

PENERAPAN DATA MINING CLUSTERING TERHADAP PEREKONOMIAN PADA KELURAHAN WONOREJO KECAMATAN MARPOYAN DAMAI PROVINSI RIAU MENGGUNAKAN METODE ALGORITMA K-MEANS

Egye Santoso¹, Susanto Hariyanto²

Fakultas Sains dan Teknologi, Universitas Buddhi Dharma
Jalan Imam Bonjol No. 41, Tangerang, Indonesia

Email: ¹EgyeSantoso0302@gmail.com, ²susanto.hariyanto@ubd.ac.id

Abstrak

Pandemi Covid-19 memberikan efek negatif yang cukup berdampak terhadap perekonomian sehingga pemerintah pusat berupaya untuk memberikan subsidi bantuan sosial untuk masyarakat tidak mampu serta kurang mampu sebagai bentuk penanggulangan bencana non alam yang terjadi. Namun, Tindakan tersebut belum tepat sasaran karena adanya faktor kedekatan sehingga berpotensi menyebabkan korupsi kolusi dan nepotisme. Untuk itu penelitian ini menerapkan pendekatan clustering data mining untuk mengidentifikasi kelompok perekonomian masyarakat, sebagai upaya pemerataan bantuan. Penerapan clustering dengan algoritma K-Means telah berhasil membentuk cluster 3 kelompok penerima bantuan di wilayah Kelurahan Wonorejo Kecamatan Marpoyan Damai kota Pekanbaru. Data yang digunakan berupa data keluarga yang terkena dampak pandemi Covid-19 tahun 2020 sebanyak 1910 data dengan mempergunakan atribut umur, penghasilan setiap bulan, jumlah tanggungan, nomor kartu keluarga, jenis pekerjaan dan nomor kategori yang selanjutnya akan diolah terlebih dahulu sesuai aturan data mining. Dari hasil Clustering tersebut menghasilkan cluster satu yaitu tidak mampu dengan data sebanyak 456 data, cluster 3 yaitu kurang mampu sebanyak 583 data, dan cluster 2 yaitu mampu sebanyak 544 data. Dari hasil tersebut maka masyarakat yang mendapatkan bantuan yaitu pada kelompok tidak mampu dan kurang mampu.

Kata Kunci

Clustering, K-Means, Covid-19, Kota Riau

Latar Belakang

Wabah Pandemi COVID-19 atau virus yang bernama ilmiah Severe acute respiratory syndrome coronavirus ini sudah muncul dan menyebar di Indonesia sejak awal 2019. Dari penyebaran wabah virus yang mengakibatkan infeksi pernapasan pada manusia tersebut menyebabkan banyak sektor dalam kegiatan manusia yang terkena dampak negative hingga berujung lumpuh. Oleh karena itu diwajibkannya pembatasan mobilitas dan jaga jarak antar individu oleh negara hingga organisasi kesehatan dunia atau WHO sebagai salah satu bentuk pencegahan penularan Virus SARS-CoV-2. Semua negara diberbagai belahan dunia terkena dampak dari menyebarnya wabah Pandemi COVID-19 ini, tidak terkecuali Negara Kesatuan Republik Indonesia. Sejak awal Maret 2019 Negara Kesatuan Republik Indonesia menghadapi wabah Pandemi COVID-19 dan mulai membuat banyak sektor kegiatan perekonomian nasional masyarakat di Indonesia mengalami perubahan dan ketidak stabilan.

Pada tanggal 5 Februari 2021 Badan Pusat Statistik merilis hasil angka pertumbuhan negara Indonesia tahun 2020 yang diukur dari Produk Domestik Bruto atau PDB atas dasar harga berlaku hingga mencapai Rp15.434,2 triliun serta PDB per kapita hingga mencapai Rp56,9 Juta atau

US\$3.911,7. Dalam rangkuman tersebut dijelaskan bahwa ekonomi negara Indonesia di tahun 2020 mengalami kontraksi pertumbuhan (Penurunan siklus ekonomi yang sangat dalam sehingga posisi angka PDB berada di kisaran minus) sebesar 2,07 persen (c-to-c) disbanding pada tahun 2019. Dari sektor produksi, kontraksi pertumbuhan yang terdalam terjadi pada Sektor Usaha Transportasi serta Pergudangan sebesar 15,04 persen. Sementara itu, dari sisi pengeluaran hampir seluruh komponen terjadi kontraksi, Komponen Ekspor Barang serta Jasa menjadi sektor dengan kontraksi yang terdalam sebesar 7,70 persen. Sementara, sektor Impor Barang dan Jasa yang menjadi faktor pengurang terkontraksi dengan nilai sebesar 14,71 persen (BPS, 2021).

