

**ANALISIS PERBANDINGAN METODE K-MEANS DAN FUZZY C-  
MEANS PADA DATA RASIO GINI DI SUMATERA UTARA**

**SKRIPSI**

**DISUSUN OLEH**

**MHD HUSEIN PASARIBU**

**NPM. 2009010112**



**UMSU**  
Unggul | Cerdas | Terpercaya

**PROGRAM STUDI SISTEM INFORMASI**

**FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI**

**UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**

**MEDAN**

**2024**

**ANALISIS PERBANDINGAN METODE K-MEANS DAN FUZZY C-  
MEANS PADA DATA RASIO GINI DI SUMATERA UTARA**

**SKRIPSI**

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana  
Komputer (S.Kom) dalam Program Studi Sistem Informasi pada Fakultas  
Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah  
Sumatera Utara**

**MHD HUSEIN PASARIBU**

**NPM. 2009010112**

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI  
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**

**MEDAN**

**2024**

## LEMBAR PENGESAHAN

Judul Skripsi : Analisis Perbandingan Metode K-Means dan Fuzzy C-Means pada Data Rasio Gini di Sumatera Utara  
Nama Mahasiswa : Mhd Husein Pasaribu  
NPM : 2009010112  
Program Studi : Sistem Informasi

Menyetujui

**Komisi Pembimbing**



**(Ferdy Riza, S.T., M.Kom)**

**NIDN. 0103068901**

**Ketua Program Studi**



**(Martiano, S.Pd., S.Kom., M.Kom)**

**NIDN. 0128029302**

**Dekan**



**(Dr. Al-Khowarizmi, S.Kom., M.Kom.)**

**NIDN. 01270992**

## PERNYATAAN ORISINALITAS

### ANALISIS PERBANDINGAN METODE K-MEANS DAN FUZZY C-MEANS PADA DATA RASIO GINI DI SUMATERA UTARA

#### SKRIPSI

Saya menyatakan bahwa karya tulis ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, 07 Mei 2024

Yang membuat pernyataan



Mhd Husein Pasaribu

NPM. 2009010112

**PERNYATAAN PERSETUJUAN PUBLIKASI  
KARYA ILMIAH UNTUK KEPENTINGAN  
AKADEMIS**

Sebagai sivitas akademika Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan dibawah ini:

Nama : Mhd Husein Pasaribu  
NPM : 2009010112  
Program Studi : Sistem Informasi  
Karya Ilmiah : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bebas Royalti Non-Eksekutif (*Non-Exclusive Royalty free Right*) atas penelitian skripsi saya yang berjudul:

**ANALISIS PERBANDINGAN METODE K-MEANS DAN FUZZY C-  
MEANS PADA DATA RASIO GINI DI SUMATERA UTARA**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk database, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya.

Medan, 07 Mei 2024

Yang membuat pernyataan



Mhd Husein Pasaribu

NPM. 2009010112

## RIWAYAT HIDUP

### DATA PRIBADI

Nama Lengkap : Mhd Husein Pasaribu  
Tempat dan Tanggal Lahir : Bandung, 15 Mei 2001  
Alamat Rumah : Jl. Jermal IV, No. 51, Medan Denai  
Telepon/Faks/HP : 0823-6232-0668  
E-mail : huseinpasaribu1945@gmail.com  
Instansi Tempat Kerja : - (Belum Bekerja)  
Alamat Kantor : - (Belum Bekerja)

### DATA PENDIDIKAN

SD : SD SWASTA MUHAMMADIYAH 02 TAMAT: 2013  
SMP : SMP SWASTA PERTIWI, MEDAN TAMAT: 2016  
SMA : SMA NEGERI 03 MEDAN TAMAT: 2019

## KATA PENGANTAR



Puji dan syukur penulis ucapkan atas kehadiran Allah SWT. yang telah memberikan rahmat dan karunianya yang tiada tara kepada kita semua dan shalawat beriring salam disampaikan kepada Nabi Muhammad SAW., sehingga penulis dapat menyelesaikan penyusunan skripsi dengan judul penelitian “Analisis Perbandingan Metode K-Means dan Fuzzy C-Means pada Data Rasio Gini di Sumatera Utara” yang dimaksudkan sebagai salah satu syarat untuk memperoleh gelar sarjana Komputer (S.Kom) di Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara. Penulis menyadari sepenuhnya bahwa dalam penyusunan proposal penelitian ini, penulis memperoleh bantuan, masukan, bimbingan, arahan, serta dukungan dari berbagai pihak. Oleh karena itu dalam kesempatan kali ini penulis bermaksud untuk berterima kasih kepada berbagai pihak atas dukungan serta doa dalam penyelesaian skripsi, adapun ucapan terima kasih penulis tujukan kepada:

1. Bapak Prof. Dr. Agussani, M.AP., Rektor Universitas Muhammadiyah Sumatera Utara (UMSU),
2. Bapak Dr. Al-Khowarizmi, S.Kom., M.Kom. Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU,
3. Bapak Martiano, S.Pd., S.Kom., M.Kom. Ketua Program Studi Sistem Informasi,
4. Ibu Yoshida Sary, S.Kom., M.Kom. Sekretaris Program Studi Sistem Informasi,
5. Bapak Ferdy Riza, S.T., M.Kom. Dosen Pembimbing penulis skripsi,
6. Bapak Prof. Dr. Fajar Pasaribu, S.E., M.Si. Selaku ayah dari penulis,
7. Ibu Prof. Dr. Widia Astuty, SE., M.Si., Ak., CA., CPAi. Selaku ibunda dari penulis, serta
8. Seluruh pihak yang terlibat baik secara langsung maupun tidak langsung yang tidak dapat penulis ucapkan satu-persatu sehingga riset ini dapat terselesaikan dengan semestinya.

## ANALISIS PERBANDINGAN METODE K-MEANS DAN FUZZY C-MEANS PADA DATA RASIO GINI DI SUMATERA UTARA

### ABSTRAK

Algoritma K-Means dan Fuzzy C-Means adalah dua algoritma yang cukup populer dikalangan peneliti karena performanya yang baik dalam mengklasterkan kumpulan data. Baik K-Means maupun Fuzzy C-Means, keduanya cukup baik untuk menampilkan pola-pola tertentu yang terkandung didalam DataSet yang dimiliki. Namun, perlu dicari tahu lebih jauh metode manakah yang paling efektif untuk mengklasterkan kumpulan data. Dalam riset ini, peneliti membandingkan kedua algoritma melalui aspek efisiensi waktu, keakuratan data, dan nilai Dunn Index. Peneliti akan melakukan pengujian kepada kedua metode tersebut dengan menggunakan data uji berupa penyusun rasio GINI yakni pengeluaran makanan dan non-makanan di kalangan masyarakat kabupaten/kota yaitu kota Medan, kota Binjai, dan Kabupaten Deli Serdang yang peneliti peroleh datanya melalui penyebaran kuesioner. Dalam mengimplementasikan algoritma untuk keperluan pengujian, peneliti memanfaatkan bahasa pemrograman VB.NET sebagai sarana pengembangan sistem. Berdasarkan analisis yang telah peneliti peroleh, didapati kesimpulan bahwa K-Means lebih unggul daripada Fuzzy C-Means baik dari aspek efisiensi waktu, serta nilai Dunn Indeks. Sedangkan pada aspek keakuratan data, K-Means dan Fuzzy C-Meansimbang dalam merepresentasikan data yang sesuai dengan kondisi lapangan yang terjadi apabila dibandingkan dengan tingkat rasio GINI yang dipublikasikan oleh BPS.

Kata Kunci: K-Means; Fuzzy C-Means; Rasio GINI; VB.NET

# COMPARATIVE ANALYSIS OF K-MEANS AND FUZZY C-MEANS METHODS ON GINI RATIO DATA IN NORTH SUMATRA

## ABSTRACT

The K-Means and Fuzzy C-Means algorithms are two popular algorithms among researchers due to their good performance in clustering data sets. Both K-Means and Fuzzy C-Means are effective in revealing certain patterns within the dataset. However, it is necessary to further investigate which method is most effective for clustering data. In this research, the researcher compares both algorithms in terms of time efficiency, data accuracy, and Dunn Index value. The researcher will test both methods using test data consisting of the GINI ratio components, namely food and non-food expenditures in the cities of Medan, Binjai, and Deli Serdang, obtained through questionnaire distribution. In implementing algorithms for testing purposes, researchers use VB.NET programming languages as a means of system development. Based on the analysis that the researcher has obtained, it is concluded that K-Means is superior to Fuzzy C-Means both in terms of time efficiency, as well as the value of the Dunn Index. Meanwhile, in terms of data accuracy, K-Means and Fuzzy C-Means are balanced in representing data that is in accordance with the field conditions that occur when compared to the GINI ratio level published by BPS.

Keywords: K-Means; Fuzzy C-Means; GINI Ratio; VB.NET

## DAFTAR ISI

<b>LEMBAR PENGESAHAN .....</b>	<b>i</b>
<b>PERNYATAAN ORISINALITAS.....</b>	<b>ii</b>
<b>PERNYATAAN PERSETUJUAN PUBLIKASI .....</b>	<b>iii</b>
<b>RIWAYAT HIDUP .....</b>	<b>iv</b>
<b>KATA PENGANTAR .....</b>	<b>v</b>
<b>ABSTRAK .....</b>	<b>vi</b>
<b>ABSTRACT .....</b>	<b>vii</b>
<b>DAFTAR ISI.....</b>	<b>viii</b>
<b>DAFTAR TABEL .....</b>	<b>x</b>
<b>DAFTAR GAMBAR.....</b>	<b>xi</b>
<b>BAB I. PENDAHULUAN.....</b>	<b>1</b>
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah .....	3
1.3. Batasan Masalah.....	4
1.4. Tujuan Penelitian.....	4
1.5. Manfaat Penelitian.....	5
<b>BAB II. LANDASAN TEORI .....</b>	<b>7</b>
2.1. Data Mining.....	7
2.2. Algoritma Klustering.....	10
2.3. Logika Fuzzy.....	14
2.4. K-Means Clustering .....	18
2.5. Fuzzy C-Means Clustering.....	23
2.6. Rasio Gini (Indeks Gini) .....	30

2.7. Microsoft Visual Basic.Net .....	31
2.8. Unified Modelling Language (UML).....	32
2.8.1. Use Case Diagram.....	33
2.9. Penelitian Terdahulu.....	35
<b>BAB III. METODOLOGI PENELITIAN .....</b>	<b>37</b>
3.1. Jenis Penelitian .....	37
3.2. Tahapan Penelitian.....	37
3.3. Flowchart Sistem.....	39
3.4. Defenisi Operasional .....	41
3.5. Teknik Pengumpulan Data .....	43
3.6. Teknik Pengambilan Sample .....	44
3.7. Tempat Dan Jadwal Penelitian .....	45
<b>BAB IV. HASIL DAN PEMBAHASAN .....</b>	<b>46</b>
4.1. Data Awal .....	46
4.2. Contoh Perhitungan Manual .....	47
4.3. Implementasi User-Interface .....	54
4.4. Hasil Uji K-Means Dan Fuzzy C-Means Clustering .....	60
4.5. Komparasi Hasil K-Means Dan Fuzzy C-Means Clustering .....	68
<b>BAB V. KESIMPULAN DAN SARAN.....</b>	<b>72</b>
5.1. Kesimpulan.....	72
5.2. Saran.....	73
<b>DAFTAR PUSTAKA.....</b>	<b>74</b>
<b>LAMPIRAN</b>	

## DAFTAR TABEL

	<b>HALAMAN</b>
TABEL 2.1. Tabulasi Hasil Penelitian Terdahulu	35
TABEL 3.1. Defenisi Operasional	42
TABEL 4.1. Data Awal Penelitian Dalam Ribuan Rupiah (000)	47
TABEL 4.2. Sample Data Contoh Perhitungan Manual K-Means	47
TABEL 4.3. Sample Data Contoh Perhitungan Fuzzy C-Means	50
TABEL 4.4. Tabulasi Jumlah Iterasi K-Means Clustering	63
TABEL 4.5. Tabulasi Nilai Dunn Index K-Means Clustering	65
TABEL 4.6. Tabulasi Jumlah Iterasi Fuzzy C-Means Clustering	66
TABEL 4.7. Tabulasi Nilai Dunn Index Fuzzy C-Means Clustering	68
TABEL 4.8. Tabulasi Perbandingan Jumlah Iterasi KM & FCM	69
TABEL 4.9. Tabulasi Perbandingan Dunn Index KM & FCM	70

## DAFTAR GAMBAR

	<b>HALAMAN</b>
GAMBAR 2.1. Visualisasi Fuzzy	17
GAMBAR 2.2 Diagram Alur K-Means	22
GAMBAR 2.3. Diagram Alur Fuzzy C-Means	29
GAMBAR 2.4. Logo Visual Basic	31
GAMBAR 2.5. Use Case Diagram	34
GAMBAR 3.1. Alur Tahapan Penelitian	38
GAMBAR 3.2. Flowchart Sistem	40
GAMBAR 4.1. Tampilan Loading Sistem	54
GAMBAR 4.2. Tampilan Menu Utama Aplikasi	55
GAMBAR 4.3. Tampilan Menu About Us	55
GAMBAR 4.4. Tampilan Menu Data Processing	56
GAMBAR 4.5. Tampilan Menu Data Processing Setelah Formulir Terisi	57
GAMBAR 4.6. Notifikasi Ketika Terdapat Formulir Yang Tidak Terisi	58
GAMBAR 4.7. Tampilan Hasil Pengolahan Algoritma K-Means	58
GAMBAR 4.8. Tampilan Hasil Pengolahan Algoritma Fuzzy C-Means	59
GAMBAR 4.9. Tampilan Hasil Pengolahan K-Means Kota Medan	63
GAMBAR 4.10. Tampilan Hasil Pengolahan K-Means Kota Binjai	64
GAMBAR 4.11. Tampilan Hasil Pengolahan K-Means Deli Serdang	64
GAMBAR 4.12. Tampilan Hasil Pengolahan FCM Kota Medan	66
GAMBAR 4.13. Tampilan Hasil Pengolahan FCM Kota Binjai	67
GAMBAR 4.14. Tampilan Hasil Pengolahan FCM Deli Serdang	67

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang Masalah

Ketidaksetaraan ekonomi adalah kondisi dimana suatu kelompok masyarakat di satu wilayah yang sama memiliki perbedaan jumlah pendapatan yang kontras antara satu individu dengan individu lainnya. Fenomena ini merupakan hal yang tidak diinginkan oleh siapapun sebab dengan adanya ketidaksetaraan ekonomi yang timbul ditengah masyarakat dapat mengindikasikan adanya hambatan dalam mensejahterakan masyarakat. Pemerintah Indonesia telah mendirikan badan pemerintahan non-kementrian yakni Badan Pusat Statistik (BPS) pada tahun 1960 untuk melaporkan hasil statistik di setiap wilayah Indonesia sebagai sarana *monitoring* agar kehidupan berbangsa dan bernegara di Indonesia dapat berjalan dengan baik dan terkendali.

Dalam mengukur tingkat ketidaksetaraan ekonomi, BPS menggunakan rasio Gini (Koefisien Gini) sebab rasio ini mampu memberi gambaran mengenai sejauh mana distribusi pendapatan di masyarakat telah merata dengan baik (Zumaeroh, Fatmawati, & Purnomo, 2023). Untuk menghitungnya, BPS terlebih dahulu mengakses data Survei Sosial Ekonomi Nasional (SUSENAS) lalu mengolahnya kedalam Kurva Lorenz. Meski Kurva Lorenz telah terbukti dapat mengukur tingkat rasio Gini di banyak implementasi, berdasarkan kajian literatur yang peneliti lakukan, prosedur yang harus dilalui terbilang panjang dan rumit sehingga rentan terhadap timbulnya *human error* selama proses perhitungan. Sebuah solusi inovatif

dibutuhkan untuk meminimalisir timbulnya hal-hal tak diinginkan selama proses pengolahan data indikator rasio Gini.

Adapun solusi yang peneliti tawarkan ialah penggunaan algoritma *data mining* dalam mengolah indikator pengukur rasio Gini berupa data pengeluaran masyarakat. Adapun algoritma *data mining* yang akan digunakan ialah *K-Means* dan *Fuzzy C-Means*. Kedua algoritma ini dipilih sebab keduanya kerap digunakan dalam mengolah dataset yang besar, dan mampu membagi anggota dataset tersebut kedalam kelompok-kelompok tertentu berdasarkan kesamaan karakteristik dengan prosedur komputasi yang singkat (Firdaus & Sofro, 2022). Meskipun keduanya cukup populer dikalangan para peneliti, beberapa hasil penelitian menunjukkan adanya perbedaan algoritma yang lebih unggul diantara keduanya. Pada penelitian yang mengelompokkan data pelanggan dengan model LRFM, metode *K-Means* terbukti lebih unggul dari *Fuzzy C-Means* (Syukron, Fayyad, Fauzan, Ikhsani, & Gurning, 2022). Namun hasil berbeda ditemukan pada penelitian yang mengelompokkan kabupaten/kota di provinsi jawa tengah, metode *Fuzzy C-Means* terbukti lebih unggul dari *K-Means* (Latifah, Sugiyarto, & Suparman, 2022).

Berdasarkan adanya fenomena ini, maka sudah sepatutnya bagi kedua algoritma tersebut untuk diuji kembali dan dibandingkan demi menemukan algoritma yang lebih baik dalam mengolah indikator pengukur rasio Gini yaitu data pengeluaran masyarakat. Algoritma *K-Means* bekerja dengan mengelompokkan data-data yang ada ke dalam sejumlah kelompok yang telah ditentukan, setiap titik data hasil pengelompokkan akan diberi label yang sesuai dengan centroid (titik pusat klaster) terdekat. Hal ini dilakukan dengan menghitung jarak antar setiap titik data terhadap setiap centroid (Wahyu & Rushendra, 2022).

Berbeda dengan *Fuzzy C-Means* yang bekerja dengan menghitung derajat keanggotaan setiap titik data terhadap setiap kelompok, hal ini memungkinkan diamatinya seberapa besar kemiripan karakteristik titik tersebut pada setiap kelompok yang ada. Didalam algoritma *Fuzzy C-Means*, derajat keanggotaan dihitung berdasarkan jarak antara titik data dan centroid menggunakan fungsi keanggotaan *fuzzy* (Syoyer & Wahyudin, 2021).

Berdasarkan penjelasan tersebut, bentuk output yang diharapkan berupa nilai yang dapat mewakili tingkat ketidaksetaraan ekonomi di suatu kelompok masyarakat namun dengan proses pengolahan data yang lebih cepat dan minim terjadinya *human error*, sehingga peneliti bermaksud untuk melakukan penelitian ini dengan judul “Analisis Perbandingan Metode K-Means dan Fuzzy C-Means pada Indikator Rasio Gini di Sumatera Utara”.

## 1.2. Rumusan Masalah

Adapun rumusan masalah yang dapat peneliti rancang berdasarkan permasalahan yang dikemukakan sebelumnya ialah:

1. Bagaimana melakukan *clustering* indikator rasio Gini masyarakat kabupaten/kota di provinsi Sumatera Utara dengan algoritma *K-Means* dan *Fuzzy C-Means*?
2. Seberapa besar tingkat ke-efektifan kinerja kedua algoritma yakni *K-Means* dan *Fuzzy C-Means* jika diukur melalui tingkat efisiensi waktu, keakuratan output terhadap perhitungan rasio Gini yang dikeluarkan oleh Badan Pusat Statistik Sumatera Utara, serta pengukuran kualitas klaster yang direfleksikan melalui *Dunn Index*?

### 1.3. Batasan Masalah

Adapun batasan masalah yang terdapat dalam penelitian ini ialah sebagai berikut:

1. Analisis hanya dilakukan pada data indikator rasio Gini tiga kabupaten/kota di provinsi Sumatera Utara,
2. Hasil analisis berupa pola pengelompokkan (*cluster*) indikator rasio Gini pada tiga kabupaten/kota di provinsi Sumatera Utara,
3. Tingkat keefektifan algoritma dinilai melalui tingkat efisiensi waktu, keakuratan data, dan penilaian *Dunn Index*, serta
4. *Prototype* program dirancang menggunakan bahasa pemrograman Microsoft Visual Basic (VB).

### 1.4. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dipaparkan sebelumnya, peneliti dapat merangkai tujuan penelitian diantaranya sebagai berikut:

1. Melakukan *clustering* data indikator rasio Gini masyarakat kabupaten/kota di provinsi Sumatera Utara dengan algoritma *K-Means* dan *Fuzzy C-Means*, serta
2. Membandingkan ke-efektifan kinerja kedua algoritma yakni *K-Means* dan *Fuzzy C-Means* melalui pengukuran efisiensi waktu, keakuratan output terhadap perhitungan rasio Gini yang dikeluarkan oleh Badan Pusat Statistik Sumatera Utara, serta pengukuran kualitas klaster yang direfleksikan melalui *Dunn Index*.

## 1.5. Manfaat Penelitian

Adapun manfaat yang diharapkan atas dilaksanakannya penelitian ini ialah sebagai berikut:

### 1.5.1. Bagi Peneliti (Penulis)

Melalui penelitian ini, peneliti berharap dapat memperoleh penambahan wawasan dan pengembangan keahlian terkhususnya terkait penggunaan metode *K-Means* dan *Fuzzy C-Means* dalam melakukan analisis data, serta sebagai penambah wawasan peneliti pula dalam memahami pengelompokan data indikator rasio Gini masyarakat kabupaten/kota di provinsi Sumatera Utara.

