

Normalisasi Data Kependudukan Dengan Model Min Max Dan Algoritma K-Means Untuk Pengelompokan Tingkat Ekonomi Masyarakat

Wenny

Fakultas Ilmu Komputer dan Teknologi Informasi, Program Studi Teknik Informatika, Universitas Budi Darma, Medan, Indonesia

Email: nazrilwenny@gmail.com

Abstrak—Data kependudukan pada Disdukcapil Deli Serdang selama ini hanya digunakan untuk keperluan administrasi seperti membuat surat keterangan, KTP, Kartu Keluarga, dan lainnya. Sedangkan data tersebut seharusnya dapat ditingkatkan fungsi kegunaannya seperti pengelompokan tingkat ekonomi masyarakat Kabupaten Deli Serdang. Selama ini, banyak masyarakat merasa dirugikan karena dalam distribusi bantuan pemerintah kepada masyarakat yang ekonomi tingkat rendah sering salah sasaran, ada masyarakat yang selayaknya memperoleh bantuan tetapi malah tidak dapat, tetapi ada masyarakat kurang layak dapat bantuan malah dapat bantuan, hal tersebut dapat dipengaruhi berbagai faktor seperti kekeluargaan, kedekatan atau hal lainnya. sangat diperlukan teknik dalam pengelompokan tingkat ekonomi masyarakat Kabupaten Deli Serdang berdasarkan kriteria tertentu. Algoritma K-Means salah satu algoritma dalam data mining untuk mengelompokkan data tertentu, untuk itu sangat cocok digunakan algoritma K-Means dalam mengelompokkan tingkat ekonomi masyarakat.

Kata Kunci: Disdukcapil; KTP; Pengelompokan; Algoritma K-Means

Abstract—Desidential data at Disdukcapil Deli Serdang has only been used for administrative purposes such as making certificates, ID cards, family cards, and others. Meanwhile, the data should be able to increase its usefulness functions such as classifying the economic level of the Deli Serdang Regency community. So far, many people feel disadvantaged because the distribution of government assistance to low-level economic communities is often misdirected, there are people who should get assistance but instead cannot, but there are people who are less deserving of assistance who can get assistance, this can be influenced by various factors such as kinship, closeness or other things. techniques are needed in classifying the economic level of the Deli Serdang Regency community based on certain criteria. The K-Means algorithm is one of the algorithms in data mining to classify certain data, for that it is very suitable to use the K-Means algorithm in classifying the economic level of the community.

Keywords: Disdukcapil; KTP; Clustering; K-Means Algorithm

1. PENDAHULUAN

Dinas Kependudukan dan Pencatatan Sipil Kabupaten Deli Serdang merupakan dinas yang bergerak dalam bidang pelayanan untuk melayani masyarakat untuk menerbitkan dokumen kependudukan berupa Akta, KIA (Kartu Identitas Anak), Surat Pindah, KTP dan Kartu Keluarga. Dalam melayani masyarakat, Disdukcapil Deli Serdang menggunakan 2 pelayanan, yaitu secara offline dan online untuk menghindari menumpuknya masyarakat yang datang secara langsung.

Data kependudukan pada Disdukcapil Deli Serdang selama ini hanya digunakan untuk keperluan administrasi seperti membuat surat keterangan, KTP, Kartu Keluarga, dan lainnya. Sedangkan data tersebut seharusnya dapat ditingkatkan fungsi kegunaannya seperti pengelompokan tingkat ekonomi masyarakat Kabupaten Deli Serdang. Selama ini, banyak masyarakat merasa dirugikan karena dalam distribusi bantuan pemerintah kepada masyarakat yang ekonomi tingkat rendah sering salah sasaran, ada masyarakat yang selayaknya memperoleh bantuan tetapi malah tidak dapat, tetapi ada masyarakat kurang layak dapat bantuan malah dapat bantuan, hal tersebut dapat dipengaruhi berbagai faktor seperti kekeluargaan, kedekatan atau hal lainnya [1].

Oleh karena itu, sangat diperlukan teknik dalam pengelompokan tingkat ekonomi masyarakat Kabupaten Deli Serdang berdasarkan kriteria tertentu. Algoritma K-Means salah satu algoritma dalam data mining untuk mengelompokkan data tertentu, untuk itu sangat cocok digunakan algoritma K-Means dalam mengelompokkan tingkat ekonomi masyarakat. Sebelum dilakukan pengelompokan data, data dinormalisasikan terlebih dahulu, tujuannya adalah agar ada kesederhanaan dalam data, keseimbangan antara kriteria dan adanya kepastian hasil pengelompokan [2].

Berdasarkan penelitian yang dilakukan oleh Efori Buulolo dan Rian Syahputra, dengan judul “Implementasi Algoritma Clustering K-Means Untuk Mengelompokkan Mahasiswa Baru Yang Berpotensi (Studi Kasus : STMIK Budi Darma)” menyimpulkan bahwa agar pengelompokan calon mahasiswa baru berdasarkan potensi yang dimiliki maka digunakan algoritma clustering K-Means. Dasar pengelompokan dengan algoritma clustering K-Means berdasarkan nilai jarak terdekat dari kriteria yang berbasis numerik dengan output dalam bentuk cluster. Cluster yang telah dibentuk digunakan sebagai dasar pengelompokan calon mahasiswa baru ke dalam satu kelas [3].