Oleh karena itu, sebagai solusi terhadap permasalahan tersebut perlu adanya penggalian data dan pengolahan data yang lebih mendalam serta akurat terhadap perekonomian masyarakat khususnya pada kalangan masyarakat yang terdampak karena pandemi tersebut dan yang benar – benar layak dan pantas mendapatkan bantuan sosial dari pemerintah Negara Kesatuan Republik Indonesia. Dan dari latar belakang tersebut timbul gagasan untuk mengolah data perekonomian masyarakat khususnya di lingkungan Kelurahan wilayah Wonorejo yang berada di Kota Pekanbaru menggunakan metode data mining clustering agar menghasilkan suatu informasi data perekonomian yang nantinya dapat menentukan penerimaan subsidi bantuan sosial sesuai kategorinya.

Metode Penelitian

Data Mining

Data mining merupakan bidang ilmu pengetahuan komputer yang digunakan untuk menangani masalah pengambilan informasi dari database yang sangat besar dengan menggabungkan sebuah teknik dari statistik, pembelajaran mesin (machine learning), visualisasi data, pengenalan pola, serta database (Indah Werdiningsih, Barry Nuqoba and Muhammadun, 2020).

Proses data mining bertujuan untuk mengekstrak informasi dengan metode cerdas dari sekumpulan data kemudian diubah menjadi struktur yang memiliki nilai tambah berbentuk informasi yang selama ini belum diketahui dan dapat dipahami untuk penggunaan lebih lanjut.

Data Mining dikelompokkan kedalam beberapa jenis sesuai cara dan kebutuhan, yaitu :

1. Klasifikasi (classification)

Klasifikasi merupakan proses pembelajaran untuk menemukannya dan memetakannya (mengklasifikasi) suatu unsur (item) data dan sekumpulan model ke dalam salah satu dari beberapa *class* yang sudah didefinisikan atau ditentukan.

2. Asosiasi (Association)

Asosiasi adalah proses yang digunakan untuk mengenali hubungan dari kejadian - kejadian khusus maupun hubungan antar item dalam data atau proses dimana hubungan asosiasi timbul pada setiap kejadian maupun data yang menampilkan pola kombinasi dari item.

3. Clustering

Clustering adalah proses pengamatan dan pengelompokan atribut didalam data, dan membentuk kelas objek-objek yang mempunyai kemiripan. Clustering memiliki perbedaan yaitu tidak adanya pendefinisian kelas objek dalam pengklusteran yang biasa disebut unsupervised learning atau pembelajaran tidak terlatih. Clustering melakukan pembagian terhadap seluruh data menjadi beberapa kelompok yang mempunyai kesamaan data (homogeny), yang dimana kesamaan didalam satu kelompok akan memiliki nilai maksimum, sedangkan kesamaan didalam kelompok yang lain akan bernilai minimum.

4. Deskripsi

Deskripsi adalah suatu cara untuk menggambarkan sebuah pola dan kecenderungan yang terdapat di dalam kumpulan data. Deskripsi dari pola dan kecenderungan didalam kumpulan data sering memberikan kemungkinan penjelasan sebagai sebuah pola atau kecenderungan yang kemudian menjadi suatu informasi yang bermanfaat.

5. Prediksi (Forecasting)

Forecasting adalah cara atau teknik memprediksi data yang digunakan yang bertujuan untuk menghasilkan gambaran tentang nilai suatu data di masa yang mendatang sesuai dengan pengumpulan informasi dari jumlah informasi yang sangat besar. Dalam hal ini dapat dicontohkan dalam hal memprediksi naik turunnya nilai sebuah saham suatu perusahaan dalam kurun waktu periode tertentu.