### 1.5.2. Bagi Peneliti Selanjutnya

Bagi para peneliti selanjutnya, penelitian ini diharapkan dapat berguna sebagai referensi dan inspirasi dalam melakukan riset terutama riset terkait *data mining* dengan topik terkait kesenjangan pendapatan dan kekayaan yang diproksikan melalui indikator rasio Gini, serta dapat pula dijadikan sebagai pedoman penelitian dan bahan komparasi hasil yang telah didapatkan pada akhir riset.

### 1.5.3. Bagi Para Pembaca

Bagi para pembaca, dengan adanya riset ini harapannya dapat digunakan sebagai sarana penambahan wawasan terkait analisis kesenjangan dan kekayaan (yang diproksikan melalui indikator rasio Gini) yang tersebar di beragam wilayah terkhususnya di provinsi Sumatera Utara, serta sebagai penambah wawasan pula bagi para pembaca terkait topik analisis data yang dilakukan menggunakan metode *data mining*.

#### 1.5.4. Bagi Entitas Pemerintahan

Bagi entitas pemerintahan, dengan adanya riset ini harapannya dapat digunakan sebagai acuan dan bahan pertimbangan dalam rangka penarikan keputusan, serta sebagai bahan referensi pembuatan regulasi dan kebijakan dalam mengupayakan evaluasi yang lebih baik di masa yang akan datang.

## BAB II

### LANDASAN TEORI

#### 2.1. Data Mining

Di dunia modern sekarang ini segala hal memiliki keterkaitan dengan penggunaan teknologi, baik itu hal-hal yang formal maupun informal. Ketergantungan dalam penggunaan teknologi tentunya tidak terlepas dari adanya pengumpulan data. Pengumpulan data umumnya dilakukan dengan tujuan agar suatu teknologi maupun hal yang terkait dengannya dapat diimprovisasi baik itu dari segi kualitas output yang dihasilkan maupun fitur-fitur yang disediakan kedepannya. Data-data yang dikumpulkan agar dapat memberikan informasi yang bermanfaat harus melalui serangkaian proses pengolahan data yang kerap disebut dengan istilah *data mining*. *Data mining* dapat dipahami sebagai suatu tahapan ekstraksi data mentah dan informasi berukuran besar yang sifatnya belum diketahui namun dapat dikaji serta digunakan untuk menarik suatu keputusan yang penting (Zai, 2022).

Selain itu, *data mining* juga dapat dipahami sebagai sebuah teknik untuk memperoleh informasi yang berguna dari sejumlah besar data yang tidak terstruktur, tidak konsisten, dan selalu berubah (dinamis). Para ahli dibidang ini kerap dipanggil dengan istilah *data miners*, umumnya berfokus pada pengidentifikasian pola yang terdapat didalam dataset (Salem et al., 2022), kemudian para *data miners* akan menjalani serangkaian prosedur *data mining* yang umumnya terdiri atas; pemahaman subjek analisis, penyiapan data, pemahaman

data, pembuatan model, evaluasi, serta ditutup dengan pelaporan hasil analisis (Akpojotor & Urhefe-Okotie, 2023).

Konsep Data Mining pada awalnya di populerkan di tahun 1900-an yang beriringan dengan berkembangnya teknologi komputasi modern di era milenial, namun meskipun demikian sejatinya konsep data mining justru telah berkembang sejak tahun 1700-an yang dimulai dengan ditemukannya Teorema Bayes tepatnya pada tahun 1763, sebuah makalah yang diterbitkan secara anumerta dengan judul “*An Essay Towards Solving a Problem in the Doctrine of Chances*” karya Thomas Bayes. Menurutny probabilitas suatu hipotesis, jikalau suatu informasi baru diperoleh, dapat dihitung ulang kembali dengan menggunakan rumus yang melibatkan probabilitas hipotesis tersebut sebelum informasi terbaru dapat diperoleh. Namun pada tahun tersebut, makalah ini tidak mendapat perhatian yang cukup luas, justru perhatian publik dapat diperoleh ketika akhir abad ke-18 disaat matematikawan asal Prancis bernama Pierre-Simon Laplace menggunakan teorema ini dalam karyanya yang membahas topik statistika dan probabilitas, hal ini tentunya memicu timbulnya penyebaran dan penerapan yang lebih luas dari konsep Teorema Bayer dalam berbagai bidang ilmu pengetahuan, seperti ilmu sosial, ilmu komputer, dan kecerdasan buatan di era modern (Ghosh, Balakrishnan, & Ng, 2021).

Konsep data mining memperoleh perkembangan pesat di abad ke-18, yang mana di abad ini konsep data mining mulai mengenal prosedur-prosedur standar yang kita kenal sekarang ini seperti pengumpulan data, penyimpanan data, manipulasi data, serta pengenalan konsep dataset. Berkat adanya perkembangan yang pesat tersebut, data mining pun pada akhirnya dapat mencapai titik dimana

jaringan saraf, pengelompokan (klustering), dan algoritma genetika dapat dilakukan (tepatnya di tahun 1950-an), konsep pohon keputusan / *decision tree* (ditahun 1960-an), dan mesin vektor pendukung / *supporting vector machines* (ditahun 1990-an) (Şahin & Yurdugul, 2020).

Di era milenial ketika teknologi komputasi telah menjadi sebuah kebutuhan publik, konsep data mining telah mencapai titik perkembangan dimana melalui data mining kita dapat melakukan *Text Mining*, *Image Mining*, dan *Graph Mining*. *Text Mining* atau penambangan teks ialah teknik yang berfokus pada analisis dan ekstraksi informasi yang terkandung dalam teks. Tujuan utama dari text mining adalah untuk mengekstrak pengetahuan yang bermanfaat dari berbagai sumber teks, seperti dokumen, artikel, pesan, posting media sosial, dan lain sebagainya. Kemudian *Image Mining* ialah proses ekstraksi informasi yang bermanfaat dari data visual, misalnya seperti gambar atau video (Khan, Khan, & Alharbi, 2020).

Tujuan utama dari *image mining* adalah untuk menemukan pola, informasi, atau pengetahuan yang terkandung dalam data visual tersebut. Teknik ini kerap digunakan untuk memahami konten visual, mengelompokkan gambar berdasarkan kesamaan visual, mengidentifikasi objek maupun fitur penting dalam gambar (Romero & Ventura, 2020). Sedangkan untuk *Graph Mining* atau penambangan grafik ialah proses mengekstraksi informasi dari data yang direpresentasikan dalam bentuk grafik atau jaringan. Data dalam bentuk grafik terdiri dari simpul (*node*) yang terhubung oleh tepi (*edge*), yang menggambarkan relasi antara entitas atau objek dalam dataset. Tujuan utama dari *graph mining* adalah untuk menemukan pola, struktur, atau informasi penting dalam data grafik yang diteliti. Tujuan ini mencakup identifikasi pola dalam jaringan, pengelompokan simpul berdasarkan

kesamaan, analisis hubungan antara simpul, dan lain sebagainya (Kang, He, Maciejewski, & Tong, 2020).

Selain itu, Data Mining juga dapat diklasifikasikan berdasarkan tujuannya, secara umum setidaknya ada 7 jenis data mining jika diklasifikasikan berdasarkan tujuannya, adapun diantaranya: Data Mining Deskriptif (Mean, Median, Modus, Varians, dan Kuartil), Data Mining Prediktif (XG Boost, LightGBM, Naïve Bayes), Data Mining Asosiatif (Apriori, FP-Growth), Data Mining Klastering (K-Means, Hierarchical Clustering, DBSCAN), Data Mining Similarity Search (Jaccard Similarity, Cosine Similarity), Data Mining Tekstual (TF-IDF, LSA, Word Embeddings), dan Data Mining Multidimensi (OLAP, Data Cube) (Harahap, Halim, Nst, & Ginting, 2023; Riza, 2022; Suhailah & Hartatik, 2023). Namun penelitian ini jenis Data Mining yang akan digunakan ialah Data Mining Klastering tepatnya pada Algoritma K-Means dan Fuzzy C-Means saja.

## **2.2. Algoritma Klastering**

Algoritma Klastering atau sering disebut *clustering algorithm* memiliki peran yang penting dalam melakukan suatu tindakan *data mining*. Melalui algoritma ini, dataset yang dimiliki dapat dikelompokkan sehingga memudahkan para pemangku kepentingan untuk memahami dan mengidentifikasi kelompok-kelompok data yang memiliki ciri sifat yang saling menyerupai satu sama lain. Proses algoritma klastering umumnya dilakukan dengan tujuan untuk mengelompokkan petunjuk-petunjuk yang ada pada suatu kelompok dataset, pengelompokannya sendiri dilakukan dengan mengaitkan hipotesis indeks sehingga objek-objek yang terdapat di dalam dataset dapat dikelompokkan sedemikian rupa, hal ini tentunya akan

memudahkan para pemangku kepentingan dalam menarik suatu keputusan atas upaya *data mining* yang telah dilakukan (Anikah, Surip, Rahayu, Al-Musa, & Tohidi, 2020).

Algoritma klastering memiliki tujuan agar data-data yang ada dapat disederhanakan berdasarkan pengelompokan yang telah dilakukan sehingga mudah untuk dipahami. Algoritma klastering memiliki beragam varian mulai dari varian klastering yang bersifat *Hard Clustering* hingga *Soft Clustering*. Meskipun memiliki beragam varian, baik itu *Hard Clustering* maupun *Soft Clustering*, setiap klastering memiliki fungsi yang sama yakni untuk mengelompokkan data-data yang telah dikumpulkan sebelumnya kedalam beberapa kelompok atau *cluster* yang anggota-anggotanya memiliki tingkat kemiripan yang maksimum, serta data antar masing-masing *cluster* memiliki kemiripan yang minimum (Rinjani, Anwar, & Herdiana, 2023). Sehingga ini mengartikan bahwa klastering dilakukan agar data-data yang memiliki kemiripan sifat dapat dikumpulkan kedalam satu kelompok, sedangkan masing-masing kelompok diupayakan agar memiliki sifat yang sangat berbeda antara satu sama lain.

Hingga saat ini algoritma klastering memiliki banyak varian yang dapat digunakan oleh para *data miners* dalam melakukan data mining. Adapun sejumlah algoritma klastering yang kerap digunakan ialah antarlain K-Means, Hierarchical Clustering, DBSCAN (Density-Based Spatial Clustering of Applications with Noise), Mean Shift Clustering, GMM (Gaussian Mixture Models), FCM (Fuzzy C-Means), Spectral Clustering, dan Affinity Propagation. Adapun penjelasannya sebagai berikut:

1. Algoritma K-Means Clustering adalah algoritma yang membagi data menjadi “k” klaster di mana setiap titik data termasuk ke dalam klaster dengan rata-rata terdekat. Huruf "k" merujuk pada jumlah klaster yang ingin dibentuk berdasarkan dataset yang dimiliki (Yudhistira & Andika, 2023).
2. Hierarchical Clustering adalah jenis klustering di mana klaster dibentuk secara hierarkis dalam bentuk pohon atau dendrogram. Ada dua jenis utama: *agglomerative* (dari bawah ke atas) dan *divisive* (dari atas ke bawah). Pendekatan *agglomerative* dimulai dengan memperlakukan setiap titik data sebagai klaster tunggal dan secara bertahap menggabungkan klaster terdekat menjadi satu klaster baru hingga semua titik data berada dalam satu klaster. Sedangkan pendekatan *divisive* dimulai dengan memperlakukan semua titik data sebagai satu klaster dan secara bertahap membagi klaster menjadi dua klaster yang semakin kecil hingga pada akhirnya setiap titik data berada dalam klaster terpisah (Septianingsih, 2022; Setyawan & Fatichah, 2020).
3. Algoritma DBSCAN ialah algoritma yang berfungsi untuk mengidentifikasi klaster berdasarkan kerapatan titik data dalam ruang. Algoritma ini bekerja dengan menemukan area dengan kepadatan tinggi sebagai klaster, sementara titik-titik yang terisolasi dianggap sebagai *noise* (Izhari, 2020).
4. Mean Shift Clustering adalah algoritma klustering yang mencari pusat-pusat klaster dengan menggerakkan titik data ke arah dimana kepadatan data paling tinggi terletak. Proses ini akan terus berulang hingga tidak ada pergeseran signifikan lagi diantara data-data tersebut, karena pusat klaster dibentuk berdasarkan letak kepadatan data yang paling tinggi, maka dalam pelaksanaannya, metode ini tidak memerlukan tahap penentuan jumlah klaster

diawal layaknya algoritma klastering lainnya (Rizuan, Haerani, Jasril, & Oktavia, 2023).

5. Gaussian Mixture Models adalah algoritma dengan model probabilistik yang menggunakan distribusi Gaussian untuk mengklasifikasikan data ke dalam beberapa klaster. Algoritma ini mencari parameter distribusi normal yang paling cocok dengan data, seperti rata-rata dan kovariansi (Chyan, 2022).
6. Fuzzy C-Means (FCM) adalah bentuk variasi dari algoritma K-Means yang mana dalam algoritma ini setiap titik data diberi bobot yang menunjukkan tingkat keanggotaannya dari setiap klaster yang ada. Dalam algoritma ini titik data dapat menjadi anggota lebih dari satu klaster dengan bobot kemiripan yang berbeda-beda antara satu klaster dengan klaster lainnya (Chyan, 2022).
7. Spectral Clustering adalah algoritma yang menggunakan representasi grafik dari data untuk mengidentifikasi klaster. Algoritma ini bekerja dengan mengekstraksi nilai-nilai *eigen* dari matriks yang mewakili jarak antara titik data dan menggunakan informasi tersebut untuk mengelompokkan data (Bai, Qi, & Liang, 2023).
8. Affinity Propagation adalah algoritma yang memodelkan data sebagai grafik dimana simpul (*node*) mewakili titik data dan tepi mewakili "pesan" antara pasangan titik data. Melalui algoritma, kita dapat menentukan klaster berdasarkan keputusan node menjadi pusat klaster atau anggota klaster (Duan, Liu, Li, Guo, & Yang, 2023).

### 2.3. Logika Fuzzy

Dalam *data mining* terdapat istilah Logika Fuzzy. Logika ini merupakan komponen yang digunakan dalam membentuk *soft-computing*. Pada mulanya komponen ini diperkenalkan oleh seorang matematikawan asal Azerbaijan bernama Lotfi Aliasker Zadeh pada tahun 1965. Logika Fuzzy memiliki dasar logika berupa “Teori Himpunan *Fuzzy*”, didalam teori ini anggota-anggota kelompok yang ada ditentukan berdasarkan “Derajat Keanggotaan” atau kerap juga disebut dengan istilah *Membership Function* (Yulianto, Rahmah, Faisol, & Amalia, 2023). Logika ini kerap digunakan untuk menjelaskan ke-ambiguan yang terdapat pada dataset yang dimiliki, selain itu dengan menggunakan logika ini kita dapat mengubah suatu pernyataan yang bersifat linguistik menjadi pernyataan yang bersifat numerik. Dalam implementasinya, Logika Fuzzy memiliki derajat keanggotaan, derajat ini memiliki nilai diantara angka 0 hingga angka 1.

Hal ini tentunya membuat logika fuzzy berbeda dengan logika digital lainnya yang hanya memiliki dua opsi nilai yakni 0 atau 1 saja, sehingga logika ini dapat menilai sejauh mana suatu data dapat dikatakan benar dan sejauh mana pula dapat dikatakan salah. Adanya perbedaan konsep logika yang berlaku pada logika fuzzy membuat tingkat toleransi yang terjadi pada data menjadi lebih fleksibel dan memungkinkan bagi setiap data untuk dinilai sebagai “sebagian benar dan sebagian salah” pada satu waktu yang sama (Audrey, Fadlil, & Sunardi, 2022).

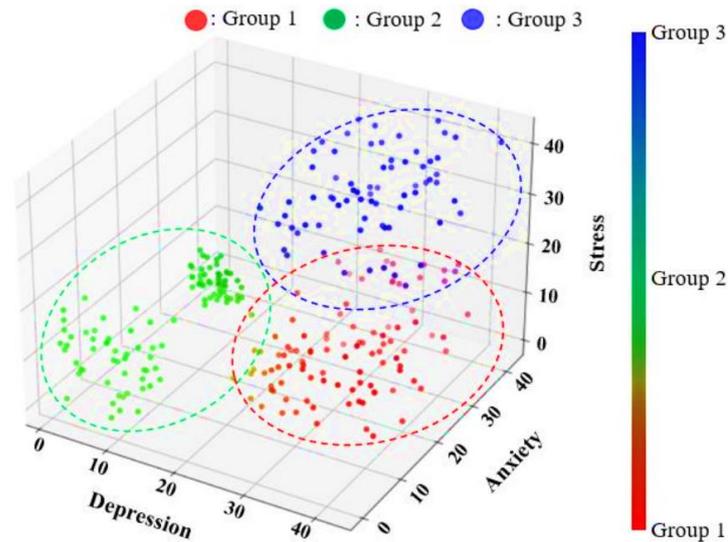
Logika Fuzzy kerap dipilih sebagai logika dalam melakukan *data mining* karena mampu menciptakan model yang bersifat *soft-computing* sehingga memudahkan para pemangku kepentingan untuk mempertimbangkan sejumlah opsi alternatif dalam menarik suatu keputusan. Di dalam Logika Fuzzy setidaknya

terdapat tiga jenis metode inferensi fuzzy yang kerap dan dapat digunakan, diantaranya ialah:

1. Fuzzy Tsukamoto, yakni metode inferensi fuzzy yang dikembangkan oleh Profesor Lotfi A. Zadeh dan Tsukamoto dimana dalam metode inferensi ini variabel input akan dirubah menjadi himpunan *fuzzy* dengan menggunakan fungsi keanggotaan, kemudian aturan-aturan inferensi pun diimplementasikan untuk menghasilkan variabel output *fuzzy*. Setelahnya, dilakukan pula proses defuzzyfikasi untuk menghasilkan output akhir dalam bentuk bilangan real (Auliana & Mansyuri, 2022).
2. Fuzzy Sugeno, yakni metode inferensi yang diperkenalkan oleh Profesor Michio Sugeno di tahun 1974. Sistem inferensi ini bekerja dengan menggabungkan aturan fuzzy yang berdasarkan nilai keanggotaan dari variabel input yang tersedia, disertai dengan sebuah fungsi linear yang memberikan hasil output berupa nilai numerik. Didalam metode inferensi ini, aturan fuzzy ditentukan berdasarkan sekumpulan proposisi fuzzy yang mengandung klausa kondisi (*antecedent*) dan klausa kesimpulan (*consequent*). Klausa kondisi dinyatakan dalam bentuk himpunan fuzzy pada variabel input, sedangkan klausa kesimpulan dinyatakan sebagai polinomial orde pertama pada variabel output (Burhanuddin, 2023).
3. Fuzzy Mamdani, yakni metode inferensi yang diperkenalkan oleh Ebrahim Mamdani di tahun 1975. Menurut teorinya, dengan menggunakan aturan-aturan linguistik dalam bentuk IF-THEN, metode inferensi ini mampu mengubah suatu bentuk input yang tidak pasti menjadi output yang tidak pasti pula. Setidaknya terdapat tiga tahapan yang terdapat didalam metode inferensi

ini yakni: fuzzifikasi, inferensi, dan defuzzifikasi. Fuzzifikasi mengacu pada proses dimana nilai numerik diubah menjadi nilai linguistik yang dapat diterima oleh sistem, kemudian tahapan inferensi ialah tahapan dimana aturan-aturan linguistik tersebut digunakan demi menciptakan nilai output yang tidak pasti, dan defuzzifikasi mengacu pada proses ketika nilai output yang diperoleh ditahap inferensi diubah menjadi suatu nilai numerik yang dapat diterima oleh manusia pada umumnya (Romansyah & Zulfikar, 2021).

Namun untuk penelitian yang akan peneliti laksanakan, tidak ada metode inferensi yang digunakan. Hal ini sebab metode data mining yang digunakan peneliti ialah K-Means, dan Fuzzy C-Means. Keduanya merupakan metode klustering yang berfokus pada pengelompokan data tanpa mengacu pada aturan berbasis IF-THEN yang digunakan dalam sistem logika fuzzy Sugeno, Tsukamoto, maupun Mamdani. K-Means Klustering bekerja dengan mengelompokkan data ke dalam “k” kelompok berdasarkan jarak antar titik data dan pusat kelompok. Fuzzy C-Means juga melakukan metode klustering serupa namun mempertimbangkan kemungkinan setiap titik data untuk menjadi bagian dari setiap kelompok dengan menghitung nilai keanggotaan fuzzy setiap data. Keduanya tidak menggunakan konsep IF-THEN sehingga dalam implementasinya tidak dibutuhkan penerapan metode inferensi baik itu Tsukamoto, Sugeno, maupun Mamdani.



**Gambar 2.1. Visualisasi Fuzzy**

*Sumber: (Wang, Lee, Do, Lee, & Chung, 2023)*

Seperti yang telah dijelaskan sebelumnya, sifat *Fuzzy* dalam logika ini memungkinkan data-data yang terkandung didalam dataset dapat dinilai sejauh mana telah menjadi anggota dari suatu klaster, dan sejauh mana pula data yang sama menjadi anggota dari klaster lainnya yang dinilai berdasarkan tinggi-rendahnya derajat keanggotaannya. Sebagai contoh dapat kita perhatikan pada gambar 2.1. yang merupakan visualisasi hasil klasterisasi dengan logika fuzzy. Pada gambar tersebut dapat kita perhatikan terdapat sejumlah data yang tergabung dalam dua klaster sekaligus, baik itu tergabung dalam klaster *Group 1-Group 2* maupun *Group 1-Group 3*. Adanya konsep derajat keanggotaan ini membuat logika *Fuzzy* menjadi lebih fleksibel dalam menentukan jumlah anggota klaster-klasternya, ini tentunya dapat memberikan informasi yang jauh lebih bermanfaat bagi para pemangku kepentingan, sebab kita dapat membentuk opsi-opsi alternatif yang lebih luas dan

lebih tepat dalam pengambilan suatu keputusan informasi, serta mampu memberikan pemahaman yang lebih kompleks pula atas suatu fenomena.

