

**KOMPARASI VALIDASI *K-MEANS* DAN *FUZZY C-MEANS*  
DALAM MENENTUKAN *CLUSTER* TERBAIK  
MENGUNAKAN *SILHOUETTE INDEX* DAN *DAVIES*  
*BOULDIN INDEX***

***COMPARISON OF MEASUREMENT OF K-MEANS AND  
FUZZY C-MEANS TO DECIDE THE BEST CLUSTER USING  
SILHOUETTE INDEX AND DAVIES BOULDIN INDEX***



**TAUFIK HIDAYAT  
15102075**

**PROGRAM STUDI S1 INFORMATIKA  
FAKULTAS TEKNOLOGI INDUSTRI DAN INFORMATIKA  
INSTITUT TEKNOLOGI TELKOM PURWOKERTO  
2019**

**TUGAS AKHIR**  
**KOMPARASI VALIDASI *K-MEANS* DAN *FUZZY C-MEANS***  
**DALAM MENENTUKAN *CLUSTER* TERBAIK**  
**MENGGUNAKAN *SILHOUETTE INDEX* DAN *DAVIES***  
***BOULDIN INDEX***

***COMPARISON OF MEASUREMENT OF K-MEANS AND***  
***FUZZY C-MEANS TO DECIDE THE BEST CLUSTER USING***  
***SILHOUETTE INDEX AND DAVIES BOULDIN INDEX***

Disusun Sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana Komputer



**TAUFIK HIDAYAT**  
**15102075**

**PROGRAM STUDI S1 TEKNIK INFORMATIKA**  
**FAKULTAS TEKNOLOGI INDUSTRI DAN INFORMATIKA**  
**INSTITUT TEKNOLOGI TELKOM PURWOKERTO**  
**2019**

Lembar Pengesahan Pembimbing

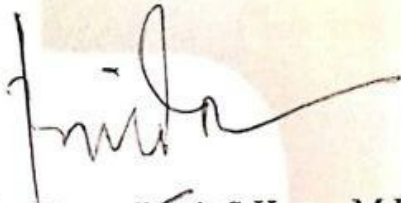
**KOMPARASI VALIDASI *K-MEANS* DAN *FUZZY C-MEANS*  
DALAM MENENTUKAN *CLUSTER* TERBAIK  
MENGUNAKAN *SILHOUETTE INDEX* DAN *DAVIES*  
*BOULDIN INDEX***

***COMPARISON OF MEASUREMENT OF K-MEANS AND  
FUZZY C-MEANS TO DECIDE THE BEST CLUSTER USING  
SILHOUETTE INDEX AND DAVIES BOULDIN INDEX***

Dipersiapkan dan Disusun Oleh  
**TAUFIK HIDAYAT**  
**15102075**

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tugas Akhir  
pada hari Selasa, 26 Maret 2019

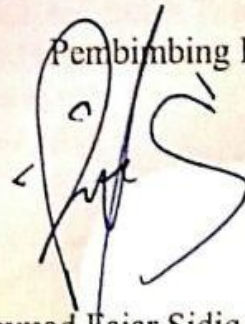
Pembimbing I



Rima Dias Ramadhani, S.Kom., M.Kom.

NIDN. 0602039301

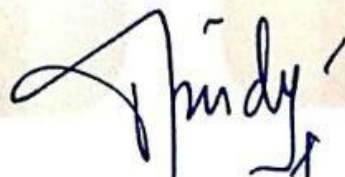
Pembimbing II



Muhammad Iqbal Sidiq, S.T., M.T.

NIDN. 0619029102

Tugas Akhir ini diterima sebagai salah satu persyaratan untuk memperoleh  
gelar Sarjana Komputer  
Tanggal 4 April 2019  
Dekan



Didi Supriyadi, S.T., M. Kom.

NIDN. 0618038404

Lembar Penetapan Penguji

**KOMPARASI VALIDASI K-MEANS DAN FUZZY C-MEANS  
DALAM MENENTUKAN CLUSTER TERBAIK  
MENGUNAKAN SILHOUETTE INDEX DAN DAVIES  
BOULDIN INDEX**

**COMPARISON OF MEASUREMENT OF K-MEANS AND  
FUZZY C-MEANS TO DECIDE THE BEST CLUSTER USING  
SILHOUETTE INDEX AND DAVIES BOULDIN INDEX**

Dipersiapkan dan Disusun Oleh


**TAUFIK HIDAYAT  
15102075**

Tugas Akhir Telah Diuji dan Dinilai Panitia Penguji Program Studi Teknik Informatika  
Fakultas Teknologi Industri dan Informatika  
Institut Teknologi Telkom Purwokerto  
Pada Tanggal: Selasa, 26 Maret 2019

Ketua  
Penguji

  
(Tri Ginanjar Laksana, M. Kom., M. Cs.)  
NIDN. 0407088502

Anggota  
Penguji I

  
(Auliya Burhanuddin, S. Si., M. Kom)  
NIDN. 0630058202

Anggota  
Penguji II

  
(Rifki Adhitama, S. Kom., M. Kom)  
NIDN. 0627089101



## HALAMAN PERNYATAAN ANTI PLAGIAT

Yang bertanda tangan dibawah ini,

Nama Mahasiswa : Taufik Hidayat  
NIM : 15102075  
Program Studi : SI Teknik Informatika

Menyatakan bahwa Tugas Akhir dengan judul berikut:

***KOMPARASI VALIDASI K-MEANS DAN FUZZY C-MEANS DALAM MENENTUKAN CLUSTER TERBAIK MENGGUNAKAN SILHOUETTE INDEX DAN DAVIES BOULDIN INDEX***

Dosen Pembimbing Utama : Rima Dias Ramadhani, S.Kom., M.Kom.  
Dosen Pembimbing Pendamping : Muhammad Fajar Sidiq, S.T., M.T.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Institut Teknologi Telkom Purwokerto maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan, dan penelitian saya sendiri, tanpa bantuan pihak lain kecuali dari Tim Dosen Pembimbing.
3. Karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab saya, bukan tanggung jawab Institut Teknologi Telkom Purwokerto.
5. Pernyataan ini saya buat dengan sesungguhnya, apabila dikemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima Sanksi Akademik dengan pencabutan gelar yang sudah diperoleh serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Purwokerto, 2 April 2019  
Yang Menyatakan,



(TAUFIK HIDAYAT)

## KATA PENGANTAR

Alhamdulillahirabbil'aalamin, segala puji bagi Allah Subhana wa Ta'ala rabb pencipta seluruh alam semesta yang telah memberikan rahmat dan hidayah-Nya, sehingga penulis mampu menyelesaikan Tugas Akhir dengan judul, "***KOMPARASI VALIDASI K-MEANS DAN FUZZY C-MEANS DALAM MENENTUKAN CLUSTER TERBAIK MENGGUNAKAN SILHOUETTE INDEX DAN DAVIES BOULDIN INDEX***" sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer (S. Kom) pada program Strata-I di Jurusan S1 Teknik Informatika, Fakultas Teknologi Industri dan Informatika, Institut Teknologi Telkom Purwokerto.

Penulis menyadari dalam penyusunan Tugas Akhir ini tidak akan selesai tanpa bantuan dari berbagai pihak. Karena itu pada kesempatan ini penulis ingin mengucapkan terima kasih kepada:

- a. Allah Subhana wa Ta'ala karena karunia dan izin-Nya Tugas Akhir ini dapat diselesaikan dengan baik;
- b. Kedua orang tua tercinta, bapak dan mama yang telah memberikan do'a, semangat, dan dukungan yang tak pernah terputus untuk penulis serta motivasi untuk menyelesaikan Tugas Akhir ini;
- c. Keluarga terdekat saya, kakak dan adik atas do'a, semangat, dan dukungan yang diberikan;
- d. Bapak Dr. Ali Rokhman, M.Si selaku Rektor IT Telkom Purwokerto;
- e. Bapak Didi Supriyadi, S.T., M.Kom selaku Dekan Fakultas Teknologi Industri dan Informatika (FTII) IT Telkom Purwokerto;
- f. Bapak M. Zidny Naf'an, Lc., M.Kom selaku Kaprodi S1 Teknik Informatika IT Telkom Purwokerto;
- g. Ibu Rima Dias Ramadhani, S.Kom., M.Kom selaku Dosen Pembimbing I atas bimbingan, saran, dan motivasi yang diberikan;
- h. Bapak Muhammad Fajar Sidiq, S.T., M.T. selaku Dosen Pembimbing II atas bimbingan, saran, dan motivasi yang diberikan;
- i. Teman-teman seperjuangan yang sedang berproses dalam mengerjakan Tugas Akhir atas doa, bimbingan, serta kasih sayang yang selalu tercurah selama ini;

- j. Teman-teman dan para warga Belajar LibreOffice Indonesia yang telah memberikan dukungan dan semangat untuk penulis;
- k. Segenap dosen Jurusan S1 Teknik Informatika FTII - IT Telkom Purwokerto yang telah memberikan ilmunya kepada penulis;
- l. Keluarga Besar Institut Teknologi Telkom Purwokerto, khususnya teman-teman seperjuangan kami di angkatan 2015 kelas B Jurusan S1 Teknik Informatika FTII – IT Telkom Purwokerto atas semua dukungan, semangat, serta kerjasamanya.

Penulis menyadari laporan Tugas Akhir ini tidak luput dari berbagai kekurangan. Penulis mengharapkan saran dan kritik demi kesempurnaan dan perbaikannya sehingga akhirnya laporan Tugas Akhir ini dapat memberikan manfaat bagi bidang pendidikan dan penerapan di lapangan serta dapat dikembangkan lagi untuk penelitian lebih lanjut.

Purwokerto, 18 Maret 2019  
Penulis

**TAUFIK HIDAYAT**  
**NIM. 15102075**

## DAFTAR ISI

Lembar Pengesahan Pembimbing.....	iii
Lembar Penetapan Penguji.....	iv
HALAMAN PERNYATAAN ANTI PLAGIAT.....	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	x
DAFTAR TABEL.....	xii
DAFTAR RUMUS.....	xv
DAFTAR LAMPIRAN.....	xvi
ABSTRAK.....	xvii
ABSTRACT.....	xviii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	5
1.3 Tujuan Penelitian.....	5
1.4 Manfaat Penelitian.....	5
1.5 Batasan Masalah.....	6
BAB II TINJAUAN PUSTAKA.....	7
2.1 Tinjauan Pustaka.....	7
2.1.1 Risma Rustiyan dan Mustakim [4].....	7
2.1.2 Ahmad Saiful dan Joko Lianto Buliali [8].....	7
2.1.3 Dila Fitriani Azuri, dkk [9].....	8
2.1.4 Christina Deni Rumiarti dan Indra Budi [10].....	8
2.1.5 Aditya Ramadhan, dkk [6].....	9
2.2 Pengertian Mortalitas.....	13
2.2.1 Angka Kematian Neonatal (AKN).....	13
2.2.2 Angka Kematian Bayi (AKB).....	14
2.2.3 Angka Kematian Balita (AKABA).....	14
2.2.4 Angka Kematian Anak Balita (AKBA).....	14
2.2.5 Angka Kematian Ibu (AKI).....	14
2.3 Pengertian <i>Data Mining</i> .....	14
2.4 Algoritma <i>K-Means</i> .....	17
2.5 Algoritma <i>Fuzzy C-Means</i> .....	19
2.6 Pengumpulan Sampel Data.....	23
2.7 <i>Preprocessing</i> Data.....	25
2.8 Bahasa Pemrograman R.....	25
2.8.1 Pengertian Bahasa Pemrograman R.....	26
2.8.2 <i>Library</i> Bahasa Pemrograman R.....	30
2.8.3 <i>Clustering</i> pada R.....	31
2.9 Uji Validasi Algoritma.....	35
BAB III METODOLOGI PENELITIAN.....	39
3.1 Tahap Penelitian.....	39
3.2 Studi Literatur.....	39
3.3 Pengumpulan Data.....	39
3.4 Tahapan <i>Preprocessing</i> .....	40
3.5 Metode Penelitian.....	40



3.6 Evaluasi.....	43
BAB IV HASIL PENGUJIAN DAN ANALISIS.....	44
4.1 Hasil Pengujian.....	44
4.1.1 Pengumpulan Data.....	44
4.1.2 Tahapan <i>Preprocessing</i> dan <i>Preprocessing</i> Data.....	45
4.1.3 <i>Clustering K-Means</i> dan <i>Fuzzy C-Means</i> .....	46
4.1.4 Pengukuran Nilai Validasi.....	109
4.1.5 Perbandingan Hasil <i>Cluster</i> .....	152
4.2 Analisis.....	153
BAB V KESIMPULAN DAN SARAN.....	156
5.1 Kesimpulan.....	156
5.2 Saran.....	156
DAFTAR PUSTAKA.....	158
LAMPIRAN.....	161

## DAFTAR GAMBAR

Gambar 1.1. Jumlah Penduduk Indonesia 2010-2035E.....	1
Gambar 2.1. Angka Kelahiran dan Kematian Penduduk Proyeksi Bappenas (2010-2035).....	13
Gambar 2.2. Visualisasi Algoritma <i>K-Means</i> .....	17
Gambar 2.3. Visualisasi <i>Exclusive Clustering</i> .....	17
Gambar 2.4. Diagram Alir Algoritma <i>K-Means</i> .....	18
Gambar 2.5. Visualisasi Algoritma <i>Fuzzy C-Means</i> .....	20
Gambar 2.6. Visualisasi <i>Overlapping Clustering</i> .....	20
Gambar 2.7. Diagram Alir <i>Fuzzy C-Means</i> .....	20
Gambar 2.8. Komputasi $s(i)$ untuk Tiap Objek, Dimana $i$ Berkaitan Terhadap <i>Cluster A</i> .....	35
Gambar 3.1. Alur Metode Penelitian.....	39
Gambar 3.2. Diagram Alir Tahapan Penelitian.....	43
Gambar 4.1. Hasil <i>Clustering</i> dengan <i>K-Means</i> .....	64
Gambar 4.2. Pengelompokan Distribusi Data pada <i>K-Means</i> .....	64
Gambar 4.3. Hasil <i>Clustering</i> dengan <i>K-Means</i> .....	65
Gambar 4.4. Pengelompokan Distribusi Data pada <i>K-Means</i> .....	66
Gambar 4.5. Hasil <i>Clustering</i> dengan <i>K-Means</i> .....	67
Gambar 4.6. Pengelompokan Distribusi Data pada <i>K-Means</i> .....	67
Gambar 4.7. Hasil <i>Clustering</i> dengan <i>K-Means</i> .....	68
Gambar 4.8. Pengelompokan Distribusi Data pada <i>K-Means</i> .....	69
Gambar 4.9. Hasil <i>Clustering</i> dengan <i>K-Means</i> .....	69
Gambar 4.10. Pengelompokan Distribusi Data pada <i>K-Means</i> .....	70
Gambar 4.11. Hasil <i>Clustering</i> dengan <i>K-Means</i> .....	71
Gambar 4.12. Pengelompokan Distribusi Data pada <i>K-Means</i> .....	72
Gambar 4.13. Hasil <i>Clustering</i> dengan <i>K-Means</i> .....	73
Gambar 4.14. Pengelompokan Distribusi Data pada <i>K-Means</i> .....	73
Gambar 4.15. Hasil <i>Clustering</i> dengan <i>K-Means</i> .....	74
Gambar 4.16. Pengelompokan Distribusi Data pada <i>K-Means</i> .....	75
Gambar 4.17. Hasil <i>Clustering</i> dengan <i>K-Means</i> .....	76
Gambar 4.18. Pengelompokan Distribusi Data pada <i>K-Means</i> .....	76
Gambar 4.19. Hasil <i>Clustering</i> dengan <i>K-Means</i> .....	77
Gambar 4.20. Pengelompokan Distribusi Data pada <i>K-Means</i> .....	78
Gambar 4.21. Hasil <i>Clustering</i> dengan <i>K-Means</i> .....	78
Gambar 4.22. Pengelompokan Distribusi Data pada <i>K-Means</i> .....	79
Gambar 4.23. Hasil <i>Clustering</i> dengan <i>K-Means</i> .....	80
Gambar 4.24. Pengelompokan Distribusi Data pada <i>K-Means</i> .....	81
Gambar 4.25. Hasil <i>Clustering</i> dengan <i>K-Means</i> .....	82
Gambar 4.26. Pengelompokan Distribusi Data pada <i>K-Means</i> .....	82
Gambar 4.27. Hasil <i>Clustering</i> dengan <i>K-Means</i> .....	83
Gambar 4.28. Pengelompokan Distribusi Data pada <i>K-Means</i> .....	84
Gambar 4.29. Hasil <i>Clustering</i> dengan <i>K-Means</i> .....	85
Gambar 4.30. Pengelompokan Distribusi Data pada <i>K-Means</i> .....	85
Gambar 4.31. Hasil <i>Clustering</i> dengan <i>Fuzzy C-Means</i> .....	86
Gambar 4.32. Pengelompokan Distribusi Data pada <i>Fuzzy C-Means</i> .....	87
Gambar 4.33. Hasil <i>Clustering</i> dengan <i>Fuzzy C-Means</i> .....	88
Gambar 4.34. Pengelompokan Distribusi Data pada <i>Fuzzy C-Means</i> .....	89
Gambar 4.35. Hasil <i>Clustering</i> dengan <i>Fuzzy C-Means</i> .....	89

Gambar 4.36. Pengelompokkan Distribusi Data pada <i>Fuzzy C-Means</i> .....	90
Gambar 4.37. Hasil <i>Clustering</i> dengan <i>Fuzzy C-Means</i> .....	91
Gambar 4.38. Pengelompokkan Distribusi Data pada <i>Fuzzy C-Means</i> .....	92
Gambar 4.39. Hasil <i>Clustering</i> dengan <i>Fuzzy C-Means</i> .....	93
Gambar 4.40. Pengelompokkan Distribusi Data pada <i>Fuzzy C-Means</i> .....	93
Gambar 4.41. Hasil <i>Clustering</i> dengan <i>Fuzzy C-Means</i> .....	94
Gambar 4.42. Pengelompokkan Distribusi Data pada <i>Fuzzy C-Means</i> .....	95
Gambar 4.43. Hasil <i>Clustering</i> dengan <i>Fuzzy C-Means</i> .....	96
Gambar 4.44. Pengelompokkan Distribusi Data pada <i>Fuzzy C-Means</i> .....	96
Gambar 4.45. Hasil <i>Clustering</i> dengan <i>Fuzzy C-Means</i> .....	97
Gambar 4.46. Pengelompokkan Distribusi Data pada <i>Fuzzy C-Means</i> .....	98
Gambar 4.47. Hasil <i>Clustering</i> dengan <i>Fuzzy C-Means</i> .....	99
Gambar 4.48. Pengelompokkan Distribusi Data pada <i>Fuzzy C-Means</i> .....	99
Gambar 4.49. Hasil <i>Clustering</i> dengan <i>Fuzzy C-Means</i> .....	100
Gambar 4.50. Pengelompokkan Distribusi Data pada <i>Fuzzy C-Means</i> .....	101
Gambar 4.51. Hasil <i>Clustering</i> dengan <i>Fuzzy C-Means</i> .....	102
Gambar 4.52. Pengelompokkan Distribusi Data pada <i>Fuzzy C-Means</i> .....	102
Gambar 4.53. Hasil <i>Clustering</i> dengan <i>Fuzzy C-Means</i> .....	103
Gambar 4.54. Pengelompokkan Distribusi Data pada <i>Fuzzy C-Means</i> .....	104
Gambar 4.55. Hasil <i>Clustering</i> dengan <i>Fuzzy C-Means</i> .....	105
Gambar 4.56. Pengelompokkan Distribusi Data pada <i>Fuzzy C-Means</i> .....	105
Gambar 4.57. Hasil <i>Clustering</i> dengan <i>Fuzzy C-Means</i> .....	106
Gambar 4.58. Pengelompokkan Distribusi Data pada <i>Fuzzy C-Means</i> .....	107
Gambar 4.59. Hasil <i>Clustering</i> dengan <i>Fuzzy C-Means</i> .....	108
Gambar 4.60. Pengelompokkan Distribusi Data pada <i>Fuzzy C-Means</i> .....	109

## DAFTAR TABEL

Tabel 2.1. Tabel Penelitian Terkait ( <i>State of The Art</i> ).....	10
Tabel 2.2. Tabel Penelitian Terkait ( <i>State of The Art</i> ).....	11
Tabel 2.3. Tabel Penelitian Terkait ( <i>State of The Art</i> ).....	12
Tabel 2.4. Perbedaan Bahasa Pemrograman R dan Python.....	27
Tabel 3.1. <i>Data set</i> Profil Kondisi Kesehatan Kabupaten/Kota Tahun 2017.....	40
Tabel 4.1. <i>Data Set</i> Profil Kondisi Kesehatan Kabupaten/Kota Tahun 2017.....	44
Tabel 4.2. <i>Data Set</i> Sebelum Proses <i>Preprocessing</i> Data.....	45
Tabel 4.3. <i>Data Set</i> Sesudah Proses <i>Preprocessing</i> Data.....	46
Tabel 4.4. Pemilihan <i>Centroid</i> pada <i>K-Means</i> .....	46
Tabel 4.5. Nilai Jarak Data ke Setiap <i>Centroid</i> .....	47
Tabel 4.6. Keanggotaan <i>Cluster</i> pada Data.....	47
Tabel 4.7. Penentuan <i>Centroid</i> Baru pada Variabel Data.....	49
Tabel 4.8. Jumlah, Jumlah Data, dan Rata-Rata.....	50
Tabel 4.9. <i>Cluster Mean</i> pada Iterasi 1.....	50
Tabel 4.10. Nilai Jarak Data ke Setiap <i>Centroid</i> .....	51
Tabel 4.11. Keanggotaan <i>Cluster</i> pada Data.....	51
Tabel 4.12. Penentuan <i>Centroid</i> Baru pada Variabel Data.....	53
Tabel 4.13. Jumlah, Jumlah Data, dan Rata-Rata.....	54
Tabel 4.14. <i>Cluster Mean</i> pada Iterasi 2.....	54
Tabel 4.15. <i>Data Set</i> Pengujian <i>Fuzzy C-Means</i> .....	55
Tabel 4.16. Keanggotaan <i>Cluster</i> Iterasi 1 pada <i>Fuzzy C-Means</i> .....	55
Tabel 4.17. Perhitungan Pusat <i>Cluster</i> 1.....	56
Tabel 4.18. Perhitungan Pusat <i>Cluster</i> 2.....	56
Tabel 4.19. Perhitungan Pusat <i>Cluster</i> 3.....	56
Tabel 4.20. Perhitungan Jumlah Pusat <i>Cluster</i> 1.....	57
Tabel 4.21. Perhitungan Jumlah Pusat <i>Cluster</i> 2.....	57
Tabel 4.22. Perhitungan Jumlah Pusat <i>Cluster</i> 3.....	57
Tabel 4.23. Hasil Pusat <i>Cluster</i> (V).....	57
Tabel 4.24. Nilai Fungsi Objektif.....	58
Tabel 4.25. Nilai Matriks Partisi U.....	58
Tabel 4.26. Nilai Keanggotaan <i>Cluster</i> Baru.....	59
Tabel 4.27. Perhitungan Pusat <i>Cluster</i> 1.....	60
Tabel 4.28. Perhitungan Pusat <i>Cluster</i> 2.....	60
Tabel 4.29. Perhitungan Pusat <i>Cluster</i> 3.....	60
Tabel 4.30. Perhitungan Jumlah Pusat <i>Cluster</i> 1.....	61
Tabel 4.31. Perhitungan Jumlah Pusat <i>Cluster</i> 2.....	61
Tabel 4.32. Perhitungan Jumlah Pusat <i>Cluster</i> 3.....	61
Tabel 4.33. Hasil Pusat <i>Cluster</i> (V).....	61
Tabel 4.34. Menghitung Fungsi Objektif.....	62
Tabel 4.35. Menghitung Matriks Partisi U.....	62
Tabel 4.36. Keanggotaan <i>Cluster</i> Baru.....	63
Tabel 4.37. <i>Cluster Means</i> dan Jumlah Anggota <i>Cluster</i> pada <i>K-Means</i> .....	64
Tabel 4.38. <i>Cluster Means</i> dan Jumlah Anggota <i>Cluster</i> pada <i>K-Means</i> .....	65
Tabel 4.39. <i>Cluster Means</i> dan Jumlah Anggota <i>Cluster</i> pada <i>K-Means</i> .....	67
Tabel 4.40. <i>Cluster Means</i> dan Jumlah Anggota <i>Cluster</i> pada <i>K-Means</i> .....	68
Tabel 4.41. <i>Cluster Means</i> dan Jumlah Anggota <i>Cluster</i> pada <i>K-Means</i> .....	70

Tabel 4.42. <i>Cluster Means</i> dan Jumlah Anggota <i>Cluster</i> pada <i>K-Means</i> .....	71
Tabel 4.43. <i>Cluster Means</i> dan Jumlah Anggota <i>Cluster</i> pada <i>K-Means</i> .....	73
Tabel 4.44. <i>Cluster Means</i> dan Jumlah Anggota <i>Cluster</i> pada <i>K-Means</i> .....	74
Tabel 4.45. <i>Cluster Means</i> dan Jumlah Anggota <i>Cluster</i> pada <i>K-Means</i> .....	76
Tabel 4.46. <i>Cluster Means</i> dan Jumlah Anggota <i>Cluster</i> pada <i>K-Means</i> .....	77
Tabel 4.47. <i>Cluster Means</i> dan Jumlah Anggota <i>Cluster</i> pada <i>K-Means</i> .....	79
Tabel 4.48. <i>Cluster Means</i> dan Jumlah Anggota <i>Cluster</i> pada <i>K-Means</i> .....	80
Tabel 4.49. <i>Cluster Means</i> dan Jumlah Anggota <i>Cluster</i> pada <i>K-Means</i> .....	82
Tabel 4.50. <i>Cluster Means</i> dan Jumlah Anggota <i>Cluster</i> pada <i>K-Means</i> .....	83
Tabel 4.51. <i>Cluster Means</i> dan Jumlah Anggota <i>Cluster</i> pada <i>K-Means</i> .....	85
Tabel 4.52. Nilai Keanggotaan, Fungsi Objektif, dan Toleransi pada <i>Fuzzy C-Means</i> .....	87
Tabel 4.53. Nilai Keanggotaan, Fungsi Objektif, dan Toleransi pada <i>Fuzzy C-Means</i> .....	88
Tabel 4.54. Nilai Keanggotaan, Fungsi Objektif, dan Toleransi pada <i>Fuzzy C-Means</i> .....	90
Tabel 4.55. Nilai Keanggotaan, Fungsi Objektif, dan Toleransi pada <i>Fuzzy C-Means</i> .....	91
Tabel 4.56. Nilai Keanggotaan, Fungsi Objektif, dan Toleransi pada <i>Fuzzy C-Means</i> .....	93
Tabel 4.57. Nilai Keanggotaan, Fungsi Objektif, dan Toleransi pada <i>Fuzzy C-Means</i> .....	95
Tabel 4.58. Nilai Keanggotaan, Fungsi Objektif, dan Toleransi pada <i>Fuzzy C-Means</i> .....	96
Tabel 4.59. Nilai Keanggotaan, Fungsi Objektif, dan Toleransi pada <i>Fuzzy C-Means</i> .....	98
Tabel 4.60. Nilai Keanggotaan, Fungsi Objektif, dan Toleransi pada <i>Fuzzy C-Means</i> .....	99
Tabel 4.61. Nilai Keanggotaan, Fungsi Objektif, dan Toleransi pada <i>Fuzzy C-Means</i> .....	101
Tabel 4.62. Nilai Keanggotaan, Fungsi Objektif, dan Toleransi pada <i>Fuzzy C-Means</i> .....	102
Tabel 4.63. Nilai Keanggotaan, Fungsi Objektif, dan Toleransi pada <i>Fuzzy C-Means</i> .....	104
Tabel 4.64. Nilai Keanggotaan, Fungsi Objektif, dan Toleransi pada <i>Fuzzy C-Means</i> .....	105
Tabel 4.65. Nilai Keanggotaan, Fungsi Objektif, dan Toleransi pada <i>Fuzzy C-Means</i> .....	107
Tabel 4.66. Nilai Keanggotaan, Fungsi Objektif, dan Toleransi pada <i>Fuzzy C-Means</i> .....	108
Tabel 4.67. Tabel Pembagian Nilai <i>Cluster</i> .....	110
Tabel 4.68. Jarak <i>Intrachuster</i> dan <i>Intercluster K-Means</i> .....	111
Tabel 4.69. Nilai SI dengan 2 <i>Cluster</i> pada <i>K-Means</i> .....	112
Tabel 4.70. Tabel Pembagian Nilai <i>Cluster</i> .....	112
Tabel 4.71. Jarak <i>Intrachuster</i> dan <i>Intercluster K-Means</i> .....	114
Tabel 4.72. Nilai SI dengan 3 <i>Cluster</i> pada <i>K-Means</i> .....	115
Tabel 4.73. Tabel Pembagian Nilai <i>Cluster</i> .....	115
Tabel 4.74. Nilai SI dengan 4 <i>Cluster</i> pada <i>K-Means</i> .....	117
Tabel 4.75. Jarak <i>Intrachuster</i> dan <i>Intercluster K-Means</i> .....	118
Tabel 4.76. Tabel Pembagian Nilai <i>Cluster</i> .....	119
Tabel 4.77. Jarak <i>Intrachuster</i> dan <i>Intercluster Fuzzy-Means</i> .....	120
Tabel 4.78. Nilai SI dengan 2 <i>Cluster</i> pada <i>Fuzzy C-Means</i> .....	121
Tabel 4.79. Tabel Pembagian Nilai <i>Cluster</i> .....	122
Tabel 4.80. Jarak <i>Intrachuster</i> dan <i>Intercluster Fuzzy-Means</i> .....	123
Tabel 4.81. Nilai SI dengan 3 <i>Cluster</i> pada <i>Fuzzy C-Means</i> .....	124
Tabel 4.82. Tabel Pembagian Nilai <i>Cluster</i> .....	124
Tabel 4.83. Nilai SI dengan 4 <i>Cluster</i> pada <i>Fuzzy C-Means</i> .....	126
Tabel 4.84. Jarak Antar Objek dalam <i>Cluster</i> dan Luar <i>Cluster</i> .....	127
Tabel 4.85. Tabel Pembagian Nilai <i>Cluster</i> .....	128
Tabel 4.86. Jarak <i>Intrachuster K-Means</i> .....	129
Tabel 4.87. Jarak <i>Intercluster K-Means</i> .....	130
Tabel 4.88. Tabel Pembagian Nilai <i>Cluster</i> .....	131
Tabel 4.89. Jarak <i>Intrachuster K-Means</i> .....	131



Tabel 4.90. Jarak <i>Intercluster K-Means</i> .....	133
Tabel 4.91. Tabel Pembagian Nilai <i>Cluster</i> .....	134
Tabel 4.92. Jarak <i>Intracluster K-Means</i> .....	134
Tabel 4.93. Jarak <i>Intercluster K-Means</i> .....	137
Tabel 4.94. Tabel Pembagian Nilai <i>Cluster</i> .....	138
Tabel 4.95. Jarak <i>Intracluster Fuzzy C-Means</i> .....	139
Tabel 4.96. Jarak <i>Intercluster Fuzzy C-Means</i> .....	140
Tabel 4.97. Tabel Pembagian <i>Cluster</i> .....	141
Tabel 4.98. Jarak <i>Intracluster Fuzzy C-Means</i> .....	142
Tabel 4.99. Jarak <i>Intercluster Fuzzy C-Means</i> .....	143
Tabel 4.100. Tabel Pembagian <i>Cluster</i> .....	144
Tabel 4.101. Jarak <i>Intracluster Fuzzy C-Means</i> .....	144
Tabel 4.102. Jarak <i>Intercluster Fuzzy C-Means</i> .....	147
Tabel 4.103. Pengukuran Nilai Validasi dengan 2 <i>Cluster</i> .....	148
Tabel 4.104. Pengukuran Nilai Validasi dengan 3 <i>Cluster</i> .....	148
Tabel 4.105. Pengukuran Nilai Validasi dengan 4 <i>Cluster</i> .....	148
Tabel 4.106. Pengukuran Nilai Validasi dengan 2 <i>Cluster</i> .....	149
Tabel 4.107. Pengukuran Nilai Validasi dengan 3 <i>Cluster</i> .....	149
Tabel 4.108. Pengukuran Nilai Validasi dengan 4 <i>Cluster</i> .....	149
Tabel 4.109. Pengukuran Nilai Validasi dengan 2 <i>Cluster</i> .....	149
Tabel 4.110. Pengukuran Nilai Validasi dengan 3 <i>Cluster</i> .....	150
Tabel 4.111. Pengukuran Nilai Validasi dengan 4 <i>Cluster</i> .....	150
Tabel 4.112. Pengukuran Nilai Validasi dengan 2 <i>Cluster</i> .....	150
Tabel 4.113. Pengukuran Nilai Validasi dengan 3 <i>Cluster</i> .....	150
Tabel 4.114. Pengukuran Nilai Validasi dengan 4 <i>Cluster</i> .....	151
Tabel 4.115. Pengukuran Nilai Validasi dengan 2 <i>Cluster</i> .....	151
Tabel 4.116. Pengukuran Nilai Validasi dengan 3 <i>Cluster</i> .....	151
Tabel 4.117. Pengukuran Nilai Validasi dengan 4 <i>Cluster</i> .....	151
Tabel 4.118. Tabel Perbandingan Hasil <i>Cluster</i> .....	152

## DAFTAR RUMUS

Rumus 2.1. Rumus Euclidean.....	19
Rumus 2.2. Perhitungan Pusat <i>Cluster</i> .....	21
Rumus 2.3. Rumus Fungsi Objektif.....	21
Rumus 2.4. Perhitungan Perubahan Matriks Partisi.....	22
Rumus 2.5. Rumus <i>Cochran</i> .....	24
Rumus 2.6. Rumus <i>Cochran</i> .....	24
Rumus 2.7. Normalisasi pada Data.....	25
Rumus 2.8. Menentukan Jumlah $s(i)$ pada Penentuan Ketidaksamaan.....	36
Rumus 2.9. Penentuan Jumlah $s(i)$ Pertimbangan Ketidaksamaan.....	36
Rumus 2.10. Penentuan Jumlah $s(i)$ Pertimbangan Kesamaan.....	36
Rumus 2.11. Rumus <i>Sum of Square within Cluster</i> .....	37
Rumus 2.12. Persamaan <i>Sum of Square Between Cluster</i> .....	37
Rumus 2.13. Persamaan Pengukuran Rasio.....	38
Rumus 2.14. Persamaan <i>Davies Bouldin Index (DBI)</i> .....	38

## DAFTAR LAMPIRAN

Lampiran 1. <i>Data Set</i> Kesehatan Kabupaten/Kota Tahun 2017.....	161
Lampiran 2. <i>Source Code</i> Algoritma <i>K-Means</i> .....	176
Lampiran 3. <i>Source Code</i> Algoritma <i>Fuzzy C-Means</i> .....	177
Lampiran 4. Pembagian Anggota <i>Cluster K-Means</i> pada Data Kesehatan Daerah.....	179
Lampiran 5. Pembagian Anggota <i>Cluster Fuzzy C-Means</i> pada Data Kesehatan Daerah.....	198

## ABSTRAK

Besarnya jumlah penduduk Indonesia harus diimbangi dengan jaminan kesehatan yang merata di setiap daerah. Pemerintah Indonesia melalui Departemen Kesehatan setiap tahunnya mengumpulkan data kesehatan penduduk mulai dari tingkat kelurahan sampai dengan provinsi. Hal ini dilaksanakan untuk menentukan peringkat daerah yang sehat dan menjadi acuan penting bagi Pemerintah Daerah (Pemda) untuk memberikan bantuan kesehatan yang lebih tepat sasaran dalam merumuskan Daerah Bermasalah Kesehatan Berat/Khusus (DBKKBK). Dalam membantu pemerintah merumuskan DBKKBK, *data mining* digunakan untuk mengelompokkan data yang besar berdasarkan kemiripan data. Algoritma *K-Means* merupakan salah satu algoritma klasterisasi sederhana, paling banyak digunakan dan mampu mengelompokkan data dalam jumlah yang besar dengan waktu yang relatif cepat dan efisien, sedangkan algoritma *Fuzzy C-Means* merupakan salah satu model pengelompokkan fuzzy dengan tingkat keanggotaan berbeda antara 0 dan 1 sehingga data dapat menjadi anggota dari klaster khusus dan setiap data memiliki jarak keanggotaan untuk setiap klaster. Penelitian ini memiliki tujuan untuk menggambarkan klaster terbaik dalam distribusi pengelompokkan kabupaten/kota berdasarkan kondisi profil kesehatan daerah dengan algoritma klaster terbaik. Data yang digunakan pada penelitian ini yaitu data kesehatan Kabupaten/Kota tahun 2017 dengan 550 sampel data dan lima buah variabel yang diperoleh dari Pusat Data dan Informasi Kementerian Kesehatan (Pusdatin Kemkes). Hasil dari penelitian ini menunjukkan bahwa algoritma *K-Means* memberikan hasil terbaik dibandingkan dengan *Fuzzy C-Means* untuk klaster  $K = 3$ , nilai validasi *Davies Bouldin Index* (DBI) sebesar 0.4571, dan nilai validasi *Silhouette Index* (SI) sebesar 0.7486.

**Kata Kunci** – *Data Mining, Davies Bouldin Index, Clustering, Silhouette Index, Kesehatan*

## ABSTRACT

*The size of Indonesia population must be balanced with health insurance that is evenly distributed in each region. The Indonesia government through the Ministry of Health annually collects population health data from the subdistrict to the provincial level. This is done to determine the ranking of healthy region and become an important reference for Regional Government (Pemda) to provide health assistance which is more targeted in formulating of Problematic Areas for Heavy/Special Health (DBKKBK). In helping the government to formulate DBKKBK, data mining is used to group the large data based on data similarities. K-Means is one of the simplest and most widely used clustering algorithm and able to group large amounts of data with relatively fast and efficient time, while the Fuzzy C-Means is one of the fuzzy grouping models with membership levels differing between 0 and 1 so that data can be a member of a special cluster and each data has a membership distance for each cluster. This study aims to describe the best clusters in the regency/city grouping distribution based on the condition of the regional health profile with the best cluster algorithm. The data used in this study are district/city health data in 2017 with 550 data samples and five variables obtained from the Ministry of Health Data and Information Center (Pusdatin Kemkes). The result of this study indicate that K-Means algorithm provides the best results compared to Fuzzy C-Means for cluster  $K=3$ , the Davies Bouldin Index (DBI) validation value is 0.4571, and the Silhouette Index (SI) validation value is 0.7486.*

**Keywords** – Data Mining, Davies Bouldin Index, Clustering, Health, Silhouette Index



# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Negara Indonesia merupakan negara dengan jumlah penduduk terbesar keempat di dunia. Penduduk Indonesia setiap tahunnya mengalami peningkatan yang cukup signifikan. Berdasarkan data yang diperoleh dari Badan Pusat Statistik tahun 2018 menunjukkan bahwa penduduk Indonesia berjumlah 261.890.900 jiwa [1]. Populasi yang besar didominasi oleh usia produktif menyebabkan angka ketergantungan yang cenderung menurun. Angka ketergantungan yaitu jumlah penduduk usia produktif terhadap usia tidak produktif, dimana pada 2016 mencapai 48.4 persen [2].



Gambar 1.1. Jumlah Penduduk Indonesia 2010-2035E

Besarnya jumlah penduduk Indonesia harus diimbangi dengan jaminan kesehatan yang merata di setiap daerah. Pemerintah Indonesia melalui Departemen Kesehatan menetapkan sejumlah indikator (Angka Kematian Neonatal, Angka Kematian Balita, Angka Kematian Anak Balita, Angka Kematian Ibu) sebagai tolak ukur kemajuan kesehatan. Departemen Kesehatan setiap tahunnya mengumpulkan data kesehatan penduduk mulai dari tingkat kelurahan sampai dengan provinsi [3]. Hal ini dilaksanakan untuk menentukan peringkat daerah yang sehat. Hasil dari data dan pemeringkatan yang sudah diolah tersebut menjadi acuan penting bagi Pemerintah Daerah (Pemda) untuk memberikan bantuan kesehatan yang lebih tepat sasaran dalam merumuskan Daerah Bermasalah Kesehatan Berat/Khusus (DBKKBK). Rumusan tersebut selanjutnya menjadi dasar dalam penentuan alokasi dana dari

pemerintah pusat ke pemerintah daerah dalam bidang kesehatan, dan juga ikut berperan dalam membantu Kementerian Negara Pembangunan Daerah Tertinggal (KNPDT) [3]. Dalam membantu pemerintah merumuskan Daerah Bermasalah Kesehatan Berat/Khusus (DBKKBK), diperlukan suatu metode untuk mengolah kumpulan data yang besar dan mengelompokkan daerah berdasarkan kemiripan karakteristik daerah yang ditinjau dari nilai lima buah indikator mortalitas derajat kesehatan daerah, yaitu angka kematian neonatal, angka kematian bayi, angka kematian balita, angka kematian anak balita, dan angka kematian ibu. Metode yang dapat membantu pemerintah untuk mengelompokkan daerah dan merumuskan Daerah Bermasalah Kesehatan Berat/Khusus (DBKKBK) yaitu menggunakan *data mining* [4].

*Data mining* merupakan analisis dari *data set* untuk menemukan hubungan tak terduga dan untuk meringkaskan data-data dengan cara baru sehingga lebih mudah dipahami dan berguna bagi pemilik data [5] dengan menemukan pola-pola yang berarti. Proses tersebut harus bersifat otomatis atau (biasanya lebih) semi-otomatis. *Data mining* dapat diukur performansinya pada berbagai jenis data dan informasi yang tersedia di *database* dan repositori, tetapi jenis-jenis pola yang bisa ditentukan oleh beragam kegunaan *data mining*, seperti: penjelasan kelas/konsep, asosiasi, analisis korelasi, klasifikasi, prediksi, analisis klaster, dan sebagainya. Diantaranya yaitu analisis klaster adalah salah satu metode analisis data utama secara luas digunakan untuk sejumlah penerapan praktis dalam daerah berkembang.

*K-Means* merupakan salah satu algoritma klasterisasi sederhana dan paling banyak digunakan dalam klasterisasi [6]. Kemampuan dalam mengelompokkan data dalam jumlah yang besar dengan waktu yang relatif cepat dan efisien menjadikan *K-Means* sebagai algoritma yang paling umum digunakan. Namun, *K-Means* memiliki kelemahan, yaitu pemilihan titik secara *random* dapat mempengaruhi hasil *clustering*, dapat terjebak pada besar kecilnya dimensi (*curse of dimensionality*), hasil klaster yang berubah-ubah, dan jumlah titik sampel yang diteliti dapat mempengaruhi pencarian dan perhitungan titik terdekat.

*Fuzzy C-Means* adalah salah satu model pengelompokkan fuzzy dengan tingkat keanggotaan berbeda antara 0 dan 1 sehingga data dapat menjadi anggota dari klaster khusus. Probabilitas dari eksistensi sebuah data ke sebuah klaster tertentu dapat dilihat dari jarak keanggotaan atau seperti klaster terdekat ke nilai dari data. Konsep dasar dari *Fuzzy C-Means*, yaitu pertama tentukan titik pusat dari klaster. Setiap data memiliki jarak keanggotaan untuk setiap klaster [7]. Dengan memperbaiki titik pusat klaster dan nilai keanggotaan setiap data

secara berulang, maka dapat ditemukan bahwa titik pusat kluster akan berada di lokasi yang tepat. Fungsi obyektif yang menggambarkan jarak dari lokasi data terhadap titik pusat kluster dapat diminimalkan dengan melakukan *looping* pemberian bobot tingkat kesamaan titik data dari himpunan lepas. *Fuzzy C-Means* memiliki kelebihan, diantaranya: sederhana, mudah diimplementasikan, memiliki kemampuan untuk mengelompokkan data yang besar, dan keluaran yang dihasilkan lebih efektif. Kelemahan dari *Fuzzy C-Means* yaitu *Fuzzy C-Means* tidak dapat melakukan penempatan objek tepat pada satu partisi, karena objek terletak diantara dua atau lebih partisi lainnya.

Penelitian sejenis mengenai penentuan kluster optimal dan penentuan nilai validasi tinggi pernah dilakukan oleh [4] yang menerapkan *Fuzzy C-Means*. Penelitian ini dilakukan kepada nasabah pada Koperasi Simpan Pinjam dan setelah ditelusuri belum ada yang melakukan penelitian dengan objek anggota PUSKUD khususnya pada bidang permodalan Simpanan Pokok dan Simpanan Wajib, padahal sedikit banyaknya hal ini dapat berpengaruh pada kelangsungan proses bisnis yang ada pada PUSKUD Riau. Dari penelitian yang telah dilakukan, diperoleh hasil yaitu terdapat 75 anggota yang tersebar pada wilayah kabupaten/kota Rokan Hulu, Kampar, Indragiri Hulu, dan Indragiri Hilir serta terdaftar pada tahun 80-an yang perlu ditinjau kembali. Hasil pengujian nilai validitas PC didapatkan sebesar 0.323732, dengan demikian kualitas *cluster* masih jauh untuk mencapai kata optimal.

Penelitian yang dilakukan oleh [6] yang menerapkan *K-Means* dan *Fuzzy C-Means*. Pada penelitian ini ditunjukkan perbandingan pengklasteran data *user knowledge modeling* menggunakan metode *K-Means* dan *Fuzzy C-Means* untuk melihat kemampuan dari setiap algoritma *clustering* sehingga mendapatkan hasil terbaik. Dari penelitian yang telah dilakukan, dapat disimpulkan bahwa penggunaan algoritma *K-Means* dan *Fuzzy C-Means* untuk pengklasteran data *user knowledge modeling* mendapatkan hasil yang berbeda. Perbedaan hasil tersebut dapat dilihat pada jumlah kluster dan nilai validitas yang diperoleh. Jika dilihat dari nilai validitas, nilai SI pada algoritma *K-Means* bernilai 0.1866, sedangkan nilai PCI pada algoritma *Fuzzy C-Means* bernilai 0.2854.