Penelitian juga pernah dilakukan oleh Fajar Nur Rahmat Fauzan Jaya Aziz, Budi Darma Setiawan, dan Issa Arwani dengan judul “Implementasi Algoritma K-Means untuk Klasterisasi Kinerja Akademik Mahasiswa” menyimpulkan bahwa dengan menggunakan metode K-Means clustering, peneliti mencoba untuk mengekstrak

pengetahuan yang bisa menggambarkan kinerja prestasi akademik mahasiswa pada akhir semester dan hasil dari penelitian tersebut menunjukkan bahwa dari semua jumlah cluster yang dimasukkan untuk cluster yang berjumlah 3 memiliki nilai silhouette coefficient yang paling mendekati nilai $S_i = 1$, yaitu dengan nilai 0, 108690751. Selain itu pendapatan orang tua tidak mempengaruhi tingkat kinerja akademik mahasiswa dan nilai akademis mahasiswa yang masuk melalui jalur reguler dan jalur prestasi akademik mempunyai nilai IPK rata-rata tertinggi. Sehingga pihak fakultas dapat mempertimbangkan untuk lebih memprioritaskan penerimaan mahasiswa baru melalui jalur reguler dan prestasi akademik.

Berdasarkan penelitian yang dilakukan oleh Efori Buulolo dan Rian Syahputra pada tahun 2019 tentang “Implementasi Algoritma Clustering K-Means Untuk Mengelompokkan Mahasiswa Baru Yang Berpotensi (Studi Kasus : Stmik Budi Darma)“, menyimpulkan bahwa agar pengelompokkan calon mahasiswa baru berdasarkan potensi yang dimiliki maka digunakan algoritma clustering K-Means. Dasar pengelompokkan dengan algoritma clustering K-Means berdasarkan nilai jarak terdekat dari kriteria yang berbasis numerik dengan output dalam bentuk cluster. Cluster yang telah dibentuk digunakan sebagai dasar pengelompokkan calon mahasiswa baru ke dalam satu kelas [3].

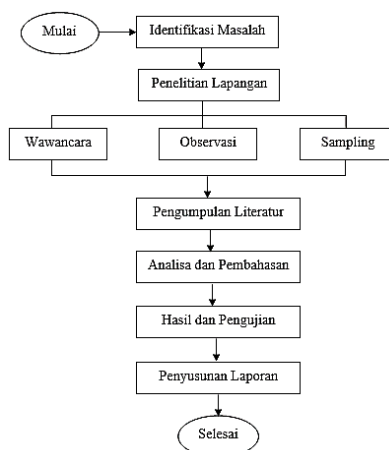
Penelitian juga pernah dilakukan oleh Fajar Nur Rahmat Fauzan Jaya Aziz, Budi Darma Setiawan, dan Issa Arwani pada tahun 2018 tentang “Implementasi Algoritma K-Means untuk Klasterisasi Kinerja Akademik Mahasiswa” menyimpulkan bahwa dengan menggunakan metode K-Means clustering, peneliti mencoba untuk mengekstrak pengetahuan yang bisa menggambarkan kinerja prestasi akademik mahasiswa pada akhir semester dan hasil dari penelitian tersebut menunjukkan bahwa dari semua jumlah cluster yang dimasukkan untuk cluster yang berjumlah 3 memiliki nilai silhouette coefficient yang paling mendekati nilai $S_i = 1$, yaitu dengan nilai 0, 108690751. Selain itu pendapatan orang tua tidak mempengaruhi tingkat kinerja akademik mahasiswa dan nilai akademis mahasiswa yang masuk melalui jalur reguler dan jalur prestasi akademik mempunyai nilai IPK rata-rata tertinggi. Sehingga pihak fakultas dapat mempertimbangkan untuk lebih memprioritaskan penerimaan mahasiswa baru melalui jalur reguler dan prestasi akademik.

Berdasarkan penelitian yang dilakukan oleh Tria Pratiwi Sutriyani, Amril Mutoi Siregar dan Dwi Sulistya Kusumaningrum pada tahun 2018 tentang “Implementasi Algoritma K-Means Terhadap Pengelompokkan Nilai Ujian Nasional Tingkat SMP di Provinsi Jawa Barat”, menyimpulkan bahwa akan dilakukan penelitian untuk menambang data menjadi informasi, pada penelitian ini data yang digunakan adalah nilai rata-rata UN tingkat SMP di Provinsi Jawa Barat menggunakan teknik data mining. Penelitian dilakukan dengan cara perhitungan manual menggunakan Microsoft Excel kemudian dievaluasi menggunakan tools Rapidminer 9.2.0. Penelitian ini menghasilkan Cluster Baik 10 Kabupaten/Kota, Cluster Cukup Baik 15 Kabupaten/Kota dan Cluster Kurang Baik 2 Kabupaten/Kota 2 Mata Pelajaran Matematika dan IPA saat ini tidak menentukan kategori, karena nilai yang didapat rata-rata lebih kecil dari Bahasa Inggris dan Bahasa Indonesia sehingga bagi Pemerintah Dinas Pendidikan untuk perlu meningkatkan lagi dalam proses pembelajaran di Provinsi Jawa Barat [4].

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Kerangka kerja penelitian adalah tahapan penelitian yang dilakukan oleh penulis dan bertujuan untuk menggambarkan prosedur sistematis yang akan diikuti selama penelitian. Adapun kerangka kerja penelitian yang digunakan penulis adalah sebagai berikut:



Gambar 1. Kerangka Kerja Penelitian

Berdasarkan pada kerangka kerja penelitian yang telah diilustrasikan sebelumnya, maka dapat dijabarkan pembahasan masing-masing prosedur dalam penelitian sebagai berikut :

1. Identifikasi Masalah

Yakni menguraikan permasalahan secara konteks pada objek penelitian menggunakan algoritma K-Means, sehingga dapat menentukan batasan dari masalah yang akan diteliti.

2. Penelitian Lapangan

Teknik yang dilakukan untuk memperoleh data secara langsung agar data yang diperoleh lebih akurat. Dalam hal ini pengambilan data dan sampel dilakukan dengan cara wawancara, observasi, dan sampling.