#### **2.4. K-Means Clustering**

Dalam melakukan suatu *data mining* terdapat algoritma-algoritma yang dapat kita gunakan, salah satunya ialah klustering. Di dalam algoritma klustering itu sendiri, ada banyak pula varian algoritma yang dapat kita gunakan salah satunya ialah *K-Means Clustering*. Algoritma ini memiliki peran yang krusial dalam melakukan suatu tindakan *data mining*. Penggunaannya yang terbilang cukup sederhana dan mudah diimplementasikan serta dieksekusi membuat metode ini cukup digemari oleh banyak *data miners* dibandingkan dengan metode klustering lainnya. Algoritma ini bekerja dengan melakukan pengklasteran data-data yang terdapat di dalam dataset berdasarkan jarak yang kemudian data-data tersebut dimasukkan ke dalam masing-masing kluster yang berbeda (Matdoan, Matdoan, & Far-Far, 2022).

Namun perlu dicatat, algoritma K-Means hanya dapat bekerja pada pernyataan numerik saja, tidak seperti logika fuzzy yang mampu mengkonversi pernyataan linguistik menjadi numerik. Hal ini tentunya menyebabkan data-data yang ada tidak mungkin bergabung ke dalam dua kluster berbeda sekaligus, sifat khas ini kerap disebut sebagai *Hard Clustering*. Dalam melakukan pengklasteran menggunakan algoritma ini, *data miners* umumnya menempuh prosedur yang diantaranya; menetapkan jumlah kluster yang akan terlibat, memastikan inisialisasi awal, serta centroid (pusat kluster) ditentukan secara *random* atau acak, menghitung jarak antara masing-masing data ke setiap centroid dengan menggunakan rumus

persamaan Euclidean Distance, mengelompokkan objek-objek berdasarkan jarak minimum ke titik pusat, menghitung rata-rata objek pada kluster untuk membentuk centroid baru yang kedudukannya baru pula, serta melakukan perulangan tahap-tahap sebelumnya hingga hasil klustering yang dilakukan tidak mengalami perubahan (Olivia, Juliantho, & Hendrik, 2023).

Sebelum menghitung algoritma K-Means, terlebih dahulu data-data yang akan diolah perlu melalui proses normalisasi. Hal ini agar nilai-nilai didalam dataset memiliki skala yang seragam atau relatif sehingga dapat dibandingkan atau diproses dengan benar tanpa adanya bias yang disebabkan adanya perbedaan skala. Dalam penelitian ini, metode normalisasi yang akan digunakan ialah metode Min-Max Scaler, nilai hasil normalisasi metode ini idealnya bernilai antara 0 hingga 1 (Permatasari, Chrisnanto, & Ningsih, 2023). Adapun rumusnya sebagai berikut:

$$X = \frac{X_{ori} - X_{min}}{X_{max} - X_{min}} \dots \dots \dots (2.1)$$

Dimana,

$X_{ori}$  = Nilai Asli Data X

$X_{min}$  = Nilai Terkecil pada Dataset

$X_{max}$  = Nilai Terbesar pada Dataset

*Sumber: (Sulistyawati & Sadikin, 2021)*

Dalam melakukan pengelompokan data menggunakan algoritma K-Means, tahap awal yang harus dilakukan ialah menentukan banyak kluster yang akan dibentuk atau kluster optimal yang akan digunakan dalam proses klustering. Adapun metode penentuan kluster optimal yang akan peneliti gunakan ialah metode Elbow. Penentuan kluster optimal dalam metode ini dilakukan berdasarkan selisih terbesar nilai *Sum of Square Error* (SSE), dalam implementasinya setidaknya

terdapat 5 tahapan yang harus dilalui diantaranya: menginisialisasi awal nilai K, menaikkan nilai K, menghitung nilai SSE dari setiap K, mengamati nilai SSE dari setiap K untuk menemukan K dengan selisih nilai SSE paling tinggi (mengalami penurunan nilai SSE paling tajam), dan diakhiri dengan menentukan nilai K yang berbentuk siku (Ekasetya & Jananto, 2020; Nurdiyanti & Yuniarti, 2023). Adapun rumus dalam menentukan nilai SSE ialah sebagai berikut:

$$SSE = \sum_{k=1}^k \sum_{x_1} |x_i - ck|^2 \dots\dots\dots (2.2)$$

Dimana,

K = Kluster ke-C

$x_i$  = Jarak Objek Data ke-i

$ck$  = Pusat Klaster ke-i

*Sumber: (Virantika, Kusnawi, & Ipmawati, 2022)*

Dalam menentukan titik pusat (*centroid*) dari setiap kluster, untuk tahap awal atau iterasi pertama, penentuan titik pusat dilakukan secara acak, sehingga untuk tahap awal (iterasi pertama) tidak ada rumus penentuan titik pusat yang akan digunakan (Saputro, Retnoningsih, & Khusnuliawati, 2024). Apabila titik pusat telah ditentukan, tahap selanjutnya ialah menghitung jarak antara titik data dan pusat setiap kluster. Adapun rumus menghitung jarak yang dapat digunakan ialah rumus *Euclidean Distance*. Adapun rumus *Euclidean Distance* ialah:

$$d = \sqrt{[(X_2 - X_1)^2 + (Y_2 - Y_1)^2]} \dots\dots\dots (2.3)$$

Dimana,

$d$  = Jarak

$x_1$  = Koordinat latitude 1

$x_2$  = Koordinat latitude 2

$y_1$  = Koordinat longitude 1

$y_2$  = Koordinat longitude 2

*Sumber: (Mughnyanti & Ginting, 2023)*

Untuk mengelompokkan titik-titik data berdasarkan jarak minimumnya ke pusat kluster dapat dilakukan dengan menyortir nilai Euclidean yang diperoleh masing-masing titik data. Nilai Euclidean terkecil mengindikasikan data tersebut merupakan anggota kluster dari titik pusat dengan jarak paling minimum. Selanjutnya untuk menghitung titik pusat kluster baru demi memastikan hasil akhir yang diperoleh telah dicapai dapat menggunakan rumus berikut ini:

$$C = \frac{\sum m}{n} \dots\dots\dots (2.4)$$

Dimana,

C = Titik Pusat Kluster Baru

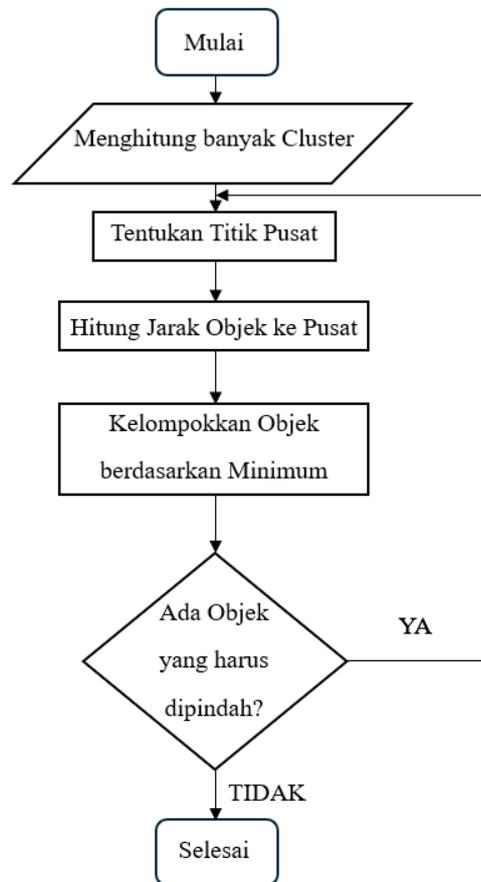
m = Nilai anggota data yang termasuk dalam centroid (titik pusat kluster) tertentu

n = Banyak data yang menjadi anggota centroid (titik pusat kluster) tertentu

*Sumber: (Bayu Lokananta, Yuana, & Puspitasari, 2024)*

Jikalau dalam menghitung titik pusat kluster baru ditemukan hasil bahwa titik pusatnya tidak berubah dari iterasi yang terakhir dilakukan (titik pusat tidak mengalami pembaharuan letak) maka hal ini mengindikasikan iterasi sudah selesai sebab jikalau iterasi selanjutnya tetap dilakukan maka hasil yang akan diperoleh ialah sama (tidak ada titik data yang berubah letaknya jika dibandingkan dengan iterasi sebelumnya) sehingga iterasi perlu dihentikan. Hal serupa juga akan berlaku jika titik pusat kluster berubah letaknya pada iterasi selanjutnya namun setelah

dihitung jaraknya dari setiap titik data justru tidak ada perubahan keanggotaan kluster, maka iterasi perlu dihentikan atau dianggap selesai (Sy, 2023). Adapun detail terkait alur pelaksanaan klustering dengan menggunakan Algoritma K-Means ialah sebagai berikut:



**Gambar 2.2. Diagram Alur K-Means**

*Sumber: (Sihananto, Puspita Sari, Khariono, Akhmad Fernanda, & Wijaya, 2022)*

Berdasarkan diagram alur pelaksanaan klustering dengan menggunakan Algoritma K-Means diatas, dapat kita perhatikan bahwa proses klustering diawali dengan menghitung banyaknya jumlah *cluster* yang akan dibuat, diikuti dengan penentuan titik pusat dari setiap *cluster* yang ada, kemudian menghitung jarak seluruh data / objek kepada pusat *cluster* terdekat dengan menggunakan formula

*Euclidean Distance* yang telah dijelaskan sebelumnya, serta mengelompokkan data-data / objek berdasarkan nilai minimum yang diperoleh. Jikalau hasil pengelompokkan telah diperoleh, maka dilakukan kembali pengujian. Jikalau hasil pengujian menunjukkan adanya objek yang berpindah, maka proses harus diulang kembali ke tahap dimana penentuan titik pusat dilakukan, namun jikalau hasil pengujian menunjukkan tidak adanya objek yang berpindah (seluruh objek tetap berada pada tempatnya) maupun titik pusat yang baru tidak mengalami perubahan posisi maka proses pengolahan Algoritma K-Means telah selesai.

## **2.5. Fuzzy C-Means Clustering**

Selain algoritma *K-Means Clustering*, *data miner* juga dapat memanfaatkan algoritma lainnya jikalau hendak menghasilkan output dengan toleransi yang lebih fleksibel, misalnya saja *Fuzzy C-Means Clustering*. Algoritma klustering yang juga kerap disingkat menjadi FCM ini bekerja dengan membagi data-data yang terdapat di dalam dataset kedalam beragam klaster berdasarkan derajat keanggotaannya, sehingga data-data yang ada tidak serta-merta hanya tergabung pada satu klaster saja melainkan dapat bergabung ke banyak klaster. Sifat unik dari algoritma ini membuat *Fuzzy C-Means Clustering* tergolong sebagai *soft clustering*. Data yang bertempat pada perbatasan dua atau lebih klaster tidak dapat dikatakan sepenuhnya sebagai anggota salah satu klaster, melainkan data tersebut dapat menjadi anggota dari beberapa klaster berbeda sekaligus dengan derajat keanggotaan menjadi indikator utama dalam melihat data tersebut paling didominasi oleh karakteristik klaster yang mana (Olivia et al., 2023).

Dalam melakukan pengklasteran menggunakan Algoritma *Fuzzy C-Means*, *data miner* akan menentukan nilai keanggotaan setiap data dengan nilai antara 0 hingga 1. Algoritma ini mampu menempatkan pusat klaster dengan lebih akurat dan tepat jikalau dibandingkan dengan metode klaster lainnya, hal ini dikarenakan algoritma *Fuzzy C-Means* akan selalu memperbaiki pusat klaster secara terus menerus hingga pusat klaster tak lagi berpindah atau dapat dipahami bahwa pusat klasternya telah sampai pada titik yang tepat (Aziz, Siregar, & Zonyfar, 2022). Umumnya *data miners* akan menempuh prosedur diantaranya; mengidentifikasi data yang ingin diklasterisasi ke dalam bentuk matriks berukuran  $n \times m$ , menentukan banyak klaster yang akan terlibat, menghasilkan angka acak guna menghasilkan elemen dari matriks partisi utama, menghitung setiap jumlah kolom, menghitung nilai elemen dari matriks partisi, menghitung nilai pusat klaster, menghitung fungsi objektif yang kiranya terdapat pada iterasi ke- $t$ , menghitung perubahan matriks partisi ke- $t$ , dan menghitung perubahan matriks partisi  $u$ , dan ditutup dengan melakukan pemeriksaan kondisi “berhenti” pada iterasi yang tengah berlangsung (Fatonah & Pancarani, 2022).

Adapun tahapan dalam melakukan pengklasteran *Fuzzy C-Means* dimulai dengan menormalisasikan data yang terdapat didalam dataset. Dalam penelitian ini, metode normalisasi yang akan digunakan dalam *Fuzzy C-Means* akan menggunakan metode Min-Max Scaler. Sehingga dengan demikian hasil ideal yang dihasilkan bernilai antara 0 dan 1 (Permatasari et al., 2023) dengan rumus sebagai berikut:

$$X = \frac{X_{ori} - X_{min}}{X_{max} - X_{min}} \dots \dots \dots (2.5)$$

Dimana,

Xori = Nilai Asli Data X

Xmin = Nilai Terkecil pada Dataset

Xmax = Nilai Terbesar pada Dataset

*Sumber: (Sulistyawati & Sadikin, 2021)*

Tahap selanjutnya yang harus dilakukan ialah menginisialisasi jumlah kluster yang akan dibentuk. Dalam menginisialisasi jumlah kluster yang akan dibentuk, peneliti menggunakan metode Elbow. Melalui metode ini, kluster optimal dapat diperoleh dengan menghitung selisih terbesar nilai *Sum of Square Error* (SSE), dalam implementasinya setidaknya terdapat 5 tahapan yang harus dilalui diantaranya: menginisialisasi awal nilai K, menaikkan nilai K, menghitung nilai SSE dari setiap K, mengamati nilai SSE dari setiap K untuk menemukan K dengan selisih nilai SSE paling tinggi (mengalami penurunan nilai SSE paling tajam), dan diakhiri dengan menentukan nilai K yang berbentuk siku (Ekasetya & Jananto, 2020; Nurdianti & Yuniarti, 2023). Adapun rumus dalam menentukan nilai SSE ialah sebagai berikut:

$$SSE = \sum_{k=1}^k \sum_{x1} |xi - ck|^2 \dots\dots\dots (2.6)$$

Dimana,

K = Kluster ke-C

xi = Jarak Objek Data ke-i

ck = Pusat Kluster ke-i

*Sumber: (Virantika et al., 2022)*

Tahapan selanjutnya ialah membangkitkan/memanggil bilangan random. Bilangan ini bernilai dibawah angka 1 dengan catatan banyaknya nilai random

ditentukan berdasarkan matriks  $y$  sebanyak jumlah data yang terlibat dalam perhitungan dan matriks  $x$  sebanyak jumlah kluster yang terlibat dalam perhitungan. Adapun total bilangan random perbarisnya harus bernilai tepat di angka 1. Jikalau bilangan random telah terbentuk secara menyeluruh, tahapan selanjutnya ialah menghitung pusat kluster. Adapun cara menghitung titik pusatnya dengan meng-pangkat dua kan nilai random yang telah diperoleh sehingga terbentuk nilai  $\mu$  (miu) kuadrat yang kemudian perkolomnya akan ditotalkan sehingga terbentuk  $\mu$  (miu) total.

Apabila tahapan ini telah tercapai, tahapan selanjutnya ialah mengkalikan setiap kolom  $\mu$  (miu) kuadrat yang telah diperoleh sebelumnya dengan data yang ada pada dataset, operasi perkalian ini terus dilakukan untuk setiap kolom  $\mu$  (miu) kuadrat. Setiap hasil perkalian antara data mentah dalam dataset dengan  $\mu$  (miu) kuadrat akan ditotal. Hasil perkalian itulah yang kemudian dibagi oleh total  $\mu$  (miu) kuadrat sehingga menghasilkan pusat kluster (*centroid*). Berdasarkan penjelasan tersebut, rumus menghitung pusat kluster secara ringkas ialah sebagai berikut:

$$V_{kj} = \sum_{i=1}^n \mu_{ik}^w (X_{ij}) / \sum_{i=1}^n (\mu_{ik})^w \dots\dots\dots (2.7)$$

Dimana,

$V_{kj}$  = Pusat cluster ke-k untuk atribut ke-j

$\mu_{ik}$  = Derajat keanggotaan untuk data sampel ke-I pada cluster ke-k

$X_{ij}$  = Data ke-i, atribut ke-j

*Sumber: (Aziz et al., 2022)*

Tahap selanjutnya yang harus dilakukan setelah titik pusat (*centroid*) telah diperoleh ialah menghitung fungsi objektif. Tahapan ini berfungsi untuk melalui rumus sebagai berikut:

$$J_m = \sum_{i=1}^n \sum_{k=1}^c u_{ik}^m \|x_i - v_k\|^2 \dots\dots\dots (2.8)$$

Dimana,

$J_m$  = fungsi objektif

$n$  = jumlah data points

$c$  = jumlah clusters

$v_{ik}$  = derajat keanggotaan data point  $i$  terhadap cluster  $k$

$m$  = parameter kefuzzyness

$x_i$  = data point ke- $i$

$v_k$  = centroid dari cluster  $k$

*Sumber:* (Latifah et al., 2022)

Tahapan selanjutnya ialah menghitung perubahan matriks partisi. Pada tahapan ini, rumus yang akan dipakai ialah:

$$\mu_{ik} = \frac{\left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{-\frac{1}{m-1}}}{\sum_{k=1}^c \left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{-\frac{1}{m-1}}} \dots\dots\dots (2.9)$$

Dimana,

$\mu_{ik}$  = Derajat keanggotaan untuk data sampel ke- $i$  pada cluster ke- $k$

$V_{kj}$  = Pusat cluster ke- $k$  untuk atribut ke- $j$

$X_{ij}$  = Data ke- $i$ , atribut ke- $j$

*Sumber:* (Aziz et al., 2022)

Jikalau nilai perubahan matriks partisi telah diperoleh, dan perhitungan masih pada iterasi pertama, maka tahapan selanjutnya ialah menghitung nilai derajat keanggotaan yang baru dengan menggunakan rumus berikut:

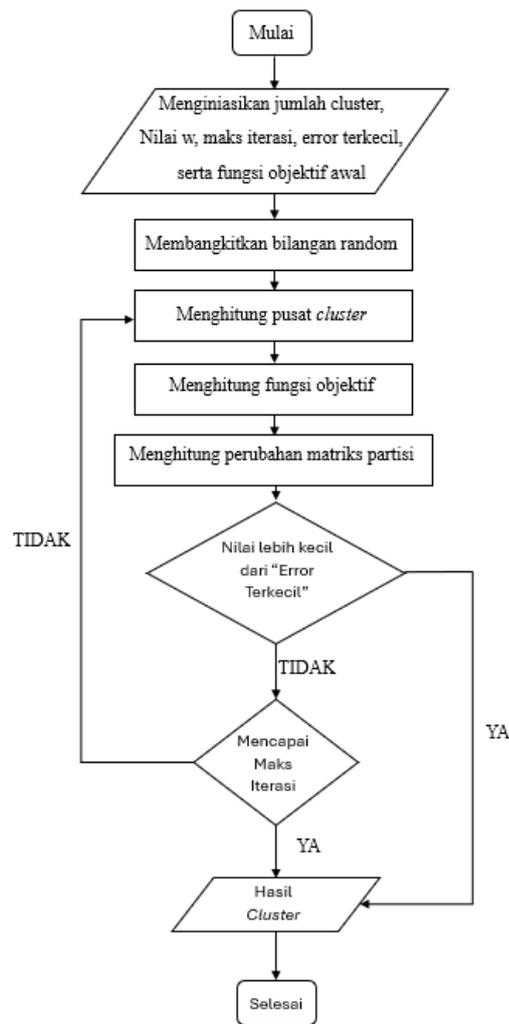
$$u_{ik} = \left( \sum_{j=1}^c \left( \frac{\|x_i - v_k\|}{\|x_i - v_j\|} \right)^{\frac{2}{m-1}} \right)^{-1} \dots\dots\dots (2.10)$$

Dimana,

- $v_{ik}$  = Derajat keanggotaan data point ke- $i$  terhadap cluster  $k$
- $\sum_{j=1}^c$  = Penjumlahan nilai di dalam tanda kurung untuk setiap cluster, mulai dari cluster pertama hingga cluster ke- $c$
- $x_i$  = Data point ke- $i$
- $v_k$  = Sentroid dari cluster  $k$
- $||x_i - v_k||$  = Jarak antara titik data  $x_i$  dan sentroid  $v_k$
- $||x_i - v_j||$  = Jarak antara titik data  $x_i$  dan sentroid  $v_j$
- $m$  = Parameter kefuzzyness

Sumber: (Latifah et al., 2022)

Namun jikalau iterasi telah berada pada iterasi kedua maupun iterasi-iterasi berikutnya, sebelum memasuki iterasi yang baru perlu dilakukan sebelumnya sebuah tahapan tambahan berupa pengamatan fungsi objektif. Dimana pada tahap ini kita akan melihat apakah nilai fungsi objektif telah lebih kecil dari nilai error terkecil (yang telah ditentukan di tahap awal perhitungan *Fuzzy C-Means*). Jikalau hasilnya lebih kecil maka iterasi dihentikan. Namun jika lebih besar dari nilai error tersebut, maka kita perlu memeriksa jumlah iterasi yang telah dilakukan. Jikalau iterasi yang dilakukan telah mencapai nilai maks iterasi (yang telah ditentukan di tahap awal perhitungan *Fuzzy C-Means*) maka iterasi dihentikan. Namun jikalau iterasi yang telah dilakukan belum mencapai maks iterasi, *data miners* dapat mengulang kembali iterasi ke tahap penentuan nilai pusat klaster. Adapun tahapan detail terkait alur pelaksanaan klustering dengan menggunakan Algoritma *Fuzzy C-Means* antarlain sebagai berikut:



**Gambar 2.3. Diagram Alur Fuzzy C-Means**

*Sumber: (Damayanti & Jakfar, 2023)*

Dapat kita perhatikan pada diagram alur diatas, proses analisis Algoritma Fuzzy C-Means diawali dengan menginisialisasikan jumlah *cluster* yang akan dibuat, nilai  $w$ , maksimal iterasi, error terkecil, serta fungsi objektif awal. Kemudian tahapan dilanjutkan dengan membangkitkan bilangan random, menghitung pusat *cluster*, menghitung fungsi objektif, dan diikuti penghitungan perubahan matriks partisi. Jikalau nilai yang diperoleh lebih kecil dari error terkecil maka tahapan selanjutnya ialah perolehan hasil *cluster*. Namun jikalau nilainya lebih besar dari

error terkecil, maka hasil tersebut akan diuji kembali terlebih dahulu untuk melihat tercapai/tidaknya maks Iterasi. Jikalau maks Iterasi tidak tercapai maka proses algoritma akan kembali pada tahap penghitungan pusat *cluster*, namun lain halnya jikalau hasil yang diperoleh telah mencapai maks Iterasi, maka hasil *cluster* sudah dapat diperoleh dan kini prosesnya telah selesai.