Penelitian yang dilakukan oleh [8] yang menerapkan *Bisecting K-Means*. Penelitian ini dilakukan karena terjadi ketimpangan pada nilai IPG (Indeks Pembangunan Gender) dan IDG (Indeks Pemberdayaan Gender) pada setiap kabupaten/kota yang menandakan belum terjadinya pemerataan pembangunan yang dirasakan oleh seluruh rakyat. Dari penelitian didapatkan 3 kluster yang terbentuk baik pada laki-laki ataupun perempuan. Pada kluster laki-

laki, klaster 1 beranggotakan 32 Kabupaten/Kota, anggota pada klaster 2 yaitu 43 Kabupaten/Kota, dan klaster 3 terdapat 44 Kabupaten/Kota. Nilai silhouette coefficient pada klaster laki-laki yaitu 0,3. Sedangkan pada klaster perempuan yaitu dimana pada klaster 1 beranggotakan 42 Kabupaten/Kota, anggota pada klaster 2 yaitu 42 Kabupaten/Kota, dan klaster 3 terdapat 35 Kabupaten/Kota. Nilai *Silhouette Coefficient* pada klaster perempuan yaitu 0,26. Baik pada klaster perempuan ataupun laki-laki, klaster yang terbentuk masih tergolong lemah.

Penelitian yang dilakukan oleh [9] yang menerapkan *K-Means*. Penelitian ini dilakukan untuk mengelompokkan rute penerbangan berdasarkan kebiasaan data dari penerbangan yang ada. Pengelompokkan dilakukan dengan menggunakan metode *clustering* berbasis jarak yaitu *K-Means clustering*. Hasil dari penelitian ini menghasilkan zona terbang berdasarkan kebiasaan sehingga dapat digunakan sebagai panduan penerbangan. Pengujian dilakukan dengan membandingkan nilai Davies Bouldin *index* dengan metode *K-Means*, *K-Medoids*, dan *Fuzzy C-Means*. Pada uji coba yang dilakukan, metode yang diusulkan menjadi kelompok metode terbaik pada lima dari enam segmen yang ada serta menghasilkan nilai Davies Bouldin *index* lebih baik pada satu segmen sebesar 0.779.

Penelitian yang dilakukan oleh [10] dengan algoritma *K-Means*. Penelitian ini dilakukan untuk menghasilkan segmentasi pelanggan pada *Customer Relationship Management* di PT Gramedia Asri Media. Data yang dikumpulkan adalah data member KGVC dan data transaksi. Jumlah cluster optimal pada segmentasi nilai pelanggan yang dihasilkan dalam penelitian ini berdasarkan pada nilai tertinggi *silhouette method* dan *calinski-harabasz index*. Nilai *average silhouette* pada dua cluster adalah 0.42. Nilai *Calinski Harabasz index* pada dua cluster adalah 1.31.

Penelitian yang dilakukan oleh [4], [6], [9], [8], dan [10] dengan algoritma *K-Means*, *Fuzzy C-Means*, dan *Bisecting K-Means*, diperoleh hasil penelitian berupa jumlah klaster yang belum optimal, kualitas klaster yang lemah, dan nilai validasi yang masih rendah. Hasil *clustering* pada *K-Means* dan *Fuzzy C-Means* sangat berpengaruh pada kualitas klaster dan nilai validasi yang dihasilkan. Dengan demikian diperlukan penelitian lebih lanjut untuk menentukan jumlah klaster terbaik dan kualitas klaster yang optimal dengan nilai validasi yang tinggi.

Penggunaan algoritma *K-Means* atau *Fuzzy C-Means* dipilih sebagai algoritma untuk pengelompokan kabupaten/kota di Indonesia dengan menggunakan lima buah indikator

(atribut) derajat kesehatan, yaitu jumlah kelahiran, jumlah neonatal, jumlah kematian bayi, angka kematian balita, dan angka kematian ibu. Dengan menggunakan Algoritma *K-Means* dan *Fuzzy C-Means*, diharapkan Pemerintah Daerah (Pemda) dapat terbantu dalam menentukan dan mengelompokkan daerah-daerah yang membutuhkan perhatian dan penanganan khusus, menentukan alokasi dana dari pemerintah pusat ke daerah lebih tepat sasaran, dan sekaligus membantu Kementerian Negara Pembangunan Daerah Tertinggal (KNPDT).

## 1.2 Rumusan Masalah

Berdasarkan permasalahan pada latar belakang di atas, maka rumusan penelitian pada Tugas Akhir ini adalah sebagai berikut:

1. Apa algoritma *cluster* yang lebih baik antara *K-Means* dan *Fuzzy C-Means* ?.
2. Berapa nilai validasi dan jumlah *cluster* yang optimal yang diperoleh dari algoritma yang lebih baik antara *K-Means* dan *Fuzzy C-Means* ?.
3. Berapa jumlah data pada setiap *cluster* yang diperoleh dari algoritma yang lebih baik antara *K-Means* dan *Fuzzy C-Means* ?.

## 1.3 Tujuan Penelitian

Dari rumusan masalah yang telah dipaparkan di atas, maka tujuan penulisan Tugas Akhir ini yaitu:

1. Menentukan algoritma *cluster* yang lebih baik antara *K-Means* dan *Fuzzy C-Means*.
2. Menentukan nilai validasi dan jumlah *cluster* yang optimal yang diperoleh dari algoritma yang lebih baik antara *K-Means* dan *Fuzzy C-Means*.
3. Menentukan jumlah data pada setiap *cluster* yang diperoleh dari algoritma yang lebih baik antara *K-Means* dan *Fuzzy C-Means*.

## 1.4 Manfaat Penelitian

Penulisan Tugas Akhir ini diharapkan dapat:

1. Menjadi acuan dalam pemilihan algoritma *cluster* yang lebih baik antara *K-Means* dan *Fuzzy C-Means*.



2. Menjadi acuan dalam menentukan nilai validasi dan jumlah *cluster* yang optimal yang diperoleh dari algoritma yang lebih baik antara *K-Means* dan *Fuzzy C-Means*.
3. Menjadi acuan dalam menentukan jumlah data pada setiap *cluster* yang diperoleh dari algoritma yang lebih baik antara *K-Means* dan *Fuzzy C-Means*

## 1.5 Batasan Masalah

Berdasarkan uraian di atas, batasan masalah pada penelitian tugas akhir ini adalah:

1. *Data set* yang digunakan yaitu data publik Profil Kesehatan Penduduk Kabupaten/Kota Tahun 2017 yang diperoleh dari *website* Departemen Kesehatan Republik Indonesia (Depkes RI).
2. Variabel penelitian yang digunakan dalam *clustering* terdiri dari lima buah indikator (atribut) derajat kesehatan, yaitu jumlah neonatal, jumlah kematian bayi, jumlah kematian balita, jumlah kematian anak balita, dan jumlah kematian ibu.
3. Percobaan yang akan dilakukan pada penelitian ini yaitu sebanyak tiga skenario dengan masing-masing skenario yaitu skenario pertama dengan dua *cluster*, skenario kedua dengan tiga *cluster*, dan skenario ketiga dengan empat *cluster*. Setiap skenario akan dilakukan lima kali percobaan mulai dari 510 data sampai dengan 550 data dengan rentang data yaitu 10 data.
4. Penelitian ini akan menggunakan dua indeks validasi, yaitu *Silhouette Index* (SI) dan *Davies Bouldin Index* (DBI).
5. Mengambarkan pengelompokkan distribusi daerah berdasarkan algoritma *K-Means* dan *Fuzzy C-Means*.
6. Bahasa pemrograman yang digunakan adalah bahasa R dan IDE yang digunakan adalah RStudio.

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Tinjauan Pustaka**

Penelitian yang berkaitan dengan penentuan jumlah kluster terbaik dan hasil validasi optimal bukan kali pertama dilakukan. Penelitian ini dilakukan dengan melakukan tinjauan pustaka untuk mengetahui pemahaman yang lebih mendalam mengenai metode yang akan dibandingkan, nilai validasi, dan objek penelitian. Berikut ini beberapa penelitian terkait dengan penelitian, “Komparasi Validasi *K-Means* dan *Fuzzy C-Means* dalam Menentukan Kluster Terbaik”.

##### **2.1.1 Risma Rustiyan dan Mustakim [4]**

Penelitian ini dilakukan kepada nasabah pada Koperasi Simpan Pinjam dan setelah ditelusuri belum ada yang melakukan penelitian dengan objek anggota PUSKUD khususnya pada bidang permodalan Simpanan Pokok dan Simpanan Wajib, padahal sedikit banyaknya hal ini dapat berpengaruh pada kelangsungan proses bisnis yang ada pada PUSKUD Riau. Oleh karena itu penelitian ini dilanjutkan dengan menggunakan algoritma *Fuzzy C-Means* (FCM). Data yang digunakan pada penelitian ini yaitu total KUD yang terdaftar menjadi anggota PUSKUD sebanyak 324 pada tahun 2014, 325 pada tahun 2015 dan 2016 yang tersebar pada 12 kabupaten/kota di provinsi Riau dan Kepulauan Riau.

Dari penelitian yang telah dilakukan, diperoleh hasil yaitu terdapat 75 anggota yang tersebar pada wilayah kabupaten/kota Rokan Hulu, Kampar, Indragiri Hulu, dan Indragiri Hilir serta terdaftar pada tahun 80-an yang perlu ditinjau kembali. Hasil pengujian nilai validitas PC didapatkan sebesar 0.323732, dengan demikian kualitas *cluster* masih jauh untuk mencapai kata optimal.

##### **2.1.2 Ahmad Saiful dan Joko Lianto Buliali [8]**

Penelitian ini dilakukan untuk mengelompokkan rute penerbangan berdasarkan kebiasaan data dari penerbangan yang ada. Pengelompokkan dilakukan dengan menggunakan metode *clustering* berbasis jarak yaitu *K-Means clustering*. Data yang digunakan pada penelitian ini adalah rute penerbangan Surabaya-Palu dengan kode penerbangan LNI860. Implementasi metode dilakukan pada setiap segmen yang terbagi berdasarkan *waypoint*. Evaluasi *cluster* berdasarkan pengukuran *internal validity cluster* menggunakan Davies Bouldin *index*. Pengujian dilakukan dengan 10 variasi dari 70% data yang dipilih secara acak,

setiap variasi akan dilakukan percobaan sebanyak 50 kali percobaan dan dibandingkan dengan metode *K-Means*, *K-Medoids* serta *Fuzzy C-Means*.

Hasil dari penelitian ini menghasilkan zona terbang berdasarkan kebiasaan sehingga dapat digunakan sebagai panduan penerbangan. Pengujian dilakukan dengan membandingkan nilai Davies Bouldin *index* dengan metode *K-Means*, *K-Medoids*, dan *Fuzzy C-Means*. Pada uji coba yang dilakukan, metode yang diusulkan menjadi kelompok metode terbaik pada lima dari enam segmen yang ada serta menghasilkan nilai Davies Bouldin *index* lebih baik pada satu segmen sebesar 0.779.

### **2.1.3 Dila Fitriani Azuri, dkk [9]**

Penelitian ini dilakukan karena terjadi ketimpangan pada nilai IPG (Indeks Pembangunan Gender) dan IDG (Indeks Pemberdayaan Gender) pada setiap kabupaten/kota yang menandakan belum terjadinya pemerataan pembangunan yang dirasakan oleh seluruh rakyat. Data yang digunakan pada penelitian ini merupakan data sekunder yaitu dari publikasi dari Badan Pusat Statistik dengan judul Pembangunan Manusia Berbasis Gender 2015. Kabupaten/kota di pulau Jawa berjumlah 119 buah dan variabel yang terlibat pada penelitian ini berjumlah enam buah variabel, yaitu angka harapan hidup, rata-rata lama sekolah, harapan lama sekolah, pengeluaran per kapita, keterwakilan di parlemen, dan profesi sebagai tenaga manajer, profesional, administrasi, dan teknisi.

Dari penelitian didapatkan 3 klaster yang terbentuk baik pada laki-laki ataupun perempuan. Pada klaster laki-laki, klaster 1 beranggotakan 32 Kabupaten/Kota, anggota pada klaster 2 yaitu 43 Kabupaten/Kota, dan klaster 3 terdapat 44 Kabupaten/Kota. Nilai silhouette coefficient pada klaster laki-laki yaitu 0,3. Sedangkan pada klaster perempuan yaitu dimana pada klaster 1 beranggotakan 42 Kabupaten/Kota, anggota pada klaster 2 yaitu 42 Kabupaten/Kota, dan klaster 3 terdapat 35 Kabupaten/Kota. Nilai *Silhouette Coefficient* pada klaster perempuan yaitu 0,26. Baik pada klaster perempuan ataupun laki-laki, klaster yang terbentuk masih tergolong lemah.

### **2.1.4 Christina Deni Rumiarti dan Indra Budi [10]**

Penelitian ini dilakukan untuk menghasilkan segmentasi pelanggan pada *Customer Relationship Management* di PT Gramedia Asri Media. Data yang dikumpulkan adalah data member KGVC dan data transaksi. Pelanggan yang masih aktif per tanggal 1 Agustus 2016 diolah untuk segmentasi. Data member KGVC diambil dari data pelanggan seluruh toko Gramedia yang masih aktif pada tanggal 31 Juli 2016. Sedangkan data transaksi merupakan

transaksi produk buku dari member KGVC pada periode 1 Januari 2016 sampai 31 Juli 2016. Data transaksi yang digunakan dalam penelitian adalah seluruh data transaksi buku dari member KGVC yang masih aktif pada tanggal 31 Juli 2016. Data yang diambil merupakan transaksi dengan jumlah nilai lebih dari nol rupiah. Dengan kata lain, transaksi pengembalian buku tidak diambil pada tahap pengumpulan data. Berdasarkan 389.686 pelanggan yang masih aktif pada tanggal 31 Juli 2016, jumlah transaksi pada periode 1 Januari – 31 Juli 2016 adalah 1.535.002 transaksi. Berdasarkan transaksi tersebut, hanya 153.347 pelanggan yang melakukan transaksi pada periode tersebut.

Dari penelitian yang telah dilakukan, dapat disimpulkan bahwa jumlah *cluster* optimal pada segmentasi nilai pelanggan yang dihasilkan dalam penelitian ini berdasarkan pada nilai tertinggi *silhouette method* dan *calinski-harabasz index*. Nilai *average silhouette* pada dua *cluster* adalah 0.42. Nilai *Calinski Harabasz index* pada dua *cluster* adalah 1.31.

#### **2.1.5 Aditya Ramadhan, dkk [6]**

Penelitian ini melakukan perbandingan *K-Means* dan *Fuzzy C-Means* untuk mengelompokkan *Data User Knowledge Modelling*. Pada penelitian ini ditunjukkan perbandingan pengklasteran data *user knowledge modeling* menggunakan metode *K-Means* dan *Fuzzy C-Means* untuk melihat kemampuan dari setiap algoritma *clustering* sehingga mendapatkan hasil terbaik. Data yang digunakan pada penelitian ini menggunakan data *user knowledge modeling* yang didapatkan dari UCI Machine Learning Repository dimana data *user knowledge modeling* tersebut memiliki 258 *record* data dan dikelompokkan menjadi 4 kelompok tingkatan siswa/pengguna.

Dari penelitian yang telah dilakukan, dapat disimpulkan bahwa penggunaan algoritma *K-Means* dan *Fuzzy C-Means* untuk pengklasteran data *user knowledge modeling* mendapatkan hasil yang berbeda. Perbedaan hasil tersebut dapat dilihat pada jumlah klaster dan nilai validitas yang diperoleh. Jika dilihat dari nilai validitas, nilai SI pada algoritma *K-Means* bernilai 0.1866, sedangkan nilai PCI pada algoritma *Fuzzy C-Means* bernilai 0.2854.

Tabel 2.1. Tabel Penelitian Terkait (*State of The Art*)

No.	Judul Penelitian, Penulis, Tahun Penelitian	Data	Masalah	Metode	Evaluasi	Perbedaan dengan Penelitian Terkait
1	Penerapan Algoritma <i>Fuzzy C-Means</i> untuk Analisis Permasalahan Simpanan Wajib Anggota Koperasi, Risma Rustiyan dan Mustakim, 2018 [4]	Total KUD yang terdaftar menjadi anggota PUSKUD provinsi Riau dan Kepulauan Riau	Partisipasi aktif dari anggota koperasi untuk menunaikan kewajiban masih kurang dari yang diharapkan	<i>Fuzzy C-Means</i>	Hasil pengujian nilai validitas PC, didapatkan sebesar 0.323732 dengan demikian kualitas <i>cluster</i> masih jauh untuk mencapai kata optimal.	Penelitian yang dilakukan merupakan perbandingan <i>K-Means</i> dan <i>Fuzzy C-Means</i> . Indeks validasi yang digunakan yaitu <i>Silhouette Index</i> (SI) dan <i>Davies Bouldin Index</i> (DBI)
2	Implementasi <i>Particle Swarm Optimization</i> pada <i>K-Means</i> untuk <i>Clustering Data Automatic Dependent Surveillance-Broadcast</i> , Achmad Saiful dan Joko Lianto Buliali, 2018 [9]	Rute penerbangan Surabaya-Palu dengan kode penerbangan LNI860	Komite Nasional Keselamatan Kerja (KNKT) melakukan investigasi pada kecelakaan penerbangan dari tahun 2010 hingga 2016 sebesar 212 investigasi. Diperlukan suatu sistem penerbangan yang dapat mengurangi kecelakaan	<i>K-Means</i> dan <i>Particle Swarm Optimization</i>	Pengujian dilakukan dengan membandingkan nilai <i>Davies Bouldin index</i> dengan metode <i>K-Means</i> , <i>K-Medoids</i> , dan <i>Fuzzy C-Means</i> . Pada uji coba yang dilakukan, metode yang diusulkan menjadi kelompok metode terbaik pada lima dari enam segmen yang ada serta menghasilkan nilai <i>Davies Bouldin index</i> lebih baik pada satu segmen sebesar 0.779.	Penelitian yang dilakukan merupakan perbandingan <i>K-Means</i> dan <i>Fuzzy C-Means</i> . Indeks validasi yang digunakan yaitu <i>Silhouette Index</i> (SI) dan <i>Davies Bouldin Index</i> (DBI)



Tabel 2.2. Tabel Penelitian Terkait (*State of The Art*)

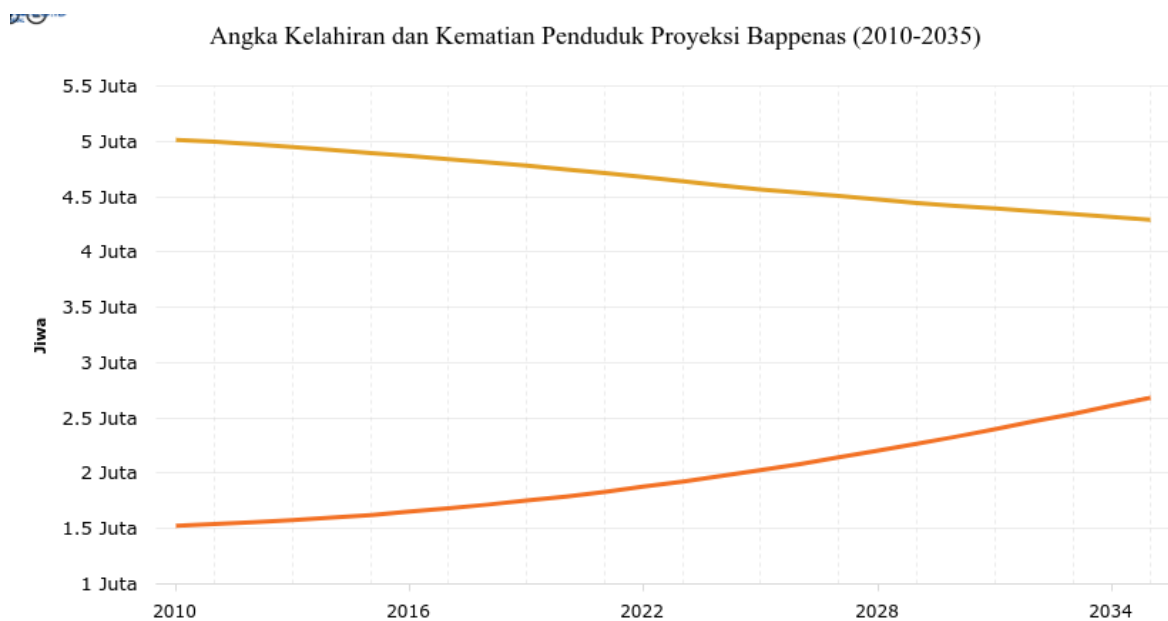
No.	Judul Penelitian, Penulis, Tahun Penelitian	Data	Masalah	Metode	Evaluasi	Perbedaan dengan Penelitian Terkait
3	Pengelompokkan Kabupaten/Kota di Pulau Jawa Berdasarkan Pembangunan Manusia Berbasis Gender Menggunakan <i>Bisecting K-Means</i> , Dila Fitriani Azuri, dkk, 2016 [8]	Pembangunan Manusia Berbasis Gender 2015 dari Badan Pusat Statistik (BPS)	Nilai IPG dan IDG pada setiap Kabupaten/Kota di pulau Jawa masih terjadi ketimpangan yang menandakan belum terjadinya pemerataan pembangunan yang dirasakan oleh seluruh rakyat	<i>Bisecting K-Means</i>	Nilai <i>Silhouette Coefficient</i> pada klaster laki-laki yaitu 0,3. Nilai <i>Silhouette Coefficient</i> pada klaster perempuan yaitu 0,26. Baik pada klaster perempuan ataupun laki-laki, klaster yang terbentuk masih tergolong lemah.	Penelitian yang dilakukan merupakan perbandingan <i>K-Means</i> dan <i>Fuzzy C-Means</i> . Indeks validasi yang digunakan yaitu <i>Silhouette Index</i> (SI) dan <i>Davies Bouldin Index</i> (DBI)
4	Segmentasi Pelanggan pada <i>Customer Relationship Management</i> di Perusahaan Ritel, Christina Deni Rumiarti dan Indra Budi, 2017 [10]	Data <i>member</i> KGVC dan data transaksi	Belum adanya segmentasi pelanggan pada CRM untuk menentukan strategi promosi yang sesuai dengan segmen pelanggan	<i>K-Means</i>	Nilai <i>average silhouette</i> pada dua <i>cluster</i> adalah 0.42. Nilai <i>Calinski Harabasz index</i> pada dua <i>cluster</i> adalah 1.31.	Penelitian yang dilakukan merupakan perbandingan <i>K-Means</i> dan <i>Fuzzy C-Means</i> . Indeks validasi yang digunakan yaitu <i>Silhouette Index</i> (SI) dan <i>Davies Bouldin Index</i> (DBI)

Tabel 2.3. Tabel Penelitian Terkait (*State of The Art*)

No.	Judul Penelitian, Penulis, Tahun Penelitian	Data	Masalah	Metode	Evaluasi	Perbedaan dengan Penelitian Terkait
5	Perbandingan <i>K-Means</i> dan <i>Fuzzy C-Means</i> untuk Pengelompokan <i>Data User Knowledge Modelling</i> , Aditya Ramadhan, dkk, 2017 [6]	<i>Data user knowledge modeling</i> yang didapatkan dari UCI <i>Machine Learning Repository</i>	Belum adanya perbandingan <i>K-Means</i> dan <i>Fuzzy C-Means</i> untuk mengelompokkan <i>Data User Knowledge Modelling</i>	<i>K-Means dan Fuzzy C-Means</i>	Nilai validasi SI dari metode <i>K-Means</i> bernilai 0.1866 dan nilai validasi PCI dari metode FCM adalah bernilai 0.2854. FCM adalah metode yang lebih baik daripada <i>K-Means</i> untuk melakukan <i>clustering</i> pada <i>Data User Knowledge Modelling</i> dikarenakan nilai validasinya mendekati 1.	Indeks validasi yang digunakan yaitu <i>Silhouette Index</i> (SI) dan <i>Davies Bouldin Index</i> (DBI)

## 2.2 Pengertian Mortalitas

Mortalitas adalah angka kematian yang terjadi pada kurun waktu dari tempat tertentu, dapat berupa penyakit maupun sebab lainnya [11]. Menurut [12] berdasarkan proyeksi penduduk Badan Perencanaan Pembangunan Nasional (Bappenas) 2010-2035, jumlah penduduk Indonesia 2010 mencapai 238, 52 juta. Adapun jumlah angka kelahiran mencapai 5 juta jiwa sementara angka kematian 1,52 juta jiwa. Sehingga jumlah penduduk Indonesia pada 2011 bertambah sekitar 3,4 juta jiwa menjadi 242 juta jiwa.



Gambar 2.1. Angka Kelahiran dan Kematian Penduduk Proyeksi Bappenas (2010-2035)

Angka kematian yang disajikan pada bab ini yaitu Angka Kematian Neonatal (AKN), Angka Kematian Bayi (AKB), Angka Kematian Balita (AKABA), Angka Kematian Anak Balita (AKBA), dan Angka Kematian Ibu (AKI) [11].

### 2.2.1 Angka Kematian Neonatal (AKN)

Angka Kematian Neonatal (AKN) adalah jumlah penduduk yang meninggal satu bulan pertama setelah kelahiran (0-28 hari) yang dinyatakan dalam 1.000 kelahiran hidup pada tahun yang sama.

### 2.2.2 Angka Kematian Bayi (AKB)

Angka Kematian Bayi (AKB) adalah jumlah penduduk yang meninggal sebelum mencapai usia 1 tahun yang dinyatakan dalam 1.000 kelahiran hidup pada tahun yang sama.

### 2.2.3 Angka Kematian Balita (AKABA)

Kematian Balita merupakan kematian yang mencakup bayi sebelum mencapai umur 1 tahun dan kematian anak balita yang terjadi dalam rentang usia 1-4 tahun. Oleh karena itu yang dimaksud dengan kematian balita adalah kematian yang terjadi pada balita sebelum usia 5 tahun (bayi + anak balita).

### 2.2.4 Angka Kematian Anak Balita (AKBA)

Angka Kematian Balita (AKABA) merupakan jumlah kematian balita 0-5 tahun per 1.000 kelahiran hidup dalam kurun waktu satu tahun. AKABA menggambarkan tingkat permasalahan kesehatan balita, tingkat pelayanan KIA/Posyandu, tingkat keberhasilan program KIA/Posyandu dan kondisi sanitasi lingkungan.

### 2.2.5 Angka Kematian Ibu (AKI)

Angka Kematian Ibu (AKI) adalah jumlah kematian wanita pada saat hamil atau kematian dalam kurun waktu 42 hari sejak terminasi kehamilan tanpa memandang lamanya kehamilan atau tempat persalinan. Kematian yang dihitung dapat terjadi karena kehamilannya, persalinannya, dan masa nifas bukan karena sebab-sebab lain seperti: kecelakaan, terjatuh, dan lain-lain.

## 2.3 Pengertian *Data Mining*

Dalam bidang informatika terdapat suatu metode yang dinilai sesuai untuk mengolah kumpulan data yang besar dan banyak, yaitu menggunakan *data mining*. *Data mining* yaitu serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual [13]. *Data mining* merupakan topik praktik yang melibatkan pembelajaran dalam praktiknya dan tidak hanya sekedar teori. *Data mining* merupakan teknik untuk menemukan dan menggambarkan pola struktural dalam data sebagai alat untuk membantu menjelaskan data tersebut dan memprediksinya [14]. Adapun pengertian *data mining* menurut para pakar, yaitu:

- a. Larose (2006) [5] menjelaskan bahwa *data mining* merupakan analisis dari *data set* yang diamati (seringnya berukuran besar) untuk menemukan hubungan tak terduga

dan untuk meringkaskan data-data dengan cara baru sehingga lebih mudah dipahami dan berguna bagi pemilik data.

- b. Witten, et Al. (2011) [15] menjelaskan bahwa *data mining* digambarkan sebagai proses menemukan pola-pola dalam data. Proses tersebut harus bersifat otomatis atau (biasanya lebih) semi-otomatis. Pola-pola yang ditemukan harus memiliki arti karena telah memberikan beberapa keuntungan, biasanya dalam hal ekonomi. Data tersebut biasanya selalu berjumlah besar.
- c. Turban, et Al. (2005) [16] menjelaskan bahwa *data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai basis data yang besar.

Banyak orang menganggap bahwa *data mining* memiliki arti yang sama dengan *Knowledge Discovery from Data* atau disingkat dengan KDD, *Knowledge discovery* sebagai sebuah proses terdiri dari tujuh langkah pelaksanaan, yaitu *data cleaning*, *data integration*, *data selection*, *data transformation*, *data mining*, *pattern evaluation*, dan *knowledge presentation* [17].

- a. *Data Cleaning* (Pembersihan Data)

Pada tahap ini data – data yang tidak relevan dan mengandung *noise* akan dihilangkan begitu juga data yang tidak konsisten, hal ini bertujuan untuk mempermudah pencarian pola.

- b. *Data Integration* (Integrasi Data)

Pada tahap ini akan dilakukan proses penggabungan data ke dalam *database* yang baru, karena pada *data mining* proses yang dilakukan tidak hanya menggunakan satu data namun beberapa data sekaligus, untuk itu perlu dilakukan data integrasi.

- c. *Data Selection* (Seleksi Data)

Seleksi data dilakukan untuk mencocokkan dimana suatu data relevan pada tugas analisis yang diambil dari *database*.

- d. *Data Transformation* (Transformasi Data)

Data yang telah dilakukan pemrosesan dengan tahap – tahap di atas kemudian dikirim dan dilakukan penggabungan dalam format tertentu yang sesuai dengan *data mining* dengan melakukan agregasi.

e. Proses *Mining*

Proses *mining* merupakan proses yang paling utama pada saat penerapan metode tertentu yang digunakan dalam *data mining* yang bertujuan untuk menemukan pola suatu data.

f. *Pattern Evaluation* (Evaluasi Pola)

Evaluasi pola dalam tahap data mining merupakan proses mengidentifikasi atau menganalisis pola – pola yang terdapat dalam data kemudian pola – pola tersebut dilakukan evaluasi untuk menilai apakah hipotesa yang didapatkan bisa tercapai atau tidak.

g. *Knowledge Presentation* (Presentasi Pengetahuan)

Proses ini adalah memformulasi keputusan atau aksi dari hasil analisis yang didapatkan dimana pemvisualisasian dan teknik representasi pengetahuan digunakan untuk menyajikan pengetahuan kepada seluruh orang.

Dalam *data mining*, terdapat sejumlah metode untuk menentukan solusi dari permasalahan yang menggunakan *data mining*. Metode-metode tersebut, yaitu:

a. Klasifikasi

Klasifikasi merupakan teknik *data mining* yang dalam prosesnya menentukan *record* dari data yang baru ke dalam salah satu atau beberapa kelas (kategori) yang terlebih dahulu didefinisikan, teknik ini juga disebut juga *supervised learning*.

b. *Clustering*

*Clustering* merupakan teknik *data mining* yang dalam prosesnya yaitu untuk mengelompokkan data berdasarkan *cluster*. Teknik ini juga disebut juga *unsupervised learning*.

c. Asosiasi

Asosiasi merupakan teknik *data mining* dengan mendeteksi suatu kumpulan atribut atau variabel yang terdapat dalam frekuensi yang sering muncul dan membentuk sejumlah kaidah dari kumpulan – kumpulan tersebut.

d. *Sequential Pattern Discovery*

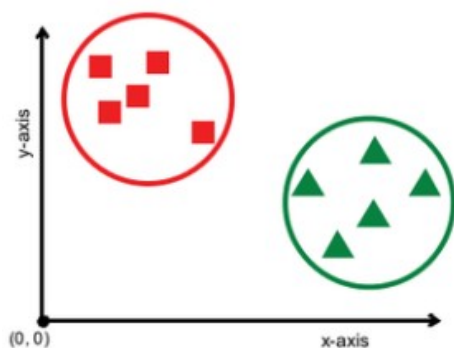
*Sequential Pattern Discovery* merupakan salah satu teknik *data mining* untuk mencari data/kejadian yang memiliki urutan dan biasanya berupa transaksi.

e. Regression

*Regresi* merupakan teknik memprediksi dari suatu variabel kontinu yang telah diberikan berdasarkan nilai dari variabel yang lain, dengan mengansumsikan sebuah model ketergantungan *linier* atau *nonlinier*. Teknik ini mirip dengan klasifikasi, namun teknik ini tidak dapat mencari pola yang dijabarkan sebagai *class* (kelas).

## 2.4 Algoritma *K-Means*

*K-Means* merupakan salah satu algoritma tertua dan paling banyak digunakan dalam klasterisasi. *K-Means* merupakan sebuah prototipe, algoritma *clustering* partisi sederhana yang mencoba untuk menemukan  $k$  buah klaster yang tidak tumpang tindih. Klaster-klaster tersebut direpresentasikan oleh masing-masing *centroid* (sebuah klaster *centroid* secara teknis berarti titik dari klaster). Tahapan pada proses *K-Means clustering* yaitu pertama,  $k$  buah inisial *centroid* dipilih, dimana  $k$  ditentukan sendiri oleh pengguna dan menandakan jumlah dari klaster. Setiap titik dalam data ditetapkan ke *centroid* terdekat, dan masing-masing dari titik-titik yang terkumpul ditetapkan ke gugus *centroid* tertentu. *Centroid* dari masing-masing klaster lalu diperbarui berdasarkan pada titik yang ditetapkan dari klaster tersebut. Proses ini diulang sampai titik tersebut tidak ada yang berubah pada klaster [18]. *K-Means* merupakan salah satu algoritma *exclusive clustering* [19] [20] / *hard clustering* [21] dan *Fuzzy C-Means* merupakan salah satu algoritma *soft clustering* [19] [21]. *Exclusive clustering/hard clustering* yaitu setiap titik pada *data set* mengacu pada satu *cluster* yang pasti. Sedangkan *overlapping clustering/soft clustering* yaitu setiap titik pada *data set* mengacu pada lebih dari satu *cluster* [22].



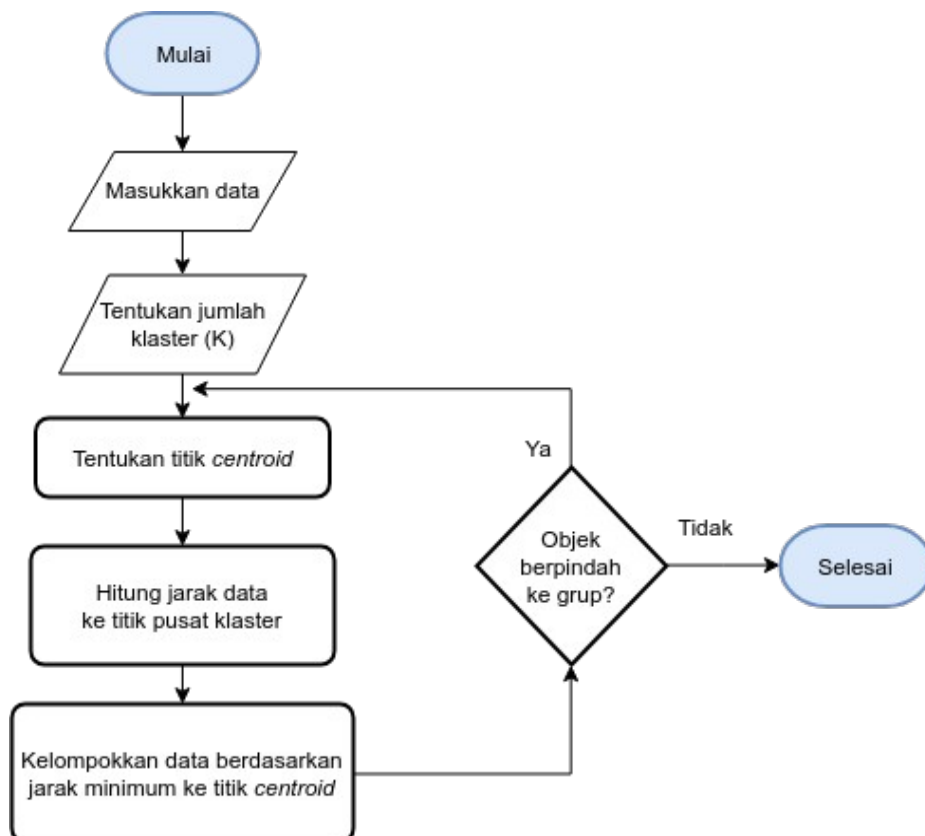
Gambar 2.2. Visualisasi Algoritma *K-Means*



Gambar 2.3. Visualisasi *Exclusive Clustering*

*K-Means* memiliki kelebihan, yaitu mudah untuk diimplementasikan dan dijalankan, waktu yang dibutuhkan untuk menjalankan pembelajaran ini relatif cepat. mudah untuk

diadaptasi, umum digunakan dalam bidang *data mining* [23], kemampuan mengelompokkan data dalam jumlah yang cukup besar dengan waktu komputasi yang relatif cepat dan efisien [24] [25], dan waktu komputasi pada *K-Means* lebih rendah dibandingkan dengan *Fuzzy C-Means* [21]. Disamping itu, *K-Means* juga memiliki kelemahan, yaitu pemilihan titik secara *random* dapat mempengaruhi hasil *clustering*, dapat terjebak pada besar kecilnya dimensi (*curse of dimensionality*), hasil klaster yang berubah-ubah, dan jumlah titik sampel yang diteliti dapat mempengaruhi pencarian dan perhitungan titik terdekat. Algoritma ini bekerja dengan cara membagi data dalam sejumlah klaster untuk dianalisis faktor kesamaan maupun ketidaksamaan yang melekat pada kumpulan data tersebut. Menurut [26] ada dua kunci keberhasilan penerapan *K-Means*, yaitu jumlah  $k$  buah klaster untuk partisi dan jarak metrik. Tanpa mengetahui jumlah  $k$  buah klaster, tak ada jaminan pemilihan  $k$  tersebut optimal. Adapun tahapan pada *K-Means* yaitu sebagai berikut:



Gambar 2.4. Diagram Alir Algoritma *K-Means*

Adapun penjelasan dari diagram alir algoritma *K-Means* sebagai berikut:

1. Masukkan *data set* yang akan diklaster.
2. Tentukan jumlah klaster.



Penentuan jumlah kluster dilakukan secara bebas dan acak berapa jumlah kluster yang akan dibentuk. Sebagai contoh kita akan menggunakan 3 buah kluster dan memberi simbol menggunakan C1, C2, dan C3. berupa C1, C2, C3.

3. Gunakan sembarang data pada *data set* sebanyak jumlah kluster yang telah ditentukan secara acak sebagai titik *centroid* (titik pusat kluster).
4. Hitung jarak antar data dengan titik pusat kluster menggunakan rumus Euclidean.

$$d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{kj})^2} \quad (2.1)$$

Keterangan:

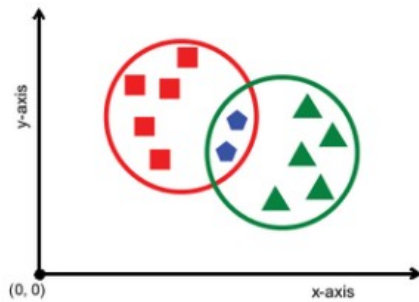
- $d_{ik}$  = Jarak antara data ke-i dengan titik pusat kluster ke-k
- $m$  = Jumlah atribut
- $x_j$  = Data ke-i
- $c_k$  = Data pusat kluster ke-k

5. Hitung kembali jarak titik pusat kluster dengan keanggotaan kluster yang baru.
6. Jarak terpendek merupakan anggota suatu kluster.
7. Lakukan iterasi sebanyak  $k$  kali sampai nilai kluster tidak berubah.

## 2.5 Algoritma *Fuzzy C-Means*

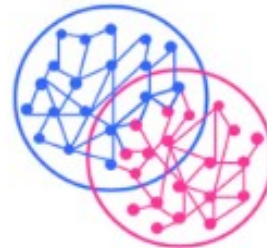
Salah satu algoritma *clustering* lainnya yaitu *Fuzzy C-Means*. *Fuzzy C-Means* adalah teknik *clustering* data dimana setiap titik data berhubungan terhadap suatu kluster pada beberapa derajat dimana dikhususkan oleh derajat keanggotaan [21]. *Fuzzy C-Means clustering* adalah salah satu model pengelompokkan fuzzy dengan tingkat keanggotaan berbeda antara 0 dan 1 sehingga data dapat menjadi anggota dari kluster khusus. Probabilitas dari eksistensi sebuah data ke sebuah kluster tertentu dapat dilihat dari jarak keanggotaan atau seperti kluster terdekat ke nilai dari data. Konsep dasar dari *Fuzzy C-Means*, yaitu pertama tentukan titik pusat dari kluster. Setiap data memiliki jarak keanggotaan untuk setiap kluster. Dengan memperbaiki titik pusat kluster dan nilai keanggotaan setiap data secara berulang, maka dapat ditemukan bahwa titik pusat kluster akan berada di lokasi yang tepat. Fungsi obyektif yang menggambarkan jarak dari lokasi data terhadap titik pusat kluster dapat diminimalkan dengan melakukan *looping* pemberian bobot tingkat kesamaan titik data dari himpunan lepas [7]. *Fuzzy C-Means* adalah salah satu metode *soft clustering/overlapping clustering* terkenal dan *K-Means* merupakan salah satu metode *hard clustering/exclusive*

*clustering* terkenal [21]. *Exclusive clustering/hard clustering* yaitu setiap titik pada data set mengacu pada satu *cluster* yang pasti. Sedangkan *overlapping clustering/soft clustering* yaitu setiap titik pada *data set* mengacu pada lebih dari satu *cluster* [20].



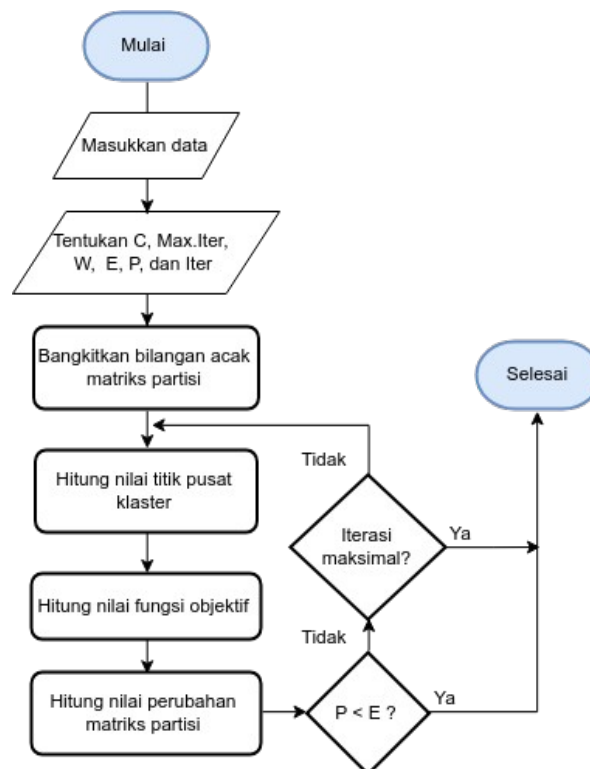
Gambar 2.5. Visualisasi Algoritma *Fuzzy C-Means*

Overlapping Clustering



Gambar 2.6. Visualisasi *Overlapping Clustering*

Kelebihan dari *Fuzzy C-Means* diantaranya: sederhana, mudah diimplementasikan, memiliki kemampuan untuk mengelompokkan data yang besar, dan keluaran yang dihasilkan lebih efektif. Kelemahan dari *Fuzzy C-Means* yaitu *Fuzzy C-Means* tidak dapat melakukan penempatan objek tepat pada satu partisi, karena objek terletak diantara dua atau lebih partisi lainnya. Adapun tahapan pada *Fuzzy C-Means* adalah sebagai berikut.



Gambar 2.7. Diagram Alir *Fuzzy C-Means*

Adapun penjelasan dari diagram alir algoritma *Fuzzy C-Means* yaitu sebagai berikut:

1. Masukkan data yang akan diklaster berupa matriks  $n \times m$  dimana  $n$  adalah banyaknya sampel data dan  $m$  adalah banyaknya atribut dari data.
2. Inisialisasi parameter awal dengan menentukan:
  - a. Jumlah klaster =  $c$
  - b. Pangkat pembobot =  $w$
  - c. Jumlah maksimum iterasi =  $\text{MaxIter}$
  - d. Error terkecil yang diharapkan =  $\epsilon$
  - e. Fungsi objektif awal =  $P_0$ , dengan nilai 0
  - f. Iterasi awal =  $t$ , dengan nilai 1
  - g. Bangkitkan bilangan acak ( $U_{ik}$ ), dengan  $i = 1, 2, \dots, n$ ;  $k = 1, 2, \dots, c$ ; sebagai elemen-elemen matriks partisi awal  $U$ .
  - h. Hitung pusat klaster ke- $k$ :  $V_{kj}$

$$V_{kj} = \frac{\sum_{i=1}^n ((U_{ik})^w * X_{ij})}{\sum_{i=1}^n (U_{ik})^w} \quad (2.2)$$

dengan  $k = 1, 2, \dots, c$ ; dan  $j = 1, 2, \dots, m$

Keterangan:

$V_{kj}$  = Pusat klaster ke- $k$  dan atribut ke- $j$

$U_{ik}$  = Derajat keanggotaan sampel data ke- $i$  dan klaster ke- $k$

$X_{ij}$  = Data sampel ke- $i$  dan atribut ke- $k$

$i$  = Indeks sampel data

$k$  = Klaster ke- $k$

$w$  = Pangkat pembobot

$n$  = Banyaknya sampel data

3. Hitung fungsi objektif pada iterasi ke- $t$ :

$$U_{ik} = \sum_{i=1}^n \sum_{k=1}^c ([\sum_{j=1}^m (X_{ij} - V_{kj})^2] (U_{ik})^w) \quad (2.3)$$

Keterangan:

$V_{kj}$  = Pusat klaster ke= $k$  dan atribut ke- $j$

$U_{ik}$  = Derajat keanggotaan sampel data ke- $i$  dan klaster ke- $k$

$X_{ij}$  = Data sampel ke- $i$  dan atribut ke- $k$

$i$  = Indeks sampel data

$k$  = Klaster ke- $k$

$w$  = Pangkat pembobot

$n$  = Banyaknya sampel dari data

$m$  = Banyaknya atribut dari data

4. Hitung perubahan matriks partisi:

$$U_{ik} = \frac{\left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{\frac{-1}{w-1}}}{\sum_{k=1}^c \left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{\frac{-1}{w-1}}} \quad (2.4)$$

dengan  $i = 1, 2, \dots, n; k = 1, 2, \dots, c$

Keterangan:

$V_{kj}$  = Pusat klaster ke= $k$  dan atribut ke- $j$

$U_{ik}$  = Derajat keanggotaan sampel data ke- $i$  dan klaster ke- $k$

$X_{ij}$  = Data sampel ke- $i$  dan atribut ke- $k$

$i$  = Indeks sampel data

$j$  = Atribut ke- $j$

$k$  = Klaster ke- $k$

$w$  = Pangkat pembobot

$m$  = Banyaknya atribut dari data

5. Cek kondisi berhenti:

- Jika:  $(|P_t - P_{t-1}| < \varepsilon)$  atau  $(t > MaxIter)$  maka berhenti;
- Jika tidak:  $t = t + 1$ , ulangi langkah ke-4.