Clustering

Clustering adalah proses pengamatan dan pengelompokan atribut didalam data, dan membentuk kelas objek-objek yang memiliki kemiripan. Clustering memiliki perbedaan yaitu tidak adanya pendefinisian kelas objek dalam pengklusteran yang biasa disebut unsupervised learning atau pembelajaran tidak terlatih. Clustering melakukan proses pembagian terhadap keseluruhan data menjadi beberapa kelompok yang mempunyai kesamaan data atau (homogeny), yang mana kesamaan dalam satu kelompok yang akan bernilai besar, sedangkan kesamaan dalam kelompok yang lain akan bernilai kecil. Pada dasarnya terdiri dua metode dalam *clustering* antaranya *Hierarchical Clustering Method* dan *NonHierarchical Clustering Method*. *Hierarchical Cluster Method* dipakai apabila belum adanya informasi jumlah cluster. Prinsipnya adalah mengatur semua objek ke dalam satu buah pohon/tree berdasarkan sebuah kriteria yang ditentukan. Sedangkan *Nonhierarchical clustering method* bertujuan untuk mengelompokkan sebuah objek ke dalam kondisi yang dapat dinyatakan jumlah kelompok di metode disampaikan oleh user.

K-Means

K-Means adalah metode algoritma clustering berbasis jarak yang melakukan proses membagi data ke dalam sejumlah cluster dan hanya bisa digunakan pada atribut numerik atau angka. Algoritma K-Means sangat terkenal dikarenakan simple serta kemampuannya untuk mengkluster data yang besar serta data outlier dengan waktu yang sangat cepat(Wahyudi *et al.*, 2020).

Algoritma ini termasuk kategori metode algoritma non-hierarki yang pada awalnya hanya mengambil beberapa komponen populasi untuk dijadikan sebagai pusat cluster awal, kemudian pusat kluster yang dinamakan Centroid dipilih secara acak dari kumpulan populasi data. Proses selanjutnya adalah menguji masing tiap komponen di dalam populasi data serta menandai komponen tersebut ke salah satu Centroid yang telah didefinisi tergantung dari nilai jarak terkecil antar komponen dengan tiap-tiap kluster. Posisi Centroid akan dihitung kembali sampai semua komponen data dapat digolongkan ke dalam tiap-tiap pusat kluster.

Metode K-Means merupakan metode clustering yang paling sederhana dan umum karena metode K-Means bisa mengidentifikasi data dalam skala besar dengan waktu perhitungan yang efisien.

Dalam menjalankan algoritma K-Means ada beberapa langkah yang harus dilakukan menurut (Irwansyah and Faisal, 2015), yaitu :

1. Tentukan jumlah cluster dari dataset yang mau dibagi.
2. pilih secara random data K sebagai pusat awal tempat cluster.

3. Untuk masing-masing tiap data, temukanlah pusat titik cluster yang terdekat. Dengan demikian, tandanya setiap pusat cluster mempunyai sebuah subset yang berasal dari dataset, sehingga akan mewakili bagian dari dataset. Oleh karena itu, sudah terbentuk cluster k: C1, C2, C3, ..., Ck. Rumus untuk menemukan pusat cluster yaitu :

$$d_{Euclidean}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

4. Untuk setiap cluster k, temukan titik pusat luasan *cluster*, dan perbaruilah lokasi dari setiap pusat ke nilai titik baru dari titik pusat tersebut.

5. Ulangilah dari langkah ke-3 sampai ke-4 hingga data yang ada pada setiap cluster akan terpusat atau selesai (centroid bernilai optimal).

Setelah dapat hasil cluster dari proses clustering menggunakan algoritma K-Means maka harus dilakukan proses evaluasi pengujian. Pengujian ini bisa dilakukan dengan memakai metode sum squared error atau (SSE). Selain menggunakan cara SSE dapat dengan menggunakan cara perhitungan melalui perbandingan antara Between - Class Variation (BCV) dan Within - Class Variation (WCV) di iterasi terakhir yang biasa disebut rasio. Apabila total hasil perhitungan pengujian yang diperoleh besar, maka semakin baik dan bagus juga tingkat kualitas clustering itu. BCV yaitu rata-rata dari nilai centroid, sedangkan untuk WCV adalah SSE nilai itu juga. Rumus perhitungan BCV sebagai berikut:

$$BCV = \frac{1}{nk} \sum_{i=0}^k d(mi, mi)$$

Sedangkan untuk rumus perhitungan BCV adalah sebagai berikut :

$$WCV = \sum_j^n \sum P \epsilon ci d(p, mi)^2$$

Dan Rumus perhitungan Rasio adalah sebagai berikut :

$$Rasio = \frac{BCV}{WCV}$$

HASIL

Dalam penelitian ini peneliti menggunakan metode Knowledge Discovery Databases (KDD) untuk mengolah data warga Kelurahan Wonorejo yang terdampak Covid-19 dengan beberapa tahap sebagai berikut :