## 2.6. Rasio Gini (Indeks Gini)

Rasio Gini atau kerap disebut sebagai Indeks Gini maupun Koefisien Gini berfungsi untuk mengukur sudah sejauh mana tingkat ketidaksetaraan ekonomi telah terjadi disuatu daerah. Rasio ini memberikan nilai antara 0 hingga 1, jikalau suatu daerah memperoleh koefisien Gini sebesar 0 maka hal ini mengindikasikan daerah tersebut telah mencapai pemerataan perekonomian yang sempurna, namun jikalau nilai yang diperoleh ialah 1 maka hal ini mengindikasikan wilayah yang bersangkutan justru telah mencapai ketidakmerataan perekonomian yang sempurna pula. Secara umum rasio ini memiliki tiga tingkatan yang diantaranya ialah; jika suatu daerah memperoleh nilai rasio Gini sebesar 0,20 hingga 0,35 mengartikan daerah tersebut memperoleh tingkat ketidakmerataan yang rendah, kemudian dengan nilai 0,36 hingga 0,49 mengartikan daerah tersebut memperoleh tingkat ketidakmerataan sedang, dan terakhir dengan nilai 0,50 hingga 0,70 mengartikan daerah tersebut memperoleh tingkat ketidakmerataan yang tinggi (Pratama, 2022).

Untuk mengukur rasio Gini, umumnya dihitung melalui kurva Lorenz dengan indikator penilaian berupa tingkat ketimpangan pengeluaran perkapita yang terdiri atas pengeluaran makanan dan pengeluaran non-makanan. Adapun jenis-jenis makanan yang tergabung dalam kelompok data pengeluaran makanan ialah

makanan berbentuk padi-padian, umbi-umbian, ikan / udang / cumi-cumi / kerang, daging, telur dan susu, sayur-sayuran, kacang-kacangan, buah-buahan, minyak dan kelapa, bahan minuman, bumbu-bumbuan, konsumsi lainnya, makanan dan minuman jadi, serta rokok. Sedangkan untuk pengeluaran non-makanan terdiri atas pengeluaran terkait perumahan dan fasilitas rumah tangga, aneka barang dan jasa, pakaian, alas kaki, dan tutup kepala, barang tahan lama, pajak, pungutan, dan asuransi, serta keperluan pesta dan upacara/kenduri (Utara, 2023).

## 2.7. Microsoft Visual Basic.NET

Microsoft Visual Basic (VB) ialah bahasa pemrograman yang diciptakan dengan orientasi berbasis objek. Bahasa pemrograman ini umum digunakan untuk mengembangkan suatu perangkat lunak terutama perangkat lunak yang dijalankan pada platform Windows (Warasto & Jalil, 2023). Bahasa pemrograman ini populer digunakan sebagai salah satu *development tools* dalam mengembangkan suatu aplikasi, kepopuleran tersebut timbul karena bahasa pemrograman ini memiliki pendekatan visual dalam merancang antar muka (UI / *User-Interface*) sedangkan untuk melakukan peng-*coding*-annya menggunakan dialek bahasa Basic yang umumnya mudah dipelajari oleh para pemula.



**Gambar 2.4. Logo Visual Basic**

*Sumber: (Microsoft, 2024)*

Bahasa pemrograman hasil pengembangan Microsoft ini memiliki pendekatan visual, kata “visual” merujuk pada fungsi utamanya yang mampu membuat Graphical User Interface (GUI), hal ini tentunya memudahkan para pengembang dalam melakukan *development* suatu program. Para pengembang tidak perlu menuliskan program kode perbaris untuk membuat desain UI yang akan ditampilkan, melainkan cukup menggunakan fitur “*Drag-and-Drop*” objek yang akan terlibat dalam program yang akan dibangun, sehingga tidak heran jikalau bahasa pemrograman ini cukup populer dalam pengembangan berbagai jenis program komputer. Namun perlu diperhatikan, Visual Basic.NET hanya dapat dijalankan pada Sistem Operasi Windows saja (Warasto & Jalil, 2023).

## **2.8. Unified Modelling Language (UML)**

Unified Modelling Language (UML) ialah suatu metode yang dapat digunakan untuk menciptakan permodelan program yang akan dirancang secara visual pada proyek pengembangan sistem berbasis objek. UML kerap dianggap sebagai standar dalam mengembangkan suatu perangkat lunak oleh para pengembang sehingga UML kerap sekali diperlakukan layaknya cetak biru / *blueprint* didalam suatu proyek. Dalam UML, proses permodelan visual yang akan dibuat oleh para pengembang umumnya meliputi penggambaran, pembangunan sistem, penciptaan spesifikasi, serta tahapan dokumentasi. Adanya kompleksitas ini tentunya membuat UML mampu memastikan proses pengembangan dapat lebih terarah dan lebih mudah dipahami sehingga perangkat lunak yang dihasilkan tidak melenceng dari rancangan awal yang telah ditetapkan (Kinanti, 2022).

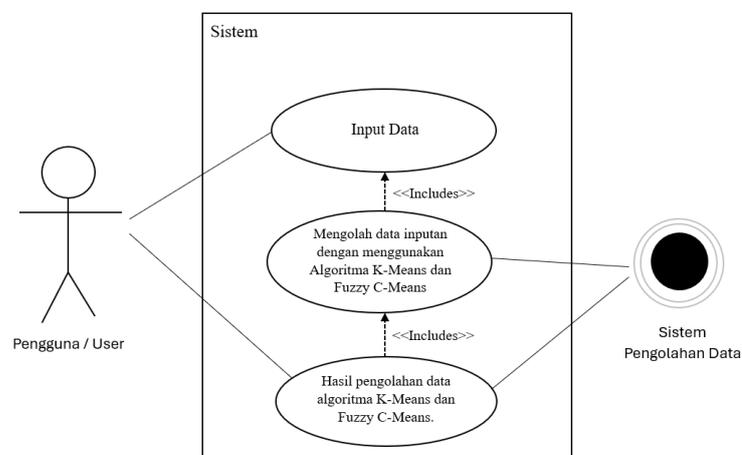
Secara umum, UML dibagi atas dua jenis yakni *Structural UML Diagram*, dan *Behavioral UML Diagram*. Adapun jenis-jenis UML yang tergabung dalam *Structural UML Diagram* secara umum ialah *Class Diagram*, *Component Diagram*, *Composite Structure Diagram*, *Deployment Diagram*, *Object Diagram*, dan *Package Diagram*. Sedangkan untuk *Behavioral UML Diagram* secara umum terdiri atas *Activity Diagram*, *Sequence Diagram*, *Communication Diagram*, *Interaction Overview Diagram*, *State Diagram*, *Timing Diagram*, dan *Use Case Diagram*. Namun dalam penelitian ini, peneliti memfokuskan pada UML berupa *Use Case Diagram* (UCD) sebagai permodelan sistem yang akan dibangun.

### **2.8.1. Use Case Diagram**

Use Case Diagram adalah salah satu bentuk UML yang dapat digunakan dalam mengembangkan suatu perangkat lunak. UML ini bekerja dengan menampilkan hubungan antara *actors* dan *Use Case*. Sehingga setidaknya terdapat tiga komponen utama dalam UML jenis ini yakni *Actors*, *Use Case*, dan *Relationship*. *Actors* atau Aktor mengacu pada pihak yang terlibat dalam mengoperasikan maupun bertransaksi informasi didalam sistem, dapat berupa user dalam bentuk individu maupun organisasi dan dapat pula berupa sistem pengolahan data. Selanjutnya *Use Case* mengacu pada *action* atau aksi yang dilakukan para Aktor kedalam sistem, sedangkan *Relationship* mengacu pada hubungan baik itu antara aktor dengan *Use Case* maupun antara *Use Case* dan *Use Case* itu sendiri, yang membedakan visualnya terdapat pada garis pehubungnya, dimana *Relationship Actors-Use Case* dihubungkan dengan garis panah lurus, sedangkan *Relationship Use Case-Use*

*Case* dihubungkan dengan garis panah putus-putus (Arianti, Fa'izi, Adam, & Wulandari, 2022).

Terkhusus *Relationship Use Case-Use Case*, terdapat dua jenis relationship yang digunakan yakni *extend* dan *includes*. *Extend* ialah bentuk *Relationship* dimana suatu *Use Case* mampu berdiri sendiri tanpa adanya *Use Case* tambahan, lain halnya dengan *includes* yang merupakan *Relationship* dimana *Use Case* tersebut tidak dapat berdiri sendiri dan membutuhkan tambahan *Use Case* yang bersandingan dengannya (Nurdianah, Hariyanto, Haeri, & Khotimah, 2021). Pemilihan *Use Case Diagram* sebagai UML dalam penelitian ini dilandasi atas kemampuan UCD yang mampu menjelaskan manfaat dari program yang akan dibangun berdasarkan sudut pandang entitas yang berkedudukan di luar sistem, dengan dirancangnya UCD pada suatu proyek para pengembang dapat lebih memahami kebutuhan maupun permintaan apa saja yang timbul saat sistem berjalan, serta dapat memberi pemahaman lebih mendalam pula terkait bagaimana sistem yang dibangun seharusnya bekerja. Adapun *Use Case Diagram* yang telah peneliti rancang untuk riset ini antarlain sebagai berikut:



**Gambar 2.5. Use Case Diagram**

## 2.9. Penelitian Terdahulu

Penelitian terdahulu kerap digunakan sebagai landasan para peneliti untuk memulai suatu riset dan memahami permasalahan serta fenomena yang timbul di lapangan. Melalui pengkajian penelitian terdahulu, suatu penelitian dapat dilaksanakan dengan lebih terarah sebab penelitian-penelitian terdahulu dapat digunakan sebagai panduan dalam teknis pelaksanaan suatu penelitian. Adapun sejumlah penelitian dengan topik pembahasan berupa perbandingan metode K-Means dan Fuzzy C-Means di beragam studi kasus dan indikator penilaian kinerja yang berhasil peneliti kumpulkan antarlain sebagai berikut:

**Tabel 2.1. Tabulasi Hasil Penelitian Terdahulu**

No	Nama Peneliti	Judul Penelitian	Metode	Hasil
1	Ummu Wachidatul Latifah, Sugiyarto, Suparman.  Jurnal: Desimal, Jurnal Matematika	K-Means and Fuzzy C-Means Algorithm Comparison On Regency/City Grouping In Central Java Province	K-Means dan Fuzzy C-Means (FCM)	FCM lebih baik daripada K-Means dengan standar deviasi 0.460093 lebih rendah dari standar deviasi yang diperoleh K-Means (0.473601).
2	Hamdi Syukron, Muhammad Fauzi Fayyad, Farin Junita Fauzan, Yulia Ikhsani, Umairah Rizkya Gurning.  Jurnal: MALCOM (Indonesian Journal of Machine Learning and Computer Science)	Perbandingan K-Means K-Medoids dan Fuzzy C-Means untuk Pengelompokan Data Pelanggan	K-Means, K-Medoids, dan Fuzzy C-Means	Algoritma K-Means lebih unggul dibanding K-Medoids dan Fuzzy C-means dibuktikannya pada nilai validitas DBI terbaik dengan perolehan nilai yaitu 0,167 pada jumlah klaster yaitu 6.

3	Lili Marlia, Putri Ngatmini, Puja Esteriani Tobing, Rahmaddeni.  Jurnal: SENTIMAS (Seminar Nasional Penelitian dan Pengabdian Masyarakat)	Analisis Segmentasi Pengunjung Mall Menggunakan Algoritma K-means dan Fuzzy C-means	K-means dan Fuzzy C-means	K-Means lebih unggul dalam mengidentifikasi segmen karakteristik pelanggan dengan tingkat akurasi lebih tinggi dari Fuzzy C-Means yakni sebesar 60%.
4	Belia Mailien, Admi Salma, Syafriandi, Dina Fitria  Jurnal: UNP, Journal of Statistics and Data Science	Comparison K-Means and Fuzzy C-Means Methods to Grouping Human Development Index Indicators in Indonesia	K-means dan Fuzzy C-means	Fuzzy C-Means lebih unggul daripada K-Means jikalau dilihat dari indikator penilaian berupa C_index and S_Dbw index. Fuzzy C-Means memiliki nilai yang lebih rendah pada kedua indikator tersebut yang mengartikan Fuzzy C-Means dapat menghasilkan pengelompokan yang lebih baik.
5	Pelsri Ramadar Noor Saputra, Ahmad Chusyairi  Jurnal: RESTI (Rekayasa Sistem dan Teknologi Informasi)	Perbandingan Metode Clustering dalam Pengelompokan Data Puskesmas pada Cakupan Imunisasi Dasar Lengkap	K-means dan Fuzzy C-means	Fuzzy C-Means menunjukkan hasil lebih unggul daripada K-Means baik itu pada indikator penilaian Silhoutte Coefficient maupun Standar Deviasinya.

*Sumber:* (Latifah et al., 2022; Mailien, Salma, Syafriandi, & Fitria, 2023; Marlia, Ngatmini, & Tobing, 2023; Saputra & Chusyairi, 2020; Syukron et al., 2022)

## BAB III

### METODOLOGI PENELITIAN

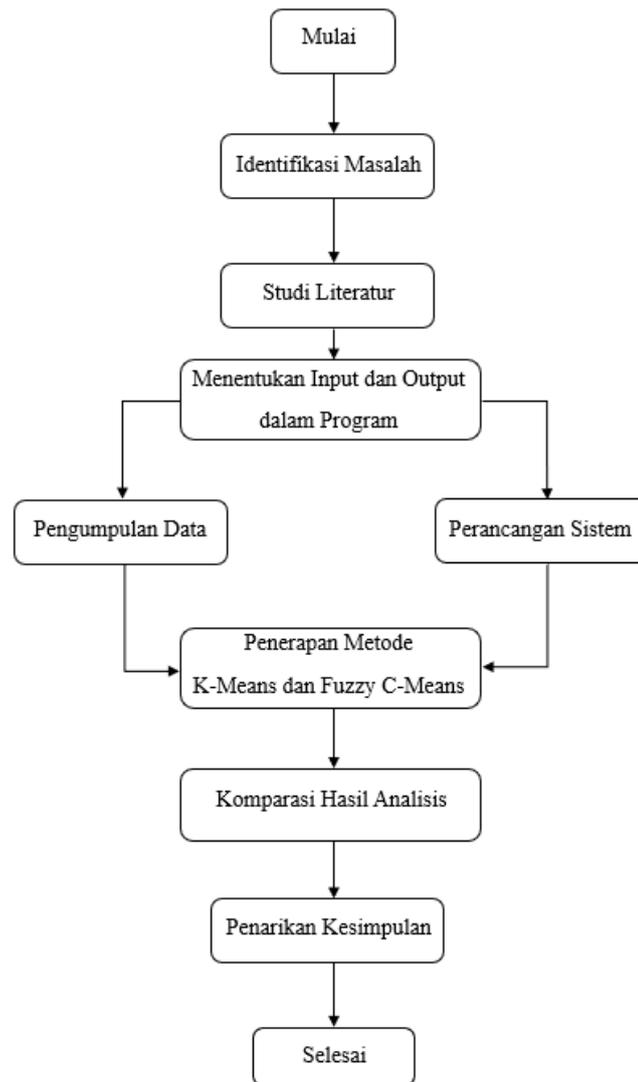
#### 3.1. Jenis Penelitian

Secara umum, penelitian dapat dipahami sebagai suatu proses pemahaman suatu perkara yang ilmiah melalui serangkaian prosedur yang terdiri atas pengumpulan data, pengolahan, analisis, serta penarikan kesimpulan berdasarkan metode, pendekatan, serta teknik-teknik tertentu untuk menyelesaikan suatu perkara ilmiah. Dalam pelaksanaannya, penelitian memiliki dua jenis varian yang dapat dipakai diantaranya penelitian kuantitatif (umumnya berbentuk penelitian eksperimen ataupun survey), dan penelitian kualitatif (umumnya berbentuk penelitian naturalistik) (Warasto & Jalil, 2023).

Dalam penelitian yang akan peneliti laksanakan, jenis penelitian yang digunakan ialah penelitian kuantitatif. Peneliti akan melakukan survey kuesioner kepada sejumlah responden, dan melalui jawaban para responden yang telah dikumpulkan, peneliti akan menggunakannya untuk membandingkan / melakukan analisis komparatif atas keakuratan kinerja metode klastering terkhususnya metode *K-Means* dan *Fuzzy C-Means* yang dibandingkan dengan perhitungan rasio Gini yang dikeluarkan oleh Badan Pusat Statistik Sumatera Utara.

#### 3.2. Tahapan Penelitian

Adapun tahapan penelitian yang akan peneliti laksanakan ialah antarlain sebagai berikut:



**Gambar 3.1. Alur Tahapan Penelitian**

Penelitian dimulai dengan pengidentifikasian masalah yang terfokus pada penggunaan metode perhitungan kurva Lorenz untuk menghitung rasio Gini. Penggunaan kurva tersebut terbilang rumit dan membutuhkan waktu lebih untuk pengolahannya, sehingga dibutuhkan sebuah solusi inovatif agar memudahkan dalam memahami ketidaksetaraan ekonomi melalui rasio Gini yang diprosikan oleh data pengeluaran perkapita masyarakat di wilayah tertentu. Kemudian penelitian dilanjutkan dengan melaksanakan studi literatur / studi pustaka. Tahap

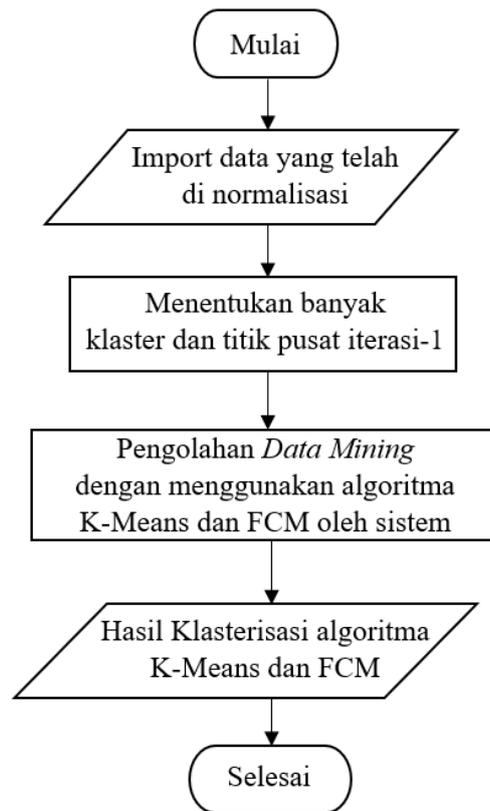
ini berguna bagi peneliti untuk memahami indikator-indikator apa saja yang akan terlibat dalam melakukan perhitungan rasio Gini yang diproksikan oleh data pengeluaran perkapita. Indikator tersebut diantaranya ialah jumlah pengeluaran perkapita yang dikategorikan berdasarkan jenis pengeluarannya.

Setelah peneliti memahami indikator-indikator yang terlibat dalam perhitungan rasio Gini, peneliti menentukan jenis data yang akan diinputkan kedalam sistem serta output yang akan dihasilkan berdasarkan pengolahan data inputan tersebut. Kemudian pada tahapan selanjutnya, peneliti akan menyebarkan kuesioner kepada responden untuk memperoleh dataset yang akan diolah, sembari menunggu pengumpulan data memenuhi syarat jumlah data yang akan diolah, peneliti akan merancang sistem program yang berfungsi untuk penerapan metode *K-means* dan *Fuzzy C-Means*. Apabila sistem yang dirancang telah siap untuk dipakai, serta data responden telah memenuhi syarat jumlah yang telah ditentukan, maka tahapan selanjutnya peneliti akan menerapkan metode *K-Means* dan *Fuzzy C-means* demi menghasilkan luaran / output berupa nilai dari masing-masing metode.