## 2.6 Pengumpulan Sampel Data

Populasi merupakan keseluruhan dari kumpulan elemen yang memiliki sejumlah karakteristik umum yang terdiri dari bidang-bidang untuk diteliti. Atau populasi adalah keseluruhan kelompok dari orang-orang, peristiwa, atau barang-barang yang diminati oleh peneliti untuk diteliti [27]. Sedangkan sampel merupakan suatu sub kelompok dari populasi yang dipilih untuk digunakan dalam penelitian [27]. Pengambilan sampel bertujuan agar sampel yang diambil dapat mewakili populasi sehingga diperoleh informasi yang cukup untuk mengestimasi populasinya. Untuk menentukan jumlah sampel yang sesuai untuk penelitian diperlukan suatu upaya untuk mendapatkan sampel yang mewakili dan menggambarkan populasi (teknik pengambilan sampel).

Teknik pengambilan sampel diantaranya memiliki keuntungan [27], yaitu:

### 1. Dapat Menghemat Biaya

Jika estimasi yang dapat diandalkan dari suatu pengukuran populasi dapat diperoleh dari bagian tertentu dari suatu populasi, maka hal ini dapat dilakukan dengan biaya yang cukup rendah dibandingkan dengan melakukan sensus secara menyeluruh.

### 2. Dapat Menghemat Waktu

Untuk beberapa alasan, data dapat dikumpulkan dan dirangkum lebih cepat dengan sampel dibandingkan dengan mengumpulkan seluruh populasi. Hal ini penting ketika informasi sangat dibutuhkan segera.

### 3. Dapat Meningkatkan yang Tinggi

Ketepatan yang lebih tinggi untuk seluruh populasi dapat diperoleh melalui jumlah sampel yang terbatas. Setiap anggota populasi dapat diperlakukan dengan pertanyaan yang tidak mendalam (*superficial*) atau suatu bagian terbatas dari populasi (suatu sampel) dapat didekati lebih mendalam. Pilihan kedua (lebih rinci dari suatu sampel) adalah pengukuran yang sering kali lebih berguna dan tepat (akurat).

Secara umum teknik pengambilan sampel dikelompokkan menjadi dua teknik, yaitu *non-probability sampling* dan *probability sampling* [27].

## 1. Non-Probability Sampling

Dalam *non-probability sampling*, peneliti dapat sesukanya atau secara sadar memutuskan elemen-elemen apa yang akan masuk ke dalam sampel. Artinya kemungkinan atau peluang seseorang atau benda untuk terpilih menjadi anggota sampel tidak diketahui. Hal ini dikarenakan pada teknik ini terlalu percaya pada pendapat pribadi peneliti daripada kesempatan untuk memilih elemen-elemen. Dalam teknik ini juga kurang mempertimbangkan penilaian secara objektif dari sampel yang diperoleh secara tepat. Adapun yang tergolong dalam teknik *non-probability sampling*, meliputi: *convenience sampling*, *judgemental sampling*, *quota sampling*, dan *snowball sampling*.

## 2. Probability Sampling

Pengambilan sampel dengan cara ini dilakukan secara *random* atau acak. Periset perlu mengetahui teknik-teknik dimana dia dapat memilih suatu sampel untuk setiap unit dalam populasi memiliki kesempatan yang sama untuk dipilih. Adapun yang tergolong dalam *probability sampling*, meliputi: *simple random sampling*, *systematic random sampling*, *stratified sampling*, dan *cluster sampling*.

Untuk menentukan jumlah data sampel pada penelitian, peneliti harus mengetahui tingkat kepercayaan (*confidence level* / 95% s/d 99%), tingkat signifikansi proporsi populasi ( $p$ ), besar *margin of error* yang dapat diterima ( $e$ ), dan banyaknya populasi ( $N$ ). Adapun rumus untuk menentukan jumlah data sampel pada penelitian dapat menggunakan rumus Cochran [28] dibawah ini.

$$\underline{n}_0 = \frac{\underline{t}^2 * \underline{s}^2}{\underline{d}^2} \quad (2.5)$$

$$\underline{n} = \frac{\underline{n}_0}{1 + \frac{\underline{n}_0}{Population}} \quad (2.6)$$

Keterangan:

$s$  = Perkiraan standar deviasi pada populasi

$t$  = Nilai seleksi tingkat alfa

$d$  = Rata-rata toleransi kesalahan yang dapat diterima

Penelitian ini merupakan penelitian yang jumlah populasinya dapat diukur (*finite population*). Pada penelitian ini diperoleh data sampel sebanyak 550 buah data puskesmas dari 6542 kecamatan di Indonesia dengan menggunakan rumus *Cochran*.

## 2.7 Preprocessing Data

*Preprocessing* data merupakan tahapan dalam normalisasi pada data. Menurut [29] normalisasi adalah sebuah proses *preprocessing* yang dilakukan untuk membentuk ulang data-data yang ada dengan rentang nol (0) hingga satu (1). Hal ini dilakukan dengan maksud agar data yang diproses memiliki rentang yang lebih sempit, namun masih tetap mengakomodir perbedaan dan nilai dari data tersebut.

$$\sum_{i=1}^n f(x) = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (2.7)$$

Keterangan:

$f(x)$  = Fungsi  $f(x)$  dengan variabel  $x$  untuk normalisasi

$x_i$  = Nilai  $x$  dengan iterasi dari  $i-1..n$

$x_{min}$  = Nilai  $x$  minimum

$x_{max}$  = Nilai  $x$  maksimum

Dengan demikian maka data yang memiliki nilai diantara maksimum dan minimum akan memiliki nilai lebih besar dari 0 ( $X > 0$ ) dan lebih kecil dari 1 ( $X < 1$ ) atau dapat dinyatakan dengan  $\{X | 0 < X < 1, X \in R\}$ .

## 2.8 Bahasa Pemrograman R

Pada pembahasan ini akan dijelaskan mengenai pengertian bahasa pemrograman R, perbedaan bahasa pemrograman R dengan bahasa Python, aplikasi-aplikasi yang dapat menjalankan bahasa pemrograman R, contoh sintaks kode, dan contoh pustaka yang digunakan pada bahasa pemrograman R.

### 2.8.1 Pengertian Bahasa Pemrograman R

R merupakan bahasa pemrograman yang penggunaannya luas, yaitu untuk manipulasi data statistika, analisis, dan grafik yang terinspirasi dan kompatibel dengan bahasa S yang dikembangkan oleh John Chambers dan rekan-rekan di AT&T Bell Laboratories. Bahasa R diciptakan oleh Ross Ihaka dan Robert Gentleman dari departemen statistika di Universitas Auckland, Selandia Baru [30]. R dilisensikan secara *free software* atau *software* bebas sehingga siapapun dapat mempelajari, mengubah, mendistribusikan, dan menggunakannya dalam tujuan apapun bahkan untuk tujuan komersil sekali *pun*. Dengan dilisensikannya R menjadi *free software*, maka R berkembang menjadi bahasa yang kaya dengan paket-paket dan dapat digunakan secara mudah oleh para pelajar dan ahli ekonomi di setiap sistem operasi.

Ada beberapa alasan mengapa R dipilih dalam penelitian data statistika, analisis, dan grafik [30]. Diantaranya adalah sebagai berikut:

1. Penerapan oleh ranah publik secara luas dari bahasa statistika S; R/S adalah standar *de facto* antar kalangan ahli statistik profesional.
2. Dapat dibandingkan, dan seringnya lebih unggul, bahkan lebih dominan untuk produk komersial dalam banyak hal.
3. Tersedia untuk *platform* Windows, Mac OS, dan GNU/Linux.
4. Dalam hal penambahan operasi statistika, bahasa R adalah bahasa pemrograman umum, sehingga kita dapat mengotomatisasi analisis kita sendiri dan membuat fungsi-fungsi baru.
5. Struktur pemrograman fungsional dan termasuk pemrograman berorientasi objek.
6. *Data set* yang sudah dibuat tersimpan diantara sesi-sesi, sehingga tidak perlu memuat ulang.
7. Bahasa yang terbuka sehingga mudah mendapatkan dukungan dan bantuan dari pengguna komunitas, dan banyak lagi fungsi-fungsi baru dikontribusikan oleh para pengguna yang secara umum merupakan ahli statistika menonjol.

Terdapat beberapa aplikasi IDE untuk menjalankan bahasa R, diantaranya yaitu RGui, R.app, Vim, Emacs, dan RStudio. RStudio merupakan salah satu IDE pihak ketiga yang menyediakan antarmuka yang konsisten untuk segala sistem operasi. RStudio direkomendasikan oleh beberapa ahli statistika karena menyediakan lingkungan *editing* yang



lebih kaya daripada RGui sehingga pengguna dapat menjalankan bahasa R dengan lebih mudah dan menyenangkan. Disamping itu, antarmuka pada RStudio mirip untuk sistem operasi Windows, Mac OS, dan GNU/Linux. Bahasa pemrograman R dapat diunduh di situs <http://cran.r-project.org>.

Menurut [30] terdapat sejumlah perbedaan antara bahasa R dan bahasa Python, diantaranya yaitu sebagai berikut:

Tabel 2.4. Perbedaan Bahasa Pemrograman R dan Python

<b>Pendahuluan</b>	
<ul style="list-style-type: none"> <li>Bahasa R merupakan implementasi dari bahasa S (Bell Labs);</li> <li>Desain R dan evolusinya dipegang oleh R-Core group dan R Foundation; dan</li> <li>Lingkungan <i>software</i> R ditulis menggunakan bahasa R, Fortran, dan R.</li> </ul>	<ul style="list-style-type: none"> <li>Python terinspirasi oleh bahasa C, Modula-3, dan terutama ABC;</li> <li>Bahasa Python diambil dari serial komedi bernama “<i>Monty Python’s Flying Circus</i>”;</li> <li>Python Software Foundation (PSF) mengembangkan bahasa Python.</li> </ul>
<b>Tujuan</b>	
Bahasa R fokus pada analisis data ramah pengguna ( <i>user friendly</i> ) statistika dan model-model grafis yang lebih baik.	Python menekankan pada produktivitas dan pembacaan kode.
<b>Penggunaan</b>	
Bahasa R utamanya digunakan oleh kalangan akademisi dan peneliti. Bagaimana <i>pun</i> bahasa R mengalami perkembangan yang pesat pada pasar <i>enterprise</i> .  “ <i>The closer you are to statistics, research, and data science, the more you might prefer R</i> ”	Python digunakan oleh para <i>programmer</i> yang ingin menjadi data analisis atau menerapkan teknik-teknik statistikal, dan oleh <i>developer</i> yang beralih ke <i>data science</i> .  “ <i>The closer you are to working in an engineering environment, the more you might prefer Python</i> ”
<b>Kegunaan (Usability)</b>	
<ul style="list-style-type: none"> <li>Pada bahasa R, model statistika dapat ditulis hanya dengan beberapa baris.</li> <li>Terdapat lembar <i>style sheet</i> pada R</li> </ul>	<ul style="list-style-type: none"> <li><i>Coding</i> dan <i>debugging</i> lebih mudah dilakukan pada Python, dikarenakan sintaks yang bagus.</li> </ul>

<p>tetapi tidak setiap orang memilikinya.</p> <ul style="list-style-type: none"> <li>Beberapa fungsionalitas dapat ditulis dalam beragam cara di R.</li> </ul>	<ul style="list-style-type: none"> <li>Indentasi dari setiap kode berdampak pada maksud kode itu sendiri.</li> <li>Beberapa fungsi selalu ditulis dengan cara yang sama di Python.</li> </ul>
<b>Fleksibilitas</b>	
R mudah digunakan untuk formula-formula yang kompleks, Semua jenis model dan tes statistika sudah tersedia dan mudah digunakan	Python bersifat fleksibel. <i>Developer</i> juga dapat menggunakannya untuk <i>scripting</i> sebuah situs <i>web</i> atau aplikasi-aplikasi lainnya.
<b>Kemudahan Dipelajari</b>	
<ul style="list-style-type: none"> <li>R terlalu sulit untuk dimulai. Namun jika sudah memahami dasar-dasarnya, maka akan mudah untuk mempelajarinya lebih dalam (<i>advance</i>).</li> <li>R tidak sulit untuk <i>programmer</i> yang berpengalaman.</li> </ul>	<ul style="list-style-type: none"> <li>Python berfokus pada kemudahan pembacaan dan kesederhanaan membuat pembelajaran Python menjadi relatif mudah dan bertahap.</li> <li>Python merupakan bahasa yang bagus untuk memulai menjadi <i>programmer</i>.</li> </ul>
<b>Kelebihan</b>	
<ul style="list-style-type: none"> <li>R sesuai untuk analisis statistika</li> <li>R dibangun dengan <i>command line</i>, namun banyak orang yang bekerja pada lingkungan RStudio atau R <i>commander</i> lainnya yang sudah termasuk <i>data editor</i>, dukungan <i>debugging</i>, dan suatu jendela tampilan untuk penggunaan grafik/gambar.</li> <li>Bahasa R merupakan <i>tool</i> terbaik untuk visualisasi data. Data visualisasi dapat dengan mudah dipahami dibandingkan dengan nomor-nomor. R dan visualisasi berjalan beriringan dan mencakup beberapa paket yang</li> </ul>	<ul style="list-style-type: none"> <li>Sejak Python menjadi bahasa pemrograman umum, maka mempelajarinya memberikan kemampuan untuk data analisis. Python banyak digunakan untuk pengembangan <i>web</i>, <i>automation testing</i>, dan ETL.</li> <li>Para <i>programmer</i> berfikir bahwa Python bertepatan dengan cara <i>programmer</i> berpikir lebih dari R, dan karena itu menerjemahkan ke bahasa lain menjadi lebih mudah. Keunikan pada R terletak pada statistika, sehingga memiliki desain yang unik.</li> </ul>

<p>sesuai dengannya.</p> <ul style="list-style-type: none"> <li>• Bahasa R menghasilkan hasil terbaik dari visualisasi yang dapat digunakan pada <i>research paper</i>. Hasil-hasil visualisai dapat di-<i>trace</i> ketika dibutuhkan dan dapat didaur ulang untuk membuat hasil struktur yang berbeda.</li> <li>• Bahasa R menyediakan dukungan komunitas luas dengan 1000 <i>developer</i> dan para ilmuwan data yang tersebar di seluruh dunia. Komunitas memasukkan paket-paket dalam beragam kebutuhan, seperti: keuangan, <i>machine learning</i>, teknologi <i>web</i>, dan farmasi.</li> </ul>	<p>Jika ingin mempelajari bahasa tujuan umum lainnya, bahasa Python adalah pilihan yang tepat untuk dipelajari.</p> <ul style="list-style-type: none"> <li>• Sebagian besar data analisis membersihkan data pada tahapan sebelumnya. Maka membersihkan data dapat akan menjadi sangat menyenangkan dengan layanan bahasa yang lengkap seperti Python karena dapat menambahkan fungsi untuk memisahkan tiap data. Jika tiap fungsi membutuhkan penyimpanan lokal atau akses web, maka akan cukup mudah untuk menyertakannya dengan Python.</li> <li>• Python berkembang seiring dengan waktu. Kode baru terus diperkenalkan dan menghapus kode lama, yang membuat Python menjadi bahasa yang hidup. Hal ini mengarah pada solusi dan kode sumber terbuka.</li> <li>• Python bergerak lebih cepat dibandingkan dengan R. Hal ini karena R dikembangkan untuk memusatkan pada kenyamanan ahli statistika, bukan pada kenyamanan komputer.</li> <li>• Python telah mendapatkan popularitas yang luas karena sintaksnya sangat jelas untuk dipahami. Ilmuwan data mendapatkan pengetahuan ahli dan pemrograman master dengan Python</li> </ul>
---	---

	untuk mendapatkan <i>output</i> yang diinginkan dengan sejumlah langkah yang telah ditentukan.
Kekurangan	
<ul style="list-style-type: none"> <li>• Untuk kalangan pengguna tanpa pengetahuan pemrograman, bahasa R akan sedikit lebih sulit untuk mulai dipelajari.</li> <li>• Mendapatkan solusi yang tepat dengan bahasa pemrograman R dapat dianggap lambat jika kode ditulis dengan buruk. Untuk mengatasi kelemahan ini, maka diwajibkan untuk menyertakan <i>library</i> untuk memperoleh hasil yang tepat.</li> </ul>	<ul style="list-style-type: none"> <li>• Performa Python lebih lambat dibandingkan dengan bahasa pemrograman lainnya sebagai bahasa <i>interpreted</i>.</li> <li>• Python mewajibkan pengujian yang teliti ketika <i>error</i> muncul.</li> <li>• Bahasa Python masih dianggap lemah pada <i>platform</i> komputasi <i>mobile</i> karena beberapa aplikasi yang dibuat dengan Python sebagai bahasa inti.</li> </ul>

### 2.8.2 Library Bahasa Pemrograman R

Dalam proses pengelompokkan data menggunakan bahasa pemrograman R, dibutuhkan *library* (pustaka) khusus untuk menampilkan hasil dari data yang telah dikelompokkan. Pada penelitian ini akan menggunakan beberapa *library* diantaranya, yaitu:

#### a. factoextra

*Library* ini menawarkan beberapa fungsi-fungsi yang mudah digunakan untuk mengekstrak dan menggambarkan analisis *multivariate data*, termasuk fungsi 'PCA' (*Principal Component Analysis*), 'CA' (*Correspondence Analysis*), 'MCA' (*Multiple Correspondence Analysis*), 'FAMD' (*Factor Analysis of Mixed Data*), 'MFA' (*Multiple Factor Analysis*) and 'HMFA' (*Hierarchical Multiple Factor Analysis*) dari paket-paket R yang berbeda [31]. Salah satu fungsi utama pada paket *factoextra* untuk visualisasi dan analisis *clustering*, yaitu *fviz\_cluster* yang berfungsi untuk menggambarkan hasil *clustering*.

#### b. ggplot2

*Library* ini merupakan salah satu paket R utama untuk grafis dan visualisasi data [31].

### c. NbClust

*Library* ini digunakan untuk menentukan jumlah klaster pada *data set* dan juga menawarkan skema klaster terbaik dari hasil yang berbeda pada *user* [32].

Untuk memasang (*install*) *library* pada RStudio menggunakan bahasa pemrograman R dapat mengetikkan perintah `install.packages("namalibrary")`. Sebagai contoh yaitu memasang *library factoextra*, maka perintah untuk memasang *library factoextra* yaitu `install.packages("factoextra")`.

### 2.8.3 Clustering pada R

Berikut ini adalah kode pemrograman (*programming code*) untuk *clustering K-Means* dan *Fuzzy C-Means* pada bahasa pemrograman R.

#### a. K-Means

##### **Input Code of K-Means**

```
kmeans(x, centers, iter.max=10, nstart=1, algorithm=c("Hartigan-
Wong", "Lloyd", "Forgy", "MacQueen"), trace=FALSE)
```

##### **Keterangan:**

*x* = Matriks numerik pada data, atau objek yang dapat digerakkan ke beberapa matriks

*centers* = Jumlah dari klaster, *k*, atau kumpulan dari titik pusat klaster

*iter.max* = Jumlah iterasi maksimum yang diizinkan

*nstart* = Banyaknya jumlah kumpulan titik pusat yang harus dipilih

*algorithm* = Algoritma yang digunakan pada *K-Means*

*object* = Objek R dari *class "kmeans"*

*method* = Metode *K-Means* yang digunakan

*trace* = Angka *logical* atau *integer*

##### **Value of K-Means**

```
"cluster", "centers", "totss", "withinss", "tot.withinss",
"betweeness", "size,iter", "ifault"
```

##### **Keterangan:**

*cluster* = Vektor dari integer (dari 1:k) menunjukkan klaster tempat setiap titik

dialokasikan

*centers* = Matriks dari titik pusat klaster  
*totss* = Jumlah total dari *sum of square*  
*withinss* = Vektor *within-cluster* dari *sum of squares*, satu komponen per klaster  
*tot.withinss* = Jumlah total *within-cluster* dari *sum of squares*  
*betweenss* = *Between-cluster* dari *sum of squares*  
*size* = Jumlah titik pada setiap klaster  
*iter* = Jumlah iterasi  
*ifault* Integer: indikator kemungkinan masalah dari suatu algoritma

#### b. *Fuzzy C-Means*

<i>Input Code of Fuzzy C-Means</i>			
fanny(x,	k,	diss=inherits(z,	"dist"), memb.exp=2,
metric=c("euclidean",	"manhattan",	"SqEuclidean"),	stand=FALSE,
iniMem.p=NULL,	cluster.only=FALSE,	keep.diss=!diss&&!	
cluster.only&&n<100,	keep.data=!diss && !cluster.only,	maxit=500,	
tol=1e-15,	trace.lev=0)		

#### Keterangan:

*x* = Matriks numerik pada data, atau *data frame*, atau matriks ketidaksamaan, bergantung pada nilai argumen *diss*  
*k* = Jumlah dari klaster, *k*, atau kumpulan dari titik pusat klaster  
*diss* = *Logical flag*; jika TRUE (*default* untuk objek ketidaksamaan), maka *x* dianggap menjadi matriks ketidaksamaan  
*memb.exp* = Jumlah *r* lebih besar dari 1 yang menentukan eksponen keanggotaan yang digunakan dalam kriteria kecocokan  
*metric* = Karakter *string* mengkhususkan metrik digunakan untuk dihitung ketidaksamaannya diantara pengujian  
*stand* = *Logical*, jika benar, pengukuran pada *x* distandarisasi sebelum dihitung ketidaksamaannya  
*iniMem.p* = Jumlah matriks *n x k* atau NULL  
*cluster.only* = *Logical*; jika benar, tidak ada informasi *Silhouette* yang akan dihitung atau dikembalikan  
*keep.diss*, = Indikator *logical* jika ketidaksamaan dan/atau *input* data *x* yang harus

*keep.data*           dijaga pada suatu hasil  
*maxit, tol*       = Jumlah maksimum iterasi dan toleransi *default* untuk konvergensi  
*trace.lev*       = *Integer* menentukan level jejak untuk mencetak diagnostik selama algoritma C-internal

### c. Validasi *Davies Bouldin Index*

<b><i>Input Code of DBI</i></b>			
NbClust(data=NULL,       distance="euclidean",       min.nc=3,       max.nc=3, method="average", index="db")			

#### **Keterangan:**

*data*           = *Data set* atau matriks  
*diss*           = Matriks ketidaksamaan yang akan digunakan  
*distance*       = Pengukuran jarak yang akan digunakan untuk menghitung matriks ketidaksamaan. Dapat dipilih salah satu dari berikut ini, "euclidean", "maximum", "manhattan", "canberra", "binary", "minkowski" or "NULL"  
*min.nc*       = Jumlah kluster minimum, antara 1 dan (jumlah objek -1)  
*max.nc*       = Jumlah kluster maksimum antara 2 dan (jumlah objek – 1), nilai lebih besar atau sama dengan nilai *min.nc*  
*method*       = Metode kluster analisis yang akan digunakan. Dapat dipilih salah satu dari berikut ini, "ward.D", "ward.D2", "single", "complete", "average", "mcquitty", "median", "centroid", "kmeans"  
*index*       = Index yang akan dihitung. Dapat dipilih salah satu dari berikut ini, "kl", "ch", "hartigan", "ccc", "scott", "marriot", "trcovw", "tracew", "friedman", "rubin", "cindex", "db", "silhouette", "duda", "pseudot2", "beale", "ratkowsky", "ball", "ptbiseria", "gap", "frey", "mcclain", "gamma", "gplus", "tau", "dunn", "hubert", "sdindex", "dindex", "sdbw", "all" (seluruh indeks kecuali GAP, Gamma, Gplus dan Tau), "alllong" (seluruh indeks, termasuk Gap, Gamma, Gplus dan Tau)  
*alphaBeale*   = Nilai signifikan untuk indeks Beale

<i>Value of DBI</i>
"All.index", "Best.nc", "Best.partition"

**Keterangan:**

*All.index* = Nilai validasi indeks  
*Best.nc* = Jumlah klaster terbaik  
*Best.partition* = Partisi data terbaik

**d. Validasi *Silhouette Index***

<i>Input Code of SI</i>
<pre>NbClust(data=NULL, distance="euclidean", min.nc=3, max.nc=3, method="average", index="silhouette")</pre>

**Keterangan:**

*data* = *Data set* atau matriks  
*diss* = Matriks ketidaksamaan yang akan digunakan  
*distance* = Pengukuran jarak yang akan digunakan untuk menghitung matriks ketidaksamaan. Dapat dipilih salah satu dari berikut ini, "euclidean", "maximum", "manhattan", "canberra", "binary", "minkowski" or "NULL"  
*min.nc* = Jumlah klaster minimum, antara 1 dan (jumlah objek -1)  
*max.nc* = Jumlah klaster maksimum antara 2 dan (jumlah objek - 1), nilai lebih besar atau sama dengan nilai *min.nc*  
*method* = Metode klaster analisis yang akan digunakan. Dapat dipilih salah satu dari berikut ini, "ward.D", "ward.D2", "single", "complete", "average", "mcquitty", "median", "centroid", "kmeans"  
*index* = Index yang akan dihitung. Dapat dipilih salah satu dari berikut ini, "kl", "ch", "hartigan", "ccc", "scott", "marriot", "trcovw", "tracew", "friedman", "rubin", "cindex", "db", "silhouette", "duda", "pseudot2", "beale", "ratkowsky", "ball", "ptbserial", "gap", "frey", "mcclain", "gamma", "gplus", "tau", "dunn", "hubert", "sdindex", "dindex", "sdbw", "all" (seluruh indeks kecuali GAP, Gamma, Gplus dan Tau), "alllong" (seluruh indeks, termasuk Gap, Gamma, Gplus dan Tau)



*alphaBeale* = Nilai signifikan untuk indeks Beale

<i>Value of SI</i>
"All.index", "Best.nc", "Best.partition"

**Keterangan:**

*All.index* = Nilai validasi indeks

*Best.nc* = Jumlah klaster terbaik

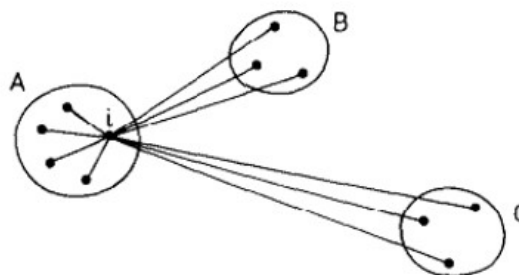
*Best.partition* = Partisi data terbaik

## 2.9 Uji Validasi Algoritma

Untuk mengetahui suatu klaster valid, maka akan dilakukan uji validitas algoritma. Pada penelitian ini akan menggunakan indeks validasi *Silhouette Index* (SI) dan *Davies Bouldin Index* (DBI). Validasi SI dan DBI digunakan untuk memperoleh jumlah *cluster* dan nilai validasi yang optimal.

### 1. *Silhouette* / *Silhouette Index*

*Silhouette* adalah suatu alat untuk mengakses validasi dari *clustering*. *Silhouette* dibangun untuk menyeleksi jumlah klaster optimal dengan rasio data skala (dalam kasus *Euclidean distance*) yang sesuai untuk klaster yang terpisah secara jelas. *Silhouette* biasa digunakan untuk mendukung evaluasi hasil *clustering* dengan nilai maksimum pada *silhouette* [33].



Gambar 2.8. Komputasi  $s(i)$  untuk Tiap Objek, Dimana  $i$  Berkaitan Terhadap *Cluster A*

### Kasus #1 Pertimbangan Ketidaksamaan

Dari gambar 2.3 diatas dijelaskan bahwa untuk memperoleh objek  $i$  pada *data set*, dan menetapkannya pada klaster A, berikut adalah langkah-langkahnya [33].

$s(i)$  = Kasus ketidaksamaan

- $i$  = Objek  $i$  berkaitan dengan klaster A  
 $a(i)$  = Rata-rata ketidaksamaan dari  $i$  pada seluruh objek di A  
 $d(i, C)$  = Rata-rata ketidaksamaan dari  $i$  pada seluruh objek di C  
 $b(i)$  = Nilai minimum  $d(i, C)$ , dimana  $SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, c_i)$   
 $B$  = Klaster B dimana nilai minimum mencapai tetangga objek  $i$

Klaster  $B$  merupakan pilihan terbaik kedua untuk objek  $i$ : jika klaster B tidak dapat mengarah pada klaster A, dimana klaster B akan menjadi kompetitor terdekat seperti pada gambar 2.3 . Jumlah  $s(i)$  dituliskan pada rumus dibawah ini.

$$SSB_{i,j} = d(c_i, c_j) \quad (2.8)$$

Jumlah  $s(i)$  ditentukan oleh perbandingan antara  $a(i)$  dan  $b(i)$  sebagai berikut.

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (2.9)$$

nilai  $s(i)$  diperoleh diantara  $-1 \leq s(i) \leq 1$ .

## Kasus #2 Pertimbangan Kesamaan

Dalam hal pertimbangan kesamaan dan menentukan  $a'(i)$ ,  $d'(i, C)$ , dan menempatkan  $b'(i)$  = nilai maksimum  $d'(i, C)$ , dimana  $C \neq A$  [33]. Jumlah  $s(i)$  ditentukan sebagai berikut.

$$s(i) = \begin{cases} 1 - \frac{a_i}{b_i}; & \text{if } a_i < b_i, \\ 0; & \text{if } a_i = b_i, \\ \frac{b_i}{a_i} - 1; & \text{if } a_i > b_i \end{cases} \quad (2.10)$$

Menurut Kaufman dan Rousseeuw, nilai SI yaitu:

- $0.7 < SI \leq 1$  *strong structure*
- $0.5 < SI \leq 0.7$  *medium structure*
- $0.25 < SI \leq 0.5$  *weak structure*
- $SI \leq 0.25$  *no structure*

Sehingga dapat disimpulkan, bahwa:

- Jika nilai SI diantara 0 dan 1 berarti data memiliki nilai validasi yang optimal
- Jika nilai SI berada pada angka 0 berarti diantara dua buah klaster
- Jika nilai SI berada pada -1 berarti data terletak pada klaster yang salah

## 2. *Davies Bouldin Index (DBI)*

*Davies Bouldin Index (DBI)* diperkenalkan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979 adalah metrik untuk mengevaluasi hasil algoritma *clustering* [34]. *Davies Bouldin Index (DBI)* merupakan salah satu metode evaluasi internal yang mengukur evaluasi *cluster* pada suatu metode pengelompokkan yang didasarkan pada nilai kohesi dan separasi [35]. Dalam suatu pengelompokkan, jumlah dari kedekatan suatu data terhadap titik pusat (*centroid*) dari *cluster* yang diikuti didefinisikan sebagai kohesi. Sedangkan jarak antar *centroid* dari *cluster*-nya didefinisikan sebagai separasi.

Menurut [35] terdapat beberapa langkah untuk mendapatkan nilai validasi *Davies Bouldin Index (DBI)*. *Sum of Square Within Cluster (SSW)* merupakan persamaan yang digunakan untuk mengetahui matriks kohesi dalam sebuah *cluster* ke-*i* dengan persamaan sebagai berikut.

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, c_i) \quad (2.11)$$

Keterangan:

- $m_i$  = Jumlah data dalam *cluster* ke-*i*  
 $c_i$  = *Centroid* klaster ke-*i*  
 $x_j$  = Objek data ke-*j*  
 $i$  = 1, 2, 3,..n  
 $d()$  = Jarak setiap data ke *centroid* yang dihitung menggunakan jarak *Euclidean*  
 $SSW_i$  = *Sum of Squared Within Cluster*

*Sum of Squared Between Cluster* merupakan persamaan yang digunakan untuk mengetahui separasi antar *cluster* yang dihitung menggunakan persamaan sebagai berikut.

$$SSB_{i,j} = d(c_i, c_j) \quad (2.12)$$

Setelah nilai kohesi dan separasi diperoleh, langkah selanjutnya yaitu melakukan pengukuran rasio ( $R_{ij}$ ) untuk mengetahui nilai perbandingan antara klaster ke- $i$  dan klaster ke- $j$ . Klaster yang baik adalah *cluster* yang memiliki nilai kohesi sekecil mungkin dan separasi yang sebesar mungkin. Nilai rasio dihitung menggunakan persamaan dibawah ini.

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (2.13)$$

Nilai rasio yang diperoleh selanjutnya digunakan untuk mencari nilai validasi *Davies Bouldin Index* (DBI) dengan menggunakan persamaan dibawah ini.

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad (2.14)$$

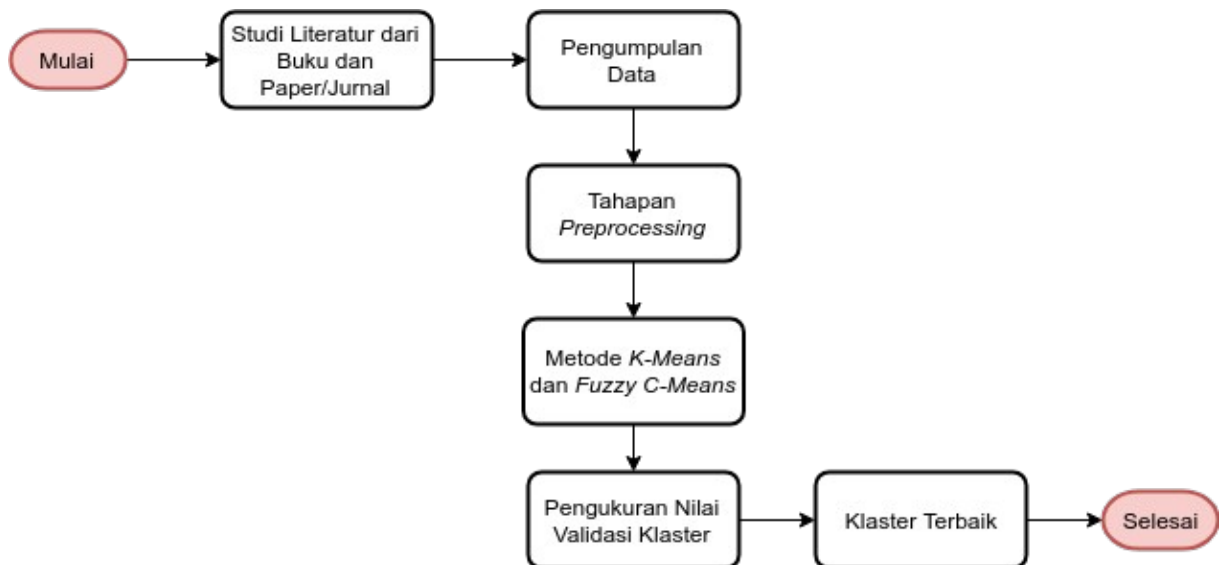
Dari persamaan tersebut,  $k$  merupakan jumlah klaster yang digunakan. Semakin kecil nilai DBI yang diperoleh (non-negatif  $\geq 0$ ), maka semakin baik klaster yang diperoleh dari pengelompokan *K-means* dan *Fuzzy C-Means* yang digunakan.

## BAB III

### METODOLOGI PENELITIAN

#### 3.1 Tahap Penelitian

Pada bab ini akan dijelaskan tahapan-tahapan penelitian yang diterapkan pada penelitian ini. Tahapan-tahapan penelitian yang akan diterapkan digambarkan pada bagan di bawah ini.



Gambar 3.1. Alur Metode Penelitian

#### 3.2 Studi Literatur

Peneliti melakukan studi literatur pada penelitian sebelumnya yang berkaitan dengan topik penelitian yang dilakukan sebagai sumber referensi bagi peneliti. Sumber studi literatur diambil dari buku dan *technical paper*. Penelitian terkait yang dijadikan rujukan utama dalam penelitian ini adalah penelitian [4], [6], [9], [8], dan [10].

#### 3.3 Pengumpulan Data

Pengumpulan data dilakukan dengan menggunakan *data set* profil kesehatan Kabupaten/Kota tahun 2017 yang diperoleh dari Pusat Data dan Informasi Kementerian Kesehatan Republik Indonesia (Pusdatin Kemkes RI) dengan menggunakan lima buah indikator (atribut) derajat kesehatan, yaitu kematian neonatal, kematian bayi, kematian anak balita, kematian balita, dan kematian ibu yang diambil dari repositori Departemen Kesehatan Republik Indonesia. Seluruh atribut *data set* ditunjukkan pada tabel di bawah ini.

Tabel 3.1. *Data set* Profil Kondisi Kesehatan Kabupaten/Kota Tahun 2017

Kecamatan	Puskesmas	Kematian Neonatal (jiwa)	Kematian Bayi (jiwa)	Kematian Anak Balita (jiwa)	Kematian Balita (jiwa)	Kematian Ibu (jiwa)
Singkil	Singkil	2	5	0	5	4
P. Banyak	P. Banyak	0	1	1	2	1
Singkil Utara	Singkil Utara	2	3	0	3	1
Gn. Meriah	Gn. Meriah	9	10	1	11	3
...	...	...	...	...	...	...
...	...	...	...	...	...	...
...	...	...	...	...	...	...
IV Nagari	Muaro Bodi	2	4	0	4	0
Kupitan	Pd. Sibusuk	1	3	0	3	0
Koto VII	Tj. Ampalu	5	10	0	10	1

### 3.4 Tahapan *Preprocessing*

Sebelum melakukan *data mining*, terdapat beberapa langkah pelaksanaan dalam melakukan *preprocessing*. Pertama, melakukan proses *cleaning* pada data yang mencakup antara lain: membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data. Integrasi data perlu dilakukan untuk menggabungkan/penyatuan data yang terpisah ke dalam *database* baru agar terhindar dari adanya duplikasi data. Selanjutnya yaitu melakukan seleksi data dengan proses meminimalkan jumlah data yang akan digunakan untuk selanjutnya dilakukan proses *mining* dengan tetap mempertahankan data aslinya. Setelah melakukan seleksi data, selanjutnya yaitu mentransformasi data atau menggabungkan data ke dalam bentuk normalisasi. Tahapan selanjutnya yaitu melakukan proses *mining* atau penambahan data menggunakan metode *K-Means* dan *Fuzzy C-Means*. *Preprocessing* pada data mencegah kualitas data rendah yang akan mempengaruhi tingkat, konsistensi, dan aksesibilitas pada proses *data mining*.

### 3.5 Metode Penelitian

Proses *data mining* dilakukan dengan menggunakan metode *clustering*. Algoritma yang digunakan adalah kombinasi *K-Means* dan *Fuzzy C-Means*. Proses pertama yang dilakukan adalah memasukkan data pada *software* RStudio. Selanjutnya dilakukan *clustering*

menggunakan algoritma *K-Means* dan *Fuzzy C-Means* untuk mendapatkan jumlah *cluster* dari kedua algoritma.

Adapun perintah *clustering* dengan *K-Means* pada RStudio yaitu dengan melakukan perintah `kmeans.proc <- kmeans(sd, centers = 3)` dimana `kmeans.proc` merupakan variabel untuk menampung hasil *clustering*, `kmeans` merupakan fungsi untuk menjalankan perintah *K-Means clustering*, `sd` merupakan data yang digunakan untuk proses *clustering* dan telah dinormalisasi, dan `centers = 3` merupakan jumlah *centroid* yang akan digunakan. Adapun perintah *clustering* dengan *Fuzzy C-Means* pada RStudio yaitu dengan memanggil pustaka *cluster* dengan perintah `library(cluster)`, selanjutnya melakukan *clustering* dengan perintah `proses.fcm <- fanny(sd, 3, metric = "euclidean", stand = FALSE)` dimana `proses.fcm` yaitu variabel untuk menampung hasil *clustering*, `sd` merupakan data yang digunakan untuk proses *clustering* dan telah dinormalisasi, `centers = 3` merupakan jumlah *centroid* yang akan digunakan, `metric = euclidean` merupakan pengukuran jarak data pada *cluster* akan menggunakan rumus Euclidean.

Setelah dilakukan proses *clustering* pada *K-Means* dan *Fuzzy C-Means*, langkah selanjutnya yaitu menghitung nilai validasi kedua algoritma. Validasi *Silhouette Index* (SI) dan *Davies Bouldin Index* (DBI) digunakan untuk menghitung nilai validasi pada algoritma *K-Means* dan *Fuzzy C-Means*. Adapun tahapan menghitung nilai validasi DBI yaitu dengan menjalankan perintah `library(NbClust)` dimana akan memanggil pustaka `NbClust`, `DBIFCM<-NbClust(sd, distance = "euclidean", min.nc = 5, max.nc = 5, index = "db", method = "average")`, dimana `DBIFCM` merupakan variabel untuk menampung nilai DBI pada *Fuzzy C-Means*, `NbClust` merupakan perintah untuk melakukan validasi *cluster*, `sd` merupakan data yang digunakan untuk proses *clustering* dan telah dinormalisasi, `distance = "euclidean"` merupakan formula untuk perhitungan jarak antar objek di dalam *cluster*, `min.nc` dan `max.nc` = jumlah *cluster* minimum dan jumlah *cluster* maksimum, `index = db` merupakan indeks validasi DBI, dan `method = average` merupakan metode perhitungan validasi DBI yang akan digunakan. Adapun perintah `DBIFCM$Best.nc` digunakan untuk mengetahui keluaran yang dihasilkan dari proses perhitungan validasi menggunakan DBI berupa jumlah *cluster terbaik* dari rentang jumlah *cluster* pada `min.nc` dan `max.nc`.

Adapun tahapan menghitung nilai validasi SI yaitu dengan menjalankan perintah `library(NbClust)` dimana akan memanggil pustaka `NbClust`, `SIFCM<-NbClust(sd, distance = "euclidean", min.nc = 5, max.nc = 5, index = "silhouette", method = "average")`, dimana `SIFCM` merupakan sebagai variabel untuk menampung nilai SI pada *Fuzzy C-Means*, `NbClust` merupakan perintah untuk melakukan validasi *cluster*, `sd` merupakan data yang digunakan untuk proses *clustering* dan telah dinormalisasi, `distance = "euclidean"` merupakan formula untuk perhitungan jarak antar objek di dalam *cluster*, `min.nc` dan `max.nc` = jumlah *cluster* minimum dan jumlah *cluster* maksimum, `index = silhouette` merupakan indeks validasi SI, dan `method = average` merupakan metode perhitungan validasi SI yang akan digunakan. Adapun perintah `SIFCM$Best.nc` digunakan untuk mengetahui keluaran yang dihasilkan dari proses perhitungan validasi menggunakan SI berupa jumlah *cluster terbaik* dari rentang jumlah *cluster* pada `min.nc` dan `max.nc`.

Hasil akhir yaitu berupa pemilihan algoritma *cluster* terbaik diantara algoritma *K-Means* dan *Fuzzy C-Means* berdasarkan nilai validasi DBI yang lebih mendekati 0 dan nilai validasi SI yang lebih mendekati 1. Adapun tahapan pada penelitian ini dapat dilihat pada Gambar 3.2 dibawah ini.





Gambar 3.2. Diagram Alir Tahapan Penelitian

### 3.6 Evaluasi

Evaluasi dari penelitian ini dilakukan dengan membandingkan *K-Means* dan *Fuzzy C-Means*. Untuk memeriksa hasil *cluster* terbaik, maka akan dibandingkan nilai validasi dan jumlah *cluster* terbaik dari algoritma *K-Means* dan *Fuzzy C-Means* yang telah dilakukan berdasarkan teori pada validasi *Silhouette Index* dan *Davies Bouldin Index*. Selanjutnya menentukan algoritma *cluster* terbaik antara *K-Means* dan *Fuzzy C-Means*.

## BAB IV

### HASIL PENGUJIAN DAN ANALISIS

#### 4.1 Hasil Pengujian

Pada bab ini akan dijelaskan tahapan pengumpulan data, proses *preprocessing*, data penelitian, perhitungan secara manual dan secara program pada proses *clustering K-Means* dan *Fuzzy C-Means*, perhitungan secara manual dan secara program pada proses indeks validasi *K-Means* dan *Fuzzy C-Means*, perbedaan antara penelitian terkait dan yang dilakukan, dan penentuan algoritma klaster terbaik.

##### 4.1.1 Pengumpulan Data

Sebelum melakukan proses pengujian, langkah pertama yaitu melakukan pengumpulan data. Pengumpulan data dilakukan dengan menggunakan *data set* profil kesehatan Kabupaten/Kota tahun 2017 yang diperoleh dari Pusat Data dan Informasi Kementerian Kesehatan Republik Indonesia (Pusdatin Kemkes RI) dengan menggunakan lima buah indikator (atribut) derajat kesehatan, yaitu kematian neonatal, kematian bayi, kematian anak balita, kematian balita, dan kematian ibu yang diambil dari repositori Departemen Kesehatan Republik Indonesia. Seluruh atribut *data set* ditunjukkan pada tabel di bawah ini. Pada tabel dibawah ini menggunakan *data set* dengan 550 data.

