{\sum_{j=1}^n (x_j - c_j)^2}$$

Dimana :

d = jarak

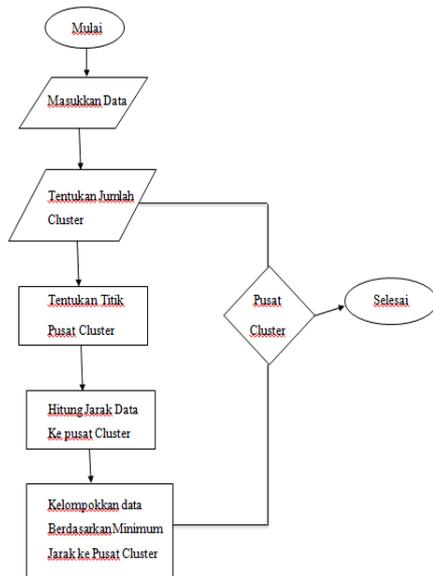
$x_j$  = Data ke-j

$c_j$  = centroid ke-j

- d. Data ditempatkan dalam *cluster* yang terdekat, dihitung dari tengah *cluster*.
- e. Pusat *cluster* baru akan ditentukan bila semua data telah ditetapkan dalam *cluster* terdekat.
- f. Proses penentuan pusat *cluster* dan penempatan data dalam *cluster* dilulangi sampai nilai *centroid* tidak berubah lagi.

Adapun alur proses data yang akan diterapkan menggunakan algoritma *K-Means* untuk

pendataan Penduduk Miskin berdasarkan data penduduk 2019 dapat dilihat pada diagram alur proses di bawah :



Gambar 1 Diagram alur proses

#### 4.2 Cluster Model (Clustering)

Gambar 1 merupakan sheet untuk menampilkan database yang telah diolah secara keseluruhan lengkap dengan clusternya. Hasil dari data 200 dimana C0 memiliki 18 items, C1 memiliki 72 items, C2 memiliki 110 items.[14]

### Cluster Model

Cluster 0: 18 items  
 Cluster 1: 72 items  
 Cluster 2: 110 items  
 Total number of items: 200

Gambar 2 Cluster Model

#### 4.3 Pengelompokkan Data/ Iterasi

1. Iterasi Pertama

Pada iterasi pertama nilai centroid awal pada penelitian ini dilakukan pemilihan dari tertinggi-Terendah dengan jumlah centroid awal dilakukan sebanyak tiga centroid awal, yaitu:

No	Nama	Pekerjaan	Penghasilan/ Bulan	Jumlah Anak SD	Jumlah Anak SMP	Jumlah Anak SMA	Jumlah Anak Kuliah	Jumlah Anak tidak Sekolah	Jumlah ART
C1	Mansur	PNS	Rp. 4.000.000	0	0	0	2	0	4
C2	Azmi	Buruh H.L	Rp. 2.000.000	0	0	2	1	0	5
C3	Hairudin	Buruh H.L	Rp. 500.000	0	0	1	0	0	3

Gambar 3 Centroid Awal/Pertama

Dibawah ini merupakan hasil iterasi/percobaan yang telah dilakukan pada Microsoft Exel :

3000000	1000000	500000	500000			1	
2500000	500000	1000000	500000		1		
2500000	500000	1000000	500000		1		
3000000	1000000	500000	500000			1	
2000000	3.741657387	1500000	3.741657387		1		
2500000	500000	1000000	500000		1		
3000000	1000000	500000	500000			1	
2000000	2.449489743	1500000	2.449489743		1		
2500000	500000	1000000	500000		1		
2000000	2.828427125	1500000	2.828427125		1		
2500000	500000	1000000	500000		1		
3000000	1000000	500000	500000			1	
3000000	1000000	500000	500000			1	
3000000	1000000	500000	500000			1	
2000000	3.16227766	1500000	3.16227766		1		
3500000	1500000	1.732050808	1.732050808			1	
3000000	1000000	500000	500000			1	
2000000	3.16227766	1500000	3.16227766		1		
1000000	1000000	2500000	1000000	1			
3000000	1000000	500000	500000			1	
3300000	1300000	200000	200000			1	
3500000	1500000	1.732050808	1.732050808			1	
3000000	1000000	500000	500000			1	
					18	72	110

Gambar 4 Iterasi Pertama

Berdasarkan proses iterasi 1 terlihat pada titik pusat cluster yang diproses sama, untuk melakukan perhitungan kembali untuk membuktikan apakah hasil iterasi 1 akan tetap sama sampai iterasi seterusnya.

2. Iterasi ke dua

Iterasi ini dilakukan untuk membuktikan apakah hasil yang diperoleh pada iterasi pertama akan tetap sama atau akan berbeda.