Hasil tersebut kemudian dibandingkan keakuratannya dengan perhitungan rasio Gini yang dikeluarkan oleh Badan Pusat Statistik Sumatera Utara dan selanjutnya akan ditarik kesimpulan terkait metode manakah diantara *K-Means* dan *Fuzzy C-means* yang dapat menghasilkan nilai output yang paling mendekati perhitungan rasio Gini yang dikeluarkan oleh Badan Pusat Statistik Sumatera Utara.

### **3.3. Flowchart Sistem**

Adapun *flowchart* atas sistem yang akan peneliti bangun dalam penelitian ini antarlain sebagai berikut:



**Gambar 3.2. Flowchart Sistem**

Pada *flowchart* diatas, dapat kita perhatikan setidaknya terdapat 6 tahapan yang harus dilalui ketika mengoperasikan sistem yang akan peneliti kembangkan. Adapun keterangannya antarlain sebagai berikut:

1. Pada tahap awal, pengguna akan menjalankan sistem terlebih dahulu,
2. Kemudian pengguna dapat mengimport data mentah yang telah dinormalisasikan kedalam sistem,
3. Pengguna selanjutnya dapat menentukan berapa banyak klaster yang hendak dibentuk, pada tahap ini pengguna tidak perlu menghitung secara manual, melainkan sistem yang akan menghitungnya dengan menggunakan metode Elbow, sedangkan untuk pusat iterasi pertama nilainya akan ditentukan secara

acak (*random*) sesuai dengan ketentuan dalam melakukan analisis dengan algoritma K-Means dan Fuzzy C-Means,

4. Selanjutnya setelah tahapan ketiga selesai, sistem akan melakukan pengolahan *data mining* dengan menggunakan algoritma K-Means dan Fuzzy C-Means. Seluruh prosedur dari awal hingga tercapainya iterasi tanpa adanya perubahan titik keanggotaan akan dilakukan oleh sistem,
5. Setelah pengolahan data telah selesai, sistem akan menampilkan hasil klusterisasi baik itu klusterisasi K-Means dan juga Fuzzy C-Means. Hasil tersebut meliputi visualisasi hasil atas kedua algoritma, nilai Dunn Index yang diperoleh masing-masing algoritma, serta lama waktu yang dibutuhkan masing-masing algoritma untuk menyelesaikan pengolahan data, dan
6. Prosedur pengolahan telah selesai.

### **3.4. Defenisi Operasional**

Defenisi operasional dapat dipahami sebagai deskriptif penjelasan yang konkrit mengenai konsep-konsep abstrak yang terdapat didalam penelitian sehingga dapat diubah menjadi suatu hal yang dapat diamati maupun diukur. Melalui Defenisi operasional, konsep-konsep abstrak tersebut dapat di-operasionalkan sehingga dapat digunakan sesuai dengan tujuan penelitian yang ada (Bertua, Panggabean, Yolanda, Naftalia, & Sitompul, 2024). Berdasarkan penjelasan yang telah dipaparkan sebelumnya, maka peneliti dapat membentuk defenisi operasional dalam penelitian ini sebagai berikut:

Tabel 3.1. Defenisi Operasional

Variable	Defenisi	Indikator
Algoritma K-Means	Algoritma K-Means ialah algoritma <i>unsupervised learning</i> (tidak diklasifikasi dan tidak diberi label) yang berfungsi untuk mengelompokkan beberapa data ke dalam kelompok sehingga data-data tersebut dikumpulkan berdasarkan kemiripan karakteristik yang terdapat pada satu sama lain.	Efisiensi Waktu, Keakuratan Data, Nilai <i>Dunn Index</i>
Algoritma Fuzzy C-Means	Algoritma Fuzzy C-Means ialah algoritma <i>unsupervised learning</i> (tidak diklasifikasi dan tidak diberi label) yang berfungsi untuk membagi data-data yang terdapat di dalam dataset kedalam beragam klaster berdasarkan derajat keanggotaannya.	Efisiensi Waktu, Keakuratan Data, Nilai <i>Dunn Index</i>
Rasio Gini	Rasio Gini ialah rasio yang berfungsi untuk mengukur sudah sejauh mana tingkat ketidaksetaraan ekonomi telah terjadi disuatu daerah. Rasio ini memberikan nilai antara 0 hingga 1, jikalau suatu daerah memperoleh koefisien Gini sebesar 0 maka hal ini mengartikan bahwa daerah tersebut telah mencapai pemerataan perekonomian yang	1. Pengeluaran Makanan, 2. Pengeluaran Non-Makanan.

	sempurna, namun sebaliknya jikalau nilai yang diperoleh ialah 1 maka hal ini mengartikan bahwa wilayah yang bersangkutan justru telah mencapai ketidakmerataan perekonomian yang sempurna pula.	
Microsoft Visual Basic.NET	Microsoft Visual Basic (VB) adalah bahasa pemrograman yang diciptakan berorientasi berbasis objek. Bahasa pemrograman ini umumnya digunakan untuk mengembangkan suatu perangkat lunak terutama perangkat lunak yang dijalankan pada platform Windows	Bahasa Pemrograman

### 3.5. Teknik Pengumpulan Data

Teknik pengumpulan data merupakan aspek penting dalam suatu penelitian, dengan adanya pengumpulan data dari narasumber yang terlibat dalam penelitian, para peneliti dapat memperoleh informasi faktual yang dapat digunakan untuk menganalisis suatu fenomena yang diamati. Dalam penelitian ini, peneliti menggunakan teknik pengumpulan data berupa survei kuesioner (angket) yang disebarkan kepada masyarakat di tiga kabupaten/kota di provinsi Sumatera Utara.

Survei kuesioner (angket) merupakan instrumen pengumpulan data dimana data-data dikumpulkan dari para responden dalam bentuk pertanyaan-pertanyaan yang telah terstandarisasi sehingga data-data yang dikumpulkan akan berbentuk

angka-angka maupun informasi lainnya yang telah terarah sesuai pertanyaan yang diajukan. Adapun tujuan utama dari pengumpulan data menggunakan instrumen penelitian kuesioner (angket) ialah untuk memperoleh informasi yang relevan dengan tujuan penelitian yang dilaksanakan, memperoleh nilai validitas yang tinggi, serta memperoleh informasi yang reliabel sesuai dengan fakta yang terjadi di lapangan (Nursalam & Djaha, 2023). Selain itu, peneliti ini juga menerapkan teknik pengumpulan data berupa studi literatur terhadap sejumlah modul terkait rasio Gini yang telah dipublikasikan oleh Badan Pusat Statistik.

### **3.6. Teknik Pengambilan Sample**

Teknik pengambilan sample merupakan aspek yang penting untuk dilakukan dalam suatu penelitian. Dengan adanya penentuan teknik pengambilan sample, peneliti dapat memilih data yang akan digunakan dalam suatu analisis sehingga data-data yang akan terlibat sesuai dengan kebutuhan analisis yang akan dilakukan. Dalam penelitian ini, peneliti akan menggunakan teknik pengambilan sample berupa *Quota Sampling*. Dalam teknik ini, data yang akan digunakan jumlahnya ditentukan oleh peneliti (Veronica, Yuliana, & Weny, 2023), adapun jumlah sample yang peneliti tentukan ialah 200 sample per-kabupaten/kota, dikarenakan kuesioner akan disebarakan kepada masyarakat di tiga kabupaten/kota di provinsi Sumatera Utara dengan demikian total keseluruhan sample yang akan terlibat dalam penelitian ini berjumlah 600 sample responden.

### **3.7. Tempat dan Jadwal Penelitian**

Dalam penelitian ini, peneliti akan mengumpulkan data pengeluaran makanan dan non-makanan kepada masyarakat di kota Medan, kota Binjai, dan kabupaten Deli Serdang, maka dari itu penelitian akan dilakukan pada tiga kabupaten/kota tersebut. Kemudian peneliti juga akan menggunakan data statistik rasio Gini yang dipublikasikan oleh Badan Pusat Statistik Sumatera Utara yang beralamatkan di Jl. Asrama No.179, Kota Medan, Sumatera Utara sebagai indikator penilaian tingkat keakuratan metode K-Means dan Fuzzy C-Means. Adapun pelaksanaan penelitian ini berlangsung dari bulan januari hingga akhir mei tahun 2024.

## BAB IV

### HASIL DAN PEMBAHASAN

#### 4.1. Data Awal

Dalam penelitian yang akan peneliti laksanakan, data yang digunakan sebagai bahan pengolahan data menggunakan algoritma K-Means dan Fuzzy C-Means ialah data pengeluaran masyarakat yaitu berupa gabungan antara pengeluaran makanan dan non-makanan. Ketentuan ini mengacu kepada tata cara BPS untuk mengukur rasio GINI, yakni menggunakan data pengeluaran masyarakat yang terdiri atas kedua jenis pengeluaran tersebut. Adapun jenis-jenis makanan yang tergabung dalam kelompok data pengeluaran makanan ialah makanan berbentuk padi-padian, umbi-umbian, ikan / udang / cumi-cumi / kerang, daging, telur dan susu, sayur-sayuran, kacang-kacangan, buah-buahan, minyak dan kelapa, bahan minuman, bumbu-bumbuan, konsumsi lainnya, makanan dan minuman jadi, serta rokok. Sedangkan untuk pengeluaran non-makanan terdiri atas pengeluaran terkait perumahan dan fasilitas rumah tangga, aneka barang dan jasa, pakaian, alas kaki, tutup kepala, barang tahan lama, pajak, pungutan, asuransi, serta keperluan pesta dan upacara/kenduri.

Namun pada riset ini, peneliti tidak merincikan makanan maupun non-makanan berdasarkan jenis barangnya, melainkan memuatnya sebagai pengeluaran makanan dan non-makanan saja. Peneliti memperoleh data pengeluaran masyarakat dengan melakukan penyebaran kuesioner daring kepada masyarakat yang berdomisili di tiga kabupaten/kota di Sumatera Utara, yakni kota Medan, kota Binjai, dan

kabupaten Deli Serdang. Adapun hasil pengumpulan data yang peneliti peroleh diantaranya:

**Tabel 4.1. Data Awal Penelitian dalam Ribuan Rupiah (000)**

No Responden	Kota Medan			Kota Binjai			Kabupaten Deli Serdang		
	Makanan	Non-Makanan	Total	Makanan	Non-Makanan	Total	Makanan	Non-Makanan	Total
1	1200	4200	5400	1000	2000	3000	6000	4000	10000
2	1250	2100	3350	5000	6000	11000	14000	8000	22000
3	1300	1200	2500	7000	12000	19000	5000	6000	11000
4	2500	1800	4300	5400	4000	9400	9000	10500	19500
5	1350	2100	3450	4500	1500	6000	1000	3000	4000
6	4500	3500	8000	2800	1200	4000	4000	3500	7500
7	2000	2000	4000	3800	2500	6300	6000	5000	11000
8	5000	10000	15000	7200	9000	16200	4000	6500	10500
9	3000	2500	5500	5000	4200	9200	4000	7000	11000
10	7000	4000	11000	2000	2500	4500	5000	6000	11000
...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...
199	2800	2500	5300	3200	2000	5200	1400	1800	3200
200	1800	2500	4300	2500	2500	5000	2200	2000	4200

#### 4.2. Contoh Perhitungan Manual

Pada perhitungan manual K-Means Clustering yang akan peneliti lakukan, sebagai contoh peneliti akan menggunakan 5 sample data yang telah ternormalisasi dengan rincian sebagai berikut:

**Tabel 4.2. Sample Data Contoh Perhitungan Manual KM**

Data Awal Pengeluaran Makanan	Data Awal Pengeluaran Non-Makanan	Data Pengeluaran Makanan Ternormalisasi	Data Pengeluaran Non-Makanan Ternormalisasi
1200000	4200000	0	1
1250000	2100000	0.038462	0.3
1300000	1200000	0.076923	0
2500000	1800000	1	0.2
1350000	2100000	0.115385	0.3

Dalam memperoleh data pengeluaran yang ternormalisasi, terlebih dahulu peneliti menentukan siapakah yang akan menjadi Data terbesar dan data terkecil. Kemudian memasukkan setiap titik data kedalam rumus normalisasi Min-Max sebagai berikut:

$$\text{Data Normalisasi} = \frac{X_{ori} - X_{min}}{X_{max} - X_{min}}$$

Sehingga didapat salah satu contoh perhitungannya sebagai berikut:

$$\text{Data Normalisasi} = \frac{1200000 - 1200000}{2500000 - 1200000} = \frac{0}{2500000 - 1200000} = \frac{0}{1300000} = 0$$

Perhitungan ini kemudian dilakukan pula terhadap titik data lainnya juga baik itu pada Data Pengeluaran Makanan dan juga Data Pengeluaran Non-Makanan, sehingga didapati hasil layaknya yang terdapat didalam tabel 4.2. Dalam percontohan hitungan manual ini, peneliti akan menggunakan 2 centroid (titik data), sehingga diperoleh perhitungan sebagai berikut:

- a. Pemilihan Centroid Awal secara acak, secara acak centroid awal akan peneliti letakkan pada titik (0,1) dan (0.115385, 0.3).
- b. Perhitungan jarak setiap titik data kepada setiap centroid yang ada dengan menggunakan rumus *Euclidean Distance*:

$$d = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}$$

b.a. Perhitungan ke Centroid 1 (0,1)

(0, 1) ke (0, 1): 0

(0.038462, 0.3) ke (0, 1):

$$\sqrt{(0.038462 - 0)^2 + (0.3 - 1)^2} = \sqrt{0.0014803 + 0.49} \approx 0.700$$

(0.076923, 0) ke (0, 1):

$$\sqrt{(0.076923 - 0)^2 + (0 - 1)^2} = \sqrt{0.0059 + 1} \approx 1.001$$

(1, 0.2) ke (0, 1):

$$\sqrt{(1 - 0)^2 + (0.2 - 1)^2} = \sqrt{1 + 0.64} \approx 1.204$$

(0.115385, 0.3) ke (0, 1):

$$\sqrt{(0.115385 - 0)^2 + (0.3 - 1)^2} = \sqrt{0.0133 + 0.49} \approx 0.711$$

b.b. Perhitungan ke Centroid 2 (0.115385, 0.3)

(0, 1) ke (0.115385, 0.3):

$$\sqrt{(0 - 0.115385)^2 + (1 - 0.3)^2} = \sqrt{0.0133 + 0.49} \approx 0.711$$

(0.038462, 0.3) ke (0.115385, 0.3):

$$\sqrt{(0.038462 - 0.115385)^2 + (0.3 - 0.3)^2} = \sqrt{0.0059} \approx 0.077$$

(0.076923, 0) ke (0.115385, 0.3):

$$\sqrt{(0.076923 - 0.115385)^2 + (0 - 0.3)^2} = \sqrt{0.00148 + 0.09} \approx 0.302$$

(1, 0.2) ke (0.115385, 0.3):

$$\sqrt{(1 - 0.115385)^2 + (0.2 - 0.3)^2} = \sqrt{0.783 + 0.01} \approx 0.794$$

(0.115385, 0.3) ke (0.115385, 0.3): 0

c. Kategorisasikan setiap titik data kedalam kluster berdasarkan jarak terdekat, berdasarkan perhitungan diatas, dapat ditarik kesimpulan anggota setiap kluster sebagai berikut:

Kluster 1 (Centroid 1): (0, 1)

Kluster 2 (Centroid 2):

(0.038462, 0.3), (0.076923, 0), (1, 0.2), (0.115385, 0.3)

d. Memperbarui titik pusat (centroid baru)

$$\text{Centroid 1} = \left( \frac{0}{1}, \frac{1}{1} \right) = (0, 1)$$

$$\text{Centroid 2} = \left( \frac{0.038462 + 0.076923 + 1 + 0.115385}{4}, \frac{0.3 + 0 + 0.2 + 0.3}{4} \right) = \left( \frac{1.23077}{4}, \frac{0.8}{4} \right) = (0.3076925, 0.2)$$

Setelah titik pusat (centroid baru) telah diperoleh, tahapan selanjutnya ialah menghitung kembali rangkaian prosedur serupa mulai dari perhitungan *Euclidean Distance* (dengan menggunakan titik pusat (centroid baru) hingga pada tahap memperoleh titik pusat (centroid baru) kembali. Siklus perhitungan (iterasi) akan dihentikan ketika titik pusat (centroid baru) yang diperoleh ialah sama dengan titik pusat (centroid baru) yang didapati pada siklus (iterasi) sebelumnya). Sedangkan untuk perhitungan manual Fuzzy C-Means Clustering yang akan peneliti lakukan, sebagai contoh peneliti akan menggunakan 5 sample data yang telah ternormalisasi dengan rincian sebagai berikut:

**Tabel 4.3. Sample Data Contoh Perhitungan Manual FCM**

Data Awal Pengeluaran Makanan	Data Awal Pengeluaran Non-Makanan	Data Pengeluaran Makanan Ternormalisasi	Data Pengeluaran Non-Makanan Ternormalisasi
1200000	4200000	0	1
1250000	2100000	0.038462	0.3
1300000	1200000	0.076923	0
2500000	1800000	1	0.2
1350000	2100000	0.115385	0.3

Dalam memperoleh data pengeluaran yang ternormalisasi, terlebih dahulu peneliti menentukan siapakah yang akan menjadi data terbesar dan data terkecil. Kemudian memasukkan setiap titik data kedalam rumus normalisasi Min-Max sebagai berikut:

$$\text{Data Normalisasi} = \frac{X_{ori} - X_{min}}{X_{max} - X_{min}}$$

Sehingga didapat perhitungan sebagai berikut:

$$\text{Data Normalisasi} = \frac{1200000 - 1200000}{2500000 - 1200000} = \frac{0}{2500000 - 1200000} = \frac{0}{1300000} = 0$$

Perhitungan dengan rumus serupa kemudian dilakukan pula terhadap titik data lainnya juga baik itu pada Data Pengeluaran Makanan dan juga Data Pengeluaran Non-Makanan, sehingga didapat hasil seperti pada tabel 4.3. Dalam percontohan hitungan manual ini, peneliti akan menggunakan 2 centroid (titik data), sehingga diperoleh perhitungan sebagai berikut:

- a. Inisialisasi Fitur Klustering, dengan rincian yakni jumlah Kluster ialah 2, sedangkan derajat Fuzziness (m) ialah 2.
- b. Pemilihan Centroid Awal secara acak, secara acak centroid awal akan peneliti letakkan pada titik (0.1, 0.4) dan (0.9, 0.1).
- c. Menghitung jarak setiap titik data terhadap setiap centroid dengan menggunakan rumus *Euclidean Distance* sebagai berikut:

$$d = \sqrt{[(X_2 - X_1)^2 + (Y_2 - Y_1)^2]}$$

- d. Perhitungan ke Centroid 1 (0.1, 0.4)

$$\sqrt{(0 - 0.1)^2 + (1 - 0.4)^2} \approx 0.608$$

$$\sqrt{(0.38462 - 0.1)^2 + (0.3 - 0.4)^2} \approx 0.122$$

$$\sqrt{(0.76923 - 0.1)^2 + (0 - 0.4)^2} \approx 0.401$$

$$\sqrt{(1 - 0.1)^2 + (0.2 - 0.4)^2} \approx 0.921$$

$$\sqrt{(0.115385 - 0.1)^2 + (0.3 - 0.4)^2} \approx 0.100$$

e. Perhitungan ke Centroid 2 (0.9, 0.1)

$$\sqrt{(0 - 0.9)^2 + (1 - 0.1)^2} \approx 1.345$$

$$\sqrt{(0.38462 - 0.9)^2 + (0.3 - 0.1)^2} \approx 0.874$$

$$\sqrt{(0.76923 - 0.9)^2 + (0 - 0.1)^2} \approx 0.825$$

$$\sqrt{(1 - 0.9)^2 + (0.2 - 0.1)^2} \approx 0.141$$

$$\sqrt{(0.115385 - 0.9)^2 + (0.3 - 0.1)^2} \approx 0.806$$

f. Menghitung Matriks Keanggotaan ( $U_{ij}$ ) Awal dengan rumus sebagai berikut:

$$U_{ij} = \frac{1}{\left(\frac{\|x_i - v_j\|}{\|x_i - v_1\|}\right)^2 + \left(\frac{\|x_i - v_j\|}{\|x_i - v_2\|}\right)^2}$$

$$u_{11} = \frac{1}{\left(\frac{0.608}{0.608}\right)^2 + \left(\frac{0.608}{1.345}\right)^2} \approx \frac{1}{1 + 0.204} \approx 0.803$$

$$u_{12} = \frac{1}{\left(\frac{1.345}{0.608}\right)^2 + \left(\frac{1.345}{1.345}\right)^2} \approx \frac{1}{4.889 + 1} \approx 0.170$$

Perhitungan serupa pun dilakukan pada setiap titik data terhadap setiap centroid. Pada bagian awal perhitungan (misalnya  $u_{11}$ ), angka digit pertama merepresentasikan titik data yang akan dihitung, sedangkan digit kedua merepresentasikan titik centroid yang dihitung. Melalui perhitungan matrik keanggotaan ( $U_{ij}$ ) diatas terhadap seluruh data, diperoleh hasil sebagai berikut:

$$U = \begin{bmatrix} 0.830 & 0.170 \\ 0.980 & 0.020 \\ 0.788 & 0.212 \\ 0.023 & 0.977 \\ 0.987 & 0.013 \end{bmatrix}$$

- g. Memperbarui Centroid Fuzzy C-Means Clustering melalui penggunaan rumus sebagai berikut:

$$V_j = \frac{\sum_{i=1}^N u^{m_{ij}} \cdot x_i}{\sum_{i=1}^N u^{m_{ij}}}$$

Pada rumus diatas,  $U_{ij}$  merupakan nilai derajat keanggotaan, sedangkan  $x_i$  merupakan titik data yang telah ternormalisasikan.