3. Pengumpulan Literatur

Dalam pengumpulan literatur, penulis memperoleh dengan cara membaca, memahami dan mengutip hal-hal yang bersangkutan dengan penelitian yang diangkat seperti pada buku, jurnal online, artikel dan media cetak online lainnya yang membahas tentang normalisasi data kependudukan dengan model min max dengan algoritma k means untuk pengelompokkan tingkat ekonomi masyarakat.

4. Analisa dan Pembahasan

Menganalisa metode yang digunakan berdasarkan kasus penelitian yang akan diteliti pada Dinas Kependudukan dan Pencatatan Sipil Kabupaten Deli Serdang. Data yang diperoleh dan dikumpulkan dikelola dengan menggunakan metode yang dikombinasikan dan menganalisis cara kerja metode tersebut.

5. Hasil dan Pengujian

Pada tahapan ini peneliti akan menjabarkan keluaran akhir yang dihasilkan sistem mulai dari perhitungan menggunakan algoritma K-Means, lalu menormalisasikan hasilnya dengan model min max dalam pengelompokkan tingkat ekonomi masyarakat.

6. Penyusunan Laporan

Peneliti melakukan penyusunan laporan sesuai dengan kerangka kerja penelitian yang telah dilakukan pada Dinas Kependudukan dan Pencatatan Sipil Kabupaten Deli Serdang.

2.2 Data Mining

Data mining atau kadang disebut juga *Knowledge Discovery in Database* (KDD) merupakan aktivitas yang berkaitan dengan pengumpulan data, pemakaian data historis untuk menemukan pengetahuan, informasi, keteraturan, pola atau hubungan dalam data yang berukuran besar[5][6][7]. Output dalam data mining dapat dipergunakan sebagai alternatif dalam pengambilan keputusan atau untuk memperbaiki pengambilan keputusan dimasa yang akan datang [6].

2.3 Algoritma K-Means

K-means merupakan algoritma clustering. K-means Clustering adalah salah satu “unsupervised machine learning algorithms” yang paling sederhana dan populer. K-Means Clustering adalah suatu metode penganalisaan data atau metode Data Mining yang melakukan proses pemodelan tanpa supervisi (unsupervised) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi[8][9].

Dengan kata lain, metode K-Means Clustering bertujuan untuk meminimalisasikan *objective function* yang diset dalam proses clustering dengan cara meminimalkan variasi antar data yang ada di dalam suatu cluster dan memaksimalkan variasi dengan data yang ada di cluster lainnya juga bertujuan untuk menemukan grup dalam data, dengan jumlah grup yang diwakili oleh variabel K[10][11]. Variabel K sendiri adalah jumlah cluster yang diinginkan. Membagi data menjadi beberapa kelompok. Algoritma ini menerima masukan berupa data tanpa label kelas. Hal ini berbeda dengan supervised learning yang menerima masukan berupa vektor $(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i)$, di mana x_i merupakan data dari suatu data pelatihan dan y_i merupakan label kelas untuk x_i [12][13].

Data clustering menggunakan metode K Means Clustering ini secara umum dilakukan dengan algoritma dasar sebagai berikut[14][15]:

1. Tentukan jumlah cluster.
2. Alokasikan data ke dalam cluster secara random.
3. Hitung centroid / rata-rata dari data yang ada di masing-masing cluster.
4. Alokasikan masing-masing data ke centroid/rata-rata terdekat.
5. Kembali ke Step 3, apabila masih ada data yang berpindah cluster atau apabila perubahan nilai centroid, ada yang di atas nilai threshold yang ditentukan atau apabila perubahan nilai pada objective function yang digunakan di atas nilai threshold yang ditentukan.

2.4 Normalisasi Data Model Min Max

Normalisasi data adalah proses membuat beberapa variabel memiliki rentang nilai yang sama, tidak ada yang terlalu besar maupun terlalu kecil sehingga dapat membuat analisis statistik menjadi lebih mudah.

Normalisasi model min max, cara kerjanya setiap nilai pada sebuah fitur dikurangi dengan nilai minimum fitur tersebut, kemudian dibagi dengan rentang nilai atau nilai maksimum dikurangi nilai minimum dari fitur tersebut.

$$X_{new} = \frac{X_{old} - X_{min}}{X_{max} - X_{min}} \tag{1}$$

Cara kerja ini juga akan menghasilkan nilai baru hasil normalisasi antara 0 sampai 1. Sama seperti method max (), method min () digunakan untuk mengambil nilai minimum dari fitur tersebut [1].

2.5 Data Kependudukan

Secara umum data dapat diartikan sebagai kumpulan informasi yang diperoleh dari suatu pengamatan berupa angka, lambang atau sifat yang dapat memberikan gambaran tentang suatu keadaan atau persoalan. Data juga dapat didefinisikan sebagai kumpulan informasi atau nilai yang diperoleh dari pengamatan (observasi) suatu objek. Oleh karena itu data yang baik adalah data yang bisa dipercaya kebenarannya (reliable), tepat waktu dan mencakup ruang lingkup yang luas atau bias memberikan gambaran tentang suatu masalah secara menyeluruh merupakan data yang relevan.

Sedangkan kependudukan atau demografi merupakan ilmu yang mempelajari dinamika kependudukan manusia. Demografi meliputi ukuran, struktur dan distribusi penduduk serta bagaimana jumlah penduduk berubah setiap waktu akibat kelahiran, kematian, migrasi, serta penuaan.