Tabel 4.1. *Data Set* Profil Kondisi Kesehatan Kabupaten/Kota Tahun 2017

Kecamatan	Puskesmas	Kematian Neonatal (jiwa)	Kematian Bayi (jiwa)	Kematian Anak Balita (jiwa)	Kematian Balita (jiwa)	Kematian Ibu (jiwa)
Singkil	Singkil	2	5	0	5	4
P. Banyak	P. Banyak	0	1	1	2	1
Singkil Utara	Singkil Utara	2	3	0	3	1
Gn. Meriah	Gn. Meriah	9	10	1	11	3
...	...	...	...	...	...	...
...	...	...	...	...	...	...
...	...	...	...	...	...	...
IV Nagari	Muaro Bodi	2	4	0	4	0
Kupitan	Pd. Sibusuk	1	3	0	3	0
Koto VII	Tj. Ampalu	5	10	0	10	1

#### 4.1.2 Tahapan *Preprocessing* dan *Preprocessing* Data

Sebelum melakukan pengujian data, terdapat beberapa langkah pelaksanaan dalam melakukan *preprocessing*. Tahapan yang akan dilakukan yaitu mengumpulkan *data set* yang diperoleh dari Pusat Data dan Informasi Kementerian Kesehatan Republik Indonesia (Pusdatin Kemkes RI) untuk diteliti. Data yang telah terkumpul akan dilakukan proses *cleaning* atau pembersihan data dengan memeriksa data inkonsisten dan memperbaiki kesalahan data. Setelah proses *cleaning* data selesai, tahapan selanjutnya yaitu dengan menggabungkan data ke dalam satu berkas utuh untuk menghindari adanya duplikasi data. Langkah selanjutnya yaitu menyeleksi data dengan menentukan jumlah sampel data menggunakan rumus *finite population*. Setelah proses seleksi data, data yang telah diseleksi selanjutnya dilakukan proses normalisasi untuk menghilangkan adanya duplikasi data, mengurangi kompleksitas, dan mempermudah dalam pemodifikasian data. Langkah terakhir adalah melakukan proses *clustering* menggunakan metode *K-Means* dan *Fuzzy C-Means*.

Sebelum memulai percobaan *K-Means* dengan RStudio dibutuhkan proses *preprocessing* data pada *K-Means* dan *Fuzzy C-Means*. Langkah ini dilakukan agar proses data dapat dilakukan dengan baik tanpa menimbulkan galat (*error*). Berikut adalah *data set* sebelum dan sesudah dilakukan *preprocessing*. Pada tabel dibawah ini menggunakan *data set* dengan 550 data.

Tabel 4.2. *Data Set* Sebelum Proses *Preprocessing* Data

No.	Kematian Neonatal (jiwa)	Kematian Bayi (jiwa)	Kematian Anak Balita (jiwa)	Kematian Balita (jiwa)	Kematian Ibu (jiwa)
1	2	5	0	5	4
2	0	1	1	2	1
3	2	3	0	3	1
4	9	10	1	11	3
...	...	...	...	...	...
548	2	4	0	4	0
549	1	3	0	3	0
550	5	10	0	10	1

Tabel 4.3. *Data Set* Sesudah Proses *Preprocessing* Data

No.	Kematian Neonatal	Kematian Bayi	Kematian Anak Balita	Kematian Balita	Kematian Ibu
1	0.0174553	0.9311569	-0.3305231	0.7612878	4.6459413
2	-0.81749	-0.4025313	1.6454302	-0.1417495	0.7526944
3	-0.0174553	0.2643128	-0.3305231	0.159263	0.7526944
4	2.782666	2.5982671	1.6454302	2.5673625	3.3481924
...	...	...	...	...	...
548	-0.0174553	0.5977348	-0.3305231	0.4602754	-0.5450546
549	-0.4174726	0.2643128	-0.3305231	0.159263	-0.5450546
550	1.1825967	2.5982671	-0.3305231	2.2663501	0.7526944

#### 4.1.3 *Clustering K-Means dan Fuzzy C-Means*

Pada penelitian ini, proses *clustering K-Means* dan *Fuzzy C-Means* dilakukan secara manual menggunakan aplikasi LibreOffice Calc dan secara program menggunakan aplikasi RStudio. Pada perhitungan manual menggunakan LibreOffice Calc jumlah iterasi yang ditentukan yaitu dua iterasi.

##### a. *Clustering K-Means dan Fuzzy C-Means dengan Spreadsheet*

Proses *clustering* dengan menggunakan *K-Means* dan *Fuzzy C-Means* akan dilakukan terhadap 550 sampel data kabupaten/kota di Indonesia. Hasil *clustering* akan diperoleh kelompok daerah yang akan digunakan untuk proses klasifikasi penentuan klaster (kelompok data) yang terdiri atas 3 buah klaster. Penelitian ini akan menggunakan LibreOffice Calc versi 6.1.5.2 yang berjalan diatas sistem operasi Linux Mint 18.1 arsitektur 64 bit.

##### 1. *Clustering pada K-Means*

- Untuk melakukan *clustering* pada *K-Means*, langkah pertama yaitu menentukan jumlah titik pusat (*centroid*) untuk pengelompokkan data secara acak. Pada penelitian ini akan menggunakan tiga buah klaster. Klaster 1 yaitu data ke-23, klaster 2 yaitu data ke-60, dan klaster 3 yaitu data ke-154.

Tabel 4.4. Pemilihan *Centroid* pada *K-Means*

Data ke-i	Puskesmas	Kematian Neonatal	Kematian Bayi	Kematian Anak Balita	Kematian Balita	Kematian Ibu
23	Kluet Utara	1	2	0	2	0
60	Meutulang	2	3	0	3	1

Data ke-i	Puskesmas	Kematian Neonatal	Kematian Bayi	Kematian Anak Balita	Kematian Balita	Kematian Ibu
154	Simpang Jaya	1	1	0	1	1

- b. Selanjutnya pada iterasi 1, setiap data dikelompokkan ke setiap *centroid* menggunakan rumus *Euclidean* untuk memperoleh jarak data ke *centroid*. Sebagai contoh, perhitungan jarak menggunakan Euclidean yaitu

$$d_{ik} = \sqrt{(2-1)^2 + (5-2)^2 + (0-0)^2 + (5-2)^2 + (4-0)^2} = 5.9160$$

Adapun perintah untuk menghitung jarak data ke *centroid* pada LibreOffice Calc, yaitu =SQRT((Sheet1.C2-Sheet2.\$J\$555)^2+(Sheet1.D2-Sheet2.\$K\$555)^2 + (Sheet1.E2-Sheet2.\$L\$555)^2 + (Sheet1.F2 - Sheet2.\$M\$555)^2 + (Sheet1.G2-Sheet2.\$N\$555)^2).

Tabel 4.5. Nilai Jarak Data ke Setiap *Centroid*

Data ke-i	C1	C2	C3
1	5.91607978309962	4.12310562561766	6.48074069840786
2	2	3.16227766016838	1.73205080756888
3	2	0	3
4	14.7986485869487	12.9228479833201	15.8113883008419
...	...	...	...
548	3	1.73205080756888	4.47213595499958
549	1.4142135623731	1.4142135623731	3
550	12.0415945787923	10.3440804327886	13.3416640641263

- c. Selanjutnya yaitu menentukan keanggotaan klaster pada data dengan menentukan nilai rata-rata terendah. Keanggotaan klaster pada data merupakan hal penting untuk mengetahui bahwa data termasuk pada anggota klaster tertentu. Adapun perintah untuk menentukan keanggotaan klaster data pada LibreOffice Calc yaitu =IF(AND(B3<C3,B3<D3),1,IF(AND(C3<B3,C3<D3),2,3)).

Tabel 4.6. Keanggotaan *Cluster* pada Data

Data ke-i	C1	C2	C3	Cluster
1	5.91607978309962	4.12310562561766	6.48074069840786	2
2	2	3.16227766016838	1.73205080756888	3

Data ke-i	C1	C2	C3	Cluster
3	2	0	3	2
4	14.7986485869487	12.9228479833201	15.8113883008419	2
...	...	...	...	...
548	3	1.73205080756888	4.47213595499958	2
549	1.4142135623731	1.4142135623731	3	3
550	12.0415945787923	10.3440804327886	13.3416640641263	2

- d. Setelah mengelompokkan data pada masing-masing *cluster*, diketahui jumlah anggota C1 = 62, C2 = 181, dan C3 = 307 pada iterasi 1.
- e. Selanjutnya yaitu menentukan *centroid* pada variabel data berdasarkan *cluster*. Hal ini dilakukan untuk nilai sebagai titik *centroid* yang baru. Ketentuan untuk menentukan titik *centroid* baru yaitu jika data ke-1 anggota *cluster* 1, maka akan menampilkan nilai *data set*, jika tidak maka akan menampilkan nilai 0. Adapun perintah untuk menentukan *centroid* yaitu =IF(\$E3=1,\$Sheet1.C2,0). Adapun simbol a, b, c, d, dan e merupakan representasi dari variabel data. Tabel 4.7 adalah hasil penentuan *centroid* baru.
- f. Setelah menentukan nilai *centroid* baru, maka akan dilakukan perhitungan jumlah seluruh nilai pada setiap variabel, jumlah data, dan nilai rata-rata. Jumlah seluruh nilai pada setiap variabel yaitu akumulasi dari data ke-1 hingga data ke-550. Jumlah data yaitu untuk setiap nilai yang lebih besar atau lebih kecil dari 0 maka akan menampilkan nilai jumlah data. Sedangkan nilai rata-rata yaitu nilai perbandingan dari jumlah seluruh nilai pada setiap variabel dan jumlah data. Pada LibreOffice Calc, perintah menentukan jumlah yaitu =SUM(J3:J552), jumlah data yaitu =COUNTIF(J3:J552,">0"), dan rata-rata (jumlah/jumlah data) yaitu =J553/J554. Pada iterasi 1, nilai rata-rata setiap data telah diperoleh dan menjadi acuan sebagai *centroid* baru untuk iterasi selanjutnya. Tabel 4.8 adalah hasil jumlah seluruh nilai pada setiap variabel, jumlah data, dan nilai rata-rata.

Tabel 4.7. Penentuan *Centroid* Baru pada Variabel Data

Data ke-i	Cluster 1					Cluster2					Cluster 3				
	a	b	c	d	e	a	b	c	d	e	a	b	c	d	e
1	0	0	0	0	0	2	5	0	5	4	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	1	1	2	1
3	0	0	0	0	0	2	3	0	3	1	0	0	0	0	0
4	0	0	0	0	0	9	10	1	11	3	0	0	0	0	0
5	0	0	0	0	0	8	11	1	12	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
7	2	2	0	2	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	2	2	0
9	0	2	1	3	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	1	2	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
546	0	0	0	0	0	4	5	1	6	0	0	0	0	0	0
547	0	0	0	0	0	2	8	0	8	0	0	0	0	0	0
548	0	0	0	0	0	2	4	0	4	0	0	0	0	0	0
549	0	0	0	0	0	0	0	0	0	0	1	3	0	3	0
550	0	0	0	0	0	5	10	0	10	1	0	0	0	0	0
Jumlah	87	115	13	122	4	815	952	58	1076	133	222	147	21	161	94
Jumlah data	51	59	11	57	4	181	178	43	179	82	141	112	17	121	76
Rata-Rata	1.70588	1.94915	1.18182	2.14035	1.00000	4.50276	5.34831	1.34884	6.01117	1.62195	1.57447	1.31250	1.23529	1.33058	1.23684

Tabel 4.8. Jumlah, Jumlah Data, dan Rata-Rata

Jumlah	Jumlah Data	Rata-Rata
87	51	1.70588
115	59	1.94915
113	11	1.18182
122	57	2.14035
4	4	1
815	181	4.50276
952	178	5.34831
58	43	1.34884
1076	179	6.01117
133	82	1.62195
222	141	1.57447
147	112	1.31250
21	17	1.23529
161	121	1.33058
94	76	1.23684

- g. Selanjutnya diperoleh rata-rata *cluster* (*cluster mean*) dari variabel penelitian, yaitu rata-rata dari Kematian Neonatal, Kematian Bayi, Kematian Anak Balita, Kematian Balita, dan Kematian Ibu. Berikut adalah tabel rata-rata *cluster* (*cluster mean*).

Tabel 4.9. *Cluster Mean* pada Iterasi 1

Cluster/Variabel	Kematian Neonatal	Kematian Bayi	Kematian Anak Balita	Kematian Balita	Kematian Ibu
C1	1.7059	1.9492	1.1818	2.1404	1.0000
C2	4.5028	5.3483	1.3488	6.0112	1.6220
C3	1.5745	1.3125	1.2353	1.3306	1.2368

- h. Langkah selanjutnya yaitu melakukan iterasi 2 dengan *centroid* awal yaitu nilai rata-rata dari iterasi 1. Setiap data dikelompokkan ke setiap *centroid* menggunakan metrik Euclidean untuk memperoleh jarak data ke *centroid*. Sebagai contoh, perhitungan jarak menggunakan Euclidean yaitu

$$d_{ik} = \sqrt{(2 - 1.70588)^2 + (5 - 1.94915)^2 + (0 - 1.18182)^2 + (5 - 2.14035)^2 + (4 - 1)^2} = 5.2885$$



Adapun perintah untuk menghitung jarak data ke *centroid* yaitu  $=\text{SQRT}((\text{Sheet1.C2}-\text{Sheet2.J\$555})^2 + (\text{Sheet1.D2}-\text{Sheet2.K\$555})^2 + (\text{Sheet1.E2}-\text{Sheet2.L\$555})^2 + (\text{Sheet1.F2}-\text{Sheet2.M\$555})^2 + (\text{Sheet1.G2}-\text{Sheet2.N\$555})^2)$ .

Tabel 4.10. Nilai Jarak Data ke Setiap *Centroid*

Data ke-i	C1	C2	C3
1	5.28852178973182	3.85773144209549	6.03360429781552
2	1.96562493134818	7.46864984194837	1.77092874890786
3	1.82386304275615	4.80122383989768	2.71988005319786
4	14.1614525196256	8.29291393823043	15.0755756592443
...	...	...	...
548	3.1855093092942	4.07144657555862	4.19348002135655
549	2.17675012452373	5.59474554038918	3.00339272898545
550	11.7829594927235	6.32474084052295	12.804211800013

- i. Selanjutnya yaitu menentukan keanggotaan *cluster* pada data dengan menentukan nilai rata-rata terendah. Keanggotaan *cluster* pada data merupakan hal penting untuk mengetahui bahwa data termasuk pada anggota *cluster* tertentu. Adapun perintah untuk menentukan keanggotaan *cluster* pada data yaitu  $=\text{IF}(\text{AND}(\text{B3}<\text{C3},\text{B3}<\text{D3}),1,\text{IF}(\text{AND}(\text{C3}<\text{B3},\text{C3}<\text{D3}),2,3))$ .

Tabel 4.11. Keanggotaan *Cluster* pada Data

Data ke-i	C1	C2	C3	Cluster
1	5.28852178973182	3.85773144209549	6.03360429781552	2
2	1.96562493134818	7.46864984194837	1.77092874890786	3
3	1.82386304275615	4.80122383989768	2.71988005319786	1
4	14.1614525196256	8.29291393823043	15.0755756592443	2
...	...	...	...	...
548	3.1855093092942	4.07144657555862	4.19348002135655	1
549	2.17675012452373	5.59474554038918	3.00339272898545	1
550	11.7829594927235	6.32474084052295	12.804211800013	2

- j. Setelah mengelompokkan data pada masing-masing *cluster*, diketahui jumlah anggota  $C1 = 143$ ,  $C2 = 116$ , dan  $C3 = 291$  pada iterasi 2.
- k. Selanjutnya yaitu menentukan *centroid* pada variabel data berdasarkan *cluster*. Hal ini dilakukan untuk nilai sebagai titik *centroid* yang baru. Ketentuan untuk menentukan titik *centroid* baru yaitu jika data ke-1 anggota *cluster* 1, maka akan menampilkan nilai *data set*, jika tidak maka akan menampilkan nilai 0 Adapun perintah untuk menentukan *centroid* yaitu `=IF($E3=1,$Sheet1.C2,0)`. Tabel 4.12 adalah hasil penentuan *centroid* baru.
- l. Setelah menentukan nilai *centroid* baru, maka akan dilakukan perhitungan jumlah seluruh nilai pada setiap variabel, jumlah data, dan nilai rata-rata. Jumlah seluruh nilai pada setiap variabel yaitu akumulasi dari data ke-1 hingga data ke-550. Jumlah data yaitu untuk setiap nilai yang lebih besar atau lebih kecil dari 0 maka akan menampilkan nilai jumlah data. Sedangkan nilai rata-rata yaitu nilai perbandingan dari jumlah seluruh nilai pada setiap variabel dan jumlah data. Pada LibreOffice Calc, perintah menentukan jumlah yaitu `=SUM(J3:J552)`, jumlah data yaitu `=COUNTIF(J3:J552,">0")`, dan rata-rata (jumlah/jumlah data) yaitu `=J553/J554`. Tabel 4.13 adalah hasil jumlah, jumlah data, dan rata-rata.

Tabel 4.12. Penentuan *Centroid* Baru pada Variabel Data

Data ke-i	Cluster 1					Cluster 2					Cluster 3				
	a	b	c	d	e	a	b	c	d	e	a	b	c	d	e
1	0	0	0	0	0	2	5	0	5	4	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	1	1	2	1
3	2	3	0	3	1	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	9	10	1	11	3	0	0	0	0	0
5	0	0	0	0	0	8	11	1	12	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
7	2	2	0	2	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	2	2	0
9	0	2	1	3	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	1	2	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
546	0	0	0	0	0	4	5	1	6	0	0	0	0	0	0
547	0	0	0	0	0	2	8	0	8	0	0	0	0	0	0
548	2	4	0	4	0	0	0	0	0	0	0	0	0	0	0
549	1	3	0	3	0	0	0	0	0	0	0	0	0	0	0
550	0	0	0	0	0	5	10	0	10	1	0	0	0	0	0
Jumlah	294	348	21	373	52	638	753	52	873	95	192	113	19	113	84
Jumlah data	133	138	18	141	39	116	114	37	116	58	124	97	16	100	65
Rata-Rata	2.21053	2.52174	1.16667	2.64539	1.33333	5.50000	6.60526	1.40541	7.52586	1.63793	1.54839	1.16495	1.18750	1.13000	1.29231

Tabel 4.13. Jumlah, Jumlah Data, dan Rata-Rata

Jumlah	Jumlah Data	Rata-Rata
294	133	2.21503
348	138	2.52174
21	18	1.66667
373	141	2.64539
52	39	1.33333
638	116	5.5
753	114	6.60526
52	37	1.40541
873	116	7.52586
95	58	1.63793
192	124	1.54839
113	97	1.16495
19	16	1.18750
113	100	1.13
84	65	1.29231

- m. Selanjutnya diperoleh rata-rata *cluster* (*cluster mean*) dari variabel penelitian, yaitu rata-rata dari Kematian Neonatal, Kematian Bayi, Kematian Anak Balita, Kematian Balita, dan Kematian Ibu. Berikut adalah tabel rata-rata *cluster* (*cluster mean*).

Tabel 4.14. *Cluster Mean* pada Iterasi 2

Cluster/Variabel	Kematian Neonatal	Kematian Bayi	Kematian Anak Balita	Kematian Balita	Kematian Ibu
C1	2.2105	2.5217	1.1667	2.6454	1.3333
C2	5.5000	6.6053	1.4054	7.5259	1.6379
C3	1.5484	1.1649	1.1875	1.1300	1.2923

- n. Pada iterasi 2, nilai rata-rata setiap data telah diperoleh dan menjadi acuan sebagai *centroid* baru untuk iterasi selanjutnya. Langkah selanjutnya yaitu melakukan iterasi 3 dengan *centroid* awal yaitu nilai rata-rata dari iterasi 2. Langkah ini dilakukan hingga nilai rata-rata setiap data tidak berubah, maka iterasi berhenti. Pada penelitian ini jumlah iterasi yang dilakukan yaitu sebanyak dua iterasi.

## 2. *Clustering* pada *Fuzzy C-Means*

- a. Untuk melakukan *clustering* pada *Fuzzy C-Means*, langkah pertama yaitu menentukan jumlah *cluster* (C), iterasi maksimum (*max iter*) pangkat pembobot (w), *error* terkecil yang diharapkan (E/Epsilon), fungsi objektif (P), dan iterasi awal. Pada penelitian ini akan menggunakan 3 *cluster*, iterasi maksimum sebanyak 100 kali, pangkat pembobot yaitu 2, *error* terkecil yang diharapkan yaitu 0.1, fungsi objektif dimulai dari 0, dan iterasi awal yaitu 1. Adapun simbol a1 – a5 merupakan simbol dari variabel yang akan diuji.

Tabel 4.15. *Data Set* Pengujian *Fuzzy C-Means*

No.	a1	a2	a3	a4	a5
1	2	5	0	5	4
2	0	1	1	2	1
3	2	3	0	3	1
...	...	...	...	...	...
549	1	3	0	3	0
550	5	10	0	10	1

- b. Langkah selanjutnya yaitu penentuan keanggotaan *cluster* dimana keanggotaan *cluster* pada iterasi 1 bersifat acak (*random*) namun harus bernilai 1 jika dijumlahkan, karena keanggotaan fuzzy berada pada rentang nilai 0 dan 1. Pada penelitian ini menggunakan tiga buah *cluster* dengan simbol U sebagai *cluster* dan nilai yang beragam pada keanggotaan *cluster*.

Tabel 4.16. Keanggotaan *Cluster* Iterasi 1 pada *Fuzzy C-Means*

No	U1	U2	U3
1	0.3	0.3	0.4
2	0.3	0.4	0.3
3	0.4	0.3	0.3
...	...	...	...
548	0.1	0.4	0.5
549	0.1	0.6	0.3
550	0.5	0.4	0.1

- c. Tahapan selanjutnya yaitu menghitung pusat *cluster* pada setiap data. Namun sebelum menghitung pusat *cluster*, setiap *cluster* harus dipangkatkan dengan pembobot. Sebagai contoh,  $U1^w = 0.3^2 = 0.09$ ,  $U1^w * a1 = 0.09 * 2 = 0.18$ . Adapun perintah untuk memangkatkan nilai *cluster* dengan pangkat pembobot yaitu **=POWER(S7,\$J\$7)**. Dan perintah untuk menghitung pusat *cluster* pada setiap data yaitu **=Y7\*T7**.

Tabel 4.17. Perhitungan Pusat *Cluster* 1

Data ke-i	U1	U1 <sup>w</sup>	U1 <sup>w.a1</sup>	U1 <sup>w.a2</sup>	U1 <sup>w.a3</sup>	U1 <sup>w.a4</sup>	U1 <sup>w.a5</sup>
1	0.3	0.09	0.18	0.45	0	0.45	0.36
2	0.3	0.09	0	0.09	0.09	0.18	0.09
3	0.4	0.16	0.32	0.48	0	0.48	0.16
...	...	...	...	...	...	...	...
549	0.1	0.01	0.01	0.03	0	0.03	0
550	0.5	0.25	1.25	2.5	0	2.5	0.25

Tabel 4.18. Perhitungan Pusat *Cluster* 2

Data ke-i	U2	U2 <sup>w</sup>	U2 <sup>w.a1</sup>	U2 <sup>w.a2</sup>	U2 <sup>w.a3</sup>	U2 <sup>w.a4</sup>	U2 <sup>w.a5</sup>
1	0.3	0.09	0.18	0.45	0	0.45	0.36
2	0.4	0.16	0	0.16	0.16	0.32	0.16
3	0.3	0.09	0.18	0.27	0	0.27	0.09
...	...	...	...	...	...	...	...
549	0.6	0.36	0.36	1.08	0	1.08	0
550	0.4	0.16	0.8	1.6	0	1.6	0.16

Tabel 4.19. Perhitungan Pusat *Cluster* 3

Data ke-i	U3	U3 <sup>w</sup>	U3 <sup>w.a1</sup>	U3 <sup>w.a2</sup>	U3 <sup>w.a3</sup>	U3 <sup>w.a4</sup>	U3 <sup>w.a5</sup>
1	0.4	0.16	0.32	0.8	0	0.8	0.64
2	0.3	0.09	0	0.09	0.09	0.18	0.09
3	0.3	0.09	0.18	0.27	0	0.27	0.09
...	...	...	...	...	...	...	...
549	0.3	0.09	0.09	0.27	0	0.27	0
550	0.1	0.01	0.05	0.1	0	0.1	0.01

- d. Langkah selanjutnya yaitu menghitung jumlah hasil pusat *cluster* dari  $U_i^w$ ,  $U_i^w.a1$ ,  $U_i^w.a2$ ,  $U_i^w.a3$ ,  $U_i^w.a4$ , dan  $U_i^w.a5$  dengan perintah SUM. Simbol hasil pusat *cluster* yaitu V. Selanjutnya menghitung masing-masing hasil pusat *cluster* dimana

nilai dari pusat *cluster* yaitu jumlah dari  $V_{ai} = \frac{U_i^w \cdot a_i}{U_i^w}$ . Sebagai contoh

$$V_{a1} = \frac{175.27}{79.24} = 2.21189$$

Adapun perintah perhitungannya pada LibreOffice Calc yaitu =Z557/\$Y\$557.

Tabel 4.20. Perhitungan Jumlah Pusat *Cluster 1*

Jumlah	79.24	175.27	193.03	13.67	207.11	34.36
Pusat <i>Cluster 1</i>	2.21189	2.43602	0.17251	2.61371	0.43362	

Tabel 4.21. Perhitungan Jumlah Pusat *Cluster 2*

Jumlah	56.30	113	122.86	10.08	145.49	22.79
Pusat <i>Cluster 2</i>	2.00675	2.18185	0.17900	2.58373	0.40472	

Tabel 4.22. Perhitungan Jumlah Pusat *Cluster 3*

Jumlah	104.21	196.91	206.25	16.73	233.02	42.25
Pusat <i>Cluster 3</i>	1.88955	1.97918	0.16054	2.23606	0.40543	

- e. Setelah memperoleh perhitungan jumlah pusat *cluster* pada masing-masing *cluster*, selanjutnya yaitu menghitung hasil pusat *cluster* (V). Adapun hasil pusat *cluster* yaitu jumlah pusat *cluster* pada masing-masing *cluster*.

Tabel 4.23. Hasil Pusat *Cluster* (V)

V	a1	a2	a3	a4	a5
	2.21189	2.43602	0.17251	2.61371	0.43362
	2.00675	2.18185	0.17900	2.58373	0.40472
	1.88955	1.97918	0.16054	2.23606	0.40543

- f. Setelah memperoleh hasil pusat *cluster* (V), langkah selanjutnya yaitu menghitung fungsi objektif dari kuadrat derajat keanggotaan. Simbol penentuan fungsi objektif adalah L. Nilai L diperoleh dari jumlah akar kuadrat hasil pusat *cluster* dan *data set*

lalu dikalikan dengan derajat keanggotaan data ( $U_i^w$ ). Sebagai contoh,

$$L1 = ((2 - 2.21189)^2 + (5 - 2.43602)^2 + (0 - 0.17251)^2 + (5 - 2.61371)^2 + (4 - 0.43362)^2) * 0.09 = 2.25559$$

Adapun perintah pada LibreOffice Calc yaitu **=SUM((T7-\$BI\$6)^2 , (U7-\$BJ\$6)^2 , (V7-\$BK\$6)^2 , (W7-\$BL\$6)^2 , (X7-\$BM\$6)^2)\*BO6**. Fungsi objektif diperoleh dari jumlah seluruh  $L1 + L2 + L3$ .

Tabel 4.24. Nilai Fungsi Objektif

Kuadrat derajat keanggotaan data ke-i			L1	L2	L3	Jumlah
U1 <sup>w</sup>	U2 <sup>w</sup>	U3 <sup>w</sup>				L1+L2+L3
0.09	0.09	0.16	2.2556	2.4065	4.7558	9.4178
0.09	0.16	0.09	0.7503	1.0869	0.5079	2.3451
0.16	0.09	0.09	0.1380	0.1106	0.1815	0.4302
...	...	...	...	...	...	...
0.01	0.36	0.09	0.0215	0.7387	0.2346	0.9949
0.25	0.16	0.01	29.9738	20.0753	1.3467	51.3958
Fungsi Objektif (P1)						6320.3171

- g. Setelah menghitung fungsi objektif, langkah selanjutnya yaitu menghitung nilai matriks partisi U. Nilai matriks partisi U diperoleh dari jumlah akar kuadrat hasil pusat *cluster* (V) dikurang dengan *data set* lalu dipangkatkan dengan negatif 1. Simbol matriks parisi U yaitu LT. Sebagai contoh,

$$L1 = ((2 - 2.21189)^2 + (5 - 2.43602)^2 + (0 - 0.17251)^2 + (5 - 2.61371)^2 + (4 - 0.43362)^2)^{-1} = 0.03990$$

Rumus menghitung matriks partisi U yaitu **=SUM((ST7-BI\$6)^2, (SU7-BJ\$6)^2, (SV7-BK\$6)^2, (SW7-BL\$6)^2, (SX7-BM\$6)^2)^-1**.

Tabel 4.25. Nilai Matriks Partisi U

L1	L2	L3	LT
0.03990082395273	0.037399376308295	0.033643272471686	0.11094347273271
0.119950850656109	0.147211806159768	0.177207121863638	0.444369778679515
1.15909123969043	0.813609789563584	0.495744910139636	2.46844593939365
...	...	...	...
0.464304396178966	0.487321161319101	0.383563678887177	1.33518923638524
0.008340610226236	0.007970005173062	0.007425750203234	0.023736365602532



- h. Setelah memperoleh nilai matriks partisi  $U$ , langkah selanjutnya yaitu menghitung keanggotaan *cluster* baru. Nilai keanggotaan *cluster* baru diperoleh dari hasil perbandingan antara jumlah  $LT$  dan  $LT_i$ . Nilai keanggotaan *cluster* baru digunakan sebagai pusat *cluster* baru pada iterasi 2. Sebagai contoh,

$$U^1 * w_{baru} = \frac{0.03990}{0.11094} = 0.35965 \text{ . Perintah untuk menghitung keanggotaan } cluster$$

baru yaitu  $=BW6/BZ6$ .

Tabel 4.26. Nilai Keanggotaan *Cluster* Baru

$U1^w$	$U2^w$	$U3^w$
0.35965003591388	0.337102989361069	0.30324697472505
0.269934762468666	0.331282218600062	0.398783018931272
0.469563145456266	0.329604054348235	0.200832800195499
...	...	...
0.347744262405812	0.364982841412372	0.287272896181817
0.35138531171539	0.335771925092518	0.312842763192093

- i. Langkah selanjutnya yaitu memeriksa kondisi selisih fungsi objektif dimana, jika  $P1 < P0$ , maka proses *clustering Fuzzy C-Means* berhenti. Pada iterasi 1, nilai fungsi objektif yang diperoleh yaitu  $P1 = 6320.3171$  dan  $P0 = 0$ , sehingga  $P1 - P0 = 6320.3171$ . Maka iterasi dilanjutkan ke iterasi 2 karena nilai  $P1 > P0$ .
- j. Proses iterasi 1 telah selesai dan memperoleh nilai fungsi objektif ( $P1$ ) sebesar 6320.3171. Langkah selanjutnya yaitu melakukan proses iterasi 2. Pada proses iterasi 2, langkah-langkah proses perhitungan yang dilakukan sebagian besar adalah sama. Yang membedakan hanya pada penentuan nilai matriks partisi awal. Nilai matriks partisi awal digunakan sebagai nilai perhitungan pusat *cluster* yang baru. Pada iterasi 1, penentuan nilai matriks partisi awal yang digunakan yaitu secara acak, sedangkan pada iterasi 2, penentuan nilai matriks partisi awal yaitu menggunakan nilai keanggotaan *cluster* baru yang diperoleh pada iterasi 1. Langkah pertama pada proses iterasi 2 dimulai dengan menentukan pusat *cluster*. Nilai matriks partisi awal akan dipangkatkan dengan nilai pangkat pembobot. Sebagai contoh,
- $$U1^w = (0.35965)^2 = 0.12935 \text{ . Berikut adalah perhitungan pusat } cluster \text{ pada iterasi 2.}$$

Tabel 4.27. Perhitungan Pusat *Cluster 1*

Data ke-i	U1	$U1^w$	$U1^w.a1$	$U1^w.a2$	$U1^w.a3$	$U1^w.a4$	$U1^w.a5$
1	0.3597	0.1293	0.2587	0.6467	0.0000	0.6467	0.5174
2	0.2699	0.8773	0.0000	0.8773	0.8773	1.7545	0.8773
3	0.4696	1.0000	2.0000	3.0000	0.0000	3.0000	1.0000
...	...	...	...	...	...	...	...
549	0.3477	1.0000	1.0000	3.0000	0.0000	3.0000	0.0000
550	0.3514	1.0000	5.0000	10.0000	0.0000	10.0000	1.0000

Tabel 4.28. Perhitungan Pusat *Cluster 2*

Data ke-i	U2	$U2^w$	$U2^w.a1$	$U2^w.a2$	$U2^w.a3$	$U2^w.a4$	$U2^w.a5$
1	0.3371	0.0900	0.1800	0.4500	0.0000	0.4500	0.3600
2	0.3313	0.1600	0.0000	0.1600	0.1600	0.3200	0.1600
3	0.3296	0.0900	0.1800	0.2700	0.0000	0.2700	0.0900
...	...	...	...	...	...	...	...
549	0.3650	1.0000	1.0000	3.0000	0.0000	3.0000	0.0000
550	0.3358	1.0000	5.0000	10.0000	0.0000	10.0000	1.0000

Tabel 4.29. Perhitungan Pusat *Cluster 3*

Data ke-i	U3	$U3^w$	$U3^w.a1$	$U3^w.a2$	$U3^w.a3$	$U3^w.a4$	$U3^w.a5$
1	0.3032	0.0920	0.1839	0.4598	0.0000	0.4598	0.3678
2	0.3988	0.9122	0.0000	0.9122	0.9122	1.8243	0.9122
3	0.2008	1.0000	2.0000	3.0000	0.0000	3.0000	1.0000
...	...	...	...	...	...	...	...
549	0.2873	1.0000	1.0000	3.0000	0.0000	3.0000	0.0000
550	0.3128	1.0000	5.0000	10.0000	0.0000	10.0000	1.0000

- k. Langkah selanjutnya yaitu menghitung jumlah pusat *cluster* dari  $U_i^w$ ,  $U_i^w.a1$ ,  $U_i^w.a2$ ,  $U_i^w.a3$ ,  $U_i^w.a4$ , dan  $U_i^w.a5$  dengan perintah SUM. Simbol hasil pusat *cluster* yaitu V. Selanjutnya menghitung masing-masing hasil pusat *cluster* dimana

nilai dari pusat *cluster* yaitu jumlah dari  $V_{ai} = \frac{U_i^w \cdot a_i}{U_i^w}$ . Sebagai contoh

$$V_{a1} = \frac{1116.40}{548.36} = 2.0359 \quad \text{Adapun perintah perhitungannya yaitu } = \mathbf{J557/SI\$557}.$$

Tabel 4.30. Perhitungan Jumlah Pusat *Cluster* 1

<b>Jumlah</b>	<b>548.36</b>	<b>1116.40</b>	<b>1203.01</b>	<b>91.23</b>	<b>1347.24</b>	<b>225.44</b>
<b>Pusat <i>Cluster</i> 1</b>		<b>2.0359</b>	<b>2.19385</b>	<b>0.16636</b>	<b>2.45687</b>	<b>0.41112</b>

Tabel 4.31. Perhitungan Jumlah Pusat *Cluster* 2

<b>Jumlah</b>	<b>548.34</b>	<b>1,116.24</b>	<b>1,202.81</b>	<b>91.23</b>	<b>1,347.04</b>	<b>225.35</b>
<b>Pusat <i>Cluster</i> 2</b>		<b>2.03566</b>	<b>2.19354</b>	<b>0.16637</b>	<b>2.45657</b>	<b>0.41097</b>

Tabel 4.32. Perhitungan Jumlah Pusat *Cluster* 3

<b>Jumlah</b>	<b>548.32</b>	<b>1,116.03</b>	<b>1,202.54</b>	<b>91.23</b>	<b>1,346.77</b>	<b>225.23</b>
<b>Pusat <i>Cluster</i> 3</b>		<b>2.03537</b>	<b>2.19313</b>	<b>0.16638</b>	<b>2.45617</b>	<b>0.41076</b>

1. Setelah memperoleh perhitungan jumlah pusat *cluster* pada masing-masing *cluster*, selanjutnya yaitu menghitung hasil pusat *cluster* (V). Adapun hasil pusat *cluster* yaitu jumlah pusat *cluster* pada masing-masing *cluster*.

Tabel 4.33. Hasil Pusat *Cluster* (V)

	<b>a1</b>	<b>a2</b>	<b>a3</b>	<b>a4</b>	<b>a5</b>
<b>V</b>	2.0359	2.19385	0.16636	2.45687	0.41112
	2.03566	2.19354	0.16637	2.45657	0.41097
	2.03537	2.19313	0.16638	2.45617	0.41076

- m. Setelah memperoleh hasil pusat *cluster* (V), langkah selanjutnya yaitu menghitung fungsi objektif dari kuadrat derajat keanggotaan. Simbol penentuan fungsi objektif adalah L. Nilai L diperoleh dari jumlah akar kuadrat hasil pusat *cluster* dan *data set* lalu dikalikan dengan derajat keanggotaan data ( $U_i^w$ ).

$$L1 = ((2 - 2.0359)^2 + (5 - 2.19385)^2 + (0 - 0.16636)^2 + (5 - 2.45687)^2 + (4 - 0.41112)^2) * 0.09 = 2.25167 \quad .$$

Adapun perintah pada LibreOffice Calc yaitu **=SUM((D7-\$A\$56)^2 , (E7-\$A\$56)^2, (F7-\$A\$56)^2 , (G7-\$A\$56)^2 , (H7-\$A\$56)^2)\*AY6**.

Tabel 4.34. Menghitung Fungsi Objektif

Kuadrat derajat keanggotaan data ke-i			L1	L2	L3	Jumlah
U1 <sup>w</sup>	U2 <sup>w</sup>	U3 <sup>w</sup>				L1+L2+L3
0.09	0.09	0.16	2.2517	2.4065	4.7558	9.4139
0.09	0.16	0.09	0.6830	1.0869	0.5079	2.2778
0.16	0.09	0.09	0.1311	0.1106	0.1815	0.4232
...	...	...	...	...	...	...
0.01	0.36	0.09	0.0176	0.7387	0.2346	0.9949
0.25	0.16	0.01	30.2269	20.0753	1.3467	51.3958
Fungsi Objektif (P1)						6322.7713

- n. Setelah menghitung fungsi objektif, langkah selanjutnya yaitu menghitung matriks partisi U. Matriks partisi U diperoleh dari jumlah akar kadrat hasil pusat *cluster* (V) dikurang dengan *data set* lalu di.pangkatkan dengan negatif 1. Simbol matriks parisi U yaitu

$$L1 = ((2 - 2.0359)^2 + (5 - 2.19385)^2 + (0 - 0.16636)^2 + (5 - 2.45687)^2 + (4 - 0.41112)^2)^{-1} = 0.03997$$

Rumus menghitung matriks partisi U yaitu =SUM(((\$R7-ASS\$6)^2,(\$E7-AT\$6)^2,(\$F7-AU\$6)^2,(\$G7-AV\$6)^2,(\$H7-AW\$6)^2)^-1.

Tabel 4.35. Menghitung Matriks Partisi U

L1	L2	L3	LT
0.039970371864261	0.037399376308295	0.033643272471686	0.111013020644241
0.131766492134546	0.147211806159768	0.177207121863638	0.456185420157952
1.22079691160299	0.813609789563584	0.495744910139636	2.53015161130621
...	...	...	...
0.568772218685905	0.487321161319101	0.383563678887177	1.43965705889218
0.008270776469664	0.007970005173062	0.007425750203234	0.02366653184596

- o. Setelah memperoleh nilai matriks partisi U, langkah selanjutnya yaitu menghitung keanggotaan *cluster* baru. Nilai keanggotaan *cluster* baru diperoleh dari hasil perbandingan antra jumlah LT dan LT<sub>i</sub>. Nilai keanggotaan *cluster* baru digunakan sebagai pusat *cluster* baru pada iterasi 3. Sebagai contoh,

$$U^1 * w_{baru} = \frac{0.03997}{0.11101} = 0.36005 \text{ . Rumus menghitung keanggotaan cluster baru}$$

yaitu = **BG6/BJ6**.

Tabel 4.36. Keanggotaan *Cluster* Baru

$U1^w$	$U2^w$	$U3^w$
0.36005120509559	0.336891799639856	0.303056995264553
0.288844154837133	0.322701690266201	0.388454154896667
0.482499509573952	0.321565627106255	0.195934863319793
...	...	...
0.395074795884775	0.338498087658525	0.266427116456699
0.349471419111871	0.336762700379466	0.313765880508664

- p. Langkah selanjutnya yaitu memeriksa kondisi selisih fungsi objektif dimana, jika  $P2 < P1$ , maka proses *clustering Fuzzy C-Means* berhenti. Pada iterasi 2, nilai fungsi objektif yang diperoleh yaitu  $P2 = 6322.77134$  dan  $P1 = 6320.3171$ . Sehingga  $P2 - P1 = 2.45424023$ . Maka iterasi dilanjutkan ke iterasi 3 karena nilai  $P2 > P1$ . Pada penelitian ini jumlah iterasi yang dilakukan yaitu sebanyak dua iterasi.

#### b. *Clustering K-Means dan Fuzzy C-Means dengan RStudio*

Proses *clustering* dengan menggunakan *K-Means* dan *Fuzzy C-Means* akan dilakukan terhadap 510 - 550 sampel data dengan rentang data sebanyak 10 data pada *data set* kabupaten/kota di Indonesia. Hasil *clustering* akan diperoleh kelompok daerah yang terdiri atas tiga sesi, yaitu 2 *cluster*, 3 *cluster*, dan 4 *cluster*. Penelitian ini akan menggunakan bahasa pemrograman R versi 3.4.4 dan *software* RStudio versi 1.1.456 yang berjalan diatas sistem operasi Linux Mint 18.1 arsitektur 64 bit.

##### 1. *Clustering pada K-Means*

###### a. *Clustering pada K-Means dengan 510 data menggunakan 2 cluster*

1. Untuk melakukan *clustering* pada *K-Means*, data yang digunakan adalah data yang sudah melalui proses *preprocessing* dengan *cluster* awal (acak) berjumlah dua *cluster*. Perintah untuk proses *K-Means* yaitu `kmeans.proc <- kmeans(sd, centers = 2)`, dimana `kmeans.proc` adalah variabel untuk menampung data pengolahan *K-Means*, `kmeans` adalah fungsi menjalankan *K-Means*, `sd` adalah

data yang telah dinormalisasi, dan `centers = 2` adalah jumlah *cluster* awal. Berikut adalah hasil *clustering* dengan *K-Means*.

K-means clustering with 2 clusters of sizes 424, 86

Cluster means:

	Kematian.Neonatal	Kematian.Bayi	Kematian.AnakBalita	Kematian.Balita	Kematian.Ibu
1	-0,3232128	-0,3299082	-0,1357180	-0,3483602	-0,1450036
2	1,5935143	1,6265241	0,6691214	1,7174966	0,7149014

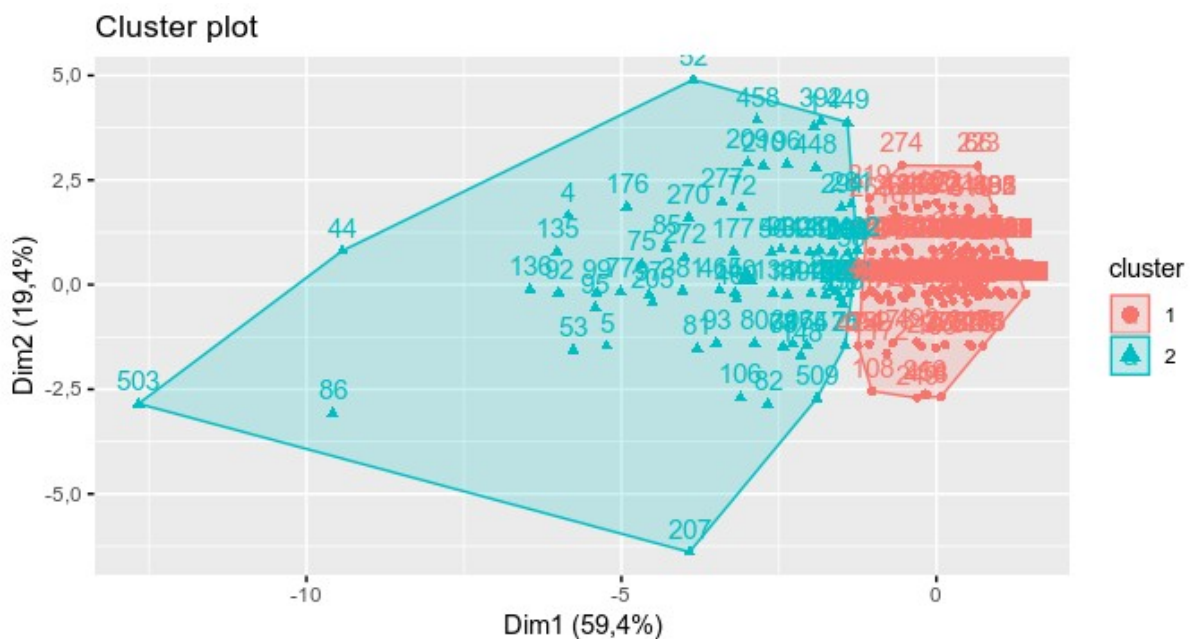
Gambar 4.1. Hasil *Clustering* dengan *K-Means*

Pada proses *K-Means* diperoleh jumlah anggota *cluster* pada *cluster* 1 = 424 data dan jumlah anggota *cluster* pada *cluster* 2 = 86 data.