Data Selection

Dalam tahapan Data selection ini penulis menggunakan data set yang berasal dari data primer yang telah diperoleh melalui cara kuesioner yang ditujukan kepada seluruh kepala keluarga di lingkungan Kelurahan Wonorejo. Pada seleksi data yang dilakukan dari data keluarga terdampak Covid-19 tersebut, penulis melakukan langkah pemilihan atribut tertentu yang akan digunakan didalam penelitian ini. Atribut yang dipilih adalah sebagai berikut :

Table 1. Atribut yang dipilih dari dataset

NO.	Atribut	Keterangan
1.	NIK KK	NIK Kepala Keluarga
2.	Nomor KK	Nomor KK
3.	Nama Lengkap KK	Jumlah anggota keluarga dalam satu KK
4.	JenisKelamin	Laki-Laki / Perempuan
5.	Usia	Usia Kepala Keluarga
6.	Alamat (KTP)	Alamat sesuai KTP
7.	Alamat (Tinggal)	Alamat sesuai tempat tinggal
8.	Pekerjaan	Pekerjaan Kepala Keluarga
9.	Penghasilan	Rata-rata Penghasilan per bulan
10.	Pengeluaran	Rata-rata pengeluaran per bulan
11.	Tanggungans	Anggota keluarga yang ditanggung

Analisis Kriteria

Proses pencarian objek yang dijadikan sebagai data didalam perhitungan yang bermanfaat dalam suatu kumpulan data pada proses pembuatan sistem yang akan dibangun. Kriteria yang akan digunakan pada studi kasus ini diambil berdasarkan dari data yang sudah didapatkan. Dari beberapa atribut tersebut maka kriteria untuk tahapan transformasi data yang dipakai adalah sebagai berikut :

Table 2. Analisis Kriteria

No.	Kriteria	DEFINISI	Pengelompokkan Kriteria	Bobot
1.	Pendapatan	Penghasilan rata-rata kepala keluarga	< 500.000	5
			500.000 – 1 Jt	4
			1 Jt – 1.5 Jt	3
			1.5 Jt – 2 Jt	2
			>2 Jt	1
2.	Pekerjaan	Jenis – Jenis Pekerjaan dan Bobot nilai terhadap Jenis pekerjaan sesuai dengan Peraturan Kementerian Ketenagakerjaan dan Badan Pusat	Pekerja Kasar	10
			Operator & Perakit mesin	9
			Pekerja Pengolah, Kerajinan	8
			Pekerja Terampil pertanian, Kehutanan, dan Perikanan	7

		Statistik.(Isnainingsih <i>et al.</i> , 2014)	Tenaga Usaha Jasa dan Tenaga Penjualan	6
			Tenaga Tata Usaha	5
			Teknisi dan Asisten Profesional	4
			Profesional	3
			Manajer	2
			TNI dan POLRI	1
3.	Usia	Usia dari setiap kepala keluarga	>61	4
			41-60	3
			18 – 40	2
			<17	1
4.	Tanggungan	Jumlah anggota keluarga yang ditanggung	>7	5
			5 – 6	4
			3 – 4	3
			1 – 2	2
			<1	1
5.	Pengeluaran	Pengeluara Rata-Rata setiap bulan	>2 Jt	5
			1.5 Jt – 2 Jt	4
			1 Jt – 1.5 Jt	3
			500.000 – 1 Jt	2
			< 500.000	1

Data Cleaning/Pre-Processing

Pada tahapan Data Cleaning ini akan dilakukan penghapusan data-data yang kosong, jika ada data ganda akan dilakukan penghapusan duplikasi data, memeriksa data yang bersifat tidak konsisten, memperbaiki setiap kesalahan pada data serta melakukan proses penghapusan beberapa buah atribut yang tidak dipakaikan pada penelitian ini. Dsalam tahapan ini juga dilakukan sebuah proses enrichment atau memperkaya data yang telah ada dengan data maupun informasi lain yang relevan dan dibutuhkan dalam proses tersebut.