- g.a. Menghitung pembaharuan centroid atas titik pusat 1 (0.1, 0.4)

$$v_1 = \frac{0.830^2 \cdot (0.1) + 0.980^2 \cdot (0.038462, 0.3) + 0.788^2 \cdot (0.076923, 0) + 0.023^2 \cdot (1, 0.2) + 0.987^2 \cdot (0.115385, 0.3)}{0.830^2 + 0.980^2 + 0.788^2 + 0.023^2 + 0.987^2}$$

$$v_1 = \frac{(0, 0.6889) + (0.0364, 0.294) + (0.0479, 0) + (0.0005, 0.0001) + (0.1126, 0.292)}{0.6889 + 0.9604 + 0.6204 + 0.0005 + 0.9742}$$

$$v_1 = \frac{(0.1974, 1.275)}{3.2444} \approx (0.061, 0.393)$$

- g.b. Menghitung pembaharuan centroid atas titik pusat 2 (0.9, 0.1)

$$v_2 = \frac{0.170^2 \cdot (0.1) + 0.020^2 \cdot (0.038462, 0.3) + 0.212^2 \cdot (0.076923, 0) + 0.977^2 \cdot (1, 0.2) + 0.013^2 \cdot (0.115385, 0.3)}{0.170^2 + 0.020^2 + 0.212^2 + 0.977^2 + 0.013^2}$$

$$v_2 = \frac{(0, 0.0289) + (0, 0.0001) + (0.035, 0) + (0.954, 0.189) + (0.00002, 0.00005)}{0.0289 + 0.0004 + 0.045 + 0.954 + 0.0002}$$

$$v_2 = \frac{(0.9575, 0.218)}{1.0285} \approx (0.931, 0.212)$$

Sehingga berdasarkan perhitungan tersebut, diperoleh centroid (titik data) baru yang dapat digunakan untuk iterasi (perulangan) selanjutnya sebagai berikut:

$$v_1^{\text{baru}} = (0.061, 0.393)$$

$$v_2^{\text{baru}} = (0.931, 0.212)$$

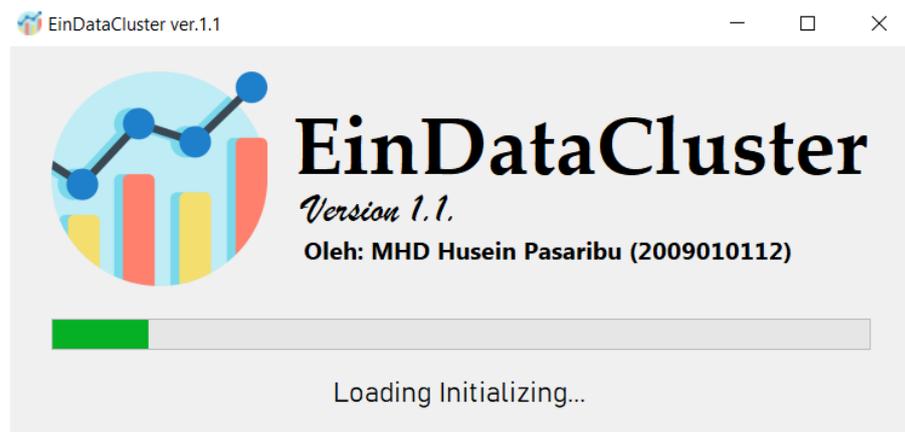
Siklus perhitungan ini akan kembali dilakukan pada iterasi (perulangan) selanjutnya mulai dari tahap perhitungan matriks keanggotaan dengan titik pusat baru yang diperoleh pada akhir iterasi sebelumnya. Siklus berhenti ketika kondisi konvergensi diperoleh, yakni ketika perubahan pada matriks

keanggotaan untuk setiap titik data ialah sama antara matriks keanggotaan iterasi baru dengan iterasi sebelumnya.

### 4.3. Implementasi *User-Interface*

Implementasi sistem dapat kita pahami sebagai tahap translasi kebutuhan perancangan dan pembangunan sistem aplikasi kedalam bentuk perangkat lunak. Tujuan utama dari implementasi *User-Interface* ialah untuk menyediakan sarana bagi para pengguna sehingga sistem aplikasi yang dibangun dan dirancang dapat digunakan dan dioperasikan dengan mudah. Singkatnya, *User-Interface* adalah bentuk sistem yang akan terlihat pada layar para pengguna. Adapun *User-Interface* yang telah peneliti rancang diantaranya sebagai berikut:

#### 4.2.1 Tampilan Loading Sistem



**Gambar 4.1. Tampilan Loading Sistem**

Tampilan ini merupakan tampilan awal yang akan terlihat pada layar pengguna. Tidak ada yang harus dilakukan oleh pengguna pada bagian ini, secara otomatis *loading-bar* akan berjalan dengan sendirinya dari sisi kiri kearah kanan, kemudian saat *loading-bar* terisi secara keseluruhan,

tampilan ini akan tertutup dan pengguna akan dibawa kepada tampilan menu utama.

#### 4.2.2 Tampilan Menu Utama Aplikasi



**Gambar 4.2. Tampilan Menu Utama Aplikasi**

Pada tampilan ini, pengguna dapat memilih dua opsi yang akan ditampilkan, yakni *Data Processing* serta *About Us*. Pengguna dapat mengarahkan *cursor* kearah salah satu gambar maupun tulisan yang tersedia diantara keduanya kemudian menekan klik kiri, maka tampilan selanjutnya akan tertampil, sedangkan tampilan menu utama aplikasi akan tertutup.

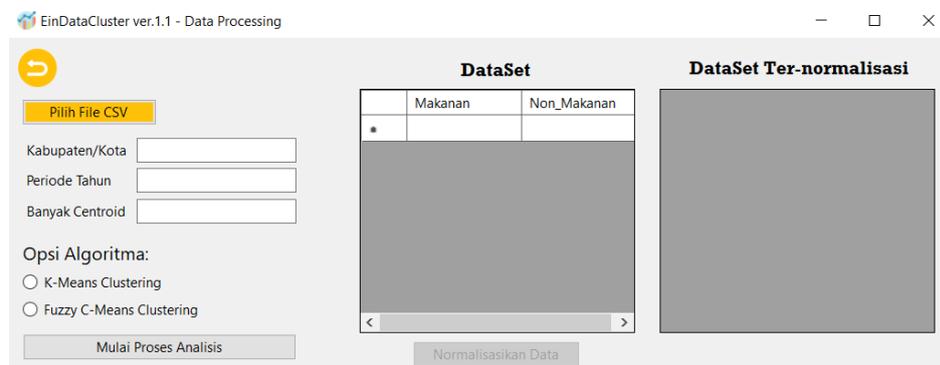
#### 4.2.3 Tampilan Menu *About Us*



**Gambar 4.3. Tampilan Menu *About Us***

Pada tampilan Menu *About us*, sistem akan menampilkan deskripsi terkait pengembang aplikasi, latar belakang perancangan aplikasi, serta kegunaan, fungsi, dan tujuan dikembangkannya aplikasi ini. Selain itu, pada sisi kanan tampilan, tertampil pula gambar gedung rektorat Universitas Muhammadiyah Sumatera Utara. Selanjutnya pada sisi kiri-atas tampilan, peneliti juga telah menyematkan sebuah tombol yang apabila pengguna tekan, maka halaman *About Us* akan segera tertutup, dan tampilan menu utama akan kembali diperlihatkan kepada para pengguna.

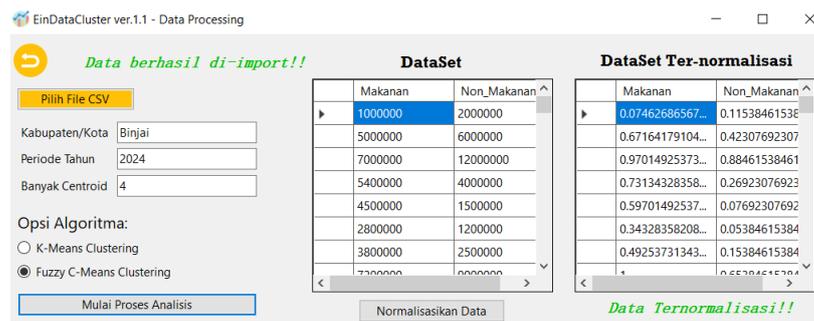
#### 4.2.4 Tampilan Menu *Data Processing*



**Gambar 4.4. Tampilan Menu *Data Processing***

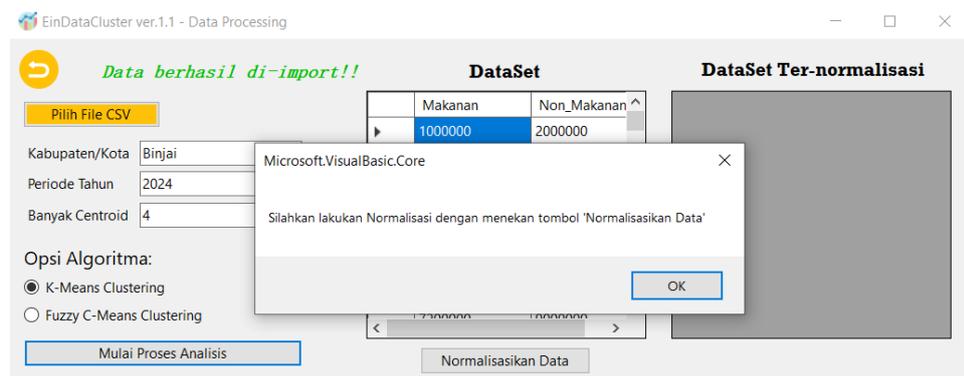
Pada tampilan ini, sistem akan memperlihatkan sebuah formulir yang dapat diisi oleh pengguna. Adapun formulir ini mencakup tombol yang dapat digunakan untuk meng-import berkas ber-ekstensi .csv untuk segera diolah. Apabila berkas ber-ekstensi .csv telah dipilih dan berhasil terbaca, maka isian dari berkas ber-ekstensi .csv tersebut akan tertampil pada tabel yang berada tepat dibawah tulisan “DataSet”, kemudian pengguna dapat menekan tombol “Normalisasikan Data” yang berada tepat dibawah tabel tersebut untuk menormalisasikan data yang dimiliki. Apabila tidak ada data yang terbaca pada tabel DataSet, tombol Normalisasi Data tidak akan bekerja.

Normalisasi Data perlu dilakukan agar skala yang terdapat didalam DataSet sama antara satu data dengan data lainnya, dalam proses ini tombol “Normalisasi Data” akan membaca DataSet yang telah diinputkan, menentukan nilai tertinggi dan juga nilai terendah dari DataSet tersebut, kemudian kepada setiap kolom yang berisikan titik data akan dinormalisasikan menggunakan metode normalisasi Min-Max. Normalisasi data juga dapat mempercepat tercapainya kondisi konvergensi dalam pengolahan data, serta menghasilkan luaran yang lebih akurat dan andal. Hasil normalisasi akan ditampilkan pada tabel yang terletak tepat dibawah tulisan “DataSet Ter-normalisasi”. Kedua tabel ini memanfaatkan kontrol DataGridView yang tersedia didalam Toolbox bawaan (*default*) yang disediakan oleh VB.NET. Formulir selanjutnya ialah pengisian kabupaten/kota yang akan dianalisis, periode tahun yang akan diolah, serta banyak *Centroid* (titik pusat) yang akan digunakan. Ketiga formulir ini memanfaatkan kontrol *TextBox* yang tersedia didalam Toolbox bawaan (*default*) yang disediakan oleh VB.NET. pengguna juga dapat memilih salah satu opsi algoritma yang akan digunakan. Fitur ini memanfaatkan kontrol *RadioButton* yang tersedia didalam Toolbox bawaan (*default*) yang disediakan oleh VB.NET.



**Gambar 4.5. Tampilan Menu *Data Processing* setelah formulir terisi**

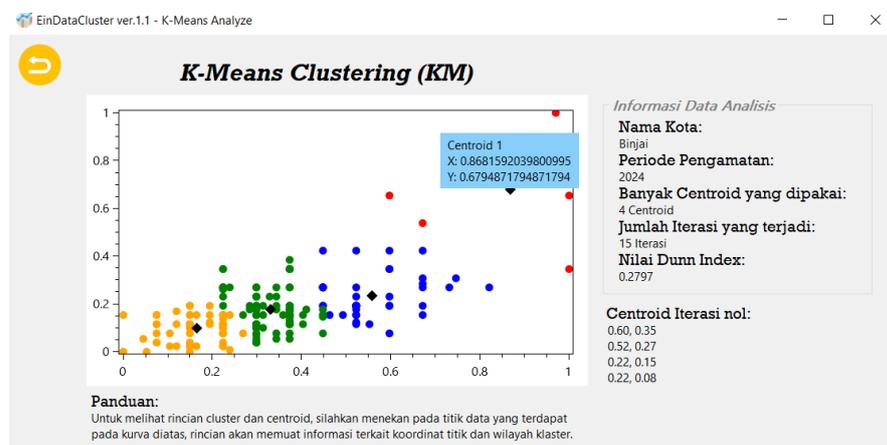
Apabila seluruh formulir telah terisi maka pengguna dapat menekan tombol “Mulai Proses Analisis” yang terletak pada sisi kiri bawah tampilan layar. Apabila masih terdapat formulir yang belum terisi, sistem tidak akan membawa pengguna kepada halaman selanjutnya, melainkan pengguna akan memperoleh notifikasi yang menyatakan bahwa terdapat field yang belum terisi dengan benar. Notifikasi yang timbul disesuaikan dengan formulir apa yang belum diisi dengan benar oleh pengguna. Adapun salah satu bentuk notifikasi tersebut ialah sebagai berikut:



**Gambar 4.6. Notifikasi ketika terdapat formulir yang tidak terisi**

Tidak lupa pula, pada tampilan ini disediakan tombol pada sisi kiri atas yang akan membawa kembali pengguna kepada tampilan menu utama.

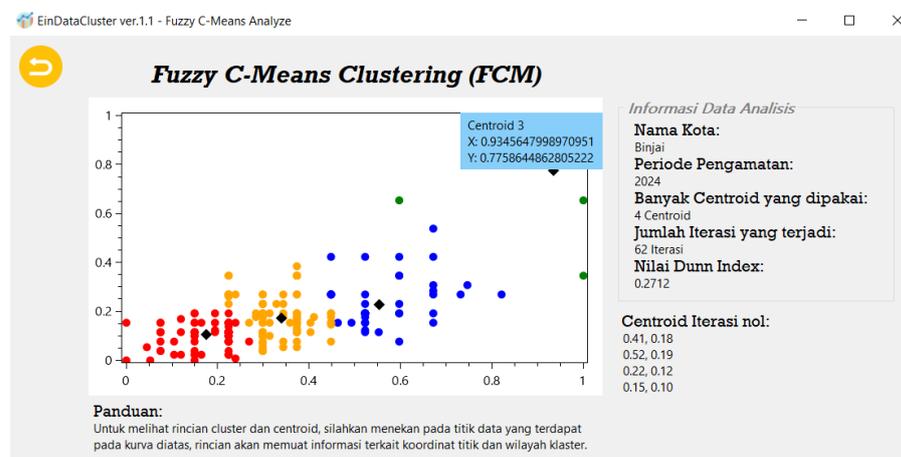
#### 4.2.5 Tampilan Hasil Pengolahan Algoritma K-Means



**Gambar 4.7. Tampilan Hasil Pengolahan Algoritma K-Means**

Pada tampilan ini, pengguna dapat melihat hasil analisis yang telah dilakukan oleh sistem. Tampilan ini meliputi kurva visualisasi *K-Means Clustering* yang dalam praktiknya menggunakan kontrol *PlotView* yang secara bawaan (*default*) tidak tersedia didalam VB.NET. Sebelum implementasi, peneliti terlebih dahulu melakukan instalasi *package* tambahan berupa *OxyPlot* pada VB.NET sehingga kontrol ini dapat diimplementasikan. Pada tampilan ini, selain kurva visualisasi *K-Means Clustering* yang memuat letak koordinat titik data beserta centroid (titik pusat kluster), sistem juga menampilkan informasi terkait nama kota yang sedang dianalisis, periode tahun yang diamati, banyak centroid yang terlibat / digunakan dalam analisis, jumlah iterasi yang terjadi hingga kondisi konvergen tercapai, serta nilai *Dunn Index* yang diperoleh atas analisis yang dilakukan. Kemudian pengguna juga dapat mengetahui letak koordinat centroid awal (iterasi ke-0) yang dipilih secara acak oleh sistem. Tidak lupa pula, pada tampilan ini disediakan tombol pada sisi kiri atas yang akan membawa kembali pengguna kepada tampilan menu utama.

#### 4.2.6 Tampilan Hasil Pengolahan Algoritma Fuzzy C-Means



**Gambar 4.8. Tampilan Hasil Pengolahan Algoritma Fuzzy C-Means**

Pada tampilan ini, pengguna dapat melihat hasil analisis yang telah dilakukan oleh sistem. Tampilan ini meliputi kurva visualisasi *Fuzzy C-Means Clustering* yang dalam praktiknya menggunakan kontrol *PlotView* yang secara bawaan (*default*) tidak tersedia didalam VB.NET. Tampilan yang disajikan kepada para pengguna hampir serupa dengan apa yang disajikan pada hasil pengolahan *K-Means Clustering*, namun yang membedakan ialah metode algoritma klaster yang digunakan. Pada tampilan ini, selain kurva visualisasi *Fuzzy C-Means Clustering* yang memuat letak koordinat titik data beserta centroid (titik pusat klaster), sistem juga menampilkan informasi-informasi lainnya yang serupa dengan apa yang disajikan pada hasil pengolahan *K-Means Clustering*. Penyajian ini meliputi informasi terkait nama kota yang sedang dianalisis, periode tahun yang diamati, banyak centroid yang terlibat / digunakan dalam analisis, jumlah iterasi yang terjadi hingga kondisi konvergen tercapai, serta nilai *Dunn Index* yang diperoleh atas analisis yang dilakukan. Kemudian, sama halnya dengan apa yang disajikan dalam hasil pengolahan *K-Means Clustering*, pengguna juga dapat mengetahui letak koordinat centroid awal (iterasi ke-0) yang dipilih secara acak oleh sistem. Tidak lupa pula, pada tampilan ini disediakan tombol pada sisi kiri atas yang akan membawa kembali pengguna kepada tampilan menu utama.

#### **4.4. Hasil Uji K-Means dan Fuzzy C-Means Clustering**

Pada penelitian ini, peneliti akan melakukan pengujian program sistem aplikasi dengan menggunakan data pengeluaran masyarakat (yang terkategori sebagai

pengeluaran makanan dan non-makanan). Adapun indikator penilaian yang akan diuji dari kedua algoritma yang terdapat dalam riset ini (K-Means dan Fuzzy C-Means) diantaranya ialah:

1. Efisiensi Waktu, indikator penilaian ini dapat dinilai dengan mengamati jumlah iterasi yang dihasilkan ketika kondisi konvergensi telah tercapai. Kondisi Konvergensi adalah kondisi dimana ketika iterasi yang dilakukan menghasilkan luaran yang tidak berbeda secara signifikan dengan luaran yang dihasilkan pada iterasi sebelumnya. Semakin sedikit jumlah iterasi yang dihasilkan, maka hal ini sama artinya dengan kondisi konvergensi semakin cepat diperoleh. Maka dari itu, apabila suatu algoritma memperoleh jumlah iterasi yang lebih kecil, maka hal ini sama artinya dengan semakin cepat pula program akan menyelesaikan tugasnya.
2. Keakuratan Data, indikator ini dapat dinilai dengan melakukan komparasi antara data kurva algoritma dengan nilai rasio GINI yang telah disediakan oleh Badan Pusat Statistik. Apabila titik data pada kurva algoritma (dalam program) menunjukkan data yang saling berdekatan, serta centroid yang tidak berjauhan maka hal ini mengindikasikan pemerataan pendapatan telah tercapai (atau setidaknya mendekati pemerataan sempurna). Hal ini akan dibandingkan dengan nilai rasio GINI yang dipublikasikan oleh BPS. Menurut BPS, rasio GINI akan semakin baik apabila nilainya semakin kecil, nilai yang semakin kecil (mendekati 0) sama artinya dengan semakin meratanya (setara-nya) pendapatan masyarakat di wilayah tertentu. Berbeda halnya apabila nilainya semakin besar (mendekati 1) maka hal ini mengindikasikan bahwa kesenjangan pendapatan semakin mudah terlihat di wilayah tersebut. Apabila kurva

algoritma (dalam program) dan nilai rasio GINI memiliki interpretasi yang serupa, maka hal ini dapat mengindikasikan bahwa algoritma data mining dapat digunakan sebagai alternatif dalam mengukur rasio GINI di suatu daerah. Namun apabila menunjukkan interpretasi yang berlawanan, maka algoritma data mining tidak dapat digunakan sebagai alternatif dalam mengukur rasio GINI di suatu daerah

3. Nilai *Dunn Index*, yakni metrik yang dapat digunakan untuk mengukur kualitas klastering dengan mengidentifikasi jarak minimum antar cluster (*cluster separation*) dan diameter maksimum dalam cluster (*cluster compactness*). Dalam mengukur metrik *Dunn Index*, kita dapat menggunakan rumus berikut:

$$Dunn Index = \frac{Dmin}{Dmax}$$

Keterangan:

$Dmin$  = Jarak terdekat antar klaster berbeda,

$Dmax$  = Jarak terjauh antar dua titik dalam satu klaster.