Analisis kependudukan dapat merujuk masyarakat secara keseluruhan atau kelompok tertentu yang didasarkan kriteria seperti pendidikan, kewarganegaraan, agama atau etnisitas tertentu. Dengan demikian data kependudukan adalah segala tampilan data penduduk dalam bentuk resmi maupun tidak resmi yang diterbitkan oleh badan-badan pencatatan kependudukan (pemerintah maupun non pemerintah), dalam berbagai bentuk baik angka, grafik, gambar dan lain-lain. Secara khusus UU No. 24 Tahun 2013 pasal 1 point 9 menyebutkan bahwa data kependudukan adalah data perseorangan dan/atau data agregat yang terstruktur sebagai hasil dari kegiatan pendaftaran penduduk dan pencatatan sipil.

3. HASIL DAN PEMBAHASAN

3.1 Analisa

Tahap analisa merupakan tahap yang penting dalam melanjutkan setiap tahapan selanjutnya pada penelitian. Dalam tahap ini akan dijelaskan langkah-langkah menormalisasikan data penduduk untuk pengelompokkan tingkat ekonomi masyarakat secara manual.

Tabel 1. Data Keluarga

No.	Nomor KK	Nama Kepala Keluarga	Jlh Anggota Keluarga	Pekerjaan	Penghasilan (Rp)	Aset (Rp)
1	1207281609092083	Agus	5	1	3.000.000	15.000.000
2	1207221008160006	Edona F	2	2	5.000.000	25.000.000
3	1207221908160002	Ardi	3	2	3.500.000	15.500.000
4	1207221709180010	Teddy Hermawan	2	2	3.500.000	15.500.000
5	1207231509095399	Sugirin	4	2	4.000.000	20.000.000
6	1207230608100008	Yen Marudut Aruan	3	2	3.000.000	15.000.000
7	1207230511150020	Heri Suandi	4	2	3.500.000	15.500.000
8	1207230110130003	Pantun Tampubolon	4	2	3.500.000	15.500.000
9	1207232406220004	Piki Prayoga	3	1	3.000.000	15.000.000
10	1207020801130002	Mohammad Haidir	2	3	2.000.000	10.000.000
11	1207262708210044	Iqbal Fadilah	2	5	4.000.000	20.000.000
12	1207262406220040	Sopan Irawan	2	2	3.000.000	15.000.000
13	1207261407110025	Ghazali Hasan	4	1	3.000.000	15.000.000
14	1207242201130002	Mhd.Yamin	5	2	3.500.000	15.500.000
15	1207241509093401	Taufik Hidayat Damanik	5	4	4.500.000	21.000.000
16	1207241709094250	Ahrun Harahap	4	2	3.500.000	15.500.000
17	1207261907100026	Pairan	4	8	2.500.000	12.000.000
18	1207020502180019	Liono	4	6	5.000.000	25.000.000
19	1207262601110025	Widodo	3	2	3.500.000	16.000.000
20	1207241408180001	Sudrajat	4	2	3.500.000	16.000.000
21	1207260912110041	Iwan E. Hutabarat	4	2	4.000.000	20.000.000
22	1207242306210011	Didit Permadi	2	7	2.500.000	12.000.000
23	1207240303210008	Arifin	2	2	3.000.000	15.000.000

No.	Nomor KK	Nama Kepala Keluarga	Jlh Anggota Keluarga	Pekerjaan	Penghasilan (Rp)	Aset (Rp)
24	1207242107160009	Januardi	3	2	3.500.000	17.000.000
25	1207241910180013	Romi Hidayat	2	5	4.000.000	21.000.000
26	1207240607110001	Misnan.S.	5	7	2.500.000	13.000.000
27	1207221409094606	Yuda Heria	4	2	3.500.000	21.000.000
28	1207220504190015	Sri Waluyo	4	5	4.000.000	23.000.000
29	1207282908170003	Muhammad Hariansyah	3	9	8.000.000	50.000.000
30	1207230412130018	Santus Parsaulian Simanjuntak	6	2	4.000.000	24.000.000

Tabel 2. Konversi Pekerjaan Dalam Bentuk Angka

Pekerjaan	Angka
Buruh Harian Lepas	1
Wiraswasta	2
Tukang Batu	3
Sopir	4
Karyawan Swasta	5
Karyawan BUMN	6
Petani	7
Tukang Listrik	8
PNS	9

3.2 Normalisasi *Min Max*

Selanjutnya dilakukan normalisasi data dengan model min max tujuannya adalah untuk menghindari ketimpangan antar kriteria data.

1. Normalisasi data jumlah anggota keluarga

Nilai paling tinggi = 6

Nilai paling rendah = 2

$$\text{Agus} = \frac{(5-2)}{(6-2)} = 0.75$$

$$\text{Edona F} = \frac{(2-2)}{(6-2)} = 0$$

$$\text{Ardi} = \frac{(3-2)}{(6-2)} = 0.25$$

$$\text{Teddy Hermawan} = \frac{(2-2)}{(6-2)} = 0$$

$$\text{Sugirin} = \frac{(4-2)}{(6-2)} = 0.5$$

Lakukan perhitungan normalisasi data jumlah anggota keluarga diatas untuk Nama Kepala Keluarga dari nomor 6 hingga selesai.