Table 3. Data yang di Cleaning

No	NIK Kepala Keluarga (Kode)	Usia	Pekerjaan (Jenis Usaha)	Rata-Rata Penghasilan Perbulan (Rp)	Rata-Rata Pengeluaran Perbulan (Rp)	Jumlah Tanggungan
1	1	53	Wiraswasta	1200000	1200000	4
2	2	49	Swasta	2000000	2000000	4

3	3	32	Swasta	1000000	1000000	3
...
1581	1581	54	Swasta	700000	700000	4
1582	1582	45	Buruh Lepas	1000000	1000000	4
1583	1583	42	Wiraswasta	750000	750000	4

Data Transformation

Pada tahapan Data Transformation bertugas dan berfungsi untuk mengurangi jumlah variabel dan data yang tidak terlalu di perlukan, mengubah data ke dalam bentuk analisis serta memodelkan data supaya sesuai dengan apa yang analisis diharapkan serta format data yang dibutuhkan oleh algoritma data mining.

Table 4. Data yang di Transformasi

1	1	2	6	3	3	4
2	2	2	5	2	4	4
3	3	2	5	4	2	3
...
1582	1582	2	8	3	3	4
1583	1583	2	6	1	5	3

Perhitungan Manual

Dalam melakukan pengujian data melalui aplikasi Microsoft Excel dengan mengambil masing – masing 3 data dari setiap cluster yang dihasilkan di aplikasi Rapid Miner sehingga analisis menghitung 9 data sample sebagai berikut :

Table 5. Sample data

Penghasilan	Pengeluaran	Tanggungan	Usia	Pekerjaan
4	2	3	3	6
4	2	2	2	5
3	3	3	2	10
2	4	3	2	10
1	5	3	2	5
1	5	4	2	6

2	4	4	2	5
4	2	3	2	3
2	4	3	2	9

Selanjutnya menentukan nilai titik centroid awal secara acak yang selanjutnya digunakan sebagai pusat awal lokasi kluster. Berikut ini adalah nilai acak K yang akan digunakan oleh analisis :

Table 6. Titik Centroid Acak

4	2	3	3	6
2	4	3	2	9
1	5	4	2	6

Tahapan selanjutnya adalah menghitung jarak dari setiap objek data dengan pusat setiap cluster dengan menggunakan rumus sebagai berikut :

$$d_{Euclidean}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Contoh Perhitungan iterasi 1 :

$$(1|C1) = \sqrt{(4 - 4)^2 + (2 - 2)^2 + (3 - 3)^2 + (3 - 3)^2 + (6 - 6)^2} \\ = 0$$

$$(1|C2) = \sqrt{(4 - 2)^2 + (2 - 4)^2 + (3 - 3)^2 + (3 - 2)^2 + (6 - 9)^2} \\ = \sqrt{4 + 4 + 0 + 1 + 9} \\ = \sqrt{18} \\ = 4,242641$$

$$(1|C3) = \sqrt{(4 - 1)^2 + (2 - 5)^2 + (3 - 4)^2 + (3 - 2)^2 + (6 - 6)^2} \\ = \sqrt{9 + 9 + 1 + 1 + 0} \\ = \sqrt{20} \\ = 4,472136$$

Table 7. Hasil perhitungan iterasi 1

Jarak C1	Jarak C2	Jarak C3	Cluster
0	4,242641	4,472136	c1

1,732051	5	4,795832	c1
4,358899	1,732051	5	c2
5	1	4,358899	c2
4,472136	4,242641	1,414214	c3
4,472136	3,464102	0	c3
3,316625	4,123106	1,732051	c3
3,162278	6,63325	5,291503	c1
4,242641	0	3,464102	c2

Dari hasil perhitungan iterasi pertama maka akan dipilih sesuai jarak minimum atau terdekat dari setiap pusat *cluster*, dan jarak tersebut menunjukkan bahwa dalam data berada didalam kelompok yang sudah ditentukan.

Selanjutnya setelah mengetahui anggota setiap *cluster* proses selanjutnya adalah menghitung centroid yang akan dipakai pada iterasi ke 2 dengan cara menjumlahkan rata-rata dari data hasil setiap *cluster* yang sebelumnya telah terbentuk dengan cara sebagai berikut :

$$C1 = (4 + 4 + 4) / 3, \quad (2 + 2 + 2) / 3, (3 + 2 + 3) / 3, (3 + 2 + 3) / 3, (6 + 5 + 3) / 3$$

$$C1 = 4,2, \quad 2,666667, \quad 2,333333, \quad 4,666667$$

$$C2 = (3 + 2 + 2) / 3, \quad (3 + 4 + 4) / 3, (3 + 3 + 3) / 3, (2 + 2 + 2) / 3, (10 + 10 + 9) / 3$$

$$C2 = 2,333333, \quad 3,666667, \quad 3, \quad 2, \quad 9,666667$$

$$C3 = (1 + 1 + 2) / 3, \quad (5 + 5 + 4) / 3, (3 + 4 + 4) / 3, (2 + 2 + 2) / 3, (5 + 6 + 5) / 3$$

$$C3 = 1,333333, \quad 4,666667, \quad 3,666667, \quad 2, \quad 5,333333$$

Table 8. Titik Centroid Baru

4	2	2,666667	2,333333	4,666667
2,333333	3,66667	3	2	9,666667
1,333333	4,6667	3,666667	2	5,333333

Selanjutnya adalah mengulangi perhitungan data pertama dengan titik centroid baru.