Sumber: (Septianingsih, 2022)

Semakin tinggi nilai *Dunn Index* maka hal ini dapat mengartikan kualitas klasterisasi semakin tinggi pula, dilihat melalui jarak antar cluster yang besar dan cluster yang lebih kompak. Hal ini mengartikan cluster-cluster tersebut terpisah dengan baik dan titik-titik dalam setiap cluster saling berdekatan. Sebaliknya apabila *Dunn Index* semakin rendah maka dapat mengartikan kualitas klasterisasinya semakin rendah pula.

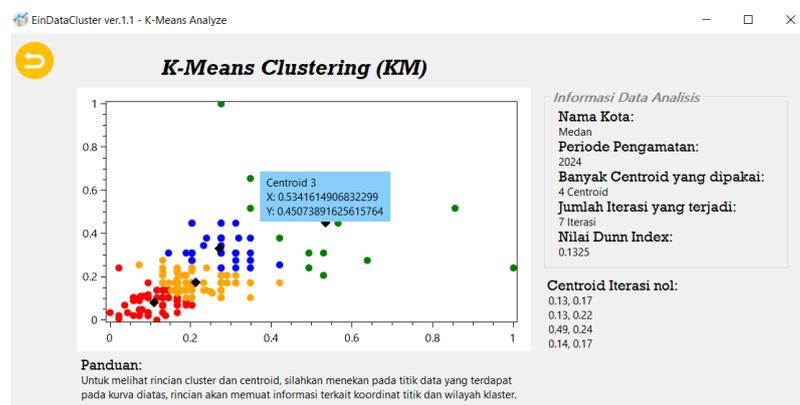
#### 4.3.1. Hasil Pengujian menggunakan K-Means Clustering

Berdasarkan pengujian yang telah peneliti lakukan, setidaknya peneliti memperoleh hasil atas pengolahan analisis pada data di 3 kabupaten/kota. Adapun hasil algoritma K-Means memperoleh jumlah iterasi sebagai berikut:

**Tabel 4.4. Tabulasi Jumlah Iterasi *K-Means Clustering***

Nama Kabupaten/Kota	Jumlah Iterasi
Kota Medan	7 Iterasi
Kota Binjai	15 Iterasi
Kabupaten Deli Serdang	8 Iterasi

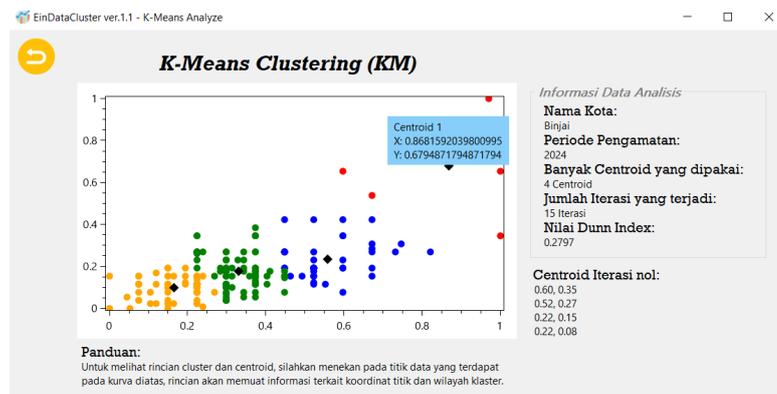
Pada tabel diatas, dapat kita perhatikan secara seksama bahwa iterasi yang dihasilkan pada analisis di tiga kabupaten/kota menunjukkan jumlah iterasi yang relatif rendah. Hal ini mengartikan bahwa algoritma K-Means mampu mencapai kondisi konvergen dengan waktu yang relatif cepat dan tidak terlalu membebani kerja komputer.



**Gambar 4.9. Tampilan Hasil Pengolahan K-Means Kota Medan**

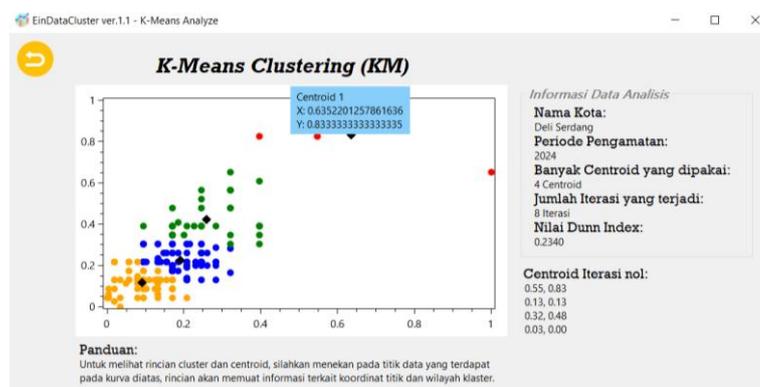
Pada kurva algoritma K-Means yang dihasilkan atas data pengeluaran kota Medan menunjukkan bahwa titik-titik data berkumpul pada sisi kiri bawah kurva, dengan beberapa titik data yang tersebar condong ke sisi kanan-atas

kurva. Hal ini menginterpretasikan bahwa pada wilayah kota Medan, pendapatan (yang diproksikan melalui pengeluaran) masyarakatnya telah merata namun tidak sempurna. Hal ini sejalan dengan apa yang dipublikasikan oleh BPS, dimana BPS mempublikasikan nilai rasio GINI kota Medan senilai 0,3730. Semakin rendah nilai rasio GINI (mendekati 0) maka dapat menginterpretasikan telah tercapainya pemerataan pendapatan (yang diproksikan melalui pengeluaran).



**Gambar 4.10. Tampilan Hasil Pengolahan K-Means Kota Binjai**

Pada kurva algoritma K-Means yang dihasilkan atas data pengeluaran kota Binjai menunjukkan bahwa titik-titik data berkumpul pada sisi kiri bawah kurva, dengan beberapa titik data yang berada condong ke sisi kanan atas kurva. Hal ini menginterpretasikan bahwa pada wilayah kota Binjai, pendapatan (yang diproksikan melalui pengeluaran) masyarakatnya telah merata namun tidak sempurna. Hal ini sejalan dengan apa yang dipublikasikan oleh BPS, dimana BPS mempublikasikan nilai rasio GINI kota Binjai senilai 0,3180.



**Gambar 4.11. Tampilan Hasil Pengolahan K-Means Kab. Deli Serdang**

Pada kurva algoritma K-Means yang dihasilkan atas data pengeluaran kabupaten Deli Serdang menunjukkan bahwa titik-titik data berkumpul pada sisi kiri bawah kurva, dengan beberapa titik data yang tersebar ke sisi kanan-atas kurva. Hal ini menginterpretasikan bahwa pada wilayah kabupaten Deli Serdang, pendapatan (yang diprosikan melalui pengeluaran) masyarakatnya telah merata namun tidak sempurna. Hal ini sejalan dengan apa yang dipublikasikan oleh BPS, dimana BPS mempublikasikan nilai rasio GINI kabupaten Deli Serdang senilai 0,2950. Temuan ini tentunya mengindikasikan pemerataan dapat lebih terlihat pada kabupaten Deli Serdang jika dibandingkan dengan dua kota lainnya yang juga turut dianalisis. Hal ini dapat dibuktikan dengan melihat titik data pada kabupaten Deli Serdang yang lebih rapat satu sama lain pada kurva K-Means.

**Tabel 4.5. Tabulasi Nilai *Dunn Index K-Means Clustering***

<b>Nama Kabupaten/Kota</b>	<b>Nilai <i>Dunn Index</i></b>
Kota Medan	0.1325
Kota Binjai	0.2797
Kabupaten Deli Serdang	0.2340

Pada tabel diatas, dapat kita perhatikan secara seksama bahwa nilai *Dunn Index* yang diperoleh pada analisis di tiga kabupaten/kota menunjukkan nilai *Dunn Index* yang relatif kecil. Namun hal ini bukan berarti analisa klasterisasi yang dilakukan buruk. Tidak ada rentang baku untuk mengukur baik/buruknya suatu klasterisasi melalui penilaian *Dunn Index*, namun semakin tinggi nilai *Dunn Index* maka hal ini mengartikan analisa klasterisasi tersebut semakin baik.

#### 4.3.2. Hasil Pengujian menggunakan Fuzzy C-Means Clustering

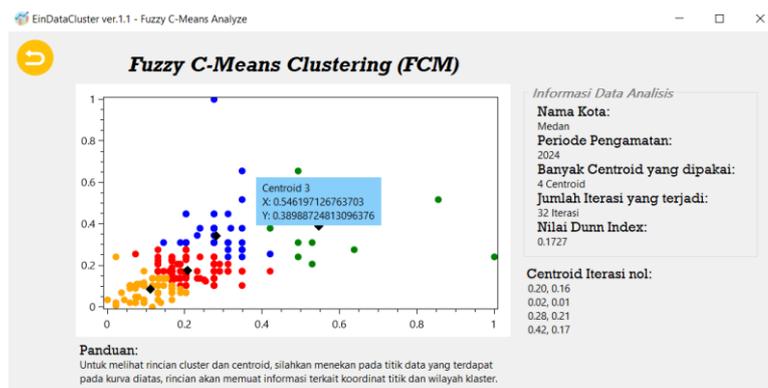
Berdasarkan pengujian yang telah peneliti lakukan, setidaknya peneliti memperoleh hasil atas pengolahan analisis pada data di 3 kabupaten/kota.

Adapun hasil algoritma Fuzzy C-Means memperoleh jumlah iterasi sebagai berikut:

**Tabel 4.6. Tabulasi Jumlah Iterasi *Fuzzy C-Means Clustering***

Nama Kabupaten/Kota	Jumlah Iterasi
Kota Medan	32 Iterasi
Kota Binjai	62 Iterasi
Kabupaten Deli Serdang	47 Iterasi

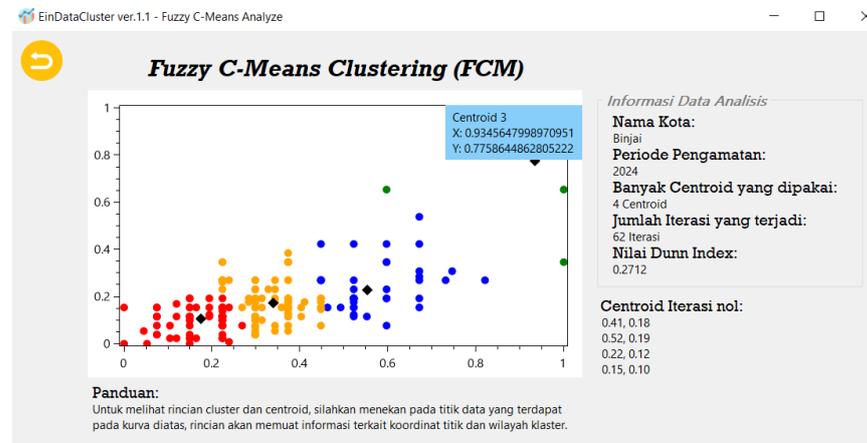
Pada tabel diatas, dapat kita perhatikan secara seksama bahwa iterasi yang dihasilkan pada analisis di tiga kabupaten/kota menunjukkan jumlah iterasi yang relatif tinggi. Hal ini mengartikan bahwa algoritma Fuzzy C-Means mampu mencapai kondisi konvergen dengan waktu yang relatif lebih lama dan pada dataset yang lebih besar tentunya dapat membebani kerja komputer.



**Gambar 4.12. Hasil Pengolahan Fuzzy C-Means Kota Medan**

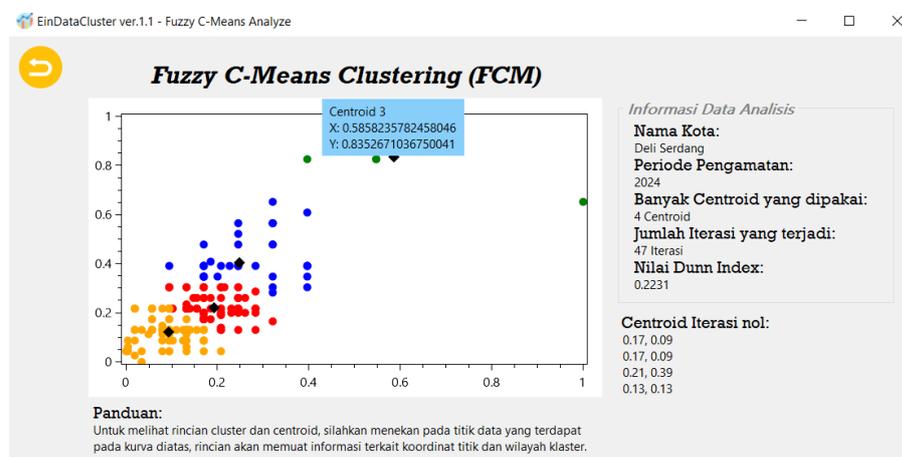
Pada kurva algoritma Fuzzy C-Means yang dihasilkan atas data pengeluaran kota Medan menunjukkan bahwa titik-titik data berkumpul pada sisi kiri bawah kurva, dengan beberapa titik data yang tersebar condong ke sisi kanan-atas kurva. Hal ini mengintepretasikan bahwa pada wilayah kota Medan, pendapatan (yang diproksikan melalui pengeluaran) masyarakatnya telah merata namun tidak sempurna. Hal ini sejalan dengan apa yang dipublikasikan oleh BPS, dimana BPS mempublikasikan nilai rasio GINI

kota Medan senilai 0,3730. Semakin rendah nilai rasio GINI (mendekati 0) maka dapat menginterpretasikan telah tercapainya pemerataan pendapatan (yang diproksikan melalui pengeluaran).



**Gambar 4.13. Hasil Pengolahan Fuzzy C-Means Kota Binjai**

Pada kurva algoritma K-Means yang dihasilkan atas data pengeluaran kota Binjai menunjukkan bahwa titik-titik data berkumpul pada sisi kiri bawah kurva, dengan beberapa titik data yang berada condong ke sisi kanan atas kurva. Hal ini menginterpretasikan bahwa pada wilayah kota Binjai, pendapatan (yang diproksikan melalui pengeluaran) masyarakatnya telah merata namun tidak sempurna. Hal ini sejalan dengan apa yang dipublikasikan oleh BPS, dimana BPS mempublikasikan nilai rasio GINI kota Binjai senilai 0,3180.



**Gambar 4.14. Hasil Pengolahan Fuzzy C-Means Kab. Deli Serdang**

Pada kurva algoritma K-Means yang dihasilkan atas data pengeluaran kabupaten Deli Serdang menunjukkan bahwa titik-titik data berkumpul pada sisi kiri bawah kurva, dengan beberapa titik data yang tersebar ke sisi kanan-atas kurva. Hal ini menginterpretasikan bahwa pada wilayah kabupaten Deli Serdang, pendapatan (yang diprosikan melalui pengeluaran) masyarakatnya telah merata namun tidak sempurna. Hal ini sejalan dengan apa yang dipublikasikan oleh BPS, dimana BPS mempublikasikan nilai rasio GINI kabupaten Deli Serdang senilai 0,2950. Temuan ini tentunya mengindikasikan pemerataan dapat lebih terlihat pada kabupaten Deli Serdang jika dibandingkan dengan dua kota lainnya yang juga turut dianalisis. Hal ini dapat dibuktikan dengan melihat titik data pada kabupaten Deli Serdang yang lebih rapat satu sama lain pada kurva K-Means.

**Tabel 4.7. Tabulasi Nilai *Dunn Index Fuzzy C-Means Clustering***

<b>Nama Kabupaten/Kota</b>	<b>Nilai Dunn Index</b>
Kota Medan	0.1727
Kota Binjai	0.2712
Kabupaten Deli Serdang	0.2231

Pada tabel diatas, dapat kita perhatikan secara seksama bahwa nilai *Dunn Index* yang diperoleh pada analisis di tiga kabupaten/kota menunjukkan nilai Dunn Index yang relatif kecil. Namun hal ini bukan berarti analisa klasterisasi yang dilakukan buruk. Tidak ada rentang baku untuk mengukur baik/buruknya suatu klasterisasi melalui penilaian Dunn Index, namun semakin tinggi nilai Dunn Index maka hal ini mengartikan analisa klasterisasi tersebut semakin baik.

#### **4.5. Komparasi Hasil K-Means dan Fuzzy C-Means Clustering**

Berdasarkan hasil yang telah peneliti peroleh, setidaknya dapat kita tarik kesimpulan bahwa dari segi penilaian jumlah iterasi, metode K-Means lebih unggul

daripada Fuzzy C-Means, hal ini dapat terlihat dari tabel tabulasi yang lebih sedikit jumlahnya pada hasil algoritma K-Means. Adapun komparasinya sebagai berikut:

**Tabel 4.8. Tabulasi Perbandingan Jumlah Iterasi KM & FCM**

Nama Kota	K-Means	Fuzzy C-Means
Kota Medan	7 Iterasi	32 Iterasi
Kota Binjai	15 Iterasi	62 Iterasi
Kabupaten Deli Serdang	8 Iterasi	47 Iterasi

Berdasarkan temuan ini, dapat ditarik kesimpulan bahwa dalam implementasi sebaiknya pengguna mempertimbangkan untuk lebih sering menggunakan opsi algoritma K-Means. Hal ini dikarenakan Algoritma ini mampu mencapai kondisi konvergen dengan sedikit perulangan / iterasi saja, dan ini tentunya tidak terlalu membebani kerja perangkat komputer. Berbeda dengan Fuzzy C-Means yang membutuhkan perulangan / iterasi hingga angka puluhan, hal ini tentunya dapat memperberat kerja komputer terlebih lagi apabila pengguna menggunakan DataSet dengan jumlah yang sangat banyak.

Peneliti menduga hal ini dapat terjadi sebab Fuzzy C-Means merupakan algoritma klasterisasi yang bersifat *soft-clustering*. Algoritma jenis ini mengharuskan setiap titik data diukur derajat keanggotaannya. Jika saja satu titik data mengalami perubahan derajat keanggotaan maka algoritma akan melakukan perulangan kembali hingga tidak ada perubahan derajat keanggotaan di setiap titik data. Hal ini tentunya berbeda dengan algoritma K-Means, yang merupakan algoritma dengan sifat *Hard-Clustering* yang dalam praktiknya hanya mengukur titik data menjadi anggota dari satu cluster, dan tidak perlu mengukur derajat keanggotaannya. Hal ini tentunya menyebabkan kondisi konvergen lebih mudah dan lebih cepat dicapai.

Dalam mengkomparasikan keakuratan data, baik K-Means maupun Fuzzy C-Means, keduanya mampu merepresentasikan hasil yang sejalan dengan apa yang dipublikasikan oleh BPS (rasio GINI). Tidak ada perbedaan antara luaran yang dihasilkan K-Means maupun Fuzzy C-Means. Hal ini dikarenakan DataSet yang digunakan adalah sama sehingga luaran klasterisasi yang dihasilkan pun akan serupa (dari segi letak data). Namun secara keseluruhan, luaran yang dihasilkan mampu merepresentasikan luaran yang sejalan dengan rasio GINI yang dipublikasikan oleh BPS. Selanjutnya untuk indikator penilaian Dunn Index, sekali lagi algoritma K-Means lebih unggul daripada algoritma Fuzzy C-Means. Temuan ini dapat dilihat pada tabel komprasi berikut:

**Tabel 4.9. Tabulasi Perbandingan Dunn Index KM & FCM**

Nama Kota	K-Means	Fuzzy C-Means
Kota Medan	0.1325	0.1727
Kota Binjai	0.2797	0.2712
Kabupaten Deli Serdang	0.2340	0.2231

Pada tabel diatas, dapat kita perhatikan bahwa secara keseluruhan, algortima K-Means memperoleh nilai *Dunn Index* yang lebih tinggi pada dua kabupaten/kota yakni kota Binjai (0.2797), dan kabupaten Deli Serdang (0.2340). Meskipun pada kota Medan memperoleh nilai *Dunn Index* yang lebih rendah (0.1325). perbandingannya ialah dengan nilai *Dunn Index* pada algoritma Fuzzy C-Means dimana DataSet kota Medan memperoleh nilai yang lebih besar (0.1727), namun pada dua kabupaten/kota lainnya justru algoritma Fuzzy C-Means tidaklah unggul dengan selisih yang tipis dengan algoritma K-Means (0.2712 pada DataSet kota Binjai, dan 0.2231 pada DataSet kabupaten Deli Serdang). Diperolehnya nilai Dunn

Index yang lebih rendah pada DataSet kota Medan dalam algoritma K-Means peneliti duga akibat adanya titik data pada klaster berwarna hijau yang dari segi tata letaknya lebih dekat kepada titik pusat klaster berwarna biru namun terklasterisasi sebagai anggota klaster berwarna hijau. Hal ini tentunya menyebabkan ketidakompakan klasterisasi dan memperbesar jarak terjauh antar dua titik dalam satu klaster ( $D_{max}$ ) yang merupakan pembagi dalam rumus Dunn Index.