2. Normalisasi data Pekerjaan

Nilai paling tinggi = 9

Nilai paling rendah = 1

$$\text{Agus} = \frac{(1-1)}{(9-1)} = 0$$

$$\text{Edona F} = \frac{(2-1)}{(9-1)} = 0.125$$

$$\text{Ardi} = \frac{(2-1)}{(9-1)} = 0.125$$

$$\text{Teddy Hermawan} = \frac{(2-1)}{(9-1)} = 0.125$$

$$\text{Sugirin} = \frac{(2-1)}{(9-1)} = 0.125$$

- Lakukan perhitungan normalisasi data pekerjaan diatas untuk Nama Kepala Keluarga dari nomor 6 hingga selesai.
3. Normalisasi data penghasilan
- Nilai paling tinggi = 8.000.000
- Nilai paling rendah = 2.000.000

Agus $= \frac{(3.000.000-2.000.000)}{(8.000.000-2.000.000)} = 0.16$

Edona F $= \frac{(5.000.000-2.000.000)}{(8.000.000-2.000.000)} = 0.5$

Ardi $= \frac{(3.500.000-2.000.000)}{(8.000.000-2.000.000)} = 0.25$

Teddy Hermawan $= \frac{(3.500.000-2.000.000)}{(8.000.000-2.000.000)} = 0.25$

Sugirin $= \frac{(4.000.000-2.000.000)}{(8.000.000-2.000.000)} = 0.33$

- Lakukan perhitungan normalisasi data penghasilan diatas untuk Nama Kepala Keluarga dari nomor 6 hingga selesai.
4. Normalisasi data aset
- Nilai paling tinggi = 50.000.000
- Nilai paling rendah = 10.000.000

Agus $= \frac{(15.000.000-10.000.000)}{(25.000.000-10.000.000)} = 0.125$

Edona F $= \frac{(25.000.000-10.000.000)}{(25.000.000-10.000.000)} = 0.375$

Ardi $= \frac{(15.500.000-10.000.000)}{(25.000.000-10.000.000)} = 0.1375$

Teddy Hermawan $= \frac{(15.500.000-10.000.000)}{(25.000.000-10.000.000)} = 0.1375$

Sugirin $= \frac{(20.000.000-10.000.000)}{(25.000.000-10.000.000)} = 0.25$

- Lakukan perhitungan normalisasi data penghasilan diatas untuk Nama Kepala Keluarga dari nomor 6 hingga selesai.

Tabel 3. Data Keluarga Setelah Normalisasi

No.	Nomor KK	Nama Kepala Keluarga	Jlh Anggota Keluarga	Pekerjaan	Penghasilan (Rp)	Aset (Rp)
1	1207281609092083	Agus	0.75	0	0.16	0.125
2	1207221008160006	Edona F	0	0.125	0.5	0.375
3	1207221908160002	Ardi	0.25	0.125	0.25	0.1375
4	1207221709180010	Teddy Hermawan	0	0.125	0.25	0.1375
5	1207231509095399	Sugirin	0.5	0.125	0.33	0.25
6	1207230608100008	Yen Marudut Aruan	0.25	0.125	0.16	0.125
7	1207230511150020	Heri Suandi	0.5	0.125	0.25	0.1375
8	1207230110130003	Pantun Tampubolon	0.5	0.125	0.25	0.1375
9	1207232406220004	Piki Prayoga	0.25	0	0.16	0.125
10	1207020801130002	Mohammad Haidir	0	0.5	0	0
11	1207262708210044	Iqbal Fadilah	0	0.25	0.33	0.25
12	1207262406220040	Sopan Irawan	0	0.125	0.16	0.125
13	1207261407110025	Ghazali Hasan	0.5	0	0.16	0.125
14	1207242201130002	Mhd. Yamin	0.75	0.125	0.25	0.1375
15	1207241509093401	Taufik Hidayat Damanik	0.75	0.375	0.41	0.275
16	1207241709094250	Ahrun Harahap	0.5	0.125	0.25	0.1375
17	1207261907100026	Pairan	0.5	0.875	0.08	0.05
18	1207020502180019	Liono	0.5	0.625	0.5	0.375

No.	Nomor KK	Nama Kepala Keluarga	Jlh Anggota Keluarga	Pekerjaan	Penghasilan (Rp)	Aset (Rp)
19	1207262601110025	Widodo	0.25	0.125	0.25	0.15
20	1207241408180001	Sudrajat	0.5	0.125	0.25	0.15
21	12072609121110041	Iwan E. Hutabarat	0.5	0.125	0.33	0.5
22	1207242306210011	Didit Permadi	0	0.75	0.08	0.05
23	1207240303210008	Arifin	0	0.125	0.16	0.125
24	1207242107160009	Januardi	0.25	0.125	0.25	0.175
25	1207241910180013	Romi Hidayat	0	0.5	0.33	0.275
26	1207240607110001	Misnan.S.	0.75	0.75	0.08	0.075
27	1207221409094606	Yuda Heria	0.5	0.125	0.25	0.275
28	1207220504190015	Sri Waluyo	0.5	0.5	0.33	0.325
29	1207282908170003	Muhammad Hariansyah	0.25	0	1	1
30	1207230412130018	Santus Parsaulian Simanjntak	1	0.125	0.33	0.35

3.3 Pengelompokan Data Dengan Algoritma K-Means

Selanjutnya adalah mengelompokkan data dengan algoritma K-Means.Untuk data tersebut diatas dikelompokan menjadi 3(tiga) kelompok /K=3.

Iterasi 1: K=3

Nilai centroid pada iterasi I ditentukan secara acak.

Tabel 4. Nilai Centroid Iterasi 1

C1	Sopan Irawan	0	0,125	0,16	0,125
C2	Edona F	0	0,125	0,5	0,375
C3	Muhammad Hariansyah	0,25	0	1	1