No	Penghasilan / Bulan	Jumlah Anak SD	Jumlah Anak SMP	Jumlah Anak SMA	Jumlah Anak Kuliah	Jumlah Anak tidak Sekolah	Jumlah ART
C1	3988888889	0.722222222	0.277777778	0.2777778	0.166666667	3.888888889	3.444444444
C2	1770833333	20834.06944	20833.63089	3.0416667	0.097222222	0.513888889	3.916666667
C3	8404545455	0.7	0.272727273	0.3272727	0.018181818	0.418181818	3.554545455

Gambar 5 Centroid Kedua

Berikut ini merupakan hal perbandingan jarak antara data dengan centroid setiap cluster yang ada. Perhitungan dilakukan terus sampai data ke-200 terhadap pusat cluster. Setelah dilakukan proses perhitungan maka akan didapatkan hasil data sebagai berikut :

298888.889	771396.219	159545.4545	159545.4545			1
198888.889	231052.5772	1159545.455	231052.5772		1	
248888.889	272431.2636	659545.4545	272431.2636		1	
298888.889	771396.219	159545.4546	159545.4546			1
198888.889	231052.7575	1159545.455	231052.7575		1	
248888.889	272431.1872	659545.4545	272431.1872		1	
198888.889	231052.8477	1159545.455	231052.8477		1	
248888.889	272431.1107	659545.4545	272431.1107		1	
298888.889	771396.192	159545.4545	159545.4545			1
298888.889	771396.192	159545.4546	159545.4546			1
298888.889	771396.138	159545.4546	159545.4546			1
198888.889	231052.9378	1159545.455	231052.9378		1	
348888.889	1271174.818	340454.5455	340454.5455			1
298888.889	771396.219	159545.4545	159545.4545			1
198888.889	231052.9378	1159545.455	231052.9378		1	
988888.8889	1229519.741	2159545.455	988888.8889	1		
298888.889	771396.219	159545.4546	159545.4546			1
328888.889	1071238.575	140454.5455	140454.5455			1
348888.889	1271174.818	340454.5455	340454.5455			1
298888.889	771396.219	159545.4546	159545.4546			1
248888.889	272431.1872	659545.4545	272431.1872		1	
				18	72	110

Gambar 6 Iterasi Kedua

Dapat terlihat pada iterasi ke 1 dan iterasi ke 2 tidak lagi mengalami perubahan pada titik cluster, sehingga dapat disimpulkan bahwa

iterasi dapat dihentikan pada iterasi ke 2 dengan hasil :

- Cluster pertama = 18 Penduduk
- Cluster kedua = 72 Penduduk
- Cluster Ketiga = 110 Penduduk

#### 4.4 Hasil Pengelompokan Masing-Masing Cluster

Berdasarkan hasil pengelompokan dari seluruh data menggunakan metode k-means clustering, didapatkan hasil akhir pengelompokan hingga iterasi ke-2, dimana titik pusat tidak lagi berubah dan tidak ada data yang berpindah antar cluster. Hasil cluster yang terbentuk dari setiap data tersebut dapat dilihat pada table 1 dibawah ini:

Table 1 Pengelompokan Data Penduduk

Cluster	Data Penduduk
Cluster 1	17, 45, 47, 54, 64, 76, 81, 91, 110, 122, 129, 135, 147, 158, 166, 172, 191, 198.
Cluster 2	2, 6, 8, 9, 12, 18, 19, 24, 27, 28, 30, 31, 33, 34, 35, 36, 38, 42, 44, 50, 55, 58, 59, 60, 61, 62, 63, 66, 67, 72, 73, 74, 75, 77, 82, 85, 87, 88, 89, 90, 93, 94, 96, 97, 103, 105, 106, 115, 125, 131, 132, 138, 139, 140, 142, 145, 148, 149, 152, 159, 174, 175, 177, 178, 180, 181, 182, 183, 187, 190, 196, 197.
Cluster 3	1, 3, 4, 5, 7, 10, 11, 13, 14, 15, 16, 20, 21, 22, 23, 25, 26, 29, 32, 37, 39, 40, 41, 43, 46, 48, 49, 51, 52, 53, 56, 57, 65, 68, 69, 70, 71, 78, 79, 80, 82, 83, 84, 86, 92, 95, 98, 99, 100, 101, 102, 104, 107, 108, 109, 111, 112, 113, 114,

116, 117, 118, 119, 120, 121, 123, 124, 126, 127, 128, 130, 133, 134, 136, 137, 140, 141, 143, 144, 146, 150, 151, 153, 154, 155, 156, 157, 160, 161, 162, 163, 164, 165, 167, 168, 170, 171, 173, 176, 179, 184, 185, 186, 188, 189, 192, 193, 194, 195, 199, 200.
---

Berdasarkan hasil pengolahan data yang telah Penulis lakukan terdapat 3 golongan kriteria penduduk yang dilihat dari segi penghasilannya yaitu :

**a. Penduduk Ekonomi Tinggi**

Kelompok penduduk yang memiliki penghasilan mulai dari Rp.3.000.000 sampai Rp.8.000.000 untuk perbulannya.

**b. Penduduk Ekonomi Sedang**

Kelompok penduduk yang memiliki penghasilan perbulannya mulai dari Rp.1.500.00 sampai Rp.2.500.000.