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1. Kesimpulan

Berdasarkan analisa yang peneliti lakukan terhadap data penyusun rasio GINI (pengeluaran makanan dan non-makanan) di tiga kabupaten/kota yakni kota Medan, Binjai, dan kabupaten Deli Serdang dengan menggunakan algoritma K-Means dan Fuzzy C-Means, peneliti dapat menarik kesimpulan diantaranya:

1. Dalam efisiensi waktu, algoritma K-Means lebih unggul dari Fuzzy C-Means. Hal ini terlihat dari jumlah iterasi yang relatif lebih sedikit untuk mencapai kondisi Konvergensi pada algoritma K-Means,
2. Dalam keakuratan data, baik K-means maupun Fuzzy C-Means, keduanya mampu menampilkan kurva yang sejalan dan mampu merepresentasikan rasio GINI yang dipublikasikan oleh BPS (Badan Pusat Statistik),
3. Dalam Indeks Dunn, K-Means lebih unggul daripada Fuzzy C-Means, hal ini dapat dilihat dari nilai Dunn Index yang relatif lebih tinggi daripada nilai Dunn Index yang diperoleh Fuzzy C-Means. Meskipun pada data kota Medan K-Means sempat memperoleh nilai Dunn Index yang lebih rendah, tetapi pada dua kabupaten/kota lainnya K-Means tetap lebih unggul. Teruntuk kota Medan, metode FCM unggul 30.3% dari K-Means, sedangkan pada kota Binjai K-Means unggul 3.13%, dan pada kabupaten Deli Serdang K-Means unggul 4.88% dari FCM.
4. Secara keseluruhan dapat ditarik kesimpulan bahwa K-Means lebih unggul dari Fuzzy C-Means dalam melakukan analisa klasterisasi data penyusun rasio

GINI (pengeluaran makanan dan non-makanan) di tiga kabupaten/kota yakni kota Medan, Binjai, dan kabupaten Deli Serdang.

## 5.2. Saran

Adapun saran yang dapat peneliti ajukan berdasarkan penelitian yang telah peneliti laksanakan diantaranya ialah:

1. Peneliti selanjutnya sebaiknya dapat mempertimbangkan untuk menggunakan metode *distance* yang berbeda (selain *Euclidean Distance*). Melalui perbedaan implementasi metode *distance* didalam algoritma, harapannya dapat memberikan hasil penelitian yang lebih variatif,
2. Peneliti selanjutnya sebaiknya mempertimbangkan penggunaan algoritma kluster yang berbeda seperti Hierarchical Clustering, DBSCAN, Mean Shift Clustering, Gaussian Mixture Models (GMM), maupun algoritma klustering lainnya untuk memperluas wawasan di bidang Data Mining,
3. Peneliti selanjutnya sebaiknya mempertimbangkan untuk melibatkan responden dari lebih banyak kabupaten/kota. Melalui keterlibatan lebih banyak responden dari beragam kabupaten/kota yang berbeda dan kondisi perekonomian yang berbeda pula, harapannya hasil penelitian yang diperoleh dapat lebih bervariasi dan lebih representatif, dan
4. Peneliti selanjutnya sebaiknya mempertimbangkan untuk menggunakan bahasa pemrograman lainnya terutama bahasa pemrograman yang dirancang sejak awal untuk melakukan analisa statistik. Hal ini perlu dilakukan agar peneliti selanjutnya tidak kesulitan dalam merancang dan juga membangun sistem pengolahan data yang dibutuhkan didalam riset.

## DAFTAR PUSTAKA

- Akpojotor, L. O., & Urhefe-Okotie, E. A. (2023). The use of data mining and data protection technology in the 21st Century library. *Journal of ICT Development, Applications and Research*, 5(1), 93–108. <https://doi.org/10.47524/jictdar.v5i1.94>
- Anikah, I., Surip, A., Rahayu, N. P., Al-Musa, M. H., & Tohidi, E. (2020). Pengelompokan Data Barang Dengan Menggunakan Metode K-Means Untuk Menentukan Stok Persediaan Barang. *KOPERTIP : Jurnal Ilmiah Manajemen Informatika Dan Komputer*, 4(2), 58–64. <https://doi.org/10.32485/kopertip.v4i2.120>
- Arianti, T., Fa'izi, A., Adam, S., & Wulandari, M. (2022). Perancangan Sistem Informasi Perpustakaan Menggunakan Diagram Uml (Unified Modelling Language). *Jurnal Ilmiah Komputer Tera[an Dan Informasi*, 1(1), 19–25. Retrieved from <https://journal.polita.ac.id/index.php/politati/article/view/110/88>
- Audrey, J., Fadlil, A., & Sunardi, S. (2022). Prediksi Jumlah Mahasiswa Baru Menggunakan Logika Fuzzy Metode Sugeno. *Informatika*, 14(1), 66. <https://doi.org/10.36723/juri.v14i1.263>
- Auliana, S., & Mansyuri, U. (2022). Penggunaan Metoda Fuzzy Tsukamoto Untuk Menentukan Produksi Barang Elektronik. *Jurnal Simasi : Jurnal Ilmiah Sistem Informasi*, 2(2), 123–129. Retrieved from <http://simasi.lppmbinabangsa.id/index.php/home>

- Aziz, A., Siregar, A. M., & Zonyfar, C. (2022). Penerapan Algoritma K-Means dan Fuzzy C-Means untuk Pengelompokan Kabupaten Kota Berdasarkan Produksi Padi di Provinsi Jawa Barat. *Scientific Student Journal for Information, Technology and Science*, *III*(1), 1–8. Retrieved from <https://journal.ubpkarawang.ac.id/mahasiswa/index.php/ssj/article/view/411>
- Bai, L., Qi, M., & Liang, J. (2023). Spectral clustering with robust self-learning constraints. *Artificial Intelligence*, *320*, 1–24. <https://doi.org/10.1016/j.artint.2023.103924>
- Bayu Lokananta, R., Yuana, H., & Puspitasari, W. D. (2024). Implementasi Algoritma K-Means Terhadap Pengelompokan Status Gizi Balita (Studi Kasus : Posyandu Melati Vii). *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(5), 3585–3592. <https://doi.org/10.36040/jati.v7i5.7377>
- Bertua, O., Panggabean, E., Yolanda, A., Naftalia, J., & Sitompul, S. (2024). PENINGKATAN BERBICARA SISWA SMP FREE METHODIST 1 MEDAN DENGAN MEMANFAATKAN MEDIA SOSIAL. *Jurnal Review Pendidikan Dan Pengajaran*, *7*(1), 2978–2983.
- Burhanuddin, A. (2023). Analisis Komparatif Inferensi Fuzzy Tsukamoto, Mamdani dan Sugeno Terhadap Produktivitas Padi di Indonesia. *LEDGER: Journal Informatic and Information Technology*, *2*(1), 48–57.
- Chyan, P. (2022). Segmentasi Kulit Manusia Dengan Ekstraksi Fitur Warna Dan Algoritma GMM-EM. *Jurnal Pendidikan Teknologi Informasi (JUKANTI)*, *5*(1), 151–156.
- Damayanti, D. K. D., & Jakfar, M. (2023). Klasifikasi Status Stunting Balita Menggunakan Algoritma Fuzzy C-Means (Studi Kasus Posyandu RW 01

- Kelurahan Jepara Surabaya). *MATHunesa: Jurnal Ilmiah Matematika*, 11(3), 524–533. <https://doi.org/10.26740/mathunesa.v11n03.p524-533>
- Duan, Y., Liu, C., Li, S., Guo, X., & Yang, C. (2023). An automatic affinity propagation clustering based on improved equilibrium optimizer and t-SNE for high-dimensional data. *Information Sciences*, 623(January), 434–454. <https://doi.org/10.1016/j.ins.2022.12.057>
- Ekasetya, V. A., & Jananto, A. (2020). Klusterisasi Optimal Dengan Elbow Method Untuk Pengelompokan Data Kecelakaan Lalu Lintas Di Kota Semarang. *Jurnal Dinamika Informatika*, 12(1), 20–28. <https://doi.org/10.35315/informatika.v12i1.8159>
- Fatonah, N. S., & Pancarani, T. K. (2022). Analisa Perbandingan Algoritma Clustering Untuk Pemetaan Status Gizi Balita Di Puskesmas Pasir Jaya. *Konvergensi*, 18(1), 1–9. <https://doi.org/10.30996/konv.v18i1.5497>
- Firdaus, H., & Sofro, A. (2022). Analisa Cluster Menggunakan K-Means Dan Fuzzy C-Means Dalam Pengelompokan Provinsi Menurut Data Intesitas Bencana Alam Di Indonesia Tahun 2017-2021. *MATHunesa: Jurnal Ilmiah Matematika*, 10(1), 50–60. <https://doi.org/10.26740/mathunesa.v10n1.p50-60>
- Ghosh, I., Balakrishnan, N., & Ng, H. K. T. (2021). *Advances in Statistics -Theory and Applications* (D.-G. Chen, ed.). Cham: Springer.
- Harahap, S. R. N., Halim, N., Nst, M. R., & Ginting, E. (2023). Penerapan Metode Apriori Pada Toko Pakaian Matahari. *UNES Journal of Scientech Research*, 8(1), 76–84.
- Izhari, F. (2020). Analisis Algoritma DbSCAN Dalam Menentukan Parameter Epsilon Pada Clustering Data Numerik. *Seminar Nasional Teknologi*

*Komputer & Sains (SAINTEKS)*, Februari, 156–158.

Kang, J., He, J., MacIejewski, R., & Tong, H. (2020). InFoRM: Individual Fairness on Graph Mining. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August*, 379–389. <https://doi.org/10.1145/3394486.3403080>

Khan, M., Khan, S., & Alharbi, Y. (2020). Text Mining Challenges and Applications, A Comprehensive Review. *International Journal of Computer Science and Network Security*, 20(12), 138–148. <https://doi.org/10.22937/IJCSNS.2020.20.12.15>

Kinanti, L. (2022). Pembangunan Aplikasi Penjualan Tiket Penerbangan dan Penerimaan Kas. *Jurnal Ilmu Data*, 2(2), 1–10. Retrieved from <http://ilmudata.org/index.php/ilmudata/article/view/74>

Latifah, U. W., Sugiyarto, S., & Suparman, S. (2022). K-means and fuzzy c-means algorithm comparison on regency/city grouping in Central Java Province. *Desimal: Jurnal Matematika*, 5(2), 155–168. <https://doi.org/10.24042/djm.v5i2.12204>

Mailien, B., Salma, A., Syafriandi, S., & Fitria, D. (2023). Comparison K-Means and Fuzzy C-Means Methods to Grouping Human Development Index Indicators in Indonesia. *UNP Journal of Statistics and Data Science*, 1(1), 23–30. <https://doi.org/10.24036/ujsds/vol1-iss1/4>

Marlia, L., Ngatmini, P., & Tobing, P. E. (2023). Analisis Segmentasi Pengunjung Mall Menggunakan Algoritma K-means dan Fuzzy C-means. *SENTIMAS: Seminar Nasional Penelitian Dan Pengabdian Masyarakat*, 240–247. Retrieved from <https://journal.irpi.or.id/index.php/sentimas%0D>

- Matdoan, M. Y., Matdoan, U. A., & Far-Far, M. S. (2022). Algoritma K-Means Untuk Klasifikasi Provinsi di Indonesia Berdasarkan Paket Pelayanan Stunting. *PANRITA Journal of Science, Technology, and Arts*, 1(2), 41–46. Retrieved from <https://journal.dedikasi.org/pjsta>
- Microsoft. (2024). Visual Basic. Retrieved March 22, 2024, from learn.microsoft website: <https://learn.microsoft.com/en-us/dotnet/visual-basic/programming-guide/?redirectedfrom=MSDN>
- Mughnyanti, M., & Ginting, S. H. N. (2023). Data Mining Manhattan Distance dan Euclidean Distance Pada Algoritma X-Means Dalam Klasifikasi Minat dan Bakat Siswa. *Remik: Riset Dan E-Jurnal Manajemen Informatika Komputer*, 7(1), 835–842. <https://doi.org/10.33395/remik.v7i1.12162>
- Nurdianah, S., Hariyanto, B., Haeri, B. R., & Khotimah, S. (2021). Analisis dan Perancangan Toko Online Berbasis Web Pada Tanaman Mini Yogyakarta. *TIN: Terapan Informatika Nusantara*, 2(5), 302–308.
- Nurdiyanti, S., & Yuniarti, D. N. (2023). Penentuan Banyak Cluster Optimal Hasil Fuzzy C-Means dengan Metode Elbow pada Klasifikasi Kecemasan Belajar Siswa (Studi Kasus: Siswa SMP Negeri 1 Tuban dan SMA Negeri 1 Tuban Jawa Timur). *MATHunesa: Jurnal Ilmiah Matematika*, 11(3), 401–413. <https://doi.org/10.26740/mathunesa.v11n3.p401-413>
- Nursalam, N., & Djaha, A. S. A. (2023). Pelatihan Pembuatan Kuesioner Penelitian Bagi Mahasiswa Prodi Administrasi Negara Fisip Universitas Nusa Cendana. *Jdistira*, 3(1), 25–31. <https://doi.org/10.58794/jdt.v3i1.433>
- Olivia, L. F., Juliantho, D. A., & Hendrik, B. (2023). Komparasi Perbandingan Algoritma K-Means dan K-Medoids dalam Clustering Penyebaran Kasus

- Covid 19. *Jised: Journal of Information System and Education Development*, 1(2), 30–32.
- Permatasari, N. A., Chrisnanto, Y. H., & Ningsih, A. K. (2023). Segmentasi Kasus Data Kematian Covid 19 Di Jawa Barat Menggunakan Algoritma DBSCAN. *IJESPG (International Journal of Engineering, Economic, Social Politic and Government)*, 1(4), 119–128.
- Pratama, R. D. (2022). Pengaruh Pertumbuhan Ekonomi, Inflasi, Pendidikan dan Rasio Gini Terhadap Pengangguran Terbuka Di Kota Sibolga Periode 2002 - 2021. *E-QIEN : Jurnal Ekonomi Dan Bisnis*, 11(2), 193 – 201.
- Rinjani, I., Anwar, S., & Herdiana, R. (2023). Pengelompokan Daerah bencana Alam Menggunakan Algoritma K-Means Clustering. *Jurnal Ilmiah Sistem Informasi Dan Ilmu Komputer*, 3(1), 31–51.
- Riza, F. (2022). Analisis dan Prediksi Data Penjualan Menggunakan Machine Learning dengan Pendekatan Ilmu Data. *Data Science Indonesia (DSI)*, 1(2), 62–68. <https://doi.org/10.47709/dsi.v1i2.1308>
- Rizuan, R., Haerani, E., Jasril, J., & Oktavia, L. (2023). Penerapan Algoritma Mean-Shift Pada Clustering Penerimaan Bantuan Pangan Non Tunai. *Journal of Computer System and Informatics (JoSYC)*, 4(4), 1019–1027. <https://doi.org/10.47065/josyc.v4i4.3876>
- Romansyah, A., & Zulfikar, A. F. (2021). Penerapan Fuzzy Inference System Metode Mamdani Dan Sugeno Untuk Menentukan Status Gizi Balita Berdasarkan Indeks Antropometri Dengan Pengujian Matlab. *Jurnal Esit (E-Bisnis, Sistem Informasi, Teknologi Informasi)*, 16(2), 62–67. Retrieved from <http://openjournal.unpam.ac.id/index.php/ESIT/article/view/13744>

- Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(3), 1–21. <https://doi.org/10.1002/widm.1355>
- Şahin, M., & Yurdugul, H. (2020). Educational Data Mining and Learning Analytics: Past, Present and Future. *Bartın University Journal of Faculty of Education*, *9*(1), 121–131. <https://doi.org/10.14686/buefad.606077>
- Salem, I. E., Mijwil, M. M., Abdulqader, A. W., Ismaeel, M. M., Alkhazraji, A., & Alaabdin, A. M. Z. (2022). Introduction to The Data Mining Techniques in Cybersecurity. *Mesopotamian Journal of Cybersecurity*, *2022*, 28–37. <https://doi.org/10.58496/mjcs/2022/004>
- Saputra, P. R. N., & Chusyairi, A. (2020). Perbandingan Metode Clustering dalam Pengelompokan Data Puskesmas pada Cakupan Imunisasi Dasar Lengkap. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *4*(6), 1077–1084. <https://doi.org/10.29207/resti.v4i6.2556>
- Saputro, R. D., Retnoningsih, D., & Khusnuliawati, H. (2024). Penggunaan Metode Rank Order Centroid dalam Penentuan Nilai Centroid ( Studi Kasus : Dataset Biji Gandum ). *JIMSTEK*, *06*(01), 18–23.
- Septianingsih, A. (2022). Pemetaan Kabupaten Kota Di Provinsi Jawa Timur Berdasarkan Tingkat Kasus Penyakit Menggunakan Pendekatan Agglomeratif Hierarchical Clustering. *Jurnal Lebesgue : Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika*, *3*(2), 367–386. <https://doi.org/10.46306/lb.v3i2.139>
- Setyawan, D. A., & Fatichah, C. (2020). PENGEMBANGAN METODE DECISION TREE DENGAN DISKRITISASI DATA DAN SPLITTING

ATRIBUT MENGGUNAKAN HIERARCHICAL CLUSTERING DAN DISPERSION RATIO. *JUTI: Jurnal Ilmiah Teknologi Informasi*, 18(2), 179–187. <https://doi.org/10.12962/j24068535.v18i2.a1005>

Sihananto, A. N., Puspita Sari, A., Khariono, H., Akhmad Fernanda, R., & Wijaya, D. C. M. (2022). Implementasi Metode K-Means Untuk Pengelompokan Kasus Covid-19 Tingkat Provinsi Di Indonesia. *Jurnal Informatika Dan Sistem Informasi*, 3(1), 76–85. <https://doi.org/10.33005/jifosi.v3i1.472>

Suhailah, E., & Hartatik, H. (2023). Pembuatan Sistem Rekomendasi Pariwisata Yogyakarta Menggunakan Triangle Multiplaying Jaccard Creating a Yogyakarta Tourism Recommendation System Using Triangle Multiplaying Jaccard. *JACIS: Journal Automation Computer Information System*, 3(2), 115–126.

Sulistiyawati, A. A. D., & Sadikin, M. (2021). SISTEMASI: Jurnal Sistem Informasi Penerapan Algoritma K-Medoids untuk Menentukan Segmentasi Pelanggan. *SISTEMASI: Jurnal Sistem Informasi*, 10(3), 516–526. Retrieved from <http://sistemasi.ftik.unisi.ac.id>

Sy, Y. S. (2023). Klasterisasi Pasien Rawat Inap Peserta BPJS Berdasarkan Jenis Penyakit Menggunakan Algoritma K-Means. *Jurnal Sistim Informasi Dan Teknologi*, 5(2), 33–37. <https://doi.org/10.37034/jsisfotek.v5i2.162>

Syoer, R. R., & Wahyudin, Y. (2021). Analisis Kelompok Dengan Algoritma Fuzzy Clustering (Studi Kasus Pengelompokan Desa Di Provinsi Kalimantan Timur). *BESTARI: Buletin Statistika Dan Aplikasi Terkini*, 1(1), 1–11. Retrieved from <https://bestari.bpskaltim.com/index.php/bestari-bpskaltim/article/view/1/2>

- Syukron, H., Fayyad, M. F., Fauzan, F. J., Ikhsani, Y., & Gurning, U. R. (2022). Perbandingan K-Means K-Medoids dan Fuzzy C-Means untuk Pengelompokan Data Pelanggan dengan Model LRFM. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 2(2), 76–83. <https://doi.org/10.57152/malcom.v2i2.442>
- Utara, B. P. S. (2023). *Statistik: Pengeluaran untuk Konsumsi Penduduk Provinsi Sumatera Utara* (M. P. Samosir & C. P. Aridesyadi, Eds.). Medan: BPS Provinsi Sumatera Utara.
- Veronica, N., Yuliana, Y., & Weny, W. (2023). Analisis Pengaruh E- Wom Terhadap Keputusan Pembelian Mahasiswa / i dari Aplikasi Tiktok Shop. *Journal of Trends Economics and Accounting Research*, 4(2), 463–468. <https://doi.org/10.47065/jtear.v4i2.1026>
- Virantika, E., Kusnawi, K., & Ipmawati, J. (2022). Evaluasi Hasil Pengujian Tingkat Clusterisasi Penerapan Metode K-Means Dalam Menentukan Tingkat Penyebaran Covid-19 di Indonesia. *Jurnal Media Informatika Budidarma*, 6(3), 1657–1666. <https://doi.org/10.30865/mib.v6i3.4325>
- Wahyu, A., & Rushendra, R. (2022). Klasterisasi Dampak Bencana Gempa Bumi Menggunakan Algoritma K-Means di Pulau Jawa. *JEPIN (Jurnal Edukasi Dan Penelitian Informatika)*, 8(1), 175–179. <https://doi.org/10.26418/jp.v8i1>
- Wang, E., Lee, H., Do, K., Lee, M., & Chung, S. (2023). Recommendation of Music Based on DASS-21 (Depression, Anxiety, Stress Scales) Using Fuzzy Clustering. *Electronics (Switzerland)*, 12(1), 1–17. <https://doi.org/10.3390/electronics12010168>
- Warasto, D., & Jalil, A. (2023). Pembuatan Program Aplikasi Inventory Berbasis

- Desktop Menggunakan Visual Basic.Net. *Jurnal Sistem Informasi Kaputama (JSIK)*, 7(1), 20–28. <https://doi.org/10.31862/9785426311961>
- Yudhistira, A., & Andika, R. (2023). Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering. *Journal of Artificial Intelligence and Technology Information (JAITI)*, 1(1), 20–28. <https://doi.org/10.58602/jaiti.v1i1.22>
- Yulianto, T., Rahmah, A. F., Faisol, F., & Amalia, R. (2023). Clustering Daerah Bencana Alam Di Indonesia Menggunakan Metode Fuzzy C-Means. *Unisda Journal of Mathematics and Computer Science*, 9(2), 29–39.
- Zai, C. (2022). Implementasi Data Mining Sebagai Pengolahan Data. *Portal Data*, 2(3), 1–12. Retrieved from <http://portaldata.org/index.php/portaldata/article/view/107>
- Zumaeroh, Z., Fatmawati, A., & Purnomo, S. D. (2023). Determinan Pembangunan Manusia di Provinsi Papua. *Majalah Ilmiah Manajemen & Bisnis (MIMB)*, 20(1), 67–75. <https://doi.org/10.30998/sosioekons.v15i1.16441>