Contoh Perhitungan iterasi 2 :

$$(1|C1) = \sqrt{(4 - 4)^2 + (2 - 2)^2 + (3 - 2,666667)^2 + (3 - 2,333333)^2 + (6 - 4,666667)^2}$$

$$= \sqrt{0 + 0 + 0,111111 + 0,444445 + 1,777778}$$

$$\begin{aligned}
&= \sqrt{2,333333} \\
&= 1,527525 \\
(1|C2) &= \sqrt{(4 - 2,333333)^2 + (2 - 3,666667)^2 + (3 - 3)^2 + (3 - 2)^2 + (6 - 9,666667)^2} \\
&= \sqrt{2,77778 + 2,77778 + 0 + 1 + 13,4444469} \\
&= \sqrt{19,0000029} \\
&= 4,472135955 \\
(1|C3) &= \sqrt{(4 - 1,333333)^2 + (2 - 4,666667)^2 + (3 - 3,66667)^2 + (3 - 2)^2 + (6 - 5,333333)^2} \\
&= \sqrt{7,111113 + 7,111113 + 0,4444449 + 1 + 0,4444449} \\
&= \sqrt{16,1111158} \\
&= 4,01386486
\end{aligned}$$

Table 9. Hasil perhitungan iterasi 2

C1	C2	C3	Cluster
1,527525	4,472136	4,013865	c1
0,816497	5,322906	4,136558	c1
5,537749	1	5,270463	c2
6,055301	0,57735	4,807402	c2
4,281744	5,033223	0,881917	c3
4,654747	4,242641	0,881917	c3
3,162278	4,795832	1,054093	c3
1,732051	7,071068	4,484541	c1
5,196152	0,816497	3,844188	c2

Kemudian proses selanjutnya adalah menghitung kembali centroid dengan cara menjumlahkan rata-rata dari data hasil setiap *cluster* yang sebelumnya telah terbentuk dengan cara sebagai berikut :

$$\begin{aligned}
C1 &= (4 + 4 + 4) / 3, \quad (2 + 2 + 2) / 3, (3 + 2 + 3) / 3, (3 + 2 + 3) / 3, (6 + 5 + 3) / 3 \\
C1 &= 4,2, \quad 2,666667, \quad 2,333333, \quad 4,666667 \\
C2 &= (3 + 2 + 2) / 3, \\
&\quad (3 + 4 + 4) / 3, (3 + 3 + 3) / 3, (2 + 2 + 2) / 3, (10 + 10 + 9) / 3 \\
C2 &= 2,333333, \quad 3,666667, \quad 3, \quad 2, \quad 9,666667 \\
C3 &= (1 + 1 + 2) / 3, \quad (5 + 5 + 4) / 3, (3 + 4 + 4) / 3, (2 + 2 + 2) / 3, (5 + 6 + 5) / 3 \\
C3 &= 1,333333, \quad 4,666667, \quad 3,666667, \quad 2, \quad 5,333333
\end{aligned}$$

Table 10. Titik Centroid Baru

4	2	2,666667	2,333333	4,666667
2,333333	3,66667	3	2	9,666667
1,333333	4,6667	3,666667	2	5,333333

Pada perhitungan ini titik centroid yang ada pada iterasi 1 dan 2 menunjukkan tidak adanya perubahan dan dapat dinyatakan perhitungan tersebut telah selesai dan dapat ditarik hasil kesimpulan yang dapat menentukan kelompok masyarakat mampu, kurang mampu dan tidak mampu.