Tabel 5. Proses Perhitungan Iterasi 1

No.	Nama Kepala Keluarga	Jlh Anggota Keluarga	Pekerjaan	Penghasilan (Rp)	Aset (Rp)	dc1	dc2	dc3	dc4
1	Agus	0.75	0	0.16	0.125	0,760345	0,869612	1,311955	1
2	Edona F	0	0.125	0.5	0.375	0,422019	0	0,847791	2
3	Ardi	0.25	0.125	0.25	0.1375	0,266	0,425918	1,149796	1
4	Teddy Hermawan	0	0.125	0.25	0.1375	0,090864	0,344828	1,176661	1
5	Sugirin	0.5	0.125	0.33	0.25	0,542702	0,542702	1,043803	2
6	Yen Marudut Aruan	0.25	0.125	0.16	0.125	0,25	0,49051	1,219365	1
7	Heri Suandi	0.5	0.125	0.25	0.1375	0,508189	0,607377	1,176661	1
8	Pantun Tampubolon	0.5	0.125	0.25	0.1375	0,508189	0,607377	1,176661	1
9	Piki Prayoga	0.25	0	0.16	0.125	0,279687	0,512957	1,219887	1
10	Mohammad Haidir	0	0.5	0	0	0,426439	0,728869	1,520691	1
11	Iqbal Fadilah	0	0.25	0.33	0.25	0,245255	0,245255	1,066021	2
12	Sopan Irawan	0	0.125	0.16	0.125	0	0,422019	1,244729	1
13	Ghazali Hasan	0.5	0	0.16	0.125	0,515388	0,666127	1,238437	1
14	Mhd.Yamin	0.75	0.125	0.25	0.1375	0,755447	0,826196	1,255528	1
15	Taufik Hidayat Damanik	0.75	0.375	0.41	0.275	0,804922	0,874414	1,286798	1
16	Ahrun Harahap	0.5	0.125	0.25	0.1375	0,508189	0,607377	1,176661	1
17	Pairan	0.5	0.875	0.08	0.05	0,908034	1,046195	1,605311	1
18	Liono	0.5	0.625	0.5	0.375	0,823468	0,707107	1,045825	2
19	Widodo	0.25	0.125	0.25	0.15	0,26688	0,419076	1,140449	1
20	Sudrajat	0.5	0.125	0.25	0.15	0,50865	0,602599	1,167529	1
21	Iwan E. Hutabarat	0.5	0.125	0.33	0.5	0,647707	0,542702	0,88149	2

No.	Nama Kepala Keluarga	Jlh Anggota Keluarga	Pekerjaan	Penghasilan (Rp)	Aset (Rp)	dc1	dc2	dc3	dc4
22	Didit Permadi	0	0.75	0.08	0.05	0,634547	0,820152	1,540747	1
23	Arifin	0	0.125	0.16	0.125	0	0,422019	1,244729	1
24	Januardi	0.25	0.125	0.25	0.175	0,27037	0,406202	1,12194	1
25	Romi Hidayat	0	0.5	0.33	0.275	0,438207	0,423704	1,134471	2
26	Misnan.S.	0.75	0.75	0.08	0.075	0,980829	1,104321	1,585725	1
27	Yuda Heria	0.5	0.125	0.25	0.275	0,529717	0,567891	1,079931	1
28	Sri Waluyo	0.5	0.5	0.33	0.325	0,677883	0,649635	1,103189	2
29	Muhammad Hariansyah Santus	0.25	0	1	1	1,244729	0,847791	0	3
30	Parsaulian Simanjntak	1	0.125	0.33	0.35	1,039002	1,014655	1,203962	2

Penentuan cluster berdasarkan nilai paling rendah antara dc1, dc2 dan dc3. Maka berdasarkan data tersebut hasil cluster sebagai berikut:

Tabel 6. Hasil Cluster Iterasi 1

Cluster	Kepala Keluarga	Jlh Anggota	Pekerjaan	Penghasilan	Aset
I	Agus	0,75	0	0,16	0,125
	Ardi	0,25	0,125	0,25	0,1375
	Teddy Hermawan	0	0,125	0,25	0,1375
	Yen Marudut Aruan	0,25	0,125	0,16	0,125
	Heri Suandi	0,5	0,125	0,25	0,1375
	Pantun Tampubolon	0,5	0,125	0,25	0,1375
	Piki Prayoga	0,25	0	0,15	0,125
	Mohammad Haidir	0	0,5	0	0
	Sopan Irawan	0	0,125	0,16	0,125
	Ghazali Hasan	0,5	0	0,16	0,125
	Mhd.Yamin	0,75	0,125	0,25	0,135
	Taufik Hidayat	0,75	0,375	0,14	0,275
	Ahrun Harahap	0,5	0,125	0,25	0,1375
	Pairan	0,5	0,875	0,08	0,05
	Widodo	0,25	0,125	0,25	0,15
	Sudrajat	0,5	0,125	0,25	0,15
	Didit Permadi	0	0,75	0,08	0,05
	Arifin	0	0,125	0,16	0,125
	Januardi	0,25	0,125	0,25	0,175
	Misnan.S.	0,75	0,75	0,08	0,075
	Yuda Heria	0,5	0,125	0,25	0,275
C1		0,37	0,23	0,18	0,13
Cluster II	Edona F	0	0,125	0,5	0,375
	Sugirin	0,5	0,125	0,33	0,25
	Iqbal Fadilah	0	0,25	0,33	0,25
	Liono	0,5	0,625	0,5	0,375
	Iwan E. Hutabarat	0,5	0,125	0,33	0,5
	Romi Hidayat	0	0,5	0,33	0,275
	Sri Waluyo	0,5	0,5	0,33	0,325
	Santus Parsaulian	1	0,125	0,33	0,35
C2		0,375	0,296875	0,3725	0,3375
Cluster III	Muhammad Hariansyah	0,25	0	1	1
C3		0,25	0	1	1

Setelah diperoleh hasil pengelompokan cluster iterasi I, maka diperoleh nilai centroid untuk digunakan pada iterasi II. Peroleh nilai centroid berdasarkan nilai rata-rata anggota setiap cluster seperti pada tabel 7.

Iterasi 2:

K=3; Nilai centroid pada iterasi II ditentukan dari nilai rata-rata hasil cluster pada iterasi 1.