Tabel 4.37. *Cluster Means* dan Jumlah Anggota *Cluster* pada *K-Means*

<i>Cluster</i>	Jumlah Anggota <i>Cluster</i>	Kematian. Neonatal	Kematian. Bayi	Kematian. AnakBalita	Kematian. Balita	Kematian. Ibu
1	424	-0.3299082	-0.3299082	-0.1357180	-0.3483602	-0.1450036
2	86	1.5935143	1.6265241	0.6691214	1.7174966	0.7149014

- Setelah memperoleh hasil *clustering*, langkah selanjutnya yaitu menggambarkan distribusi data dengan jumlah *cluster* = 2.



Gambar 4.2. Pengelompokan Distribusi Data pada *K-Means*

Setelah dilakukan proses *clustering* menggunakan *K-Means*, diperoleh sebaran distribusi data dimana warna merah merupakan daerah *cluster* 1 dengan jumlah data sebanyak 424 data dan warna biru merupakan daerah *cluster* 2 dengan jumlah data sebanyak 86 data.

**b. Clustering pada *K-Means* dengan 510 data menggunakan 3 cluster**

1. Untuk melakukan *clustering* pada *K-Means*, data yang digunakan adalah data yang sudah melalui proses *preprocessing* dengan *cluster* awal (acak) berjumlah tiga *cluster*. Perintah untuk proses *K-Means* yaitu `kmeans.proc <- kmeans(sd, centers = 3)`, dimana `kmeans.proc` adalah variabel untuk menampung data pengolahan *K-Means*, `kmeans` adalah fungsi menjalankan *K-Means*, `sd` adalah data yang telah dinormalisasi, dan `centers = 3` adalah jumlah *cluster* awal. Berikut adalah hasil *clustering* dengan *K-Means*.

K-means clustering with 3 clusters of sizes 355, 117, 38

Cluster means:

	Kematian.Neonatal	Kematian.Bayi	Kematian.AnakBalita	Kematian.Balita	Kematian.Ibu
1	-0,4234955	-0,4161455	-0,1171221	-0,4395787	-0,3568792
2	0,5553430	0,4781361	-0,1654208	0,4877254	0,9977725
3	2,2464678	2,4155191	1,6034892	2,6049095	0,2619143

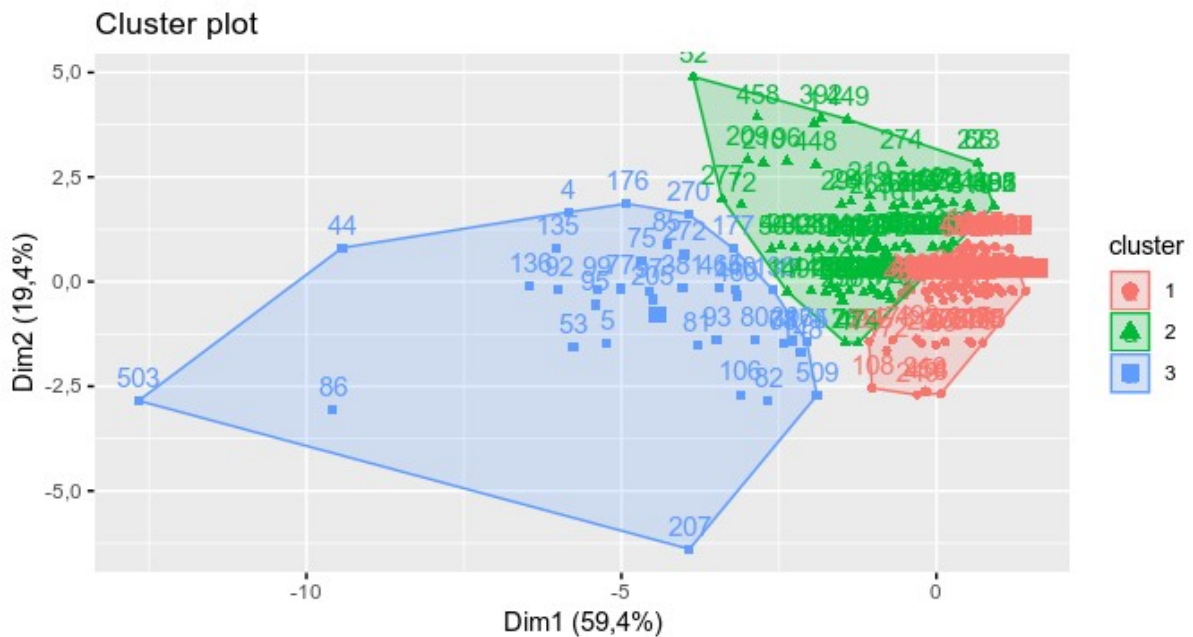
Gambar 4.3. Hasil *Clustering* dengan *K-Means*

Pada proses *K-Means* diperoleh jumlah anggota *cluster* pada *cluster* 1 = 355 data, jumlah anggota *cluster* pada *cluster* 2 = 117 data, dan jumlah anggota *cluster* pada *cluster* 3 = 38 data.

Tabel 4.38. *Cluster Means* dan Jumlah Anggota *Cluster* pada *K-Means*

<i>Cluster</i>	Jumlah Anggota <i>Cluster</i>	Kematian. Neonatal	Kematian. Bayi	Kematian. AnakBalita	Kematian. Balita	Kematian. Ibu
1	355	-0.4234955	-0.4161455	-0.1171221	-0.4395787	-0.3568792
2	117	0.5553430	0.4781361	-0.1654208	0.4877254	0.9977725
	38	2.2464678	2.4155191	1.6034892	2.6049095	0.2619143

2. Setelah memperoleh hasil *clustering*, langkah selanjutnya yaitu menggambarkan distribusi data dengan jumlah *cluster* = 3.



Gambar 4.4. Pengelompokan Distribusi Data pada *K-Means*

Setelah dilakukan proses *clustering* menggunakan *K-Means*, diperoleh sebaran distribusi data dimana warna merah merupakan daerah *cluster* 1 dengan jumlah data sebanyak 355 data, warna hijau merupakan daerah *cluster* 2 dengan jumlah data sebanyak 117 data, dan warna biru merupakan daerah *cluster* 3 dengan jumlah data sebanyak 38 data.

### c. *Clustering* pada *K-Means* dengan 510 data menggunakan 4 *cluster*

1. Untuk melakukan *clustering* pada *K-Means*, data yang digunakan adalah data yang sudah melalui proses *preprocessing* dengan *cluster* awal (acak) berjumlah empat *cluster*. Perintah untuk proses *K-Means* yaitu `kmeans.proc <- kmeans(sd, centers = 4)`, dimana `kmeans.proc` adalah variabel untuk menampung data pengolahan *K-Means*, `kmeans` adalah fungsi menjalankan *K-Means*, `sd` adalah data yang telah dinormalisasi, dan `centers = 4` adalah jumlah *cluster* awal. Berikut adalah hasil *clustering* dengan *K-Means*.



K-means clustering with 4 clusters of sizes 317, 127, 30, 36

Cluster means:

	Kematian.Neonatal	Kematian.Bayi	Kematian.AnakBalita	Kematian.Balita	Kematian.Ibu
1	-0,4990205	-0,51549456	-0,1491464	-0,5220461	-0,32436243
2	0,5079327	0,61128715	0,0711423	0,5586194	-0,07475291
3	2,6223223	2,75278096	1,5238949	2,9402517	0,42851373
4	0,4170105	0,08874665	-0,2075694	0,1760107	2,76280827

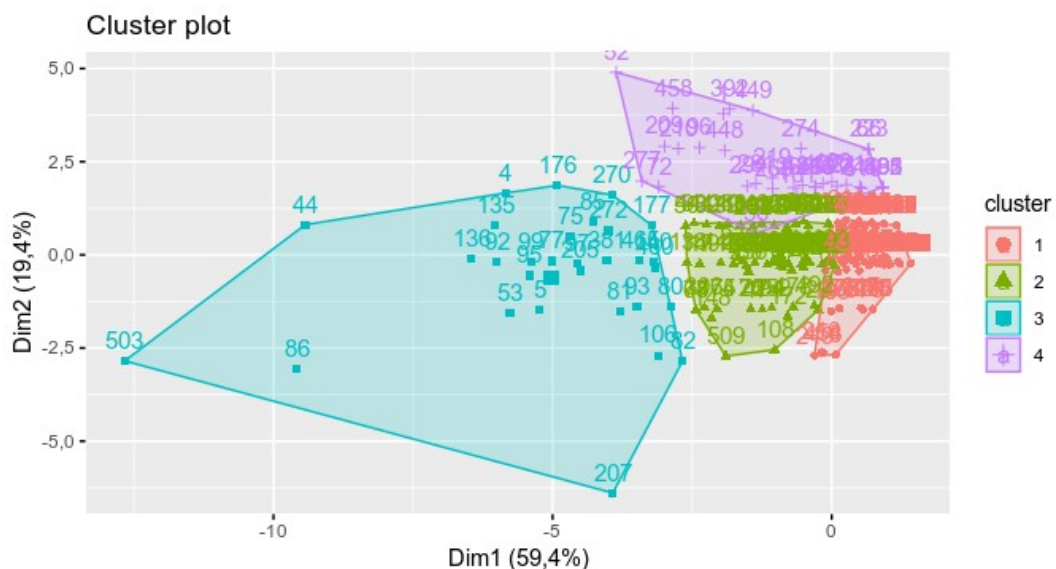
Gambar 4.5. Hasil *Clustering* dengan *K-Means*

Pada proses *K-Means* diperoleh jumlah anggota *cluster* pada *cluster* 1 = 317 data, jumlah anggota *cluster* pada *cluster* 2 = 127 data, jumlah anggota *cluster* pada *cluster* 3 = 30 data, dan jumlah anggota *cluster* pada *cluster* 4 = 36 data.

Tabel 4.39. *Cluster Means* dan Jumlah Anggota *Cluster* pada *K-Means*

<i>Cluster</i>	Jumlah Anggota <i>Cluster</i>	Kematian.Neonatal	Kematian.Bayi	Kematian.AnakBalita	Kematian.Balita	Kematian.Ibu
1	317	-0.4990205	-0.51549456	-0.1491464	-0.5220461	-0.32436243
2	127	0.5079327	0.61128715	0.0711423	0.5586194	-0.07475291
3	30	2.6223223	2.75278096	1.5238949	2.9402517	0.42851373
4	36	0.4170105	0.08874665	-0.2075694	0.1760107	2.76280827

- Setelah memperoleh hasil *clustering*, langkah selanjutnya yaitu menggambarkan distribusi data dengan jumlah *cluster* = 4.



Gambar 4.6. Pengelompokan Distribusi Data pada *K-Means*

Setelah dilakukan proses *clustering* menggunakan *K-Means*, diperoleh sebaran distribusi data dimana warna merah merupakan daerah *cluster* 1 dengan jumlah data sebanyak 317 data, warna hijau merupakan daerah *cluster* 2 dengan jumlah data sebanyak 127 data, warna biru merupakan daerah *cluster* 3 dengan jumlah data sebanyak 30 data, dan warna ungu merupakan daerah *cluster* 4 dengan jumlah data sebanyak 36 data.

**d. *Clustering* pada *K-Means* dengan 520 data menggunakan 2 *cluster***

1. Untuk melakukan *clustering* pada *K-Means*, data yang digunakan adalah data yang sudah melalui proses *preprocessing* dengan *cluster* awal (acak) berjumlah dua *cluster*. Perintah untuk proses *K-Means* yaitu `kmeans.proc2 <- kmeans(sd2, centers = 2)`, dimana `kmeans.proc2` adalah variabel untuk menampung data pengolahan *K-Means*, `kmeans` adalah fungsi menjalankan *K-Means*, `sd2` adalah data yang telah dinormalisasi, dan `centers = 2` adalah jumlah *cluster* awal. Berikut adalah hasil *clustering* dengan *K-Means*.

K-means clustering with 2 clusters of sizes 89, 431

Cluster means:

	Kematian.Neonatal	Kematian.Bayi	Kematian.AnakBalita	Kematian.Balita	Kematian.Ibu
1	1,5719610	1,5945086	0,7034016	1,698779	0,6959962
2	-0,3246045	-0,3292605	-0,1452500	-0,350792	-0,1437208

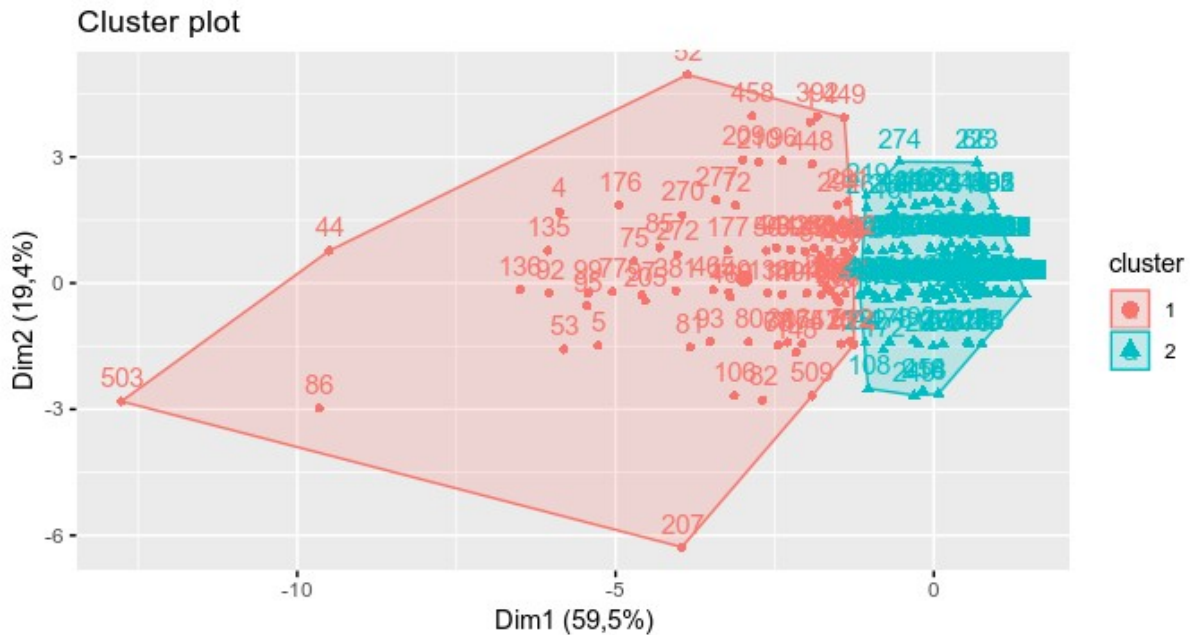
Gambar 4.7. Hasil *Clustering* dengan *K-Means*

Pada proses *K-Means* diperoleh jumlah anggota *cluster* pada *cluster* 1 = 89 data dan jumlah anggota *cluster* pada *cluster* 2 = 431 data.

Tabel 4.40. *Cluster Means* dan Jumlah Anggota *Cluster* pada *K-Means*

<i>Cluster</i>	Jumlah Anggota <i>Cluster</i>	Kematian. Neonatal	Kematian. Bayi	Kematian. AnakBalita	Kematian. Balita	Kematian. Ibu
1	89	1.5719610	1.5945086	0.7034016	1.698779	0.6959962
2	431	-0.3246045	-0.3292605	-0.1452500	-0.350792	-0.1437208

2. Setelah memperoleh hasil *clustering*, langkah selanjutnya yaitu menggambarkan distribusi data dengan jumlah *cluster* = 2.



Gambar 4.8. Pengelompokan Distribusi Data pada *K-Means*

Setelah dilakukan proses *clustering* menggunakan *K-Means*, diperoleh sebaran distribusi data dimana warna merah merupakan daerah *cluster* 1 dengan jumlah data sebanyak 89 data dan warna biru merupakan daerah *cluster* 2 dengan jumlah data sebanyak 431 data.

**e. *Clustering* pada *K-Means* dengan 520 data menggunakan 3 *cluster***

1. Untuk melakukan *clustering* pada *K-Means*, data yang digunakan adalah data yang sudah melalui proses *preprocessing* dengan *cluster* awal (acak) berjumlah tiga *cluster*. Perintah untuk proses *K-Means* yaitu `kmeans.proc2 <- kmeans(sd2, centers = 3)`, dimana `kmeans.proc2` adalah variabel untuk menampung data pengolahan *K-Means*, `kmeans` adalah fungsi menjalankan *K-Means*, `sd2` adalah data yang telah dinormalisasi, dan `centers = 3` adalah jumlah *cluster* awal. Berikut adalah hasil *clustering* dengan *K-Means*.

K-means clustering with 3 clusters of sizes 125, 330, 65

Cluster means:

	Kematian.Neonatal	Kematian.Bayi	Kematian.AnakBalita	Kematian.Balita	Kematian.Ibu
1	-0,04360283	-0,1270039	-0,1139062	-0,08678281	1,3320281
2	-0,34610306	-0,3325873	-0,1338227	-0,35841015	-0,5336731
3	1,84099020	1,9327582	0,8984582	1,98651076	0,1478250

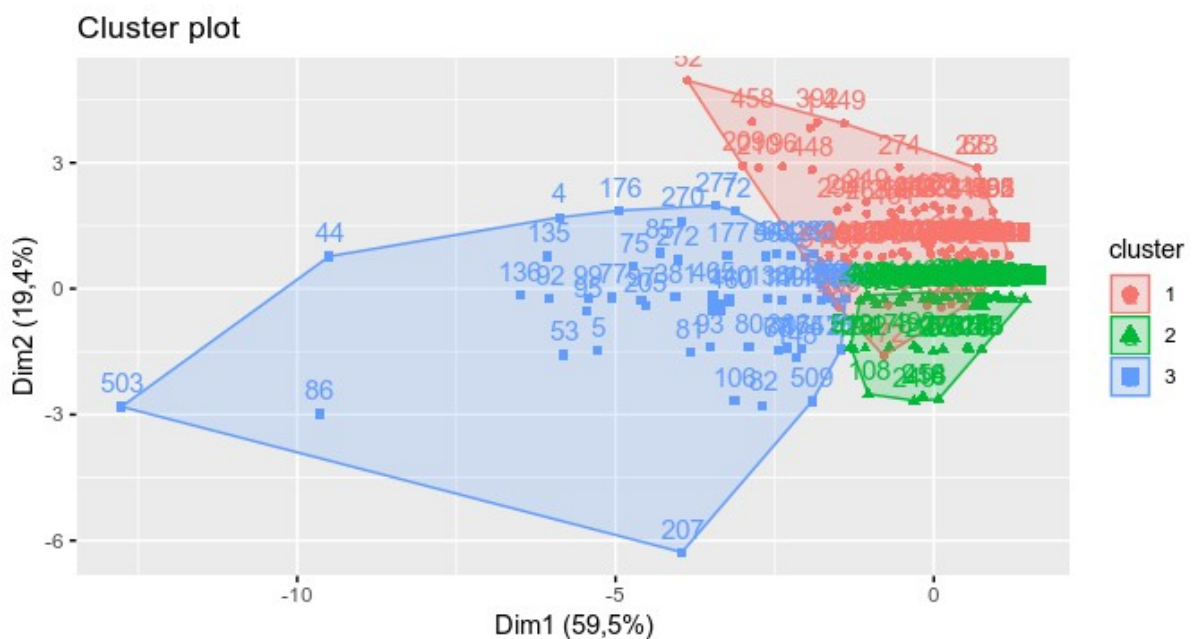
Gambar 4.9. Hasil *Clustering* dengan *K-Means*

Pada proses *K-Means* diperoleh jumlah anggota *cluster* pada *cluster* 1 = 125 data, jumlah anggota *cluster* pada *cluster* 2 = 330 data, dan jumlah anggota *cluster* pada *cluster* 3 = 65 data.

Tabel 4.41. *Cluster Means* dan Jumlah Anggota *Cluster* pada *K-Means*

<i>Cluster</i>	Jumlah Anggota <i>Cluster</i>	Kematian. Neonatal	Kematian. Bayi	Kematian. Anak Balita	Kematian. Balita	Kematian. Ibu
1	125	-0.04360283	-0.1270039	-0.1139062	-0.08678281	1.3320281
2	330	-0.34610306	-0.3325873	-0.1338227	-0.35841015	-0.5336731
3	65	1.84099020	1.9327582	0.8984582	1.98651076	0.1478250

- Setelah memperoleh hasil *clustering*, langkah selanjutnya yaitu menggambarkan distribusi data dengan jumlah *cluster* = 3.



Gambar 4.10. Pengelompokan Distribusi Data pada *K-Means*

Setelah dilakukan proses *clustering* menggunakan *K-Means*, diperoleh sebaran distribusi data dimana warna merah merupakan daerah *cluster* 1 dengan jumlah data sebanyak 125 data, warna hijau merupakan daerah *cluster* 2 dengan jumlah data sebanyak 330 data, dan warna biru merupakan daerah *cluster* 3 dengan jumlah data sebanyak 65 data.

**f. Clustering pada K-Means dengan 520 data menggunakan 4 cluster**

1. Untuk melakukan *clustering* pada *K-Means*, data yang digunakan adalah data yang sudah melalui proses *preprocessing* dengan *cluster* awal (acak) berjumlah empat *cluster*. Perintah untuk proses *K-Means* yaitu `kmeans.proc2 <- kmeans(sd2, centers = 4)`, dimana `kmeans.proc2` adalah variabel untuk menampung data pengolahan *K-Means*, `kmeans` adalah fungsi menjalankan *K-Means*, `sd2` adalah data yang telah dinormalisasi, dan `centers = 4` adalah jumlah *cluster* awal. Berikut adalah hasil *clustering* dengan *K-Means*.

K-means clustering with 4 clusters of sizes 80, 100, 25, 315

Cluster means:

	Kematian.Neonatal	Kematian.Bayi	Kematian.AnakBalita	Kematian.Balita	Kematian.Ibu
1	0,4732442	0,3078148	-0,2455532	0,3067824	1,6811956
2	0,4464257	0,5427789	0,7690193	0,6235384	-0,4164154
3	2,9394780	3,1859002	0,9471882	3,1772810	0,5086181
4	-0,4952034	-0,5033352	-0,2569441	-0,5280268	-0,3351415

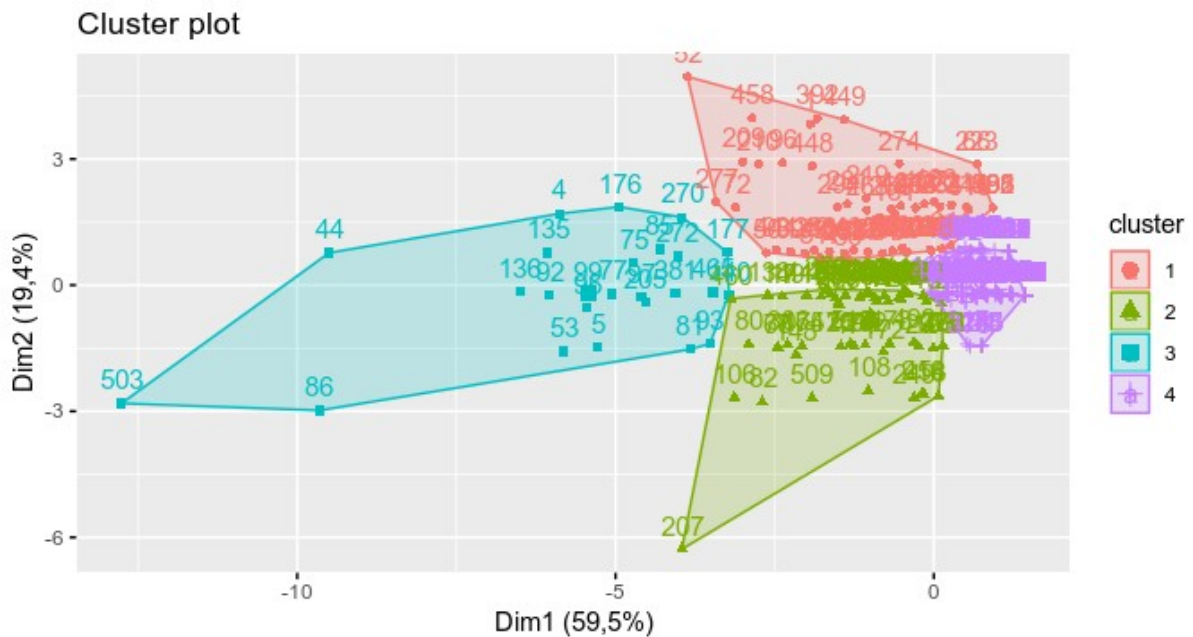
Gambar 4.11. Hasil *Clustering* dengan *K-Means*

Pada proses *K-Means* diperoleh jumlah anggota *cluster* pada *cluster* 1 = 80 data, jumlah anggota *cluster* pada *cluster* 2 = 100 data, jumlah anggota *cluster* pada *cluster* 3 = 25 data, dan jumlah anggota *cluster* pada *cluster* 4 = 315 data.

Tabel 4.42. *Cluster Means* dan Jumlah Anggota *Cluster* pada *K-Means*

<i>Cluster</i>	Jumlah Anggota <i>Cluster</i>	Kematian. Neonatal	Kematian. Bayi	Kematian. AnakBalita	Kematian. Balita	Kematian. Ibu
1	80	0.4732442	0.3078148	-0.2455532	0.3067824	1.6877956
2	100	0.4464257	0.5427789	0.7690193	0.6235384	-0.4164154
3	25	2.9394780	3.1859002	0.9471882	3.1772810	0.5086181
4	315	-0.4952034	-0.5033352	-0.2569441	-0.5280268	-0.3351415

2. Setelah memperoleh hasil *clustering*, langkah selanjutnya yaitu menggambarkan distribusi data dengan jumlah *cluster* = 4.



Gambar 4.12. Pengelompokan Distribusi Data pada *K-Means*

Setelah dilakukan proses *clustering* menggunakan *K-Means*, diperoleh sebaran distribusi data dimana warna merah merupakan daerah *cluster* 1 dengan jumlah data sebanyak 80 data, warna hijau merupakan daerah *cluster* 2 dengan jumlah data sebanyak 100 data, warna biru merupakan daerah *cluster* 3 dengan jumlah data sebanyak 25 data, dan warna ungu merupakan daerah *cluster* 4 dengan jumlah data sebanyak 315 data.

**g. *Clustering* pada *K-Means* dengan 530 data menggunakan 2 *cluster***

1. Untuk melakukan *clustering* pada *K-Means*, data yang digunakan adalah data yang sudah melalui proses *preprocessing* dengan *cluster* awal (acak) berjumlah dua *cluster*. Perintah untuk proses *K-Means* yaitu `kmeans.proc3 <- kmeans(sd3, centers = 2)`, dimana `kmeans.proc3` adalah variabel untuk menampung data pengolahan *K-Means*, `kmeans` adalah fungsi menjalankan *K-Means*, `sd3` adalah data yang telah dinormalisasi, dan `centers = 2` adalah jumlah *cluster* awal. Berikut adalah hasil *clustering* dengan *K-Means*.



K-means clustering with 2 clusters of sizes 93, 437

Cluster means:

	Kematian.Neonatal	Kematian.Bayi	Kematian.AnakBalita	Kematian.Balita	Kematian.Ibu
1	1,5654660	1,5930345	0,6816136	1,6777762	0,7229177
2	-0,3331541	-0,3390211	-0,1450573	-0,3570553	-0,1538475

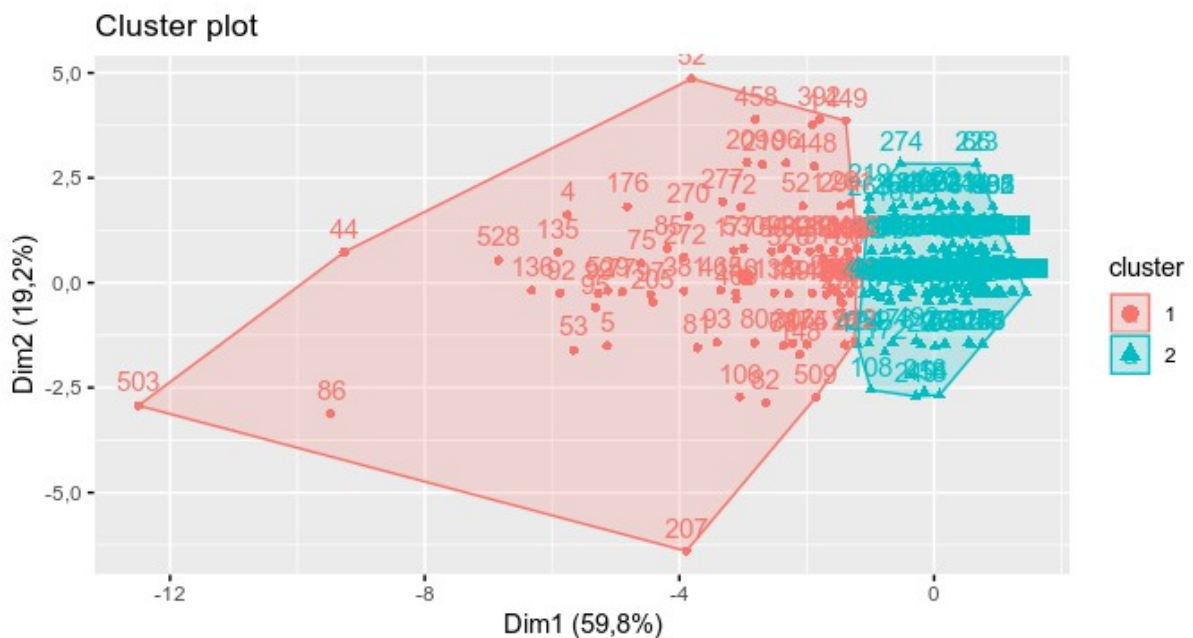
Gambar 4.13. Hasil *Clustering* dengan *K-Means*

Pada proses *K-Means* diperoleh jumlah anggota *cluster* pada *cluster* 1 = 93 data dan jumlah anggota *cluster* pada *cluster* 2 = 437 data.

Tabel 4.43. *Cluster Means* dan Jumlah Anggota *Cluster* pada *K-Means*

<i>Cluster</i>	Jumlah Anggota <i>Cluster</i>	Kematian.Neonatal	Kematian.Bayi	Kematian.AnakBalita	Kematian.Balita	Kematian.Ibu
1	93	1.5654660	1.5930345	0.6816136	1.6777762	0.7229177
2	437	-0.3331541	-0.3390211	-0.1450573	-0.3570553	-0.1538475

- Setelah memperoleh hasil *clustering*, langkah selanjutnya yaitu menggambarkan distribusi data dengan jumlah *cluster* = 2.



Gambar 4.14. Pengelompokan Distribusi Data pada *K-Means*

Setelah dilakukan proses *clustering* menggunakan *K-Means*, diperoleh sebaran distribusi data dimana warna merah merupakan daerah *cluster* 1 dengan jumlah data sebanyak 93 data dan warna biru merupakan daerah *cluster* 2 dengan jumlah data sebanyak 437 data.

#### h. *Clustering* pada *K-Means* dengan 530 data menggunakan 3 *cluster*

1. Untuk melakukan *clustering* pada *K-Means*, data yang digunakan adalah data yang sudah melalui proses *preprocessing* dengan *cluster* awal (acak) berjumlah empat *cluster*. Perintah untuk proses *K-Means* yaitu `kmeans.proc3 <- kmeans(sd3, centers = 3)`, dimana `kmeans.proc3` adalah variabel untuk menampung data pengolahan *K-Means*, `kmeans` adalah fungsi menjalankan *K-Means*, `sd3` adalah data yang telah dinormalisasi, dan `centers = 3` adalah jumlah *cluster* awal. Berikut adalah hasil *clustering* dengan *K-Means*.

K-means clustering with 3 clusters of sizes 69, 410, 51

Cluster means:

	Kematian.Neonatal	Kematian.Bayi	Kematian.AnakBalita	Kematian.Balita	Kematian.Ibu
1	1,9026627	1,87145080	0,2534672	1,8969364	0,97535644
2	-0,3103970	-0,31888703	-0,3223478	-0,3632984	-0,15355188
3	-0,0788421	0,03163877	2,2484969	0,3541906	-0,08516323

Gambar 4.15. Hasil *Clustering* dengan *K-Means*

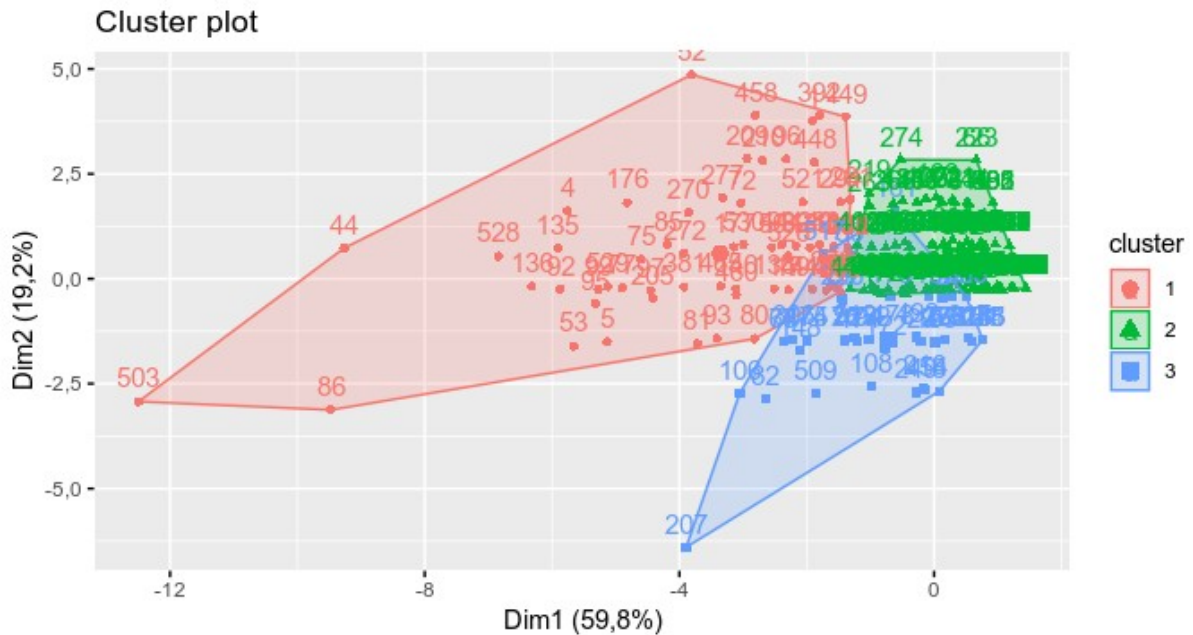
Pada proses *K-Means* diperoleh jumlah anggota *cluster* pada *cluster* 1 = 69 data, jumlah anggota *cluster* pada *cluster* 2 = 410 data, dan jumlah anggota *cluster* pada *cluster* 3 = 51 data,.

Tabel 4.44. *Cluster Means* dan Jumlah Anggota *Cluster* pada *K-Means*

<i>Cluster</i>	Jumlah Anggota <i>Cluster</i>	Kematian. Neonatal	Kematian. Bayi	Kematian. AnakBalita	Kematian. Balita	Kematian. Ibu
1	69	1.9026627	1.87145080	0.2534672	1.8969364	0.97535644
2	410	-0.3103970	-0.31888703	-0.3223478	-0.3632984	-0.15355188
3	51	-0.0788421	0.03163877	2.2484969	0.3541906	-0.08516323

2. Setelah memperoleh hasil *clustering*, langkah selanjutnya yaitu menggambarkan distribusi data dengan jumlah *cluster* = 3.





Gambar 4.16. Pengelompokan Distribusi Data pada *K-Means*

Setelah dilakukan proses *clustering* menggunakan *K-Means*, diperoleh sebaran distribusi data dimana warna merah merupakan daerah *cluster* 1 dengan jumlah data sebanyak 69 data, warna hijau merupakan daerah *cluster* 2 dengan jumlah data sebanyak 410 data, dan warna biru merupakan daerah *cluster* 3 dengan jumlah data sebanyak 51 data.

#### i. *Clustering* pada *K-Means* dengan 530 data menggunakan 4 *cluster*

1. Untuk melakukan *clustering* pada *K-Means*, data yang digunakan adalah data yang sudah melalui proses *preprocessing* dengan *cluster* awal (acak) berjumlah empat *cluster*. Perintah untuk proses *K-Means* yaitu `kmeans.proc3 <- kmeans(sd3, centers = 4)`, dimana `kmeans.proc3` adalah variabel untuk menampung data pengolahan *K-Means*, `kmeans` adalah fungsi menjalankan *K-Means*, `sd3` adalah data yang telah dinormalisasi, dan `centers = 4` adalah jumlah *cluster* awal. Berikut adalah hasil *clustering* dengan *K-Means*.

K-means clustering with 4 clusters of sizes 26, 146, 11, 347

Cluster means:

	Kematian.Neonatal	Kematian.Bayi	Kematian.AnakBalita	Kematian.Balita	Kematian.Ibu
1	2,95343904	3,1717503	0,97655800	3,1685799	0,5022778
2	0,63405973	0,5783280	-0,02300283	0,5768177	0,6268282
3	-0,09453018	0,1938907	4,37316182	0,5796302	-0,1887824
4	-0,48507868	-0,4871302	-0,20212356	-0,4984853	-0,2953877

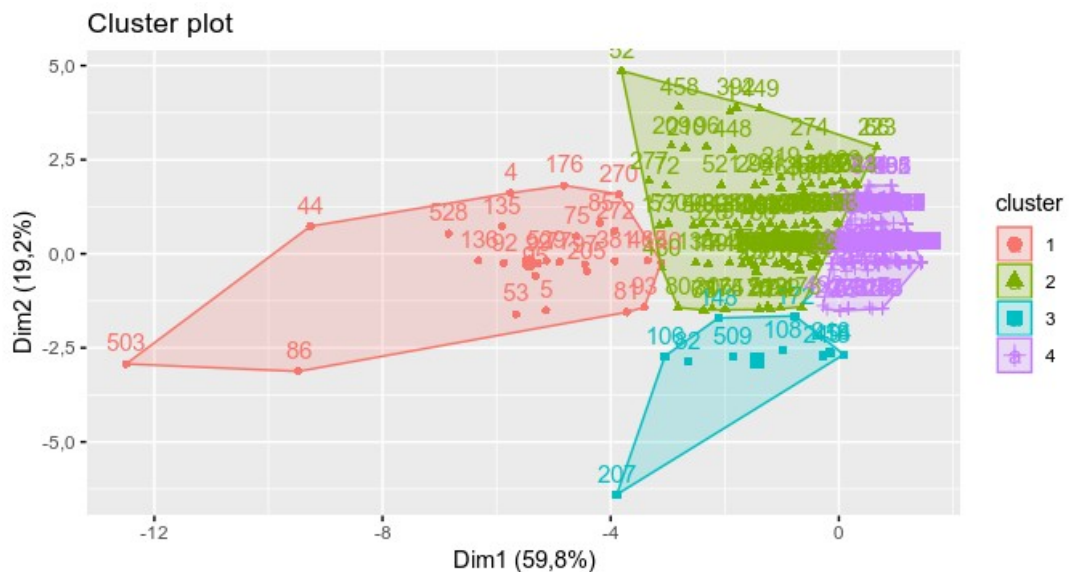
Gambar 4.17. Hasil *Clustering* dengan *K-Means*

Pada proses *K-Means* diperoleh jumlah anggota *cluster* pada *cluster* 1 = 26 data, jumlah anggota *cluster* pada *cluster* 2 = 146 data, jumlah anggota *cluster* pada *cluster* 3 = 11 data, dan jumlah anggota *cluster* pada *cluster* 4 = 347 data.

Tabel 4.45. *Cluster Means* dan Jumlah Anggota *Cluster* pada *K-Means*

<i>Cluster</i>	Jumlah Anggota <i>Cluster</i>	Kematian.Neonatal	Kematian.Bayi	Kematian.A nakBalita	Kematian.Balita	Kematian.Ibu
1	26	2.95343904	3.1717503	0.97655800	3.1685799	0.5022778
2	146	0.63405973	0.5783280	-0.02300283	0.5768177	0.6268282
3	11	-0.09453018	0.1938907	4.37316482	0.5796302	-0.1887824
4	437	-0.48507868	-0.4871302	-0.20212356	-0.4984853	-0.2953877

- Setelah memperoleh hasil *clustering*, langkah selanjutnya yaitu menggambarkan distribusi data dengan jumlah *cluster* = 4.



Gambar 4.18. Pengelompokan Distribusi Data pada *K-Means*

Setelah dilakukan proses *clustering* menggunakan *K-Means*, diperoleh sebaran distribusi data dimana warna merah merupakan daerah *cluster* 1 dengan jumlah data sebanyak 26 data, warna hijau merupakan daerah *cluster* 2 dengan jumlah data sebanyak 146 data, warna biru merupakan daerah *cluster* 3 dengan jumlah data sebanyak 11 data, dan warna ungu merupakan daerah *cluster* 4 dengan jumlah data sebanyak 437 data.

**j. *Clustering* pada *K-Means* dengan 540 data menggunakan 2 *cluster***

1. Untuk melakukan *clustering* pada *K-Means*, data yang digunakan adalah data yang sudah melalui proses *preprocessing* dengan *cluster* awal (acak) berjumlah dua *cluster*. Perintah untuk proses *K-Means* yaitu `kmeans.proc4 <- kmeans(sd4, centers = 2)`, dimana `kmeans.proc4` adalah variabel untuk menampung data pengolahan *K-Means*, `kmeans` adalah fungsi menjalankan *K-Means*, `sd4` adalah data yang telah dinormalisasi, dan `centers = 2` adalah jumlah *cluster* awal. Berikut adalah hasil *clustering* dengan *K-Means*.

```
K-means clustering with 2 clusters of sizes 97, 443

Cluster means:
  Kematian.Neonatal Kematian.Bayi Kematian.AnakBalita Kematian.Balita
1      1.5401245      1.5462098      0.6567134      1.6453747
2     -0.3372282     -0.3385606     -0.1437950     -0.3602739
  Kematian.Ibu
1      0.7263036
2     -0.1590326
```

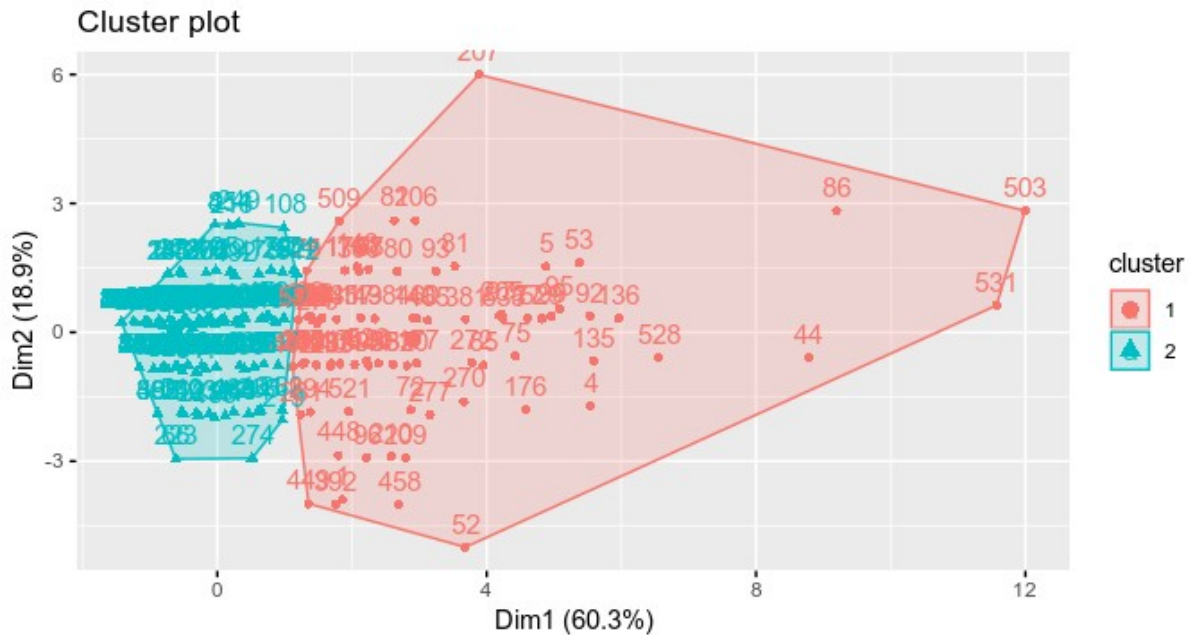
Gambar 4.19. Hasil *Clustering* dengan *K-Means*

Pada proses *K-Means* diperoleh jumlah anggota *cluster* pada *cluster* 1 = 97 data dan jumlah anggota *cluster* pada *cluster* 2 = 443 data.

Tabel 4.46. *Cluster Means* dan Jumlah Anggota *Cluster* pada *K-Means*

<i>Cluster</i>	Jumlah Anggota <i>Cluster</i>	Kematian. Neonatal	Kematian. Bayi	Kematian. AnakBalita	Kematian. Balita	Kematian. Ibu
1	97	1.5401245	1.5462098	0.6567134	1.6453747	0.7263036
2	443	-0.3372282	-0.3385606	-0.1437950	-0.3602739	-0.1590326

2. Setelah memperoleh hasil *clustering*, langkah selanjutnya yaitu menggambarkan distribusi data dengan jumlah *cluster* = 2.

Gambar 4.20. Pengelompokan Distribusi Data pada *K-Means*

Setelah dilakukan proses *clustering* menggunakan *K-Means*, diperoleh sebaran distribusi data dimana warna merah merupakan daerah *cluster* 1 dengan jumlah data sebanyak 97 data dan warna biru merupakan daerah *cluster* 2 dengan jumlah data sebanyak 443 data.