Table 11. Hasil Perhitungan Manual

C1	C2	C3	Cluster	Status
1,527525	4,472136	4,013865	c1	Mampu
0,816497	5,322906	4,136558	c1	Mampu
5,537749	1	5,270463	c2	Kurang Mampu
6,055301	0,57735	4,807402	c2	Kurang Mampu
4,281744	5,033223	0,881917	c3	Tidak Mampu
4,654747	4,242641	0,881917	c3	Tidak Mampu
3,162278	4,795832	1,054093	c3	Tidak Mampu
1,732051	7,071068	4,484541	c1	Mampu
5,196152	0,816497	3,844188	c2	Kurang Mampu

Tampilan Program

Dibawah ini merupakan tampilan aplikasi berbasis desktop yang telah dirancang untuk mengelompokkan data ekonomi sesuai clusternya :



Gambar 1 Halaman Home Aplikasi

Halaman ini menampilkan tombol yang difungsikan untuk berpindah ke halaman sesuai yang diinginkan.

The image shows a web application window with a red and white background. The form has a pink sidebar on the left with labels: "NIK", "Nama Lengkap", "Jenis Kelamin", "Penghasilan", "Pengeluaran", "Tanggungan", "Usia", and "Pekerjaan". To the right of each label is an input field. The "Jenis Kelamin" field has a dropdown menu with "L" selected. The "Penghasilan" field has a dropdown menu with ">2jt" selected. The "Pengeluaran" field has a dropdown menu with "<500000" selected. The "Tanggungan" field has a dropdown menu with "<1" selected. The "Usia" field has a dropdown menu with "<17" selected. The "Pekerjaan" field has a dropdown menu with "Pekerja Kasar" selected. At the top right of the form area is a "Kembali" button. Below the form are two buttons: "Input Data" and "Cek Penerima Bantuan sosial".

Gambar 2 Halaman Input Data

Pada halaman ini user akan melakukan proses penginputan data yang selanjutnya bisa langsung disimpan kedalam database yang ada pada *Xampp* dan user juga dapat langsung melakukan proses pengelompokkan data sesuai cluster menggunakan tombol "Cek Penerima Bantuan Sosial" kemudian akan muncul hasil dari cluster ekonomi tersebut.

NIK: 3871100211000005

Nama Lengkap: Sugeng Pratama

Jenis Kelamin: L

Penghasilan: >2jt

Pengeluaran: >2jt

Tanggungan: 1 - 2

Usia: 18-40

Pekerjaan: TNI dan POLRI

Message: Hasil Perhitungan : 3.0
Anda masuk kategori Mampu

Buttons: Kembali, Input Data, Cek Penerima Bantuan sosial

Gambar 3 Fungsi Clustering

NIK	Nama	Jenis Kelamin	Penghasilan	Pengeluaran	Tanggungan	Usia	Pekerjaan
3671006...	Ahmad S...	L	>2jt	>2jt	3 - 4	18-40	Pekerja ...
3671006...	Dewi Latif	P	1.5jt-2jt	>2jt	3 - 4	18-40	Tenaga ...
3671100...	Safrudin ...	L	1.5jt-2jt	>2jt	3 - 4	18-40	Tenaga ...
3671101...	Ahmad S...	L	>2jt	>2jt	1 - 2	18-40	TNI dan ...
3671101...	Maharani...	P	>2jt	1.5jt-2jt	<1	18-40	Profesional
3671102...	Nanang ...	L	1.5jt-2jt	>2jt	3 - 4	18-40	Operator...
3671102...	Praditya ...	L	>2jt	>2jt	3 - 4	41-60	Tenaga ...
3671201...	Fredi Kur...	L	>2jt	>2jt	<1	<17	Teknisi d...
3867001...	Siti Abibah	P	1.5jt-2jt	>2jt	1 - 2	41-60	Pekerja ...
3871004...	Abdul Ha...	L	>2jt	>2jt	5 - 6	41-60	Tenaga ...

Buttons: Kembali, HAPUS DATA

Gambar 4 Halaman Database Input Data

Dalam halaman database ini analis menambahkan fungsi button hapus untuk menghapus data yang salah input oleh user jika dibutuhkan dalam penyimpanan database. Berikut ini adalah tampilan ketika menghapus data yang ada dalam database.

NIK	Nama	Status
3671006100001543	Ahmad Sugandi	kurang mampu
3671006100003522	Dewi Latif	kurang mampu
3871004100000058	Abdul Hartono	kurang mampu
3867001500000051	Siti Abibah	kurang mampu
3671100610000001	Safrudin Mahu	kurang mampu
3671101905000001	Maharani Dwiyantri	Mampu
3671102812000004	Praditya Yoga	kurang mampu
3671101102000003	Ahmad Sutoyo	Mampu
3671102501000005	Nanang Kurniawan	kurang mampu
3671201012000002	Fredi Kurniawan	kurang mampu
3671202205000003	Komang Baharudin	kurang mampu
3671107110000002	Denamika	kurang mampu

Gambar 5 Halaman Database Clustering.