Tabel 7. Nilai Centroid Iterasi 2

C1	0,37	0,23	0,18	0,13
C2	0,375	0,296875	0,3725	0,3375
C3	0,25	0	1	1

Tabel 8. Proses Perhitungan Iterasi 2

No.	Nama Kepala Keluarga	Jlh Anggota Keluarga	Pekerjaan	Penghasilan (Rp)	Aset (Rp)	dc1	dc2	dc3	dc4
1	Agus	0.75	0	0.16	0.125	0,446727	0,564865	1,311955	1
2	Edona F	0	0.125	0.5	0.375	0,554613	0,433392	0,847791	2
3	Ardi	0.25	0.125	0.25	0.1375	0,173937	0,3165	1,149796	1
4	Teddy Hermawan	0	0.125	0.25	0.1375	0,390228	0,474523	1,176661	1
5	Sugirin	0.5	0.125	0.33	0.25	0,253649	0,233727	1,043803	2
6	Yen Marudut Aruan	0.25	0.125	0.16	0.125	0,161871	0,368074	1,219365	1
7	Heri Suandi	0.5	0.125	0.25	0.1375	0,182292	0,3165	1,176661	1
8	Pantun Tampubolon	0.5	0.125	0.25	0.1375	0,182292	0,3165	1,176661	1
9	Piki Prayoga	0.25	0	0.16	0.125	0,262984	0,445446	1,219887	1
10	Mohammad Haidir	0	0.5	0	0	0,508563	0,659202	1,520691	1
11	Iqbal Fadilah	0	0.25	0.33	0.25	0,415	0,390237	1,066021	2
12	Sopan Irawan	0	0.125	0.16	0.125	0,385001	0,510371	1,244729	1
13	Ghazali Hasan	0.5	0	0.16	0.125	0,267561	0,440536	1,238437	1
14	Mhd.Yamin	0.75	0.125	0.25	0.1375	0,401479	0,475582	1,255528	1
15	Taufik Hidayat Damanik	0.75	0.375	0.41	0.275	0,433326	0,452428	1,286798	1
16	Ahrun Harahap	0.5	0.125	0.25	0.1375	0,182292	0,3165	1,176661	1
17	Pairan	0.5	0.875	0.08	0.05	0,669047	0,719768	1,605311	1
18	Liono	0.5	0.625	0.5	0.375	0,575677	0,375438	1,045825	2
19	Widodo	0.25	0.125	0.25	0.15	0,174778	0,308753	1,140449	1
20	Sudrajat	0.5	0.125	0.25	0.15	0,183095	0,308753	1,167529	1
21	Iwan E. Hutabarat	0.5	0.125	0.33	0.5	0,431075	0,270885	0,88149	2
22	Didit Permadi	0	0.75	0.08	0.05	0,649293	0,717049	1,540747	1
23	Arifin	0	0.125	0.16	0.125	0,385001	0,510371	1,244729	1
24	Januardi	0.25	0.125	0.25	0.175	0,179084	0,294242	1,12194	1
25	Romi Hidayat	0	0.5	0.33	0.275	0,500177	0,433125	1,134471	2
26	Misnan.S.	0.75	0.75	0.08	0.075	0,653479	0,707396	1,585725	1
27	Yuda Heria	0.5	0.125	0.25	0.275	0,231609	0,253137	1,079931	1
28	Sri Waluyo	0.5	0.5	0.33	0.325	0,384613	0,242585	1,103189	2
29	Muhammad Hariansyah Santus	0.25	0	1	1	1,220634	0,967689	0	3
30	Parsaulian Simanjntak	1	0.125	0.33	0.35	0,692016	0,649714	1,203962	2

Penentuan cluster berdasarkan nilai paling rendah antara dc1, dc2 dan dc3. Maka berdasarkan data tersebut hasil cluster sebagai berikut:

Tabel 9. Hasil Cluster Iterasi 2

Cluster	Kepala Keluarga	Jlh Anggota	Pekerjaan	Penghasilan	Aset
I	Agus	0,75	0	0,16	0,125

Cluster	Kepala Keluarga	Jlh Anggota	Pekerjaan	Penghasilan	Aset
	Ardi	0,25	0,125	0,25	0,1375
	Teddy Hermawan	0	0,125	0,25	0,1375
	Yen Marudut Aruan	0,25	0,125	0,16	0,125
	Heri Suandi	0,5	0,125	0,25	0,1375
	Pantun Tampubolon	0,5	0,125	0,25	0,1375
	Piki Prayoga	0,25	0	0,15	0,125
	Mohammad Haidir	0	0,5	0	0
	Sopan Irawan	0	0,125	0,16	0,125
	Ghazali Hasan	0,5	0	0,16	0,125
	Mhd.Yamin	0,75	0,125	0,25	0,135
	Taufik Hidayat	0,75	0,375	0,14	0,275
	Ahrun Harahap	0,5	0,125	0,25	0,1375
	Pairan	0,5	0,875	0,08	0,05
	Widodo	0,25	0,125	0,25	0,15
	Sudrajat	0,5	0,125	0,25	0,15
	Didit Permadi	0	0,75	0,08	0,05
	Arifin	0	0,125	0,16	0,125
	Januardi	0,25	0,125	0,25	0,175
	Misnan.S.	0,75	0,75	0,08	0,075
	Yuda Heria	0,5	0,125	0,25	0,275
C1		0,37	0,23	0,18	0,13
Cluster II	Edona F	0	0,125	0,5	0,375
	Sugirin	0,5	0,125	0,33	0,25
	Iqbal Fadilah	0	0,25	0,33	0,25
	Liono	0,5	0,625	0,5	0,375
	Iwan E. Hutabarat	0,5	0,125	0,33	0,5
	Romi Hidayat	0	0,5	0,33	0,275
	Sri Waluyo	0,5	0,5	0,33	0,325
	Santus Parsaulian	1	0,125	0,33	0,35
C2		0,375	0,296875	0,3725	0,3375
Cluster III	Muhammad Hariansyah	0,25	0	1	1
C3		0,25	0	1	1

Cluster yang terbentuk pada iterasi II sama dengan cluster yang terbentuk pada cluster I oleh karena itu proses pembentukan cluster berhenti pada iterasi II. Sehingga, Kelompok I / Cluster I masyarakat dengan tingkat ekonomi rendah, Kelompok II / Cluster II masyarakat dengan tingkat ekonomi tinggi dan Kelompok III / Cluster III kelompok masyarakat tingkat menengah.