#### k. *Clustering* pada *K-Means* dengan 540 data menggunakan 3 *cluster*

1. Untuk melakukan *clustering* pada *K-Means*, data yang digunakan adalah data yang sudah melalui proses *preprocessing* dengan *cluster* awal (acak) berjumlah tiga *cluster*. Perintah untuk proses *K-Means* yaitu `kmeans.proc4 <- kmeans(sd4, centers = 3)`, dimana `kmeans.proc4` adalah variabel untuk menampung data pengolahan *K-Means*, `kmeans` adalah fungsi menjalankan *K-Means*, `sd4` adalah data yang telah dinormalisasi, dan `centers = 3` adalah jumlah *cluster* awal. Berikut adalah hasil *clustering* dengan *K-Means*.

K-means clustering with 3 clusters of sizes 45, 421, 74

Cluster means:

	Kematian.Neonatal	Kematian.Bayi	Kematian.AnakBalita	Kematian.Balita	Kematian.Ibu
1	2,2105325	2,3285070	1,4364284	2,4907119	0,2637912
2	-0,3184295	-0,3136063	-0,1297583	-0,3321998	-0,3420343
3	0,4673629	0,3681814	-0,1352841	0,3753254	1,7854843

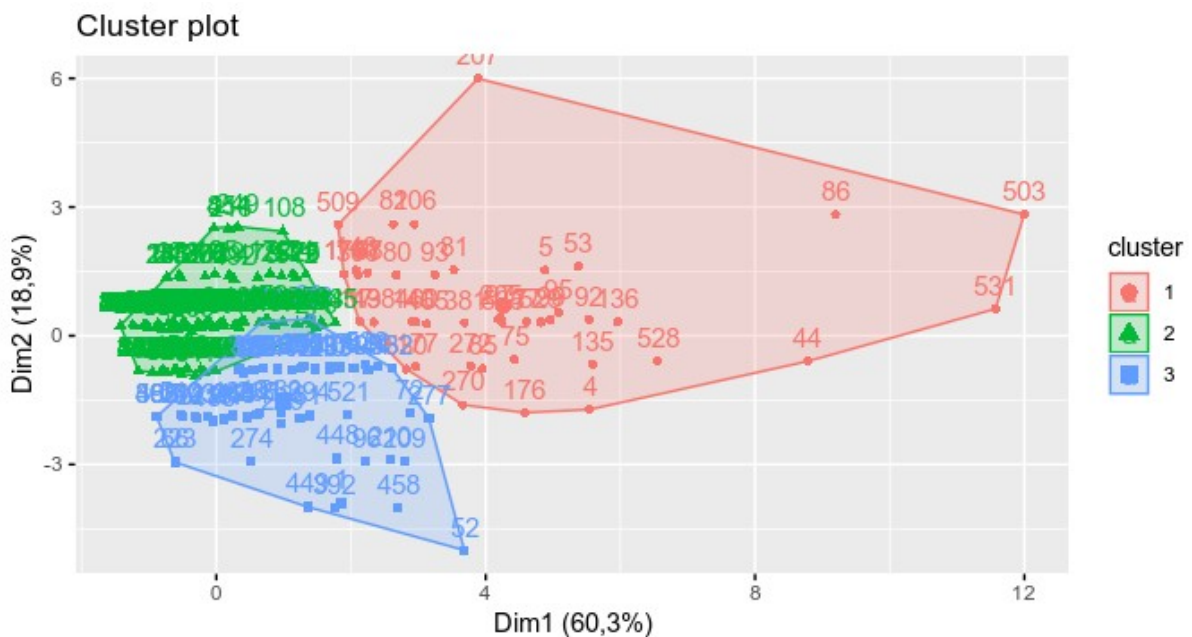
Gambar 4.21. Hasil *Clustering* dengan *K-Means*

Pada proses *K-Means* diperoleh jumlah anggota *cluster* pada *cluster* 1 = 45 data, jumlah anggota *cluster* pada *cluster* 2 = 421 data, dan jumlah anggota *cluster* pada *cluster* 3 = 74 data.

Tabel 4.47. *Cluster Means* dan Jumlah Anggota *Cluster* pada *K-Means*

<i>Cluster</i>	Jumlah Anggota <i>Cluster</i>	Kematian. Neonatal	Kematian. Bayi	Kematian. Anak Balita	Kematian. Balita	Kematian. Ibu
1	45	2.2105325	2.3285070	1.4364284	2.4907119	0.2637912
2	421	-0.3184295	-0.3136063	-0.1297583	-0.3321998	-0.3420343
3	74	0.4673629	0.3681814	-0.1352841	0.3753254	1.7854843

- Setelah memperoleh hasil *clustering*, langkah selanjutnya yaitu menggambarkan distribusi data dengan jumlah *cluster* = 3.



Gambar 4.22. Pengelompokan Distribusi Data pada *K-Means*

Setelah dilakukan proses *clustering* menggunakan *K-Means*, diperoleh sebaran distribusi data dimana warna merah merupakan daerah *cluster* 1 dengan jumlah data sebanyak 45 data, warna hijau merupakan daerah *cluster* 2 dengan jumlah data sebanyak 421 data, dan warna biru merupakan daerah *cluster* 3 dengan jumlah data sebanyak 74 data.

### 1. *Clustering* pada *K-Means* dengan 540 data menggunakan 4 *cluster*

1. Untuk melakukan *clustering* pada *K-Means*, data yang digunakan adalah data yang sudah melalui proses *preprocessing* dengan *cluster* awal (acak) berjumlah empat *cluster*. Perintah untuk proses *K-Means* yaitu `kmeans.proc4 <- kmeans(sd4, centers = 4)`, dimana `kmeans.proc4` adalah variabel untuk menampung data pengolahan *K-Means*, `kmeans` adalah fungsi menjalankan *K-Means*, `sd4` adalah data yang telah dinormalisasi, dan `centers = 4` adalah jumlah *cluster* awal. Berikut adalah hasil *clustering* dengan *K-Means*.

K-means clustering with 4 clusters of sizes 115, 49, 348, 28

Cluster means:

	Kematian.Neonatal	Kematian.Bayi	Kematian.AnakBalita	Kematian.Balita	Kematian.Ibu
1	0,56888183	0,4680060	-0,2364308	0,4741006	1,1020843
2	-0,07257294	0,0348484	2,2620307	0,3328153	-0,2512624
3	-0,41465011	-0,4068849	-0,3224635	-0,4556159	-0,3707962
4	2,94403215	3,0738455	1,0202620	3,1330291	0,5217585

Gambar 4.23. Hasil *Clustering* dengan *K-Means*

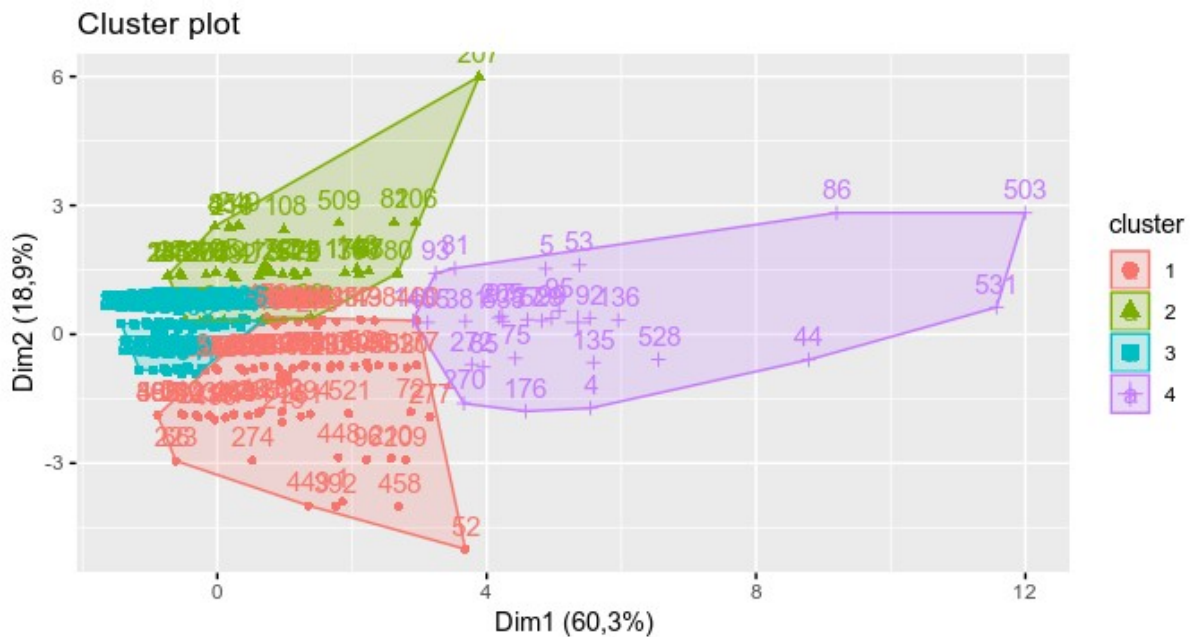
Pada proses *K-Means* diperoleh jumlah anggota *cluster* pada *cluster* 1 = 115 data, jumlah anggota *cluster* pada *cluster* 2 = 49 data, jumlah anggota *cluster* pada *cluster* 3 = 348 data, dan jumlah anggota *cluster* pada *cluster* 4 = 28 data.

Tabel 4.48. *Cluster Means* dan Jumlah Anggota *Cluster* pada *K-Means*

<i>Cluster</i>	Jumlah Anggota <i>Cluster</i>	Kematian.Neo natal	Kematian. Bayi	Kematian. AnakBalita	Kematian. Balita	Kematian. Ibu
1	115	0.56888183	0.4680060	-0.2364308	0.4741006	1.1020843
2	49	-0.07257294	0.0348484	2.2620307	0.3328153	-0.2512624
3	348	-0.41465011	-0.4068849	-0.3224635	-0.4556159	0.3707962
4	28	2.94403215	3.0738455	1.0202620	3.1330291	0.5217585

2. Setelah memperoleh hasil *clustering*, langkah selanjutnya yaitu menggambarkan distribusi data dengan jumlah *cluster* = 4.





Gambar 4.24. Pengelompokan Distribusi Data pada *K-Means*

Setelah dilakukan proses *clustering* menggunakan *K-Means*, diperoleh sebaran distribusi data dimana warna merah merupakan daerah *cluster* 1 dengan jumlah data sebanyak 115 data, warna hijau merupakan daerah *cluster* 2 dengan jumlah data sebanyak 49 data, warna biru merupakan daerah *cluster* 3 dengan jumlah data sebanyak 348 data, dan warna ungu merupakan daerah *cluster* 4 dengan jumlah data sebanyak 28 data.

#### m. *Clustering* pada *K-Means* dengan 550 data menggunakan 2 *cluster*

1. Untuk melakukan *clustering* pada *K-Means*, data yang digunakan adalah data yang sudah melalui proses *preprocessing* dengan *cluster* awal (acak) berjumlah dua *cluster*. Perintah untuk proses *K-Means* yaitu `kmeans.proc5 <- kmeans(sd5, centers = 2)`, dimana `kmeans.proc5` adalah variabel untuk menampung data pengolahan *K-Means*, `kmeans` adalah fungsi menjalankan *K-Means*, `sd5` adalah data yang telah dinormalisasi, dan `centers = 2` adalah jumlah *cluster* awal. Berikut adalah hasil *clustering* dengan *K-Means*.

K-means clustering with 2 clusters of sizes 446, 104

Cluster means:

	Kematian.Neonatal	Kematian.Bayi	Kematian.AnakBalita	Kematian.Balita	Kematian.Ibu
1	-0.3466175	-0.3539384	-0.1533076	-0.3752704	-0.1638772
2	1.4864560	1.5178514	0.6574535	1.6093325	0.7027810

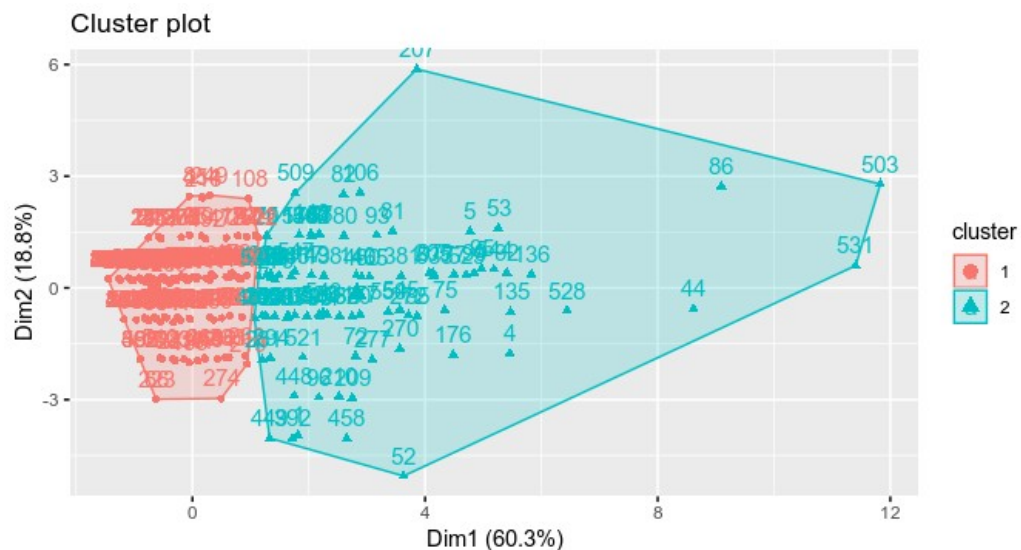
Gambar 4.25. Hasil *Clustering* dengan *K-Means*

Pada proses *K-Means* diperoleh jumlah anggota *cluster* pada *cluster* 1 = 446 data dan jumlah anggota *cluster* pada *cluster* 2 = 104 data.

Tabel 4.49. *Cluster Means* dan Jumlah Anggota *Cluster* pada *K-Means*

<i>Cluster</i>	Jumlah Anggota <i>Cluster</i>	Kematian.Neonatal	Kematian.Bayi	Kematian.AnakBalita	Kematian.Balita	Kematian.Ibu
1	446	-0.3466175	-0.3539384	-0.1533076	-0.3752704	-0.1638772
2	104	1.486456	1.5178514	0.6574535	1.6093325	0.702781

- Setelah memperoleh hasil *clustering*, langkah selanjutnya yaitu menggambarkan distribusi data dengan jumlah *cluster* = 2.



Gambar 4.26. Pengelompokan Distribusi Data pada *K-Means*

Setelah dilakukan proses *clustering* menggunakan *K-Means*, diperoleh sebaran distribusi data dimana warna merah merupakan daerah *cluster* 1 dengan jumlah data sebanyak 446 data dan warna biru merupakan daerah *cluster* 2 dengan jumlah data sebanyak 104 data.



**n. Clustering pada K-Means dengan 550 data menggunakan 3 cluster**

1. Untuk melakukan *clustering* pada *K-Means*, data yang digunakan adalah data yang sudah melalui proses *preprocessing* dengan *cluster* awal (acak) berjumlah tiga *cluster*. Perintah untuk proses *K-Means* yaitu `kmeans.proc5 <- kmeans(sd5, centers = 3)`, dimana `kmeans.proc5` adalah variabel untuk menampung data pengolahan *K-Means*, `kmeans` adalah fungsi menjalankan *K-Means*, `sd5` adalah data yang telah dinormalisasi, dan `centers = 3` adalah jumlah *cluster* awal. Berikut adalah hasil *clustering* dengan *K-Means*.

K-means clustering with 3 clusters of sizes 32, 161, 357

Cluster means:

	Kematian.Neonatal	Kematian.Bayi	Kematian.AnakBalita	Kematian.Balita	Kematian.Ibu
1	2,6451600	2,8587531	1,39843600	2,9718480	0,5499211
2	0,5689055	0,5335355	0,09903197	0,5313218	0,5673017
3	-0,4936664	-0,4968608	-0,17001148	-0,5059999	-0,3051346

Gambar 4.27. Hasil *Clustering* dengan *K-Means*

Pada proses *K-Means* diperoleh jumlah anggota *cluster* pada *cluster* 1 = 32 data, jumlah anggota *cluster* pada *cluster* 2 = 161 data, dan jumlah anggota *cluster* pada *cluster* 3 = 357 data.

Tabel 4.50. *Cluster Means* dan Jumlah Anggota *Cluster* pada *K-Means*

<i>Cluster</i>	Jumlah Anggota <i>Cluster</i>	Kematian. Neonatal	Kematian. Bayi	Kematian. Anak Balita	Kematian. Balita	Kematian. Ibu
1	32	2.6451600	2.8587531	1.39843600	2.9718480	0.5499811
2	161	0.5689055	0.5335355	0.09903197	0.5313218	0.5673017
3	357	-0.4936664	-0.4968608	-0.17001148	-0.5059999	-0.3051346

2. Setelah memperoleh hasil *clustering*, langkah selanjutnya yaitu menggambarkan distribusi data dengan jumlah *cluster* = 3.





Setelah dilakukan proses *clustering* menggunakan *K-Means*, diperoleh sebaran distribusi data dimana warna merah merupakan daerah *cluster* 1 dengan jumlah data sebanyak 129 data, warna hijau merupakan daerah *cluster* 2 dengan jumlah data sebanyak 40 data, warna biru merupakan daerah *cluster* 3 dengan jumlah data sebanyak 30 data, dan warna ungu merupakan daerah *cluster* 4 dengan jumlah data sebanyak 351 data.

## 2. *Clustering* pada *Fuzzy C-Means*

### a. *Clustering* pada *Fuzzy C-Means* dengan 510 data menggunakan 2 *cluster*

1. Untuk melakukan *clustering* pada *Fuzzy C-Means*, data yang digunakan adalah data yang sudah melalui proses *preprocessing* dengan *cluster* awal (acak) berjumlah tiga *cluster*. Perintah untuk proses *Fuzzy C-Means* yaitu `proses.fcm <- fanny(sd, 2, metric = "euclidean", stand = FALSE)`, dimana `proses.fcm` adalah variabel untuk menampung data pengolahan *Fuzzy C-Means*, `fanny` adalah fungsi menjalankan *Fuzzy C-Means*, `sd` adalah data yang telah dinormalisasi, 2 adalah jumlah *cluster* awal, `metric = "euclidean"` adalah metode perhitungan jarak pada *Fuzzy C-Means*, `stand = FALSE` adalah logika, jika bernilai TRUE maka data `sd` akan dinormalisasi sebelum dihitung ketidaksamaannya. Berikut adalah hasil *clustering* dengan *Fuzzy C-Means*.

```
Fuzzy Clustering object of class 'fanny' :
m.ship.expon.      2
objective      304,7926
tolerance        1e-15
iterations        31
converged         1
maxit            500
n                510
```

Gambar 4.31. Hasil *Clustering* dengan *Fuzzy C-Means*

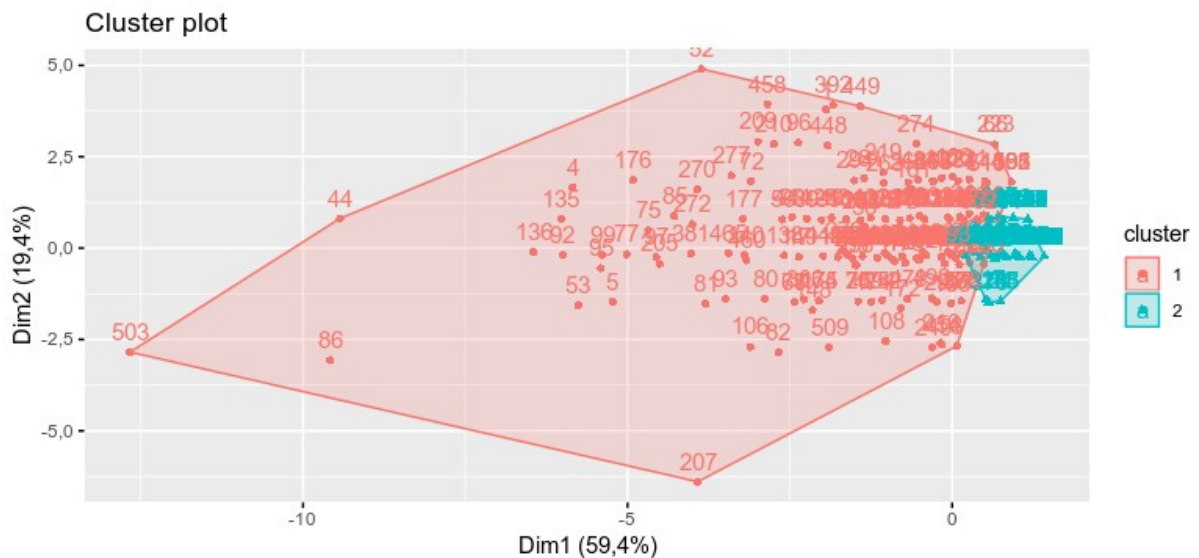
2. Pada proses *Fuzzy C-Means* diperoleh *membership exponential*, fungsi objektif, dan toleransi. *Objective function* adalah penentuan nilai maksimum atau nilai minimum dari sebuah fungsi linear. Proses *Fuzzy C-Means* dapat dilihat pada Tabel . *Membership exponential* yaitu angka  $r$  benar-benar lebih besar dari satu yang menentukan eksponen keanggotaan yang digunakan dalam kriteria kecocokan. *Tolerance* yaitu tingkat kesalahan yang diizinkan pada penelitian. *Iter*

yaitu jumlah iterasi untuk menjalankan proses algoritma *Fuzzy C-Means*. *Max Iter* yaitu iterasi maksimum yang diizinkan untuk menjalankan proses algoritma *Fuzzy C-Means*. Konvergen yaitu jumlah yang berhingga (*finite*).

Tabel 4.52. Nilai Keanggotaan, Fungsi Objektif, dan Toleransi pada *Fuzzy C-Means*

<i>Membership Exponential</i>	<i>Objective Function</i>	<i>Tolerance</i>	<i>Iter</i>	<i>Converged</i>	<i>Max Iter</i>
2	304.7926	1.00E-15	31	1	500

- Setelah memperoleh hasil *clustering*, langkah selanjutnya yaitu menggambarkan distribusi data dengan jumlah *cluster* = 2.



Gambar 4.32. Pengelompokan Distribusi Data pada *Fuzzy C-Means*

Setelah dilakukan proses *clustering* menggunakan *Fuzzy C-Means*, diperoleh sebaran distribusi data dimana warna merah merupakan daerah *cluster* 1 dan warna biru merupakan daerah *cluster* 2.

#### b. *Clustering* pada *Fuzzy C-Means* dengan 510 data menggunakan 3 *cluster*

- Untuk melakukan *clustering* pada *Fuzzy C-Means*, data yang digunakan adalah data yang sudah melalui proses *preprocessing* dengan *cluster* awal (acak) berjumlah tiga *cluster*. Perintah untuk proses *Fuzzy C-Means* yaitu `proses.fcm <- fanny(sd, 3, metric = "euclidean", stand = FALSE)`, dimana `proses.fcm` adalah variabel untuk menampung data pengolahan *Fuzzy C-Means*, `fanny` adalah fungsi menjalankan *Fuzzy C-Means*, `sd` adalah data yang

telah dinormalisasi, 3 adalah jumlah *cluster* awal, *metric* = "euclidean" adalah metode perhitungan jarak pada *Fuzzy C-Means*, *stand* = FALSE adalah logika, jika bernilai TRUE maka data *sd* akan dinormalisasi sebelum dihitung ketidaksamaannya. Berikut adalah hasil *clustering* dengan *Fuzzy C-Means*.

```
Fuzzy Clustering object of class 'fanny' :
m.ship.expon.      2
objective      199,2887
tolerance        1e-15
iterations       112
converged         1
maxit            500
n                510
```

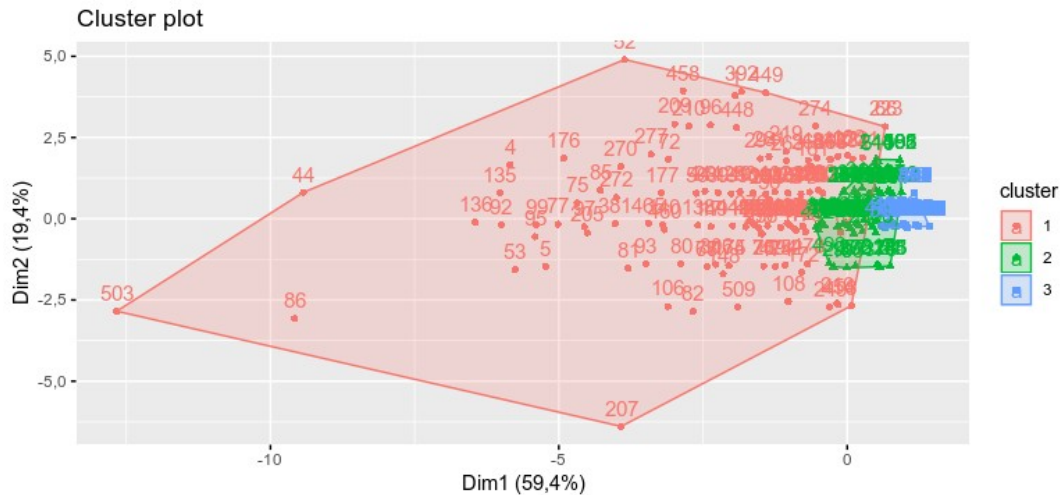
Gambar 4.33. Hasil *Clustering* dengan *Fuzzy C-Means*

2. Pada proses *Fuzzy C-Means* diperoleh *membership exponential*, fungsi objektif, dan toleransi. *Objective function* adalah penentuan nilai maksimum atau nilai minimum dari sebuah fungsi linear. Proses *Fuzzy C-Means* dapat dilihat pada Tabel . *Membership exponential* yaitu angka *r* benar-benar lebih besar dari satu yang menentukan eksponen keanggotaan yang digunakan dalam kriteria kecocokan. *Tolerance* yaitu tingkat kesalahan yang diizinkan pada penelitian. *Iter* yaitu jumlah iterasi untuk menjalankan proses algoritma *Fuzzy C-Means*. *Max Iter* yaitu iterasi maksimum yang diizinkan untuk menjalankan proses algoritma *Fuzzy C-Means*. Konvergen yaitu jumlah yang berhingga (*finite*).

Tabel 4.53. Nilai Keanggotaan, Fungsi Objektif, dan Toleransi pada *Fuzzy C-Means*

<i>Membership Exponential</i>	<i>Objective Function</i>	<i>Tolerance</i>	<i>Iter</i>	<i>Converged</i>	<i>Max Iter</i>
2	199.2887	1.00E-15	112	1	500

3. Setelah memperoleh hasil *clustering*, langkah selanjutnya yaitu menggambarkan distribusi data dengan jumlah *cluster* = 3.



Gambar 4.34. Pengelompokan Distribusi Data pada *Fuzzy C-Means*

Setelah dilakukan proses *clustering* menggunakan *Fuzzy C-Means*, diperoleh sebaran distribusi data dimana warna merah merupakan daerah *cluster 1*, warna hijau merupakan daerah *cluster 2*, dan warna biru merupakan daerah *cluster 3*.

### c. *Clustering* pada *Fuzzy C-Means* dengan 510 data menggunakan 4 *cluster*

1. Untuk melakukan *clustering* pada *Fuzzy C-Means*, data yang digunakan adalah data yang sudah melalui proses *preprocessing* dengan *cluster* awal (acak) berjumlah tiga *cluster*. Perintah untuk proses *Fuzzy C-Means* yaitu `proses.fcm <- fanny(sd, 4, metric = "euclidean", stand = FALSE)`, dimana `proses.fcm` adalah variabel untuk menampung data pengolahan *Fuzzy C-Means*, `fanny` adalah fungsi menjalankan *Fuzzy C-Means*, `sd` adalah data yang telah dinormalisasi, 4 adalah jumlah *cluster* awal, `metric = "euclidean"` adalah metode perhitungan jarak pada *Fuzzy C-Means*, `stand = FALSE` adalah logika, jika bernilai TRUE maka data `sd` akan dinormalisasi sebelum dihitung ketidaksamaannya. Berikut adalah hasil *clustering* dengan *Fuzzy C-Means*.

```
Fuzzy Clustering object of class 'fanny' :
m.ship.expon.      2
objective      146,5821
tolerance        1e-15
iterations       463
converged         1
maxit           500
n              510
```

Gambar 4.35. Hasil *Clustering* dengan *Fuzzy C-Means*

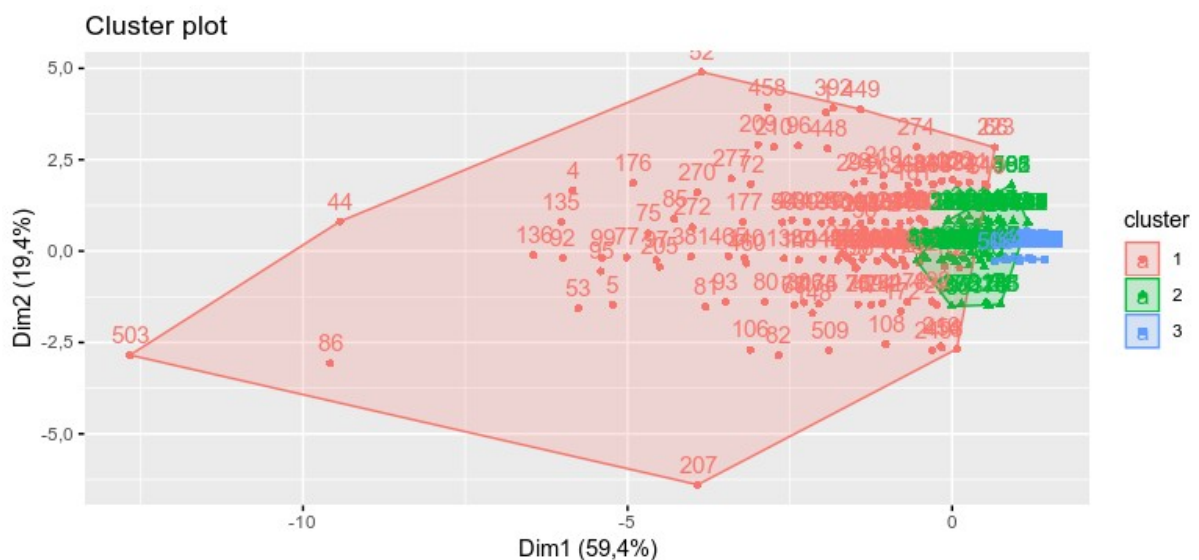


2. Pada proses *Fuzzy C-Means* diperoleh *membership exponential*, fungsi objektif, dan toleransi. *Objective function* adalah penentuan nilai maksimum atau nilai minimum dari sebuah fungsi linear. Proses *Fuzzy C-Means* dapat dilihat pada Tabel . *Membership exponential* yaitu angka  $r$  benar-benar lebih besar dari satu yang menentukan eksponen keanggotaan yang digunakan dalam kriteria kecocokan. *Tolerance* yaitu tingkat kesalahan yang diizinkan pada penelitian. *Iter* yaitu jumlah iterasi untuk menjalankan proses algoritma *Fuzzy C-Means*. *Max Iter* yaitu iterasi maksimum yang diizinkan untuk menjalankan proses algoritma *Fuzzy C-Means*. Konvergen yaitu jumlah yang berhingga (*finite*).

Tabel 4.54. Nilai Keanggotaan, Fungsi Objektif, dan Toleransi pada *Fuzzy C-Means*

<i>Membership Exponential</i>	<i>Objective Function</i>	<i>Tolerance</i>	<i>Iter</i>	<i>Converged</i>	<i>Max Iter</i>
2	146.5821	1.00E-15	463	1	500

3. Setelah memperoleh hasil *clustering*, langkah selanjutnya yaitu menggambarkan distribusi data dengan jumlah *cluster* = 4.



Gambar 4.36. Pengelompokan Distribusi Data pada *Fuzzy C-Means*

Setelah dilakukan proses *clustering* menggunakan *Fuzzy C-Means*, diperoleh sebaran distribusi data dimana warna merah merupakan daerah *cluster* 1, warna hijau merupakan daerah *cluster* 2, warna biru merupakan daerah *cluster* 3. Untuk jumlah *cluster* = 4, RStudio melakukan *clustering* secara terbatas dengan jumlah *cluster* yang dihasilkan hanya 3 *cluster*.



**d. Clustering pada Fuzzy C-Means dengan 520 data menggunakan 2 cluster**

1. Untuk melakukan *clustering* pada *Fuzzy C-Means*, data yang digunakan adalah data yang sudah melalui proses *preprocessing* dengan *cluster* awal (acak) berjumlah tiga *cluster*. Perintah untuk proses *Fuzzy C-Means* yaitu `proses.fcm2 <- fanny(sd2, 2, metric = "euclidean", stand = FALSE)`, dimana `proses.fcm2` adalah variabel untuk menampung data pengolahan *Fuzzy C-Means*, `fanny` adalah fungsi menjalankan *Fuzzy C-Means*, `sd2` adalah data yang telah dinormalisasi, 2 adalah jumlah *cluster* awal, `metric = "euclidean"` adalah metode perhitungan jarak pada *Fuzzy C-Means*, `stand = FALSE` adalah logika, jika bernilai TRUE maka data `sd2` akan dinormalisasi sebelum dihitung ketidaksamaannya. Berikut adalah hasil *clustering* dengan *Fuzzy C-Means*.

```
Fuzzy Clustering object of class 'fanny' :
m.ship.expon.      2
objective      311,8918
tolerance        1e-15
iterations        29
converged         1
maxit            500
n                520
```

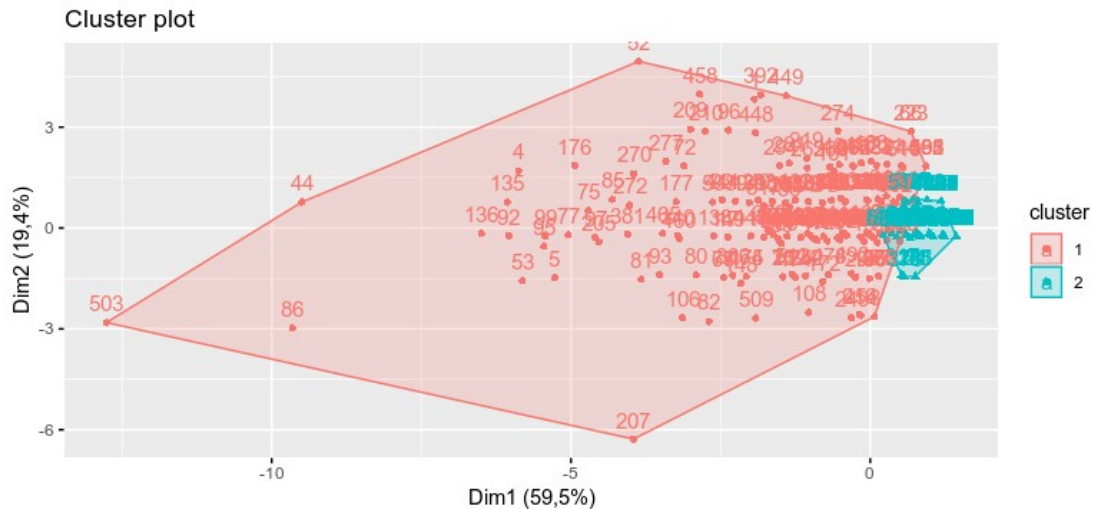
Gambar 4.37. Hasil *Clustering* dengan *Fuzzy C-Means*

2. Pada proses *Fuzzy C-Means* diperoleh *membership exponential*, fungsi objektif, dan toleransi. *Objective function* adalah penentuan nilai maksimum atau nilai minimum dari sebuah fungsi linear. Proses *Fuzzy C-Means* dapat dilihat pada Tabel . *Membership exponential* yaitu angka  $r$  benar-benar lebih besar dari satu yang menentukan eksponen keanggotaan yang digunakan dalam kriteria kecocokan. *Tolerance* yaitu tingkat kesalahan yang diizinkan pada penelitian. *Iter* yaitu jumlah iterasi untuk menjalankan proses algoritma *Fuzzy C-Means*. *Max Iter* yaitu iterasi maksimum yang diizinkan untuk menjalankan proses algoritma *Fuzzy C-Means*. Konvergen yaitu jumlah yang berhingga (*finite*).

Tabel 4.55. Nilai Keanggotaan, Fungsi Objektif, dan Toleransi pada *Fuzzy C-Means*

<i>Membership Exponential</i>	<i>Objective Function</i>	<i>Tolerance</i>	<i>Iter</i>	<i>Converged</i>	<i>Max Iter</i>
2	311.8918	1.00E-15	29	1	500

- Setelah memperoleh hasil *clustering*, langkah selanjutnya yaitu menggambarkan distribusi data dengan jumlah *cluster* = 2.



Gambar 4.38. Pengelompokan Distribusi Data pada *Fuzzy C-Means*

Setelah dilakukan proses *clustering* menggunakan *Fuzzy C-Means*, diperoleh sebaran distribusi data dimana warna merah merupakan daerah *cluster* 1 dan warna biru merupakan daerah *cluster* 2.

#### e. *Clustering* pada *Fuzzy C-Means* dengan 520 data menggunakan 3 *cluster*

- Untuk melakukan *clustering* pada *Fuzzy C-Means*, data yang digunakan adalah data yang sudah melalui proses *preprocessing* dengan *cluster* awal (acak) berjumlah tiga *cluster*. Perintah untuk proses *Fuzzy C-Means* yaitu `proses.fcm2 <- fanny(sd2, 3, metric = "euclidean", stand = FALSE)`, dimana `proses.fcm2` adalah variabel untuk menampung data pengolahan *Fuzzy C-Means*, `fanny` adalah fungsi menjalankan *Fuzzy C-Means*, `sd2` adalah data yang telah dinormalisasi, 3 adalah jumlah *cluster* awal, `metric = "euclidean"` adalah metode perhitungan jarak pada *Fuzzy C-Means*, `stand = FALSE` adalah logika, jika bernilai TRUE maka data `sd2` akan dinormalisasi sebelum dihitung ketidaksamaannya. Berikut adalah hasil *clustering* dengan *Fuzzy C-Means*.

```

Fuzzy Clustering object of class 'fanny' :
m.ship.expon.      2
objective          203,988
tolerance          1e-15
iterations         93
converged          1
maxit              500
n                  520

```

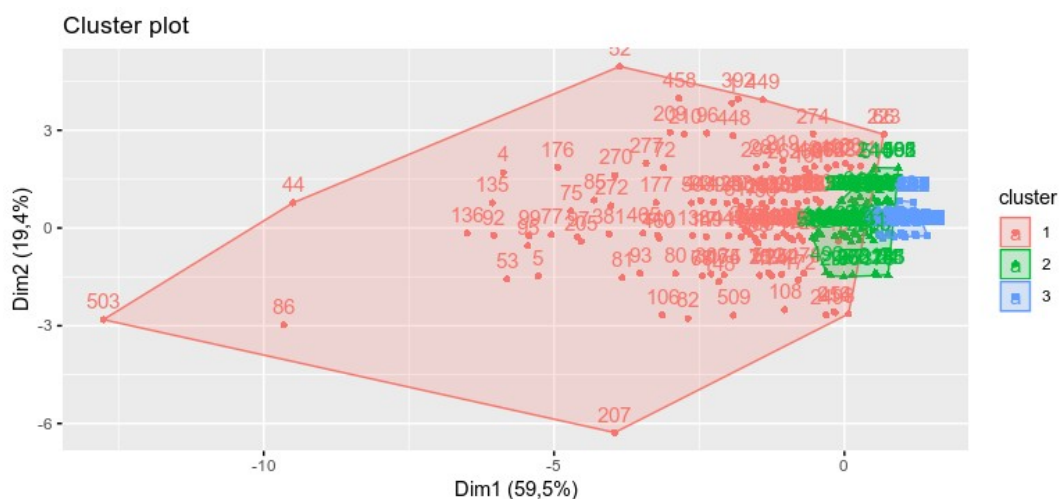
Gambar 4.39. Hasil *Clustering* dengan *Fuzzy C-Means*

2. Pada proses *Fuzzy C-Means* diperoleh *membership exponential*, fungsi objektif, dan toleransi. *Objective function* adalah penentuan nilai maksimum atau nilai minimum dari sebuah fungsi linear. Proses *Fuzzy C-Means* dapat dilihat pada Tabel . *Membership exponential* yaitu angka  $r$  benar-benar lebih besar dari satu yang menentukan eksponen keanggotaan yang digunakan dalam kriteria kecocokan. *Tolerance* yaitu tingkat kesalahan yang diizinkan pada penelitian. *Iter* yaitu jumlah iterasi untuk menjalankan proses algoritma *Fuzzy C-Means*. *Max Iter* yaitu iterasi maksimum yang diizinkan untuk menjalankan proses algoritma *Fuzzy C-Means*. Konvergen yaitu jumlah yang berhingga (*finite*).

Tabel 4.56. Nilai Keanggotaan, Fungsi Objektif, dan Toleransi pada *Fuzzy C-Means*

<i>Membership Exponential</i>	<i>Objective Function</i>	<i>Tolerance</i>	<i>Iter</i>	<i>Converged</i>	<i>Max Iter</i>
2	203.988	1.00E-15	93	1	500

3. Setelah memperoleh hasil *clustering*, langkah selanjutnya yaitu menggambarkan distribusi data dengan jumlah *cluster* = 3.



Gambar 4.40. Pengelompokan Distribusi Data pada *Fuzzy C-Means*

Setelah dilakukan proses *clustering* menggunakan *Fuzzy C-Means*, diperoleh sebaran distribusi data dimana warna merah merupakan daerah *cluster* 1, warna hijau merupakan daerah *cluster* 2, dan warna biru merupakan daerah *cluster* 3.

**f. Clustering pada Fuzzy C-Means dengan 520 data menggunakan 4 cluster**

1. Untuk melakukan *clustering* pada *Fuzzy C-Means*, data yang digunakan adalah data yang sudah melalui proses *preprocessing* dengan *cluster* awal (acak) berjumlah tiga *cluster*. Perintah untuk proses *Fuzzy C-Means* yaitu `proses.fcm2 <- fanny(sd2, 4, metric = "euclidean", stand = FALSE)`, dimana `proses.fcm2` adalah variabel untuk menampung data pengolahan *Fuzzy C-Means*, `fanny` adalah fungsi menjalankan *Fuzzy C-Means*, `sd2` adalah data yang telah dinormalisasi, 4 adalah jumlah *cluster* awal, `metric = "euclidean"` adalah metode perhitungan jarak pada *Fuzzy C-Means*, `stand = FALSE` adalah logika, jika bernilai TRUE maka data `sd2` akan dinormalisasi sebelum dihitung ketidaksamaannya. Berikut adalah hasil *clustering* dengan *Fuzzy C-Means*.

```
Fuzzy Clustering object of class 'fanny' :
m.ship.expon.      2
objective      150,0297
tolerance         1e-15
iterations        418
converged          1
maxit             500
n                 520
```

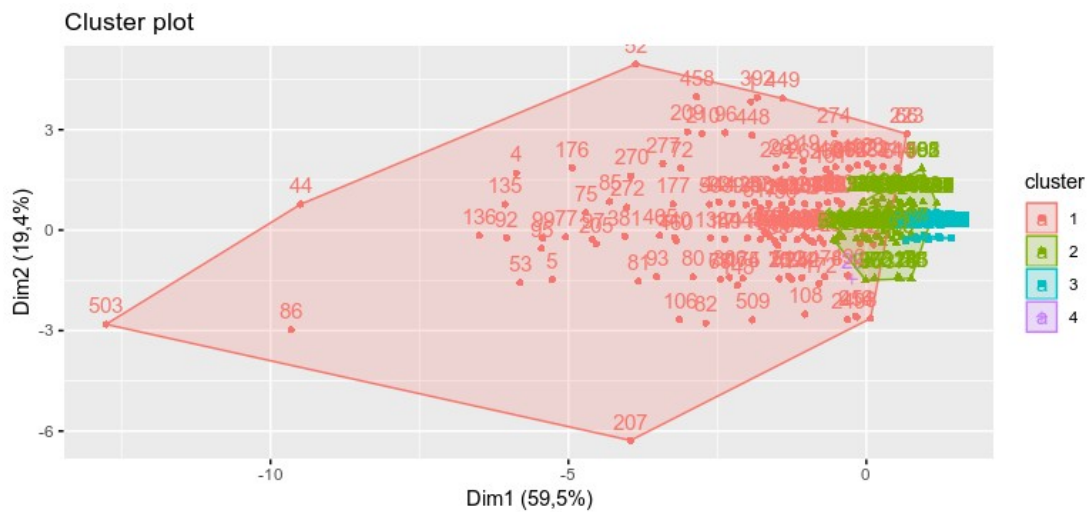
Gambar 4.41. Hasil *Clustering* dengan *Fuzzy C-Means*

2. Pada proses *Fuzzy C-Means* diperoleh *membership exponential*, fungsi objektif, dan toleransi. *Objective function* adalah penentuan nilai maksimum atau nilai minimum dari sebuah fungsi linear. Proses *Fuzzy C-Means* dapat dilihat pada Tabel . *Membership exponential* yaitu angka  $r$  benar-benar lebih besar dari satu yang menentukan eksponen keanggotaan yang digunakan dalam kriteria kecocokan. *Tolerance* yaitu tingkat kesalahan yang diizinkan pada penelitian. *Iter* yaitu jumlah iterasi untuk menjalankan proses algoritma *Fuzzy C-Means*. *Max Iter* yaitu iterasi maksimum yang diizinkan untuk menjalankan proses algoritma *Fuzzy C-Means*. Konvergen yaitu jumlah yang berhingga (*finite*).

Tabel 4.57. Nilai Keanggotaan, Fungsi Objektif, dan Toleransi pada *Fuzzy C-Means*

<i>Membership Exponential</i>	<i>Objective Function</i>	<i>Tolerance</i>	<i>Iter</i>	<i>Converged</i>	<i>Max Iter</i>
2	150.0297	1.00E-15	418	1	500

- Setelah memperoleh hasil *clustering*, langkah selanjutnya yaitu menggambarkan distribusi data dengan jumlah *cluster* = 4.