Kesimpulan

Berdasarkan penelitian analisis serta rancangan aplikasi sederhana yang telah dibuat, analisis dapat menarik kesimpulan untuk Penerapan Data Mining Clustering terhadap perekonomian pada Kelurahan Wonorejo Kecamatan Marpoyan Damai Provinsi Riau menggunakan metode algoritma K-Means adalah sebagai berikut :

1. Peneliti berhasil melakukan proses pengelompokan data keluarga di lingkungan masyarakat Kelurahan Wonorejo Kecamatan Marpoyan Damai Provinsi Riau,
2. Peneliti berhasil merancang sebuah system aplikasi sederhana untuk mengolah dan mengelompokkan data perekonomian masyarakat.
3. Terbentuknya 3 Cluster ekonomi masyarakat yang terdiri dari
 - a. C1 dengan nilai rata - rata jarak dalam nilai centroid 3.868 sebagai kategori mampu.
 - b. C2 dengan nilai rata - rata jarak dalam nilai centroid 2.657 sebagai kategori kurang mampu
 - c. C3 dengan nilai rata - rata jarak dalam nilai centroid 2.520 sebagai kategori tidak mampu
4. Dan dari ketiga Cluster tersebut maka yang layak diberikan bantuan sosial adalah Cluster pertama dan Cluster ketiga sebagai kategori penerima bantuan sosial masyarakat kurang mampu dan tidak mampu di kelurahan Wonorejo Kecamatan Marpoyan Damai Provinsi Riau,
5. Dari 1583 data yang diolah menggunakan Rapid miner dengan menggunakan algoritma Clustering K-Means mendapatkan jumlah data disetiap kelompok sebagai berikut :
 - a. Cluster Mampu sebanyak 544 data kepala keluarga
 - b. Cluster Kurang Mampu sebanyak 583 data kepala keluarga
 - c. Cluster Tidak Mampu sebanyak 456 data kepala keluarga.

Saran

Penelitian ini dilakukan oleh analisis dan analisis sadar bahwa penelitian ini tidak luput dari berbagai bentuk kekurangan yang masih harus banyak diperbaiki dan ditambahkan, Berikut ini adalah beberapa saran yang dimiliki analisis terhadap penelitian :

1. Menambahkan metode perhitungan kualitas cluster, karena masih kurangnya perhitungan manual terhadap kualitas cluster yang dibuat oleh analisis
2. Mengembangkan serta menambahkan berbagai fitur dan fungsi tambahan dalam aplikasi pengolahan data perekonomian masyarakat serta memperbaharui tampilan agar lebih menarik dan fungsional sesuai kebutuhan.

3. Terkait aplikasi yang dirancang memerlukan banyak revisi dalam logic code terutama perhitungan untuk menentukan cluster pada data baru yang baru diinput oleh user.

Ucapan Terima Kasih

Terima Kasih kepada seluruh teman dan orang tua atas doa serta dukungan yang telah diberikan.

Referensi :

[1] BPS (2021) 'Pertumbuhan Ekonomi Indonesia Triwulan IV-2020', *Www.Bps.Go.Id*, (13), p. 12. Available at: <https://www.bps.go.id/pressrelease/2021/02/05/1811/ekonomi-indonesia-2020-turun-sebesar-2-07-persen--c-to-c-.html>.

[2] Indah Werdiningsih, S. S. M. K., Barry Nuqoba, S. S. M. K. and Muhammadun, S. S. M. S. (2020) *Data Mining Menggunakan Android, Weka, dan SPSS*. Airlangga University Press. Available at: <https://books.google.co.id/books?id=xEwCEAAAQBAJ>.

[3] Isnaningsih, T. R. *et al.* (2014) 'No Title'.

[4] Setiawan, I. (2018) 'Knowledge Discovery In Databases (KDD) Terhadap Customer Reviews Pada Situs E-Commerce Oleh Program Studi Sistem Informasi', (09031281621045).

[5] Wahyudi, M. *et al.* (2020) *Data Mining: Penerapan Algoritma K-Means Clustering dan K-Medoids Clustering*. Yayasan Kita Menulis. Available at: <https://books.google.co.id/books?id=wQnhDwAAQBAJ>.