4. KESIMPULAN

Penelitian ini membuktikan bahwa Algoritma K-Means efektif untuk mengelompokkan tingkat ekonomi masyarakat berdasarkan kriteria penghasilan, pekerjaan, dan penghasilan aset. Penggunaan normalisasi data memastikan hasil pengelompokan yang akurat karena kriteria data tidak saling mendominasi. Hasil pengelompokan membagi masyarakat menjadi 3 kelompok: tingkat ekonomi rendah, menengah, dan tinggi. Informasi ini bermanfaat untuk memahami ketimpangan ekonomi dan merancang kebijakan yang tepat sasaran untuk meningkatkan kesejahteraan masyarakat.

REFERENCES

- [1] D. A. Nasution, H. H. Khotimah, and N. Chamidah, "Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN," *Comput. Eng. Sci. Syst. J.*, vol. 4, no. 1, p. 78, 2019, doi: 10.24114/cess.v4i1.11458.
- [2] M. M. K-means, A. Solichin, and K. Khairunnisa, "Klasterisasi Persebaran Virus Corona (Covid-19) Di DKI Jakarta," vol. 5, no. 2, 2020.
- [3] E. Buulolo and R. Syahputra, "Implementasi Algoritma Clustering K-Means Untuk Mengelompokkan Mahasiswa Baru Yang Berpotensi (Studi Kasus : Stmik Budi Darma)," vol. 2, no. September, pp. 17–24, 2019.
- [4] T. P. Sutriyani, A. M. Siregar, and D. S. Kusumaningrum, "Implementasi Algoritma K-Means terhadap Pengelompokan Nilai Ujian Nasional Tingkat SMP di Provinsi Jawa Barat," *Techno Xplore J. Ilmu Komput. dan Teknol. Inf.*, vol. 3, no. 1, pp. 30–36, 2018, doi: 10.36805/technoxplore.v3i1.797.

- [5] A. D. Salman and B. A. Hussain, "Gene Expression Analysis via Spatial Clustering and Evaluation Indexing," *Iraqi J. Comput. Sci. Math.*, vol. 4, no. 1, pp. 24–34, 2023, doi: 10.52866/ijcsm.2023.01.01.004.
- [6] E. Buulolo, *Data Mining Untuk Perguruan Tinggi*. Yogyakarta: deepublish, 2020.
- [7] J. Suntoro, *DATA MINING: Algoritma dan Implementasi dengan Pemrograman php*. Elex Media Komputindo, 2019.
- [8] D. N. P. Sari and Y. L. Sukestiyarno, "Analisis cluster dengan metode K-Means pada persebaran kasus COVID-19 berdasarkan Provinsi di Indonesia," in *PRISMA, Prosiding Seminar Nasional Matematika*, 2021, pp. 602–610.
- [9] M. R. Kusnaldi, T. Gulo, and S. Aripin, "Penerapan Normalisasi Data Dalam Mengelompokkan Data Mahasiswa Dengan Menggunakan Metode K-Means Untuk Menentukan Prioritas Bantuan Uang Kuliah Tunggal," *J. Comput. Syst. Informatics*, vol. 3, no. 4, pp. 330–338, 2022.
- [10] W. Andriyani, A. H. Nasyuha, Y. Syahra, and B. Triaji, "Clustering Analysis of Poverty Levels in North Sumatra Province Using the Application of Data Mining with the K-Means Algorithm," *J. MEDIA Inform. BUDIDARMA*, vol. 7, no. 4, pp. 1971–1979, 2023.
- [11] M. Herviany, S. P. Delima, T. Nurhidayah, and K. Kasini, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Daerah Rawan Tanah Longsor Pada Provinsi Jawa Barat: Comparison of K-Means and K-Medoids Algorithms for Grouping Landslide Prone Areas in West Java Province," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 34–40, 2021.
- [12] I. Ramadhani and M. Megawati, "Implementasi Algoritma K-Means Untuk Klustering Data Produktivitas Kelapa Sawit: Implementation Of K-Means Algorithm For Palm Oil Productivity Data Clustering," *Indones. J. Inform. Res. Softw. Eng.*, vol. 3, no. 1, pp. 56–64, 2023.
- [13] F. Juliawati, R. Buaton, and R. Saragih, "Pengelompokan Data Mining Penerimaan Bantuan Pangan Non Tunai (BPNT) Menggunakan Metode Clustering (Studi Kasus: Kantor Desa Payabakung Hamparan Perak)," *Explorer (Hayward)*, vol. 3, no. 2, pp. 69–76, 2023.
- [14] Z. Nabila, A. R. Isnain, P. Permata, and Z. Abidin, "Analisis data mining untuk clustering kasus covid-19 di Provinsi Lampung dengan algoritma k-means," *J. Teknol. Dan Sist. Inf.*, vol. 2, no. 2, pp. 100–108, 2021.
- [15] A. Yulistira and R. Andika, "Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering," *J. Artif. Intell. Technol. Inf.*, vol. 1, no. 1, pp. 20–28, 2023.