Gambar 4.42. Pengelompokan Distribusi Data pada *Fuzzy C-Means*

Setelah dilakukan proses *clustering* menggunakan *Fuzzy C-Means*, diperoleh sebaran distribusi data dimana warna merah merupakan daerah *cluster* 1, warna hijau merupakan daerah *cluster* 2, warna biru merupakan daerah *cluster* 3, dan warna ungu merupakan daerah *cluster* 4.

**g. *Clustering* pada *Fuzzy C-Means* dengan 530 data menggunakan 2 *cluster***

- Untuk melakukan *clustering* pada *Fuzzy C-Means*, data yang digunakan adalah data yang sudah melalui proses *preprocessing* dengan *cluster* awal (acak) berjumlah tiga *cluster*. Perintah untuk proses *Fuzzy C-Means* yaitu `proses.fcm3 <- fanny(sd3, 2, metric = "euclidean", stand = FALSE)`, dimana `proses.fcm5` adalah variabel untuk menampung data pengolahan *Fuzzy C-Means*, `fanny` adalah fungsi menjalankan *Fuzzy C-Means*, `sd3` adalah data yang telah dinormalisasi, 2 adalah jumlah *cluster* awal, `metric = "euclidean"` adalah metode perhitungan jarak pada *Fuzzy C-Means*, `stand = FALSE` adalah

logika, jika bernilai TRUE maka data sd3 akan dinormalisasi sebelum dihitung ketidaksamaannya. Berikut adalah hasil *clustering* dengan *Fuzzy C-Means*.

```
Fuzzy Clustering object of class 'fanny' :
m.ship.expon.      2
objective      318.5439
tolerance        1e-15
iterations        30
converged         1
maxit           500
n                530
```

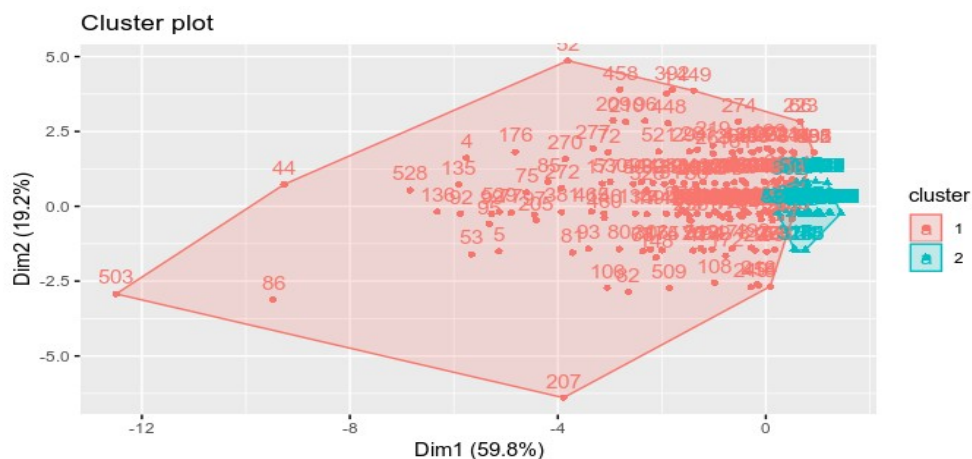
Gambar 4.43. Hasil *Clustering* dengan *Fuzzy C-Means*

2. Pada proses *Fuzzy C-Means* diperoleh *membership exponential*, fungsi objektif, dan toleransi. *Objective function* adalah penentuan nilai maksimum atau nilai minimum dari sebuah fungsi linear. Proses *Fuzzy C-Means* dapat dilihat pada Tabel . *Membership exponential* yaitu angka  $r$  benar-benar lebih besar dari satu yang menentukan eksponen keanggotaan yang digunakan dalam kriteria kecocokan. *Tolerance* yaitu tingkat kesalahan yang diizinkan pada penelitian. *Iter* yaitu jumlah iterasi untuk menjalankan proses algoritma *Fuzzy C-Means*. *Max Iter* yaitu iterasi maksimum yang diizinkan untuk menjalankan proses algoritma *Fuzzy C-Means*. Konvergen yaitu jumlah yang berhingga (*finite*).

Tabel 4.58. Nilai Keanggotaan, Fungsi Objektif, dan Toleransi pada *Fuzzy C-Means*

<i>Membership Exponential</i>	<i>Objective Function</i>	<i>Tolerance</i>	<i>Iter</i>	<i>Converged</i>	<i>Max Iter</i>
2	318.5439	1.00E-15	30	1	500

3. Setelah memperoleh hasil *clustering*, langkah selanjutnya yaitu menggambarkan distribusi data dengan jumlah *cluster* = 2.



Gambar 4.44. Pengelompokan Distribusi Data pada *Fuzzy C-Means*

Setelah dilakukan proses *clustering* menggunakan *Fuzzy C-Means*, diperoleh sebaran distribusi data dimana warna merah merupakan daerah *cluster* 1 dan warna biru merupakan daerah *cluster* 2.

#### h. *Clustering* pada *Fuzzy C-Means* dengan 530 data menggunakan 3 *cluster*

1. Untuk melakukan *clustering* pada *Fuzzy C-Means*, data yang digunakan adalah data yang sudah melalui proses *preprocessing* dengan *cluster* awal (acak) berjumlah tiga *cluster*. Perintah untuk proses *Fuzzy C-Means* yaitu `proses.fcm3 <- fanny(sd3, 3, metric = "euclidean", stand = FALSE)`, dimana `proses.fcm5` adalah variabel untuk menampung data pengolahan *Fuzzy C-Means*, `fanny` adalah fungsi menjalankan *Fuzzy C-Means*, `sd3` adalah data yang telah dinormalisasi, 3 adalah jumlah *cluster* awal, `metric = "euclidean"` adalah metode perhitungan jarak pada *Fuzzy C-Means*, `stand = FALSE` adalah logika, jika bernilai TRUE maka data `sd3` akan dinormalisasi sebelum dihitung ketidaksamaannya. Berikut adalah hasil *clustering* dengan *Fuzzy C-Means*.

```
Fuzzy Clustering object of class 'fanny' :
m.ship.expon.      2
objective      208,4563
tolerance        1e-15
iterations       108
converged         1
maxit           500
n               530
```

Gambar 4.45. Hasil *Clustering* dengan *Fuzzy C-Means*

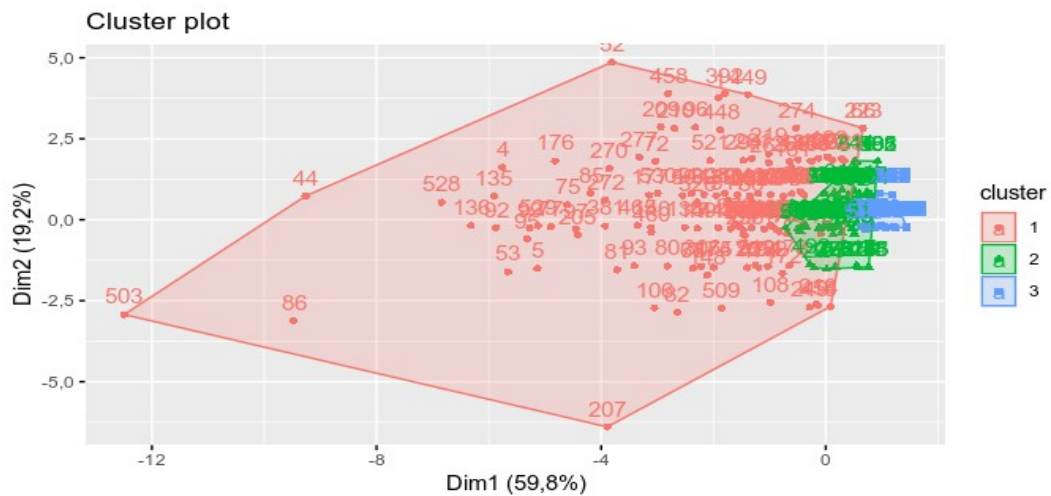
2. Pada proses *Fuzzy C-Means* diperoleh *membership exponential*, fungsi objektif, dan toleransi. *Objective function* adalah penentuan nilai maksimum atau nilai minimum dari sebuah fungsi linear. Proses *Fuzzy C-Means* dapat dilihat pada Tabel . *Membership exponential* yaitu angka  $r$  benar-benar lebih besar dari satu yang menentukan eksponen keanggotaan yang digunakan dalam kriteria kecocokan. *Tolerance* yaitu tingkat kesalahan yang diizinkan pada penelitian. *Iter* yaitu jumlah iterasi untuk menjalankan proses algoritma *Fuzzy C-Means*. *Max Iter* yaitu iterasi maksimum yang diizinkan untuk menjalankan proses algoritma *Fuzzy C-Means*. Konvergen yaitu jumlah yang berhingga (*finite*).



Tabel 4.59. Nilai Keanggotaan, Fungsi Objektif, dan Toleransi pada *Fuzzy C-Means*

<i>Membership Exponential</i>	<i>Objective Function</i>	<i>Tolerance</i>	<i>Iter</i>	<i>Converged</i>	<i>Max Iter</i>
2	208.4563	1.00E-15	108	1	500

- Setelah memperoleh hasil *clustering*, langkah selanjutnya yaitu menggambarkan distribusi data dengan jumlah *cluster* = 3.

Gambar 4.46. Pengelompokan Distribusi Data pada *Fuzzy C-Means*

Setelah dilakukan proses clustering menggunakan *Fuzzy C-Means*, diperoleh sebaran distribusi data dimana warna merah merupakan daerah *cluster* 1, warna hijau merupakan daerah *cluster* 2, dan warna biru merupakan daerah *cluster* 3.

#### i. *Clustering* pada *Fuzzy C-Means* dengan 530 data menggunakan 4 *cluster*

- Untuk melakukan *clustering* pada *Fuzzy C-Means*, data yang digunakan adalah data yang sudah melalui proses *preprocessing* dengan *cluster* awal (acak) berjumlah tiga *cluster*. Perintah untuk proses *Fuzzy C-Means* yaitu `proses.fcm3 <- fanny(sd3, 4, metric = "euclidean", stand = FALSE)`, dimana `proses.fcm5` adalah variabel untuk menampung data pengolahan *Fuzzy C-Means*, `fanny` adalah fungsi menjalankan *Fuzzy C-Means*, `sd3` adalah data yang telah dinormalisasi, 4 adalah jumlah *cluster* awal, `metric = "euclidean"` adalah metode perhitungan jarak pada *Fuzzy C-Means*, `stand = FALSE` adalah logika, jika bernilai TRUE maka data `sd3` akan dinormalisasi sebelum dihitung ketidaksamaannya. Berikut adalah hasil *clustering* dengan *Fuzzy C-Means*.



```

Fuzzy Clustering object of class 'fanny' :
m.ship.expon.      2
objective      153,3862
tolerance        1e-15
iterations        -1
converged         0
maxit           500
n               530

```

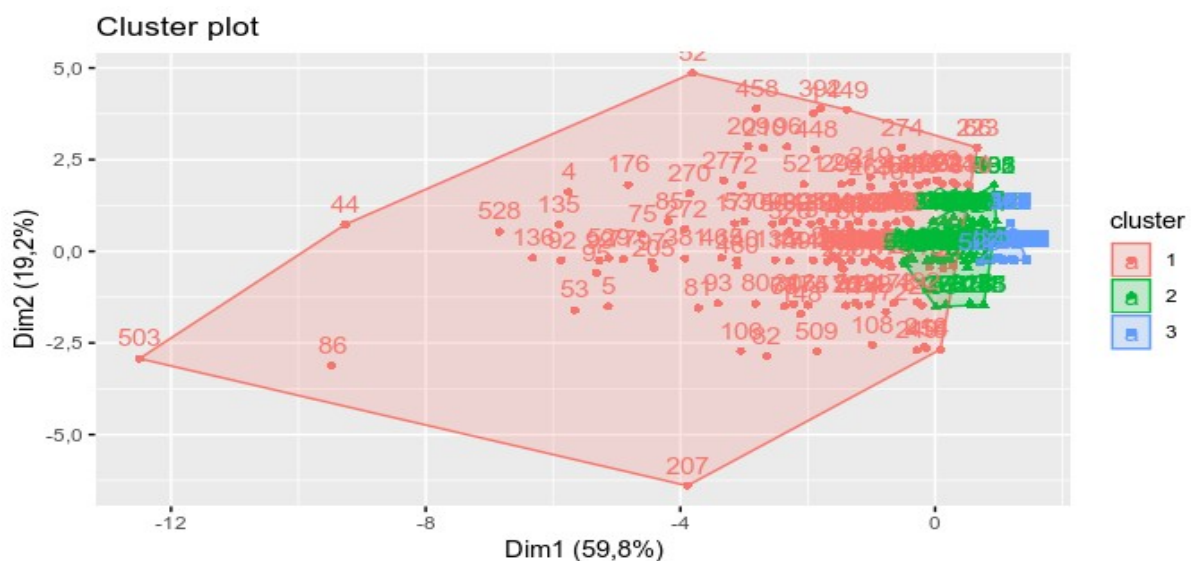
Gambar 4.47. Hasil *Clustering* dengan *Fuzzy C-Means*

2. Pada proses *Fuzzy C-Means* diperoleh *membership exponential*, fungsi objektif, dan toleransi. *Objective function* adalah penentuan nilai maksimum atau nilai minimum dari sebuah fungsi linear. Proses *Fuzzy C-Means* dapat dilihat pada Tabel . *Membership exponential* yaitu angka  $r$  benar-benar lebih besar dari satu yang menentukan eksponen keanggotaan yang digunakan dalam kriteria kecocokan. *Tolerance* yaitu tingkat kesalahan yang diizinkan pada penelitian. *Iter* yaitu jumlah iterasi untuk menjalankan proses algoritma *Fuzzy C-Means*. *Max Iter* yaitu iterasi maksimum yang diizinkan untuk menjalankan proses algoritma *Fuzzy C-Means*. Konvergen yaitu jumlah yang berhingga (*finite*).

Tabel 4.60. Nilai Keanggotaan, Fungsi Objektif, dan Toleransi pada *Fuzzy C-Means*

<i>Membership Exponential</i>	<i>Objective Function</i>	<i>Tolerance</i>	<i>Iter</i>	<i>Converged</i>	<i>Max Iter</i>
2	153.3862	1.00E-15	-1	0	500

3. Setelah memperoleh hasil *clustering*, langkah selanjutnya yaitu menggambarkan distribusi data dengan jumlah *cluster* = 3.



Gambar 4.48. Pengelompokkan Distribusi Data pada *Fuzzy C-Means*

Setelah dilakukan proses clustering menggunakan *Fuzzy C-Means*, diperoleh sebaran distribusi data dimana warna merah merupakan daerah *cluster* 1, warna hijau merupakan daerah *cluster* 2, dan warna biru merupakan daerah *cluster* 3. Untuk jumlah *cluster* = 4, RStudio melakukan *clustering* secara terbatas dengan jumlah *cluster* yang dihasilkan hanya 3 *cluster*.

**j. Clustering pada Fuzzy C-Means dengan 540 data menggunakan 2 cluster**

1. Untuk melakukan *clustering* pada *Fuzzy C-Means*, data yang digunakan adalah data yang sudah melalui proses *preprocessing* dengan *cluster* awal (acak) berjumlah tiga *cluster*. Perintah untuk proses *Fuzzy C-Means* yaitu `proses.fcm4 <- fanny(sd4, 2, metric = "euclidean", stand = FALSE)`, dimana `proses.fcm4` adalah variabel untuk menampung data pengolahan *Fuzzy C-Means*, `fanny` adalah fungsi menjalankan *Fuzzy C-Means*, `sd4` adalah data yang telah dinormalisasi, `2` adalah jumlah *cluster* awal, `metric = "euclidean"` adalah metode perhitungan jarak pada *Fuzzy C-Means*, `stand = FALSE` adalah logika, jika bernilai TRUE maka data `sd4` akan dinormalisasi sebelum dihitung ketidaksamaannya. Berikut adalah hasil *clustering* dengan *Fuzzy C-Means*.

```
Fuzzy Clustering object of class 'fanny' :
m.ship.expon.      2
objective      322,5109
tolerance         1e-15
iterations        30
converged         1
maxit            500
n                540
```

Gambar 4.49. Hasil *Clustering* dengan *Fuzzy C-Means*

2. Pada proses *Fuzzy C-Means* diperoleh *membership exponential*, fungsi objektif, dan toleransi. *Objective function* adalah penentuan nilai maksimum atau nilai minimum dari sebuah fungsi linear. Proses *Fuzzy C-Means* dapat dilihat pada Tabel . *Membership exponential* yaitu angka  $r$  benar-benar lebih besar dari satu yang menentukan eksponen keanggotaan yang digunakan dalam kriteria kecocokan. *Tolerance* yaitu tingkat kesalahan yang diizinkan pada penelitian. *Iter* yaitu jumlah iterasi untuk menjalankan proses algoritma *Fuzzy C-Means*. *Max Iter*





Setelah dilakukan proses clustering menggunakan *Fuzzy C-Means*, diperoleh sebaran distribusi data dimana warna merah merupakan daerah *cluster* 1, warna hijau merupakan daerah *cluster* 2, dan warna biru merupakan daerah *cluster* 3.

#### 1. *Clustering pada Fuzzy C-Means dengan 540 data menggunakan 4 cluster*

1. Untuk melakukan *clustering* pada *Fuzzy C-Means*, data yang digunakan adalah data yang sudah melalui proses *preprocessing* dengan *cluster* awal (acak) berjumlah tiga *cluster*. Perintah untuk proses *Fuzzy C-Means* yaitu `proses.fcm4 <- fanny(sd4, 4, metric = "euclidean", stand = FALSE)`, dimana `proses.fcm4` adalah variabel untuk menampung data pengolahan *Fuzzy C-Means*, `fanny` adalah fungsi menjalankan *Fuzzy C-Means*, `sd4` adalah data yang telah dinormalisasi, 4 adalah jumlah *cluster* awal, `metric = "euclidean"` adalah metode perhitungan jarak pada *Fuzzy C-Means*, `stand = FALSE` adalah logika, jika bernilai TRUE maka data `sd4` akan dinormalisasi sebelum dihitung ketidaksamaannya. Berikut adalah hasil *clustering* dengan *Fuzzy C-Means*.

```
Fuzzy Clustering object of class 'fanny' :
m.ship.expon.      2
objective      155,6359
tolerance        1e-15
iterations        -1
converged         0
maxit            500
n                540
```

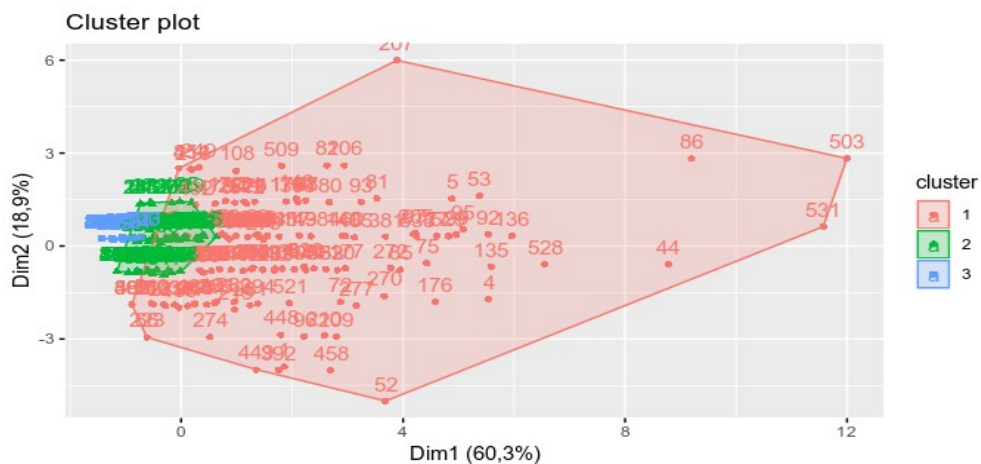
Gambar 4.53. Hasil *Clustering* dengan *Fuzzy C-Means*

2. Pada proses *Fuzzy C-Means* diperoleh *membership exponential*, fungsi objektif, dan toleransi. *Objective function* adalah penentuan nilai maksimum atau nilai minimum dari sebuah fungsi linear. Proses *Fuzzy C-Means* dapat dilihat pada Tabel . *Membership exponential* yaitu angka  $r$  benar-benar lebih besar dari satu yang menentukan eksponen keanggotaan yang digunakan dalam kriteria kecocokan. *Tolerance* yaitu tingkat kesalahan yang diizinkan pada penelitian. *Iter* yaitu jumlah iterasi untuk menjalankan proses algoritma *Fuzzy C-Means*. *Max Iter* yaitu iterasi maksimum yang diizinkan untuk menjalankan proses algoritma *Fuzzy C-Means*. Konvergen yaitu jumlah yang berhingga (*finite*).

Tabel 4.63. Nilai Keanggotaan, Fungsi Objektif, dan Toleransi pada *Fuzzy C-Means*

<i>Membership Exponential</i>	<i>Objective Function</i>	<i>Tolerance</i>	<i>Iter</i>	<i>Converged</i>	<i>Max Iter</i>
2	155,6359	1.00E-15	-1	0	500

- Setelah memperoleh hasil *clustering*, langkah selanjutnya yaitu menggambarkan distribusi data dengan jumlah *cluster* = 4.

Gambar 4.54. Pengelompokan Distribusi Data pada *Fuzzy C-Means*

Setelah dilakukan proses clustering menggunakan *Fuzzy C-Means*, diperoleh sebaran distribusi data dimana warna merah merupakan daerah *cluster* 1, warna hijau merupakan daerah *cluster* 2, dan warna biru merupakan daerah *cluster* 3. Untuk jumlah *cluster* = 4, RStudio melakukan *clustering* secara terbatas dengan jumlah *cluster* yang dihasilkan hanya 3 *cluster*.

#### m. Clustering pada *Fuzzy C-Means* dengan 550 data menggunakan 2 *cluster*

- Untuk melakukan *clustering* pada *Fuzzy C-Means*, data yang digunakan adalah data yang sudah melalui proses *preprocessing* dengan *cluster* awal (acak) berjumlah tiga *cluster*. Perintah untuk proses *Fuzzy C-Means* yaitu `proses.fcm5 <- fanny(sd5, 2, metric = "euclidean", stand = FALSE)`, dimana `proses.fcm5` adalah variabel untuk menampung data pengolahan *Fuzzy C-Means*, `fanny` adalah fungsi menjalankan *Fuzzy C-Means*, `sd5` adalah data yang telah dinormalisasi, 2 adalah jumlah *cluster* awal, `metric = "euclidean"` adalah metode perhitungan jarak pada *Fuzzy C-Means*, `stand = FALSE` adalah

logika, jika bernilai TRUE maka data `sd5` akan dinormalisasi sebelum dihitung ketidaksamaannya. Berikut adalah hasil *clustering* dengan *Fuzzy C-Means*.

```
Fuzzy Clustering object of class 'fanny' :
m.ship.expon.      2
objective      330,7679
tolerance      1e-15
iterations      29
converged        1
maxit          500
n              550
```

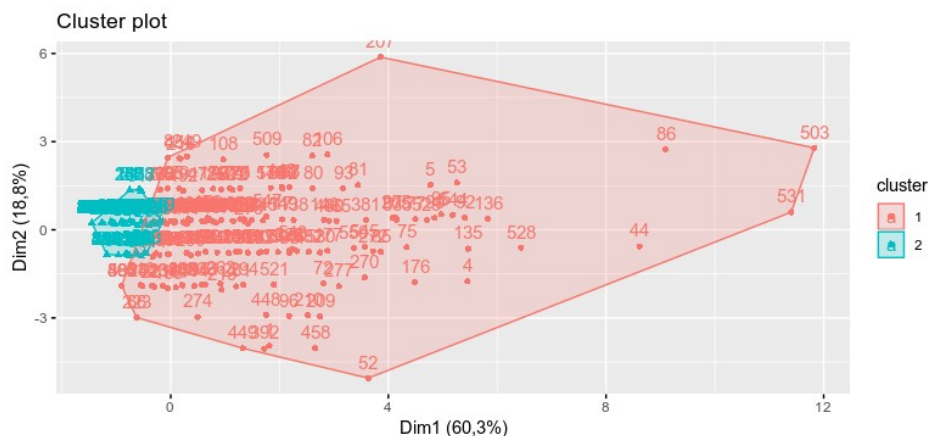
Gambar 4.55. Hasil *Clustering* dengan *Fuzzy C-Means*

2. Pada proses *Fuzzy C-Means* diperoleh *membership exponential*, fungsi objektif, dan toleransi. *Objective function* adalah penentuan nilai maksimum atau nilai minimum dari sebuah fungsi linear. Proses *Fuzzy C-Means* dapat dilihat pada Tabel . *Membership exponential* yaitu angka  $r$  benar-benar lebih besar dari satu yang menentukan eksponen keanggotaan yang digunakan dalam kriteria kecocokan. *Tolerance* yaitu tingkat kesalahan yang diizinkan pada penelitian. *Iter* yaitu jumlah iterasi untuk menjalankan proses algoritma *Fuzzy C-Means*. *Max Iter* yaitu iterasi maksimum yang diizinkan untuk menjalankan proses algoritma *Fuzzy C-Means*. Konvergen yaitu jumlah yang berhingga (*finite*).

Tabel 4.64. Nilai Keanggotaan, Fungsi Objektif, dan Toleransi pada *Fuzzy C-Means*

<i>Membership Exponential</i>	<i>Objective Function</i>	<i>Tolerance</i>	<i>Iter</i>	<i>Converged</i>	<i>Max Iter</i>
2	330.7679	1.00E-15	29	1	500

3. Setelah memperoleh hasil *clustering*, langkah selanjutnya yaitu menggambarkan distribusi data dengan jumlah *cluster* = 2.



Gambar 4.56. Pengelompokan Distribusi Data pada *Fuzzy C-Means*



Setelah dilakukan proses *clustering* menggunakan *Fuzzy C-Means*, diperoleh sebaran distribusi data dimana warna merah merupakan daerah *cluster* 1 dan warna biru merupakan daerah *cluster* 2.

**n. Clustering pada Fuzzy C-Means dengan 550 data menggunakan 3 cluster**

1. Untuk melakukan *clustering* pada *Fuzzy C-Means*, data yang digunakan adalah data yang sudah melalui proses *preprocessing* dengan *cluster* awal (acak) berjumlah tiga *cluster*. Perintah untuk proses *Fuzzy C-Means* yaitu `proses.fcm5 <- fanny(sd5, 3, metric = "euclidean", stand = FALSE)`, dimana `proses.fcm5` adalah variabel untuk menampung data pengolahan *Fuzzy C-Means*, `fanny` adalah fungsi menjalankan *Fuzzy C-Means*, `sd5` adalah data yang telah dinormalisasi, 3 adalah jumlah *cluster* awal, `metric = "euclidean"` adalah metode perhitungan jarak pada *Fuzzy C-Means*, `stand = FALSE` adalah logika, jika bernilai TRUE maka data `sd5` akan dinormalisasi sebelum dihitung ketidaksamaannya. Berikut adalah hasil *clustering* dengan *Fuzzy C-Means*.

```
Fuzzy Clustering object of class 'fanny' :
m.ship.expon.      2
objective      216,8453
tolerance        1e-15
iterations       112
converged         1
maxit            500
n                550
```

Gambar 4.57. Hasil *Clustering* dengan *Fuzzy C-Means*

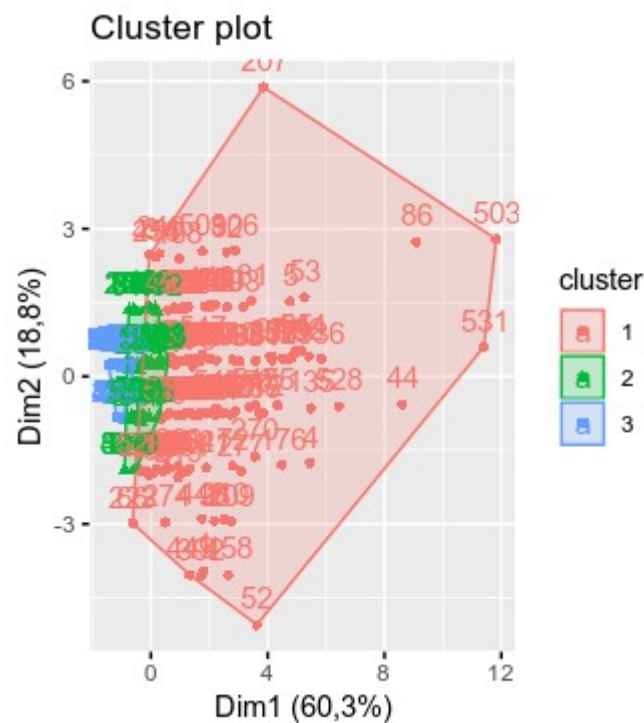
2. Pada proses *Fuzzy C-Means* diperoleh *membership exponential*, fungsi objektif, dan toleransi. *Objective function* adalah penentuan nilai maksimum atau nilai minimum dari sebuah fungsi linear. Proses *Fuzzy C-Means* dapat dilihat pada Tabel . *Membership exponential* yaitu angka  $r$  benar-benar lebih besar dari satu yang menentukan eksponen keanggotaan yang digunakan dalam kriteria kecocokan. *Tolerance* yaitu tingkat kesalahan yang diizinkan pada penelitian. *Iter* yaitu jumlah iterasi untuk menjalankan proses algoritma *Fuzzy C-Means*. *Max Iter* yaitu iterasi maksimum yang diizinkan untuk menjalankan proses algoritma *Fuzzy C-Means*. Konvergen yaitu jumlah yang berhingga (*finite*).



Tabel 4.65. Nilai Keanggotaan, Fungsi Objektif, dan Toleransi pada *Fuzzy C-Means*

<i>Membership Exponential</i>	<i>Objective Function</i>	<i>Tolerance</i>	<i>Iter</i>	<i>Converged</i>	<i>Max Iter</i>
2	216.8453	1.00E-15	112	1	500

- Setelah memperoleh hasil *clustering*, langkah selanjutnya yaitu menggambarkan distribusi data dengan jumlah *cluster* = 3.

Gambar 4.58. Pengelompokan Distribusi Data pada *Fuzzy C-Means*

Setelah dilakukan proses *clustering* menggunakan *Fuzzy C-Means*, diperoleh sebaran distribusi data dimana warna merah merupakan daerah *cluster* 1, warna hijau merupakan daerah *cluster* 2, dan warna biru merupakan daerah *cluster* 3.

#### o. *Clustering* pada *Fuzzy C-Means* dengan 550 data menggunakan 4 *cluster*

- Untuk melakukan *clustering* pada *Fuzzy C-Means*, data yang digunakan adalah data yang sudah melalui proses *preprocessing* dengan *cluster* awal (acak) berjumlah empat *cluster*. Perintah untuk proses *Fuzzy C-Means* yaitu `proses.fcm5 <- fanny(sd5, 4, metric = "euclidean", stand = FALSE)`, dimana `proses.fcm5` adalah variabel untuk menampung data

pengolahan *Fuzzy C-Means*, *fanny* adalah fungsi menjalankan *Fuzzy C-Means*, *sd5* adalah data yang telah dinormalisasi, 4 adalah jumlah *cluster* awal, *metric* = "euclidean" adalah metode perhitungan jarak pada *Fuzzy C-Means*, *stand* = FALSE adalah logika, jika bernilai TRUE maka data *sd5* akan dinormalisasi sebelum dihitung ketidaksamaannya. Berikut adalah hasil *clustering* dengan *Fuzzy C-Means*.

```
Fuzzy Clustering object of class 'fanny' :
m.ship.expon.      2
objective    159.7848
tolerance      1e-15
iterations      -1
converged       0
maxit         500
n             550
```

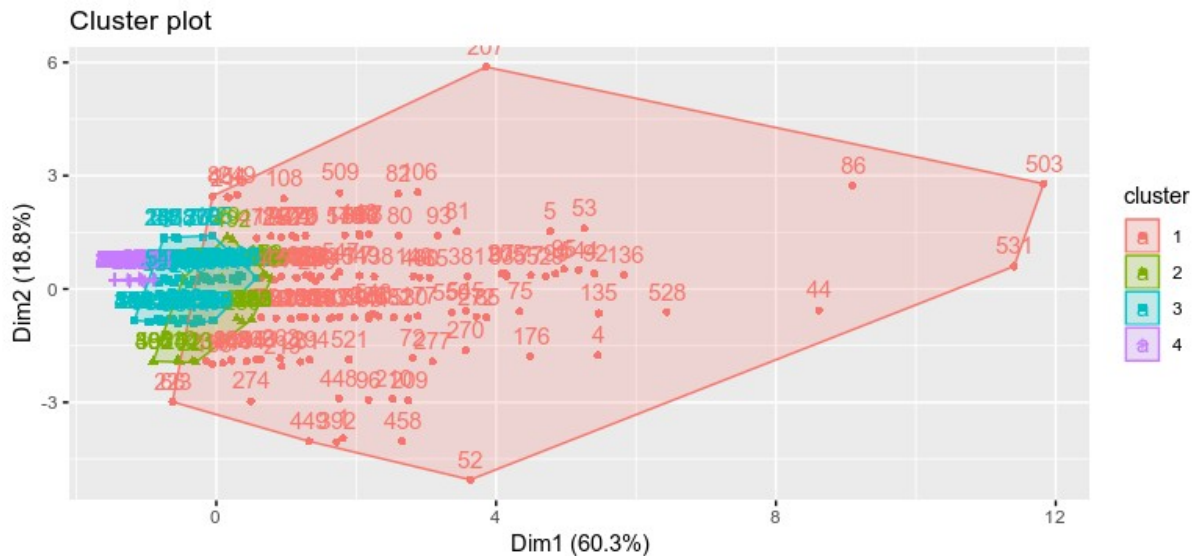
Gambar 4.59. Hasil *Clustering* dengan *Fuzzy C-Means*

2. Pada proses *Fuzzy C-Means* diperoleh *membership exponential*, fungsi objektif, dan toleransi. *Objective function* adalah penentuan nilai maksimum atau nilai minimum dari sebuah fungsi linear. Proses *Fuzzy C-Means* dapat dilihat pada Tabel . *Membership exponential* yaitu angka *r* benar-benar lebih besar dari satu yang menentukan eksponen keanggotaan yang digunakan dalam kriteria kecocokan. *Tolerance* yaitu tingkat kesalahan yang diizinkan pada penelitian. *Iter* yaitu jumlah iterasi untuk menjalankan proses algoritma *Fuzzy C-Means*. *Max Iter* yaitu iterasi maksimum yang diizinkan untuk menjalankan proses algoritma *Fuzzy C-Means*. Konvergen yaitu jumlah yang berhingga (*finite*).

Tabel 4.66. Nilai Keanggotaan, Fungsi Objektif, dan Toleransi pada *Fuzzy C-Means*

<i>Membership Exponential</i>	<i>Objective Function</i>	<i>Tolerance</i>	<i>Iter</i>	<i>Converged</i>	<i>Max Iter</i>
2	159.7848	1.00E-15	-1	0	500

3. Setelah memperoleh hasil *clustering*, langkah selanjutnya yaitu menggambarkan distribusi data dengan jumlah *cluster* = 4.



Gambar 4.60. Pengelompokan Distribusi Data pada *Fuzzy C-Means*

Setelah dilakukan proses *clustering* menggunakan *Fuzzy C-Means*, diperoleh sebaran distribusi data dimana warna merah merupakan daerah *cluster* 1, warna hijau merupakan daerah *cluster* 2, warna biru merupakan daerah *cluster* 3, dan warna ungu merupakan daerah *cluster* 4.

#### 4.1.4 Pengukuran Nilai Validasi

Setelah melakukan proses *clustering* dan mengelompokkan distribusi data pada *K-Means* dan *Fuzzy C-Means*, selanjutnya yaitu menentukan jumlah *cluster* terbaik berdasarkan nilai validasi DBI dan SI pada *K-Means* dan *Fuzzy C-Means*. Adapun rentang jumlah *cluster* yang ditentukan yaitu 2 *cluster*, 3 *cluster*, dan 4 *cluster*.

##### a. Pengukuran Nilai Validasi dengan *Spreadsheet*

##### 1. SI pada *K-Means*

Nilai validasi *Silhouette Index* (SI) pada *K-Means* dapat diukur menggunakan aplikasi *spreadsheet*. Aplikasi *spreadsheet* yang digunakan yaitu LibreOffice Calc 6.1.5.2. Berikut adalah langkah-langkah perhitungan pengukuran nilai validasi *Silhouette Index* (SI) pada aplikasi *spreadsheet*.

##### a) Nilai Validasi SI *K-Means* dengan 2 *Cluster*

1. Untuk melakukan pengukuran SI pada *K-Means*, langkah pertama yaitu membagi dua nilai *clustering*. Sebagai contoh jika nilai *cluster* yang diperoleh 5.84 pada *cluster* 1, maka nilai menjadi 5 dan 8. Hal ini dilakukan untuk

memperoleh nilai **a** dan **b** pada saat proses penentuan nilai validasi *Silhouette Index*. Berikut adalah tabel pembagian nilai *cluster*.

Tabel 4.67. Tabel Pembagian Nilai *Cluster*

C1		C2	
8	0	7	1
5	3	2	8
4	6	3	4
...	...	...	...
5	7	4	5
5	2	3	1
13	4	13	4

- Langkah selanjutnya yaitu melakukan perhitungan *intracuster* atau jarak antar objek di dalam *cluster* setiap objek ke objek lainnya. Jarak antar objek di dalam *cluster* disimbolkan sebagai **a**. Perhitungan jarak antar objek dalam *cluster* menggunakan metrik Euclidean. Sebagai contoh, pada Tabel 4.37 data ke-1 pada C1, yaitu 8 dan 0, maka akan dihitung jarak 8 dan 0 (jarak dirinya sendiri/jarak data ke-1), jarak data ke-2, dan jarak data ke-3 hingga selesai pada C1. Lalu data ke-2 pada C1 yaitu 5 dan 3, maka akan dihitung jarak antar 8 dan 0 pada data ke-1, jarak 5 dan 3 (jarak dirinya sendiri/jarak data ke-2), dan jarak data ke-3 hingga selesai pada C1. Hal ini juga berlaku pada C2. Penentuan titik awal objek untuk menghitung jarak antar objek di dalam *cluster* yaitu secara acak. Contohnya, jika data ke-1 pada C1 adalah 8 dan 0, maka jarak antar data yaitu  $d(x_1, c_1) = \sqrt{(8-8)^2 + (0-0)^2} = 0$ . Hasil perhitungan jarak *intracuster* dapat dilihat pada Tabel 4.38.
- Tahapan selanjutnya yaitu melakukan perhitungan *intercluster* atau jarak antar objek di luar *cluster* setiap objek ke objek lainnya. Metrik yang digunakan yaitu Euclidean. Jarak antar objek di luar *cluster* disimbolkan sebagai **b**. Sebagai contoh, pada Tabel 4.37 data ke-1 pada C1 yaitu 8 dan 0, maka akan dihitung jarak antar 7 dan 1 pada data ke-1, jarak data ke-2, dan jarak data ke-3 hingga selesai pada C2. Lalu data ke-2 pada C1 yaitu 5 dan 3, maka akan

dihitung jarak antar 7 dan 1 pada data ke-1, jarak data ke-2, dan jarak data ke-3 hingga selesai pada C2. Hal ini juga berlaku pada C2. Contohnya, jika nilai data ke-1 pada C1 adalah 8 dan 0 dan nilai data ke-1 pada C2 adalah 7 dan 1, maka jarak antar data yaitu  $d(c_1, x_1; c_2, x_2) = \sqrt{(8-7)^2 + (0-1)^2} = 1.4142$ . Hasil perhitungan jarak *intercluster* dapat dilihat pada Tabel 4.38.

Tabel 4.68. Jarak *Intracuster* dan *Intercluster K-Means*

Dist. Within C1	Dist. C1 to C2	Dist. Within C2	Dist. C2 to C1
0.0000	1.4142	0.0000	1.4142
4.2426	10.0000	8.6023	2.8284
7.2111	6.4031	5.0000	5.8310
...	...	...	...
7.6158	6.4031	5.0000	6.3246
3.6056	5.0990	4.0000	2.2361
6.4031	6.4031	6.7082	6.7082

- Selanjutnya yaitu menghitung nilai rata-rata pada *intracuster* atau jarak antar objek di dalam *cluster*. Sebagai contoh, jumlah data yaitu 550 dan jumlah total jarak antar objek pada C1 yaitu 3867.0261. Maka nilai rata-rata jarak antar objek pada C1 yaitu 7.0310 (**a**).
- Selanjutnya yaitu menghitung nilai rata-rata pada *intercluster* atau jarak antar objek di luar *cluster*. Sebagai contoh, jumlah data yaitu 550 dan jumlah total jarak antar objek pada C1 ke C2 yaitu 3780.8312. Maka nilai rata-rata jarak antar objek pada C1 ke C2 yaitu 6.9435 (**b**).
- Selanjutnya yaitu menghitung nilai validasi *Silhouette Index*. Perhitungan nilai

validasi menggunakan aturan, yaitu jika  $a < b$ , maka  $SI = 1 - \frac{a_i}{b_i}$ , namun jika

$a > b$ , maka  $SI = \frac{a_i}{b_i} - 1$ . Sebagai contoh, nilai **a** (rata-rata jarak antar objek pada C1) yaitu 6.9628 dan nilai **b** (rata-rata jarak antar objek C1 ke C2) yaitu

6.8742, maka  $SI = \frac{7.0310}{6.9435} = 0.0126$  Adapun nilai *Silhouette Index* pada  $K=2$

dapat dilihat pada tabel dibawah ini.

Tabel 4.69. Nilai SI dengan 2 *Cluster* pada *K-Means*

b1 (minimum distance C1 to C2)	6.9435
a1 (average distance within C1)	7.0310
<b>Silhouette C1</b>	<b>0.0126</b>
b2 (minimum distance C2 to C1)	5.9289
a2 (average distance within C2)	5.8320
<b>Silhouette C2</b>	<b>0.0166</b>
<b>Total Silhouette (C1 + C2)</b>	<b>0.0292</b>
<b>Nilai Validasi Silhouette dengan RStudio</b>	<b>0.7713</b>

7. Nilai SI pada *K-Means* dengan perhitungan manual  $K=2$  tergolong rendah yaitu 0.0292, sedangkan nilai SI pada *K-Means* dengan RStudio  $K=2$  tergolong tinggi yaitu 0.7713.

**b) Nilai Validasi SI *K-Means* dengan 3 *Cluster***

1. Untuk melakukan pengukuran SI pada *K-Means*, langkah pertama yaitu membagi dua nilai *clustering*. Sebagai contoh jika nilai *cluster* yang diperoleh 5.84 pada *cluster* 1, maka nilai menjadi 5 dan 8. Hal ini dilakukan untuk memperoleh nilai **a** dan **b** pada saat proses penentuan nilai validasi *Silhouette Index*. Berikut adalah tabel pembagian nilai *cluster*.

Tabel 4.70. Tabel Pembagian Nilai *Cluster*

C1		C2		C3	
5	2	3	8	6	0
1	9	7	4	1	7
1	8	4	8	2	7
C1		C2		C3	
...	...	...	...	...	...
3	1	4	0	4	1
2	1	5	5	3	0

C1		C2		C3	
11	7	6	3	12	8

2. Langkah selanjutnya yaitu melakukan perhitungan *intracluster* atau jarak antar objek di dalam *cluster* setiap objek ke objek lainnya. Jarak antar objek di dalam *cluster* disimbolkan sebagai **a**. Perhitungan jarak antar objek dalam *cluster* menggunakan metrik Euclidean. Sebagai contoh, pada Tabel 4.40 data ke-1 pada C1, yaitu 5 dan 2, maka akan dihitung jarak 5 dan 2 (jarak dirinya sendiri/jarak data ke-1), jarak data ke-2, dan jarak data ke-3 hingga selesai pada C1. Lalu data ke-2 pada C1 yaitu 1 dan 9, maka akan dihitung jarak antar 5 dan 2 pada data ke-1, jarak 1 dan 9 (jarak dirinya sendiri/jarak data ke-2), dan jarak data ke-3 hingga selesai pada C1. Hal ini juga berlaku pada C2 dan C3. Penentuan titik awal objek untuk menghitung jarak antar objek di dalam *cluster* yaitu secara acak. Contohnya, jika data ke-1 pada C1 adalah 5 dan 2, maka jarak antar data yaitu  $d(x_1, c_1) = \sqrt{(5-5)^2 + (2-2)^2} = 0$ . Hasil perhitungan jarak *intracluster* dapat dilihat pada Tabel 4.41.
3. Tahapan selanjutnya yaitu melakukan perhitungan *intercluster* atau jarak antar objek di luar *cluster* setiap objek ke objek lainnya. Metrik yang digunakan yaitu Euclidean. Jarak antar objek di luar *cluster* disimbolkan sebagai **b**. Sebagai contoh, pada Tabel 4.40 data ke-1 pada C1 yaitu 5 dan 2, maka akan dihitung jarak antar 3 dan 8 pada data ke-1, jarak data ke-2, dan jarak data ke-3 hingga selesai pada C2. Lalu data ke-2 pada C1 yaitu 1 dan 9, maka akan dihitung jarak antar 5 dan 2 pada data ke-1, jarak data ke-2, dan jarak data ke-3 hingga selesai pada C2. Hal ini juga berlaku pada C2 dan C3. Contohnya, jika nilai data ke-1 pada C1 adalah 5 dan 2 dan nilai data ke-1 pada C2 adalah 3 dan 8, maka jarak antar data yaitu  $d(c_1, x_1; c_2, x_2) = \sqrt{(5-3)^2 + (2-8)^2} = 6.3246$ . Hasil perhitungan jarak *intercluster* dapat dilihat pada Tabel 4.41.