**c. Penduduk Ekonomi Rendah**

Kelompok penduduk ini memiliki penghasilan yang di dapatkan perbulannya yaitu Rp.1.000.000 kebawah.[15]

**5. Kesimpulan**

Berdasarkan penelitian yang telah dilakukan maka dapat ditarik kesimpulan sebagai berikut :

- Menentukan centroid (titik pusat) pada tahap awal algoritma K-Means sangat berpengaruh pada hasil cluster seperti pada hasil pengujian yang dilakukan dengan menggunakan 200 dataset dengan centroid yang berbeda menghasilkan hasil cluster yang berbeda pula. [16]
- Berdasarkan 3 cluster yang telah dilakukan pengujian menggunakan Rapid Miner bahwa Data Penduduk terbanyak terdapat pada cluster ke-3 yang tergolong pada penduduk ekonomi rendah (Miskin).

**6. Daftar Pustaka**

[1] S. Kasus, T. Usaha, and M. Barabai, "PENERAPAN DATA MINING UNTUK KLASIFIKASI PRODUK MENGGUNAKAN ALGORITMA K-MEANS," vol. 2, no. 2, pp. 191–198, 2018.

[2] W. P. Hidayanti, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Efektivitas Penjualan Vape ( Rokok El ektrik ) pada ' Lombok Vape On ' *Infotek J. Inform. dan Teknol.*, vol. 3, no. 2, 2020.

[3] V. No, W. Purba, and W. Siawin, "IMPLEMENTASI DATA MINING UNTUK PENGELOMPOKKAN DAN PREDIKSI KARYAWAN YANG BERPOTENSI PHK DENGAN ALGORITMA K-MEANS CLUSTERING," vol. 2, no. 2, 2019.

- [4] J. I. Komputer, F. Matematika, D. A. N. Ilmu, and P. Alam, "PENERAPAN METODE CLUSTERING K-MEANS TERHADAP DOSEN BERDASARKAN PUBLIKASI," 2016.
- [5] R. W. Sari, A. Wanto, and A. P. Windarto, "IMPLEMENTASI RAPIDMINER DENGAN METODE K-MEANS ( STUDY KASUS : IMUNISASI CAMPAK PADA BALITA BERDASARKAN PROVINSI )," vol. 2, pp. 224–230, 2018.
- [6] D. Safira, E. D. Lestari, M. Iffa, and S. Annisa, "BERDASARKAN ANGKATAN KERJA MENGGUNAKAN ALGORITMA K-MEANS," vol. 6, no. 1, pp. 26–31, 2020.
- [7] I. Method, K. C. Based, S. Value, W. Interface, C. Study, and I. U. M. M. Magelang, "Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang," vol. 18, no. 1, pp. 76–82, 2015.
- [8] F. Sembiring and S. Saepudin, "IMPLEMENTASI METODE K-MEANS DALAM PENGKLASTERAN DAERAH PUNGUTAN LIAR DI KABUPATEN SUKABUMI ( STUDI KASUS : DINAS KEPENDUDUKAN DAN PENCATATAN SIPIL )," vol. 14, no. 1, pp. 40–47, 2020.
- [9] T. Noviana and Y. Novianto, "PENERAPAN DATA MINING MENENTUKAN KELOMPOK PRIORITAS PENERIMA BANTUAN BERAS RASTRA DENGAN CLUSTERING K-MEANS," pp. 159–174.
- [10] A. M. Nur and B. Harianto, "Komparasi Algoritma SVM Dan SVM Berbasis PSO Dalam Menganalisis Kinerja Guru SMAN 3 Selong," *Infotek J. Inform. dan Teknol.*, vol. 2, no. 2, pp. 86–94, 2019.
- [11] I. Nasution, A. P. Windarto, and M. Fauzan, "Penerapan Algoritma K-Means Dalam Pengelompokan Data Penduduk Miskin Menurut Provinsi," vol. 2, no. 2, pp. 76–83, 2020.
- [12] L. Listiani, Y. H. Agustin, and M. Z. Ramdhani, "Implementasi algoritma k-means cluster untuk rekomendasi pekerjaan berdasarkan pengelompokan data penduduk," pp. 761–769, 2017.
- [13] F. S. Napitupulu, I. S. Damanik, I. S. Saragih, and A. Wanto, "Algoritma K-Means untuk Pengelompokan Dokumen Akta Kelahiran pada Tiap Kecamatan di Kabupaten Simalungun," vol. 2, no. 1, pp. 55–63, 2020.
- [14] S. T. Informatika *et al.*, "e-ISSN 2614-8773 92," vol. 3, no. 2, pp. 92–103, 2020.
- [15] N. Rofiqo *et al.*, "PENERAPAN CLUSTERING PADA PENDUDUK YANG MEMPUNYAI KELUHAN KESEHATAN



---

DOI : 10.29408/jit.v4i1.2986    Link : <https://dx.doi.org/10.29408/jit.v4i1.2986>

DENGAN DATAMINING K-MEANS,” vol. 2,  
pp. 216–223, 2018.

- [16] Y. P. Sari, A. Primajaya, A. Susilo, and Y. Irawan, “Implementasi Algoritma K-Means untuk Clustering Penyebaran Tuberkulosis di Kabupaten Karawang,” pp. 229–239, 2020.