Tabel 4.71. Jarak *Intracuster* dan *Intercluster K-Means*

Dist. Within C1	Dist. C1 to C2	Dist. C1 to C3	Dist. Within C2	Dist. C2 to C1	Dist. C2 to C3	Dist. Within C3	Dist. C3 to C1	Dist. C3 to C2
0.0000	6.3246	2.2361	0.0000	6.3246	8.5440	0.0000	2.2361	8.5440
8.0623	2.8284	6.4031	5.6569	2.2361	2.2361	8.6023	10.2956	4.1231
7.2111	6.0828	5.8310	1.0000	2.0000	1.4142	8.0623	9.4340	8.2462
...	...	...	...	...	...	...	...	...
2.2361	2.2361	1.4142	8.0623	7.0000	7.0711	2.2361	3.1623	2.0000
3.1623	3.0000	2.8284	3.6056	7.0711	8.0000	3.0000	4.1231	5.0990
7.8102	1.4142	9.2195	5.8310	8.0623	9.0000	10.0000	8.6023	3.0000

4. Selanjutnya yaitu menghitung nilai rata-rata pada *intracuster* atau jarak antar objek di dalam *cluster*. Sebagai contoh, jumlah data yaitu 550 dan jumlah total jarak antar objek pada C1 yaitu 2456.8242. Maka nilai rata-rata jarak antar objek pada C1 yaitu 4.4670 (**a**).
5. Selanjutnya yaitu menghitung nilai rata-rata pada *intercluster* atau jarak antar objek di luar *cluster*. Sebagai contoh, jumlah data yaitu 550 dan jumlah total jarak antar objek pada C1 ke C2 yaitu 2323.8242. Maka nilai rata-rata jarak antar objek pada C1 ke C2 yaitu 4.2250 (**b**).
6. Selanjutnya yaitu menghitung nilai validasi *Silhouette Index*. Perhitungan nilai

validasi menggunakan aturan, yaitu jika  $a < b$ , maka  $SI = 1 - \frac{a_i}{b_i}$ , namun jika

$a > b$ , maka  $SI = \frac{a_i}{b_i} - 1$ . Sebagai contoh, nilai **a** (rata-rata jarak antar objek pada C1) yaitu 4.4670 dan nilai **b** (rata-rata jarak antar objek C1 ke C2) yaitu

4.2250, maka  $SI = \frac{4.4670}{4.2250} - 1 = 0.0573$  Adapun nilai *Silhouette Index* pada

K=2 dapat dilihat pada tabel dibawah ini.



Tabel 4.72. Nilai SI dengan 3 *Cluster* pada *K-Means*

b1 (minimum distance C1 to C2 & C1 to C3)	4.2250
a1 (average distance within C1)	4.4670
<b>Silhouette C1</b>	<b>0.0573</b>
b2 (minimum distance C2 to C1 & C2 to C3)	4.4467
a2 (average distance within C2)	5.8426
<b>Silhouette C2</b>	<b>0.3139</b>
b3 (minimum distance C3 to C1 & C3 to C2)	5.1115
a3 (average distance within C3)	5.7936
<b>Silhouette C3</b>	<b>0.1335</b>
<b>Total Silhouette (C1 + C2 + C3)</b>	<b>0.5046</b>
<b>Nilai Validasi Silhouette dengan RStudio</b>	<b>0.7486</b>

7. Nilai SI pada *K-Means* dengan perhitungan manual  $K=3$  tergolong masih rendah yaitu 0.5046, sedangkan nilai SI pada *K-Means* dengan RStudio  $K=3$  tergolong tinggi yaitu 0.7486.

**c) Nilai Validasi SI *K-Means* dengan 4 *Cluster***

1. Untuk melakukan pengukuran SI pada *K-Means*, langkah pertama yaitu membagi dua nilai *clustering*. Sebagai contoh jika nilai *cluster* yang diperoleh 5.84 pada *cluster* 1, maka nilai menjadi 5 dan 8. Hal ini dilakukan untuk memperoleh nilai **a** dan **b** pada saat proses penentuan nilai validasi *Silhouette Index*. Berikut adalah tabel pembagian nilai *cluster*.

Tabel 4.73. Tabel Pembagian Nilai *Cluster*

C1		C2		C3		C4	
5	9	3	1	6	4	5	1
1	0	7	1	1	7	1	8
1	0	4	0	2	0	1	4
...	...	...	...	...	...	...	...
3	0	4	7	4	4	3	7
2	4	5	4	3	0	2	2
11	0	6	3	12	3	11	7

2. Langkah selanjutnya yaitu melakukan perhitungan *intracluster* atau jarak antar objek di dalam *cluster* setiap objek ke objek lainnya. Jarak antar objek di dalam *cluster* disimbolkan sebagai **a**. Perhitungan jarak antar objek dalam *cluster* menggunakan metrik Euclidean. Sebagai contoh, pada Tabel 4.43 data ke-1 pada C1, yaitu 5 dan 9, maka akan dihitung jarak 5 dan 9 (jarak dirinya sendiri/jarak data ke-1), jarak data ke-2, dan jarak data ke-3 hingga selesai pada C1. Lalu data ke-2 pada C1 yaitu 1 dan 0, maka akan dihitung jarak antar 5 dan 9 pada data ke-1, jarak 1 dan 0 (jarak dirinya sendiri/jarak data ke-2), dan jarak data ke-3 hingga selesai pada C1. Hal ini juga berlaku pada C2, C3 dan C4. Penentuan titik awal objek untuk menghitung jarak antar objek di dalam *cluster* yaitu secara acak. Contohnya, jika data ke-1 pada C1 adalah 5 dan 9, maka jarak antar data yaitu  $d(x_1, c_1) = \sqrt{(5-5)^2 + (9-9)^2} = 0$ . Hasil perhitungan jarak *intracluster* dapat dilihat pada Tabel 4.45.
3. Tahapan selanjutnya yaitu melakukan perhitungan *intercluster* atau jarak antar objek di luar *cluster* setiap objek ke objek lainnya. Metrik yang digunakan yaitu Euclidean. Jarak antar objek di luar *cluster* disimbolkan sebagai **b**. Sebagai contoh, pada Tabel 4.43 data ke-1 pada C1 yaitu 5 dan 9, maka akan dihitung jarak antar 3 dan 1 pada data ke-1, jarak data ke-2, dan jarak data ke-3 hingga selesai pada C2. Lalu data ke-2 pada C1 yaitu 1 dan 0, maka akan dihitung jarak antar 5 dan 9 pada data ke-1, jarak data ke-2, dan jarak data ke-3 hingga selesai pada C2. Hal ini juga berlaku pada C2, C3 dan C4. Contohnya, jika nilai data ke-1 pada C1 adalah 5 dan 9 dan nilai data ke-1 pada C2 adalah 3 dan 1, maka jarak antar data yaitu  $d(c_1, x_1; c_2, x_2) = \sqrt{(5-3)^2 + (9-1)^2} = 8.2462$ . Hasil perhitungan jarak *intercluster* dapat dilihat pada Tabel 4.45.
4. Selanjutnya yaitu menghitung nilai rata-rata pada *intracluster* atau jarak antar objek di dalam *cluster*. Sebagai contoh, jumlah data yaitu 550 dan jumlah total jarak antar objek pada C1 yaitu 3894.6930. Maka nilai rata-rata jarak antar objek pada C1 yaitu 7.0813 (**a**).
5. Selanjutnya yaitu menghitung nilai rata-rata pada *intercluster* atau jarak antar objek di luar *cluster*. Sebagai contoh, jumlah data yaitu 550 dan jumlah total

jarak antar objek pada C1 ke C2 yaitu 3180.2057. Maka nilai rata-rata jarak antar objek pada C1 ke C2 yaitu 5.7822 (**b**).

6. Selanjutnya yaitu menghitung nilai validasi *Silhouette Index*. Perhitungan nilai

validasi menggunakan aturan, yaitu jika  $a < b$ , maka  $SI = 1 - \frac{a_i}{b_i}$ , namun jika

$a > b$ , maka  $SI = \frac{a_i}{b_i} - 1$ . Sebagai contoh, nilai **a** (rata-rata jarak antar objek pada C1) yaitu 7.0813 dan nilai **b** (rata-rata jarak antar objek C1 ke C2) yaitu

5.782, maka  $SI = \frac{7.0813}{5.7822} - 1 = 0.2247$ . Adapun nilai *Silhouette Index* pada

K=4 dapat dilihat pada tabel dibawah ini.

Tabel 4.74. Nilai SI dengan 4 *Cluster* pada *K-Means*

b1 (minimum distance C1 to C2, C1 to C3, & C1 to C4)	5.7822
a1 (average distance within C1)	7.0813
<b>Silhouette C1</b>	<b>0.2247</b>
b2 (minimum distance C2 to C1, C2 to C3, & C2 to C4)	3.5239
a2 (average distance within C2)	5.7176
<b>Silhouette C2</b>	<b>0.6225</b>
b3 (minimum distance C3 to C1, C3 to C2, & C3 to C4)	3.6654
a3 (average distance within C3)	4.9934
<b>Silhouette C3</b>	<b>0.3623</b>
b4 (minimum distance C4 to C1, C4 to C2, & C4 to C3)	4.3048
a4 (average distance within C4)	4.3452
<b>Silhouette C4</b>	<b>0.0094</b>
<b>Total Silhouette (C1 + C2 + C3 + C4)</b>	<b>1.2189</b>
<b>Nilai Validasi Silhouette dengan RStudio</b>	<b>0.5694</b>

7. Nilai SI pada *K-Means* dengan perhitungan manual K=4 tergolong rendah yaitu 1.2189, sedangkan nilai SI pada *K-Means* dengan RStudio K=4 tergolong masih rendah yaitu 0.5694.

Tabel 4.75. Jarak *Intrachuster* dan *Intercluster K-Means*

Dist. Within C1	Dist. C1 to C2	Dist. C1 to C3	Dist. C1 to C4	Dist. Within C2	Dist. C2 to C1	Dist. C2 to C3	Dist. C2 to C4	Dist. Within C3	Dist. C3 to C1	Dist. C3 to C2	Dist. C3 to C4	Dist. Within C4	Dist. C4 to C1	Dist. C4 to C2	Dist. C4 to C3
0.0000	8.2462	5.0990	8.0000	0.0000	8.2462	4.2426	2.0000	0.0000	5.0990	4.2426	3.1623	0.0000	8.0000	2.0000	3.1623
9.8489	8.2462	4.4721	4.1231	4.0000	2.2361	6.3246	7.2801	5.8310	6.4031	3.1623	6.4031	8.0623	4.1231	2.0000	7.2111
9.8489	9.0554	9.4868	6.4031	1.4142	2.2361	1.4142	3.6056	5.6569	6.4031	4.4721	5.0000	5.0000	4.1231	1.4142	3.1623
9.2195	3.0000	10.0499	12.0416	9.4340	12.5300	13.8924	11.0000	9.8489	8.5440	5.3852	8.5440	9.0000	10.8167	8.5440	12.2066
12.0416	5.6569	11.1803	9.2195	7.2111	11.0000	12.3693	12.5300	9.0000	8.5440	3.1623	8.5440	10.8167	9.0000	5.6569	10.4403
3.6056	4.4721	9.4868	7.2801	8.4853	5.0000	1.4142	1.0000	5.6569	3.6056	4.2426	3.6056	2.2361	5.3852	7.2111	3.1623
9.8489	2.2361	9.4868	9.8489	6.7082	2.2361	1.4142	2.2361	5.6569	6.4031	3.0000	6.4031	4.1231	4.1231	6.0828	3.1623
9.4868	6.7082	3.1623	9.4868	5.3852	1.4142	7.0711	1.4142	5.6569	5.6569	2.2361	5.6569	3.1623	3.1623	3.6056	7.6158
3.6056	3.1623	3.1623	3.1623	5.8310	6.0828	7.0711	7.0711	5.6569	5.0000	2.0000	5.6569	7.6158	6.7082	5.0990	7.6158
9.4868	5.8310	3.6056	9.4868	5.8310	1.4142	6.0828	1.4142	5.0000	5.6569	2.0000	5.6569	3.1623	3.1623	4.2426	6.7082
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3.6056	8.2462	9.8489	5.0000	0.0000	7.8102	6.0828	6.4031	5.0000	3.6056	4.2426	2.2361	5.0000	6.7082	2.0000	4.1231
9.8489	7.2111	12.5300	10.1980	6.3246	2.2361	13.1529	13.4164	10.0499	6.4031	3.1623	9.4868	11.6619	4.1231	4.4721	11.1803
10.2956	6.0000	9.2195	5.3852	2.8284	7.0711	8.0623	9.2195	5.3852	5.6569	1.4142	5.0000	7.8102	5.0990	2.0000	6.0828
0.0000	7.2111	7.0711	5.0000	2.8284	8.2462	3.1623	3.6056	2.0000	5.0990	5.0990	1.0000	3.0000	8.0000	4.4721	1.4142
5.0000	8.0623	9.8489	6.7082	1.0000	6.4031	6.0828	5.3852	5.0000	2.2361	3.6056	2.2361	3.6056	5.0000	1.0000	4.1231
9.2195	2.2361	5.0990	2.8284	6.0828	1.0000	3.1623	6.0000	2.0000	5.0000	3.6056	4.2426	6.3246	2.2361	6.0828	3.1623
5.8310	5.0000	9.2195	7.6158	3.6056	3.1623	1.0000	1.4142	5.0000	4.0000	1.0000	4.4721	3.1623	4.2426	3.0000	2.2361
10.8167	6.0828	9.2195	6.3246	3.6056	8.0623	9.2195	10.0000	6.0828	6.4031	1.0000	5.8310	8.4853	6.0828	2.2361	7.2801

## 2. SI pada *Fuzzy C-Means*

Nilai validasi *Silhouette Index* (SI) pada *Fuzzy C-Means* dapat diukur menggunakan aplikasi *spreadsheet*. Aplikasi *spreadsheet* yang digunakan yaitu LibreOffice Calc 6.1.5.2. Berikut adalah langkah-langkah perhitungan pengukuran nilai validasi *Silhouette Index* (SI) pada aplikasi *spreadsheet*.

### a) Nilai Validasi SI dengan 2 *Cluster*

1. Untuk melakukan pengukuran SI pada *Fuzzy C-Means*, langkah pertama yaitu membagi dua nilai *clustering*. Sebagai contoh jika nilai *cluster* yang diperoleh 5.84 pada *cluster* 1, maka nilai menjadi 5 dan 8. Hal ini dilakukan untuk memperoleh nilai **a** dan **b** pada saat proses penentuan nilai validasi *Silhouette Index*. Berikut adalah tabel pembagian nilai *cluster*.

Tabel 4.76. Tabel Pembagian Nilai *Cluster*

U1		U2	
0	5	0	4
0	4	0	5
0	6	0	3
...	...	...	...
0	5	0	4
0	5	0	4
0	5	0	4

2. Langkah selanjutnya yaitu melakukan perhitungan *intracluster* atau jarak antar objek di dalam *cluster* setiap objek ke objek lainnya. Jarak antar objek di dalam *cluster* disimbolkan sebagai **a**. Perhitungan jarak antar objek dalam *cluster* menggunakan metrik Euclidean. Sebagai contoh, pada Tabel 4.46 data ke-1 pada U1, yaitu 0 dan 5, maka akan dihitung jarak 0 dan 5 (jarak dirinya sendiri/jarak data ke-1), jarak data ke-2, dan jarak data ke-3 hingga selesai pada U1. Lalu data ke-2 pada U1 yaitu 0 dan 4, maka akan dihitung jarak antar 0 dan 5 pada data ke-1, jarak 0 dan 4 (jarak dirinya sendiri/jarak data ke-2), dan jarak data ke-3 hingga selesai pada U1. Hal ini juga berlaku pada U2. Penentuan titik awal objek untuk menghitung jarak antar objek di dalam

*cluster* yaitu secara acak. Contohnya, jika data ke-1 pada U1 adalah 0 dan 5, maka jarak antar data yaitu  $d(x_1, c_1) = \sqrt{(0-0)^2 + (5-5)^2} = 0$ . Hasil perhitungan jarak *intracluster* dapat dilihat pada Tabel 4.47.

3. Tahapan selanjutnya yaitu melakukan perhitungan *intercluster* atau jarak antar objek di luar *cluster* setiap objek ke objek lainnya. Metrik yang digunakan yaitu Euclidean. Jarak antar objek di luar *cluster* disimbolkan sebagai **b**. Sebagai contoh, pada Tabel 4.46 data ke-1 pada U1 yaitu 0 dan 5, maka akan dihitung jarak antar 0 dan 4 pada data ke-1, jarak data ke-2, dan jarak data ke-3 hingga selesai pada U2. Lalu data ke-2 pada U1 yaitu 0 dan 4, maka akan dihitung jarak antar 0 dan 4 pada data ke-1, jarak data ke-2, dan jarak data ke-3 hingga selesai pada U2. Hal ini juga berlaku pada U2. Contohnya, jika nilai data ke-1 pada U1 adalah 0 dan 5 dan nilai data ke-1 pada U2 adalah 0 dan 4, maka jarak antar data yaitu  $d(c_1, x_1; c_2, x_2) = \sqrt{(0-5)^2 + (0-4)^2} = 1$ . Hasil perhitungan jarak *intercluster* dapat dilihat pada Tabel 4.47.

Tabel 4.77. Jarak *Intracluster* dan *Intercluster Fuzzy-Means*

Dist. Within U1	Dist. U1 to U2	Dist. Within U2	Dist. U2 to U1
0	1	0	1
1	0	1	0
1	2	1	4
...	...	...	...
0	1	0	1
0	1	0	1
0	1	0	1

4. Selanjutnya yaitu menghitung nilai rata-rata pada *intracluster* atau jarak antar objek di dalam *cluster*. Sebagai contoh, jumlah data yaitu 550 dan jumlah total jarak antar objek pada U1 yaitu 378. Maka nilai rata-rata jarak antar objek pada U1 yaitu 0.6873 (**a**).
5. Selanjutnya yaitu menghitung nilai rata-rata pada *intercluster* atau jarak antar objek di luar *cluster*. Sebagai contoh, jumlah data yaitu 550 dan jumlah total

jarak antar objek pada U1 ke U2 yaitu 192. Maka nilai rata-rata jarak antar objek pada U1 ke U2 yaitu 0.3491 (**b**).

6. Selanjutnya yaitu menghitung nilai validasi *Silhouette Index*. Perhitungan nilai validasi menggunakan aturan, yaitu jika  $a < b$ , maka  $SI = 1 - \frac{a_i}{b_i}$ , namun jika  $a > b$ , maka  $SI = \frac{a_i}{b_i} - 1$ . Sebagai contoh, nilai **a** (rata-rata jarak antar objek pada C1) yaitu 6.9628 dan nilai **b** (rata-rata jarak antar objek C1 ke C2) yaitu 6.8742, maka  $SI = \frac{0.6873}{0.3491} - 1 = 0.9688$ . Adapun nilai *Silhouette Index* pada K=2 dapat dilihat pada tabel dibawah ini.

Tabel 4.78. Nilai SI dengan 2 *Cluster* pada *Fuzzy C-Means*

b1 (minimum distance U1 to U2)	0.3491
a1 (average distance within U1)	0.6873
<b>Silhouette U1</b>	<b>0.9688</b>
b2 (minimum distance U2 to U1)	0.3855
a2 (average distance within U2)	0.6873
<b>Silhouette U2</b>	<b>0.7830</b>
<b>Total Silhouette (U1 + U2)</b>	<b>1.7518</b>
<b>Nilai Validasi Silhouette dengan RStudio</b>	<b>0.7713</b>

7. Nilai SI pada *Fuzzy C-Means* dengan perhitungan manual K=2 tergolong rendah yaitu 1.7518, sedangkan nilai SI pada *K-Means* dengan RStudio K=2 tergolong tinggi yaitu 0.7713.

#### b) Nilai Validasi SI dengan 3 *Cluster*

1. Untuk melakukan pengukuran SI pada *Fuzzy C-Means*, langkah pertama yaitu membagi dua nilai *clustering*. Sebagai contoh jika nilai *cluster* yang diperoleh 5.84 pada *cluster* 1, maka nilai menjadi 5 dan 8. Hal ini dilakukan untuk memperoleh nilai **a** dan **b** pada saat proses penentuan nilai validasi *Silhouette Index*. Berikut adalah tabel pembagian nilai *cluster*.

Tabel 4.79. Tabel Pembagian Nilai *Cluster*

U1		U2		U3	
0	3	0	3	0	3
0	2	0	3	0	3
0	4	0	3	0	2
...	...	...	...	...	...
0	4	0	3	0	2
0	3	0	3	0	2
0	3	0	3	0	3

2. Langkah selanjutnya yaitu melakukan perhitungan *intracuster* atau jarak antar objek di dalam *cluster* setiap objek ke objek lainnya. Jarak antar objek di dalam *cluster* disimbolkan sebagai **a**. Perhitungan jarak antar objek dalam *cluster* menggunakan metrik Euclidean. Sebagai contoh, pada Tabel 4.49 data ke-1 pada U1, yaitu 0 dan 3, maka akan dihitung jarak 0 dan 3 (jarak dirinya sendiri/jarak data ke-1), jarak data ke-2, dan jarak data ke-3 hingga selesai pada U1. Lalu data ke-2 pada U1 yaitu 0 dan 2, maka akan dihitung jarak antar 0 dan 3 pada data ke-1, jarak 0 dan 2 (jarak dirinya sendiri/jarak data ke-2), dan jarak data ke-3 hingga selesai pada U1. Hal ini juga berlaku pada U2 dan U3. Penentuan titik awal objek untuk menghitung jarak antar objek di dalam *cluster* yaitu secara acak. Contohnya, jika data ke-1 pada U1 adalah 0 dan 3, maka jarak antar data yaitu  $d(x_1, c_1) = \sqrt{(5-5)^2 + (2-2)^2} = 0$ . Hasil perhitungan jarak *intracuster* dapat dilihat pada Tabel 4.50.
3. Tahapan selanjutnya yaitu melakukan perhitungan *intercluster* atau jarak antar objek di luar *cluster* setiap objek ke objek lainnya. Metrik yang digunakan yaitu Euclidean. Jarak antar objek di luar *cluster* disimbolkan sebagai **b**. Sebagai contoh, pada Tabel 4.49 data ke-1 pada U1 yaitu 0 dan 3, maka akan dihitung jarak antar 0 dan 3 pada data ke-1, jarak data ke-2, dan jarak data ke-3 hingga selesai pada U2. Lalu data ke-2 pada U1 yaitu 0 dan 2, maka akan dihitung jarak antar 0 dan 3 pada data ke-1, jarak data ke-2, dan jarak data ke-3 hingga selesai pada U2. Hal ini juga berlaku pada U2 dan U3. Contohnya, jika nilai data ke-1 pada U1 adalah 0 dan 3 dan nilai data ke-1 pada U2 adalah 0



dan 3, maka jarak antar data yaitu  $d(c_1, x_1; c_2, x_2) = \sqrt{(0-0)^2 + (3-3)^2} = 0$ .

Hasil perhitungan jarak *intercluster* dapat dilihat pada Tabel 4.50.

Tabel 4.80. Jarak *Intracuster* dan *Intercluster Fuzzy-Means*

Dist. Within U1	Dist. U1 to U2	Dist. U1 to U3	Dist. Within U2	Dist. U2 to U1	Dist. U2 to U3	Dist. Within U3	Dist. U3 to U1	Dist. U3 to U2
0	0	0	0	0	0	0	0	0
1	0	0	0	1	0	0	1	0
1	0	1	0	1	1	1	1	0
...	...	...	...	...	...	...	...	...
1	0	1	0	1	1	1	1	0
0	0	1	0	0	1	1	0	0
0	0	0	0	0	0	0	0	0

- Selanjutnya yaitu menghitung nilai rata-rata pada *intracuster* atau jarak antar objek di dalam *cluster*. Sebagai contoh, jumlah data yaitu 550 dan jumlah total jarak antar objek pada U1 yaitu 425. Maka nilai rata-rata jarak antar objek pada U1 yaitu 0.7727 (**a**).
- Selanjutnya yaitu menghitung nilai rata-rata pada *intercluster* atau jarak antar objek di luar *cluster*. Sebagai contoh, jumlah data yaitu 550 dan jumlah total jarak antar objek pada U1 ke U2 yaitu 40. Maka nilai rata-rata jarak antar objek pada U1 ke U2 yaitu 0.0727 (**b**).
- Selanjutnya yaitu menghitung nilai validasi *Silhouette Index*. Perhitungan nilai

validasi menggunakan aturan, yaitu jika  $a < b$ , maka  $SI = 1 - \frac{a_i}{b_i}$ , namun jika

$a > b$ , maka  $SI = \frac{a_i}{b_i} - 1$ . Sebagai contoh, nilai **a** (rata-rata jarak antar objek pada C1) yaitu 4.4670 dan nilai **b** (rata-rata jarak antar objek C1 ke C2) yaitu

4.2250, maka  $SI = \frac{0.7727}{0.0727} - 1 = 9.6250$ . Adapun nilai *Silhouette Index* pada

K=2 dapat dilihat pada tabel dibawah ini.

Tabel 4.81. Nilai SI dengan 3 *Cluster* pada *Fuzzy C-Means*

b1 (minimum distance U1 to U2 & U1 to U3)	0.0727
a1 (average distance within U1)	0.7727
<b>Silhouette U1</b>	<b>9.6250</b>
b2 (minimum distance U2 to U1 & U2 to U3)	0.5855
a2 (average distance within U2)	0.0727
<b>Silhouette U2</b>	<b>-7.0500</b>
b3 (minimum distance U3 to U1 & U3 to U2)	0.0727
a3 (average distance within U3)	0.5855
<b>Silhouette U3</b>	<b>7.0500</b>
<b>Total Silhouette (U1 + U2 + U3)</b>	<b>9.6250</b>
<b>Nilai Validasi Silhouette dengan RStudio</b>	<b>0.6987</b>

7. Nilai SI pada *Fuzzy C-Means* dengan perhitungan manual K=3 tergolong masih rendah yaitu 9.6250, sedangkan nilai SI pada *K-Means* dengan RStudio K=3 tergolong tinggi yaitu 0.7486.

**c) Nilai Validasi SI dengan 4 *Cluster***

1. Untuk melakukan pengukuran SI pada *K-Means*, langkah pertama yaitu membagi dua nilai *clustering*. Sebagai contoh jika nilai *cluster* yang diperoleh 5.84 pada *cluster* 1, maka nilai menjadi 5 dan 8. Hal ini dilakukan untuk memperoleh nilai **a** dan **b** pada saat proses penentuan nilai validasi *Silhouette Index*. Berikut adalah tabel pembagian nilai *cluster*.

Tabel 4.82. Tabel Pembagian Nilai *Cluster*

U1		U2		U3		U4	
0	3	0	3	0	3	0	0
0	2	0	3	0	3	0	0
0	4	0	2	0	2	0	0
...	...	...	...	...	...	...	...
0	3	0	3	0	3	0	0
0	3	0	3	0	3	0	0
0	3	0	3	0	3	0	4

2. Langkah selanjutnya yaitu melakukan perhitungan *intracluster* atau jarak antar objek di dalam *cluster* setiap objek ke objek lainnya. Jarak antar objek di dalam *cluster* disimbolkan sebagai **a**. Perhitungan jarak antar objek dalam *cluster* menggunakan metrik Euclidean. Sebagai contoh, pada Tabel 4.52 data ke-1 pada U1, yaitu 0 dan 3, maka akan dihitung jarak 0 dan 3 (jarak dirinya sendiri/jarak data ke-1), jarak data ke-2, dan jarak data ke-3 hingga selesai pada U1. Lalu data ke-2 pada U1 yaitu 0 dan 2, maka akan dihitung jarak antar 0 dan 3 pada data ke-1, jarak 0 dan 2 (jarak dirinya sendiri/jarak data ke-2), dan jarak data ke-3 hingga selesai pada U1. Hal ini juga berlaku pada U2, U3 dan U4. Penentuan titik awal objek untuk menghitung jarak antar objek di dalam *cluster* yaitu secara acak. Contohnya, jika data ke-1 pada U1 adalah 0 dan 3, maka jarak antar data yaitu  $d(x_1, c_1) = \sqrt{(0-0)^2 + (3-3)^2} = 0$ . Hasil perhitungan jarak *intracluster* dapat dilihat pada Tabel 4.54.
3. Tahapan selanjutnya yaitu melakukan perhitungan *intercluster* atau jarak antar objek di luar *cluster* setiap objek ke objek lainnya. Metrik yang digunakan yaitu Euclidean. Jarak antar objek di luar *cluster* disimbolkan sebagai **b**. Sebagai contoh, pada Tabel 4.52 data ke-1 pada U1 yaitu 0 dan 3, maka akan dihitung jarak antar 0 dan 3 pada data ke-1, jarak data ke-2, dan jarak data ke-3 hingga selesai pada U2. Lalu data ke-2 pada C1 yaitu 0 dan 2, maka akan dihitung jarak antar 0 dan 3 pada data ke-1, jarak data ke-2, dan jarak data ke-3 hingga selesai pada U2. Hal ini juga berlaku pada U2, U3 dan U4. Contohnya, jika nilai data ke-1 pada U1 adalah 0 dan 3 dan nilai data ke-1 pada C2 adalah 0 dan 3, maka jarak antar data yaitu  $d(c_1, x_1; c_2, x_2) = \sqrt{(0-0)^2 + (3-3)^2} = 0$ . Hasil perhitungan jarak *intercluster* dapat dilihat pada Tabel 4.54.
4. Selanjutnya yaitu menghitung nilai rata-rata pada *intracluster* atau jarak antar objek di dalam *cluster*. Sebagai contoh, jumlah data yaitu 550 dan jumlah total jarak antar objek pada U1 yaitu 351. Maka nilai rata-rata jarak antar objek pada U1 yaitu 0.6382 (**a**).
5. Selanjutnya yaitu menghitung nilai rata-rata pada *intercluster* atau jarak antar objek di luar *cluster*. Sebagai contoh, jumlah data yaitu 550 dan jumlah total

jarak antar objek pada U1 ke U2 yaitu 32. Maka nilai rata-rata jarak antar objek pada U1 ke U2 yaitu 0.0582 (**b**).

6. Selanjutnya yaitu menghitung nilai validasi *Silhouette Index*. Perhitungan nilai

validasi menggunakan aturan, yaitu jika  $a < b$ , maka  $SI = 1 - \frac{a_i}{b_i}$ , namun jika

$a > b$ , maka  $SI = \frac{a_i}{b_i} - 1$ . Sebagai contoh, nilai **a** (rata-rata jarak antar objek pada C1) yaitu 7.0813 dan nilai **b** (rata-rata jarak antar objek C1 ke C2) yaitu

5.782, maka  $SI = \frac{0.6382}{0.0582} - 1 = 9.9688$ . Adapun nilai *Silhouette Index* pada

K=4 dapat dilihat pada tabel dibawah ini.

Tabel 4.83. Nilai SI dengan 4 *Cluster* pada *Fuzzy C-Means*

b1 (minimum distance U1 to U2, U1 to U3, & U1 to U4)	0.0582
a1 (average distance within U1)	0.6382
<b>Silhouette U1</b>	<b>9.9688</b>
b2 (minimum distance U2 to U1, U2 to U3, & U2 to U4)	0.1436
a2 (average distance within U2)	0.0582
<b>Silhouette U2</b>	<b>-1.4688</b>
b3 (minimum distance U3 to U1, U3 to U2, & U3 to U4)	0.0582
a3 (average distance within U3)	0.1436
<b>Silhouette U3</b>	<b>1.4688</b>
b4 (minimum distance U4 to U1, U4 to U2, & U4 to U3)	2.4745
a4 (average distance within U4)	0.3491
<b>Silhouette U4</b>	<b>-6.0885</b>
<b>Total Silhouette (U1 + U2 + U3 + U4)</b>	<b>3.8802</b>
<b>Nilai Validasi Silhouette dengan RStudio</b>	<b>0.6826</b>

7. Nilai SI pada *Fuzzy C-Means* dengan perhitungan manual K=4 tergolong rendah yaitu 3.8802, sedangkan nilai SI pada *Fuzzy C-Means* dengan RStudio K=4 tergolong tinggi yaitu 0.6826.

Tabel 4.84. Jarak Antar Objek dalam *Cluster* dan Luar *Cluster*

Dist. Within U1	Dist. U1 to U2	Dist. U1 to U3	Dist. U1 to U4	Dist. Within U2	Dist. U2 to U1	Dist. U2 to U3	Dist. U2 to U4	Dist. Within U3	Dist. U3 to U1	Dist. U3 to U2	Dist. U3 to U4	Dist. Within U4	Dist. U4 to U1	Dist. U4 to U2	Dist. U4 to U3
0	0	0	3	0	0	0	3	0	0	0	3	0	3	3	3
1	0	0	3	0	1	0	3	0	1	0	3	0	2	3	3
1	1	1	3	1	1	1	3	1	1	1	3	0	4	2	2
0	0	0	1	0	0	0	1	0	0	0	1	4	3	3	3
0	0	0	1	0	0	0	1	0	0	0	1	4	3	3	3
0	0	0	3	0	0	0	3	0	0	0	3	0	3	3	3
1	0	1	3	0	1	1	3	1	1	0	3	0	2	3	4
1	0	0	3	0	1	0	3	0	1	0	3	0	2	3	3
0	0	0	3	0	0	0	3	0	0	0	3	0	3	3	3
1	0	0	3	0	1	0	3	0	1	0	3	0	2	3	3
1	0	0	3	0	1	0	3	0	1	0	3	0	2	3	3
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
0	0	0	2	0	0	0	2	0	0	0	2	5	3	3	3
0	0	0	1	0	0	0	1	0	0	0	1	4	3	3	3
0	0	0	2	0	0	0	2	0	0	0	2	1	3	3	3
0	0	0	1	0	0	0	1	0	0	0	1	2	3	3	3
0	0	0	3	0	0	0	3	0	0	0	3	0	3	3	3
0	0	0	3	0	0	0	3	0	0	0	3	0	3	3	3
0	0	0	1	0	0	0	1	0	0	0	1	4	3	3	3

### 3. DBI pada *K-Means*

Nilai validasi *Davies Bouldin Index* (DBI) pada *K-Means* dapat diukur menggunakan aplikasi *spreadsheet*. Aplikasi *spreadsheet* yang digunakan yaitu LibreOffice Calc 6.1.5.2. Berikut adalah langkah-langkah perhitungan pengukuran nilai validasi *Davies Bouldin Index* (DBI) pada aplikasi *spreadsheet*.

#### a) Nilai Validasi DBI dengan 2 *Cluster*

1. Untuk melakukan pengukuran DBI pada *K-Means*, langkah pertama yaitu membagi dua nilai *clustering*. Sebagai contoh jika nilai *cluster* yang diperoleh 5.84 pada *cluster* 1, maka nilai menjadi 5 dan 8. Hal ini dilakukan untuk memperoleh nilai **a** dan **b** pada saat proses penentuan nilai validasi *Davies Bouldin Index*. Berikut adalah tabel pembagian nilai *cluster*.

Tabel 4.85. Tabel Pembagian Nilai *Cluster*

C1		C2	
8	0	7	1
5	3	2	8
4	6	3	4
...	...	...	...
5	6	3	6
5	8	3	3
3	9	0	0

2. Langkah selanjutnya yaitu melakukan perhitungan jarak *intracluster* atau jarak antar objek di dalam *cluster* setiap objek ke objek lainnya. Metrik yang digunakan yaitu Euclidean. Sebagai contoh, pada Tabel 4.55 data ke-1 pada C1 yaitu 8 dan 0, maka akan dihitung jarak antar 8 dan 0 (jarak dirinya sendiri/jarak data ke-1), jarak data ke-2, dan jarak data ke-3 hingga selesai pada C1. Lalu data ke-2 pada C1 yaitu 5 dan 3, maka akan dihitung jarak antar 8 dan 0 pada data ke-1, jarak 5 dan 3 (jarak dirinya sendiri/jarak data ke-2), dan jarak data ke-3 hingga selesai pada C1. Hal ini juga berlaku pada C2. Contohnya, jika nilai data ke-1 pada C1 adalah 8 dan 0, maka jarak antar data

yaitu  $d(x_1, c_1) = \sqrt{(8-8)^2 + (0-0)^2} = 0$ . Berikut adalah tabel hasil perhitungan jarak *intracluster*.

Tabel 4.86. Jarak *Intracluster K-Means*

Dari Data ke-i	Ke Data ke-i	Intracluster Within C1	Intracluster Within C2
1	1	0.0000	0.0000
1	2	4.2426	8.6023
1	3	7.2111	5.0000
...	...	...	...
10	8	3.6056	6.7082
10	9	2.2361	4.2426
10	10	0.0000	0.0000

- Setelah melakukan perhitungan jarak *intracluster*, selanjutnya menghitung jarak maksimal masing-masing *cluster*. Jarak maksimal yang diperoleh pada C1 dan C2, yaitu 11.4018 dan 8.6023. Jumlah nilai dari C1 dan C2 yaitu 20.0041.
- Langkah selanjutnya yaitu menghitung nilai total SSW. Adapun rumus SSW,

yaitu  $SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, c_i)$ . Contohnya, jika  $m$  (banyak data) = 100 maka

$$SSW = \frac{1}{100} (11.4018 + 8.6023) = 0.2000.$$

- Tahapan selanjutnya yaitu melakukan perhitungan SSB (*Sum of Square Between Cluster*) atau perhitungan jarak antar objek di luar *cluster* setiap objek ke objek lainnya. Metrik yang digunakan yaitu Euclidean. Sebagai contoh, pada Tabel 4.55 data ke-1 pada C1 yaitu 8 dan 0, maka akan dihitung jarak antar 7 dan 1 pada data ke-1, jarak data ke-2, dan jarak data ke-3 hingga selesai pada C2. Lalu data ke-2 pada C1 yaitu 5 dan 3, maka akan dihitung jarak antar 7 dan 1 pada data ke-1, 2 dan 8 pada data ke-2, dan jarak data ke-3 hingga selesai pada C2. Hal ini juga berlaku pada C2. Contohnya, jika nilai data ke-1 pada C1 adalah 8 dan 0 dan nilai data ke-1 pada C2 adalah 7 dan 1, maka jarak

antar data yaitu  $d(c_1, x_1; c_2, x_2) = \sqrt{(8-7)^2 + (0-1)^2} = 1.4142$  Berikut adalah tabel hasil perhitungan jarak SSB.

Tabel 4.87. Jarak *Intercluster K-Means*

Dari Data ke-i	Ke Data ke-i	Dist. C1 to C2	Dist. C2 to C1
1	1	1.4142	1.4142
1	2	10.0000	2.8284
1	3	6.4031	5.8310
...	...	...	...
10	8	3.0000	7.8102
10	9	6.0000	9.4340
10	10	9.4868	9.4868

6. Setelah melakukan perhitungan jarak *intercluster*, langkah selanjutnya yaitu menghitung jarak minimal masing-masing *cluster*. Jarak minimal masing-masing *cluster* disebut sebagai total SSB. Adapun rumus SSB, yaitu  $SSB_{i,j} = d(c_i, c_j)$ . Jarak minimal yang diperoleh pada C1 dan C2, yaitu 0.0000. Jumlah nilai minimum dari C1 dan C2 yaitu 0.0000.

7. Langkah selanjutnya yaitu menghitung  $R_{i,j}$  atau rasio. Adapun rumusnya, yaitu

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_i} \text{ . Nilai SSW yang diperoleh yaitu 0.200 dan nilai SSB}$$

yang diperoleh yaitu 0.0000. Maka  $R_{i,j} = \frac{0.2000}{0.000} = \infty$  .

8. Langkah terakhir yaitu menghitung nilai DBI. Adapun rumus perhitungannya

$$\text{yaitu, } DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \text{ . Contohnya jika } k=3, \text{ maka}$$

$$DBI = \frac{1}{3} \sum_1^3 (11.4018 + 8.6023) = 0.0666 \text{ . Diperoleh nilai DBI yaitu 0.0666.}$$



**b) Nilai Validasi DBI dengan 3 Cluster**

1. Untuk melakukan pengukuran DBI pada *K-Means*, langkah pertama yaitu membagi dua nilai *clustering*. Sebagai contoh jika nilai *cluster* yang diperoleh 5.84 pada *cluster* 1, maka nilai menjadi 5 dan 8. Hal ini dilakukan untuk memperoleh nilai **a** dan **b** pada saat proses penentuan nilai validasi *Davies Bouldin Index*. Berikut adalah tabel pembagian nilai *cluster*.

Tabel 4.88. Tabel Pembagian Nilai *Cluster*

C1		C2		C3	
5	2	3	8	6	0
1	9	7	4	1	7
1	8	4	8	2	7
...	...	...	...	...	...
2	8	8	2	2	6
2	1	6	5	2	7
2	7	8	0	2	3

2. Langkah selanjutnya yaitu melakukan perhitungan jarak *intracluster* atau jarak antar objek di dalam *cluster* setiap objek ke objek lainnya. Metrik yang digunakan yaitu Euclidean. Sebagai contoh, pada Tabel 4.58 data ke-1 pada C1 yaitu 5 dan 2, maka akan dihitung jarak antar 5 dan 2 (jarak dirinya sendiri/jarak data ke-1), jarak data ke-2, dan jarak data ke-3 hingga selesai pada C1. Lalu data ke-2 pada C1 yaitu 1 dan 9, maka akan dihitung jarak antar 5 dan 2 pada data ke-1, jarak 1 dan 9 (jarak dirinya sendiri/jarak data ke-2), dan jarak data ke-3 hingga selesai pada C1. Hal ini juga berlaku pada C2 dan C3. Contohnya, jika nilai data ke-1 pada C1 adalah 5 dan 2, maka jarak antar data yaitu  $d(x_1, c_1) = \sqrt{(5-5)^2 + (2-2)^2} = 0$ . Berikut adalah tabel hasil perhitungan jarak *intracluster*.

Tabel 4.89. Jarak *Intracluster K-Means*

Dari Data ke-i	Ke Data ke-i	Intracluster Within C1	Intracluster Within C2	Intracluster Within C3
1	1	0.0000	0.0000	0.0000

Dari Data ke-i	Ke Data ke-i	Intracluster Within C1	Intracluster Within C2	Intracluster Within C3
1	2	8.0623	5.6569	8.6023
1	3	7.2111	1.0000	8.0623
...	...	...	...	...
10	8	1.0000	2.0000	3.0000
10	9	6.0000	5.3852	4.0000
10	10	0.0000	0.0000	0.0000

- Setelah melakukan perhitungan jarak *intracluster*, selanjutnya menghitung jarak maksimal masing-masing *cluster*. Jarak maksimal yang diperoleh pada C1, C2, dan C3, yaitu 15.2643, 10.0000, dan 15.6525. Jumlah nilai dari C1, C2, dan C3 yaitu 40.9168.
- Langkah selanjutnya yaitu menghitung nilai total SSW. Adapun rumus SSW,

yaitu  $SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, c_i)$ . Contohnya, jika  $m$  (banyak data) = 100 maka

$$SSW = \frac{1}{100} (15.2643 + 10.0000 + 15.6525) = 0.4091$$

- Tahapan selanjutnya yaitu melakukan perhitungan SSB (*Sum of Square Between Cluster*) atau perhitungan jarak antar objek di luar *cluster* setiap objek ke objek lainnya. Metrik yang digunakan yaitu Euclidean. Sebagai contoh, pada Tabel 4.58 data ke-1 pada C1 yaitu 5 dan 2, maka akan dihitung jarak antar 3 dan 8 pada data ke-1, jarak data ke-2, dan jarak data ke-3 hingga selesai pada C2. Lalu data ke-2 pada C1 yaitu 1 dan 9, maka akan dihitung jarak antar 3 dan 8 pada data ke-1, 7 dan 4 pada data ke-2, dan jarak data ke-3 hingga selesai pada C2. Hal ini juga berlaku pada C2 dan C3. Contohnya, jika nilai data ke-1 pada C1 adalah 5 dan 2 dan nilai data ke-1 pada C2 adalah 3 dan 8, maka jarak antar data yaitu  $d(c_1, x_1; c_2, x_2) = \sqrt{(5-3)^2 + (2-8)^2} = 6.3246$ . Berikut adalah tabel hasil perhitungan jarak SSB.

Tabel 4.90. Jarak *Intercluster K-Means*

Dari Data ke-i	Ke Data ke-i	Dist. C1 to C2	Dist. C1 to C3	Dist. C2 to C1	Dist. C2 to C3	Dist. C3 to C1	Dist. C3 to C2
1	1	6.3246	2.2361	6.3246	8.5440	2.2361	8.5440
1	2	2.8284	6.4031	2.2361	2.2361	10.2956	4.1231
1	3	6.0828	5.8310	2.0000	1.4142	9.4340	8.2462
...	...	...	...	...	...	...	...
10	8	7.8102	1.0000	10.0000	8.4853	5.0000	6.0828
10	9	4.4721	0.0000	6.0828	9.2195	2.0000	4.4721
10	10	9.2195	4.0000	9.2195	6.7082	4.0000	6.7082

6. Setelah melakukan perhitungan jarak *intercluster*, langkah selanjutnya yaitu menghitung jarak minimal masing-masing *cluster*. Jarak minimal masing-masing *cluster* disebut sebagai total SSB. Adapun rumus SSB, yaitu  $SSB_{i,j} = d(c_i, c_j)$ . Jarak minimal yang diperoleh pada C1, C2, dan C3, yaitu 0.0000. Jumlah nilai minimum dari C1, C2, dan C3 yaitu 0.0000.
7. Langkah selanjutnya yaitu menghitung  $R_{i,j}$  atau rasio. Adapun rumusnya, yaitu

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_i} \text{ . Nilai SSW yang diperoleh yaitu 0.4091 dan nilai SSB}$$

yang diperoleh yaitu 0.0000. Maka  $R_{i,j} = \frac{0.4091}{0.000} = \infty$  .

8. Langkah terakhir yaitu menghitung nilai DBI. Adapun rumus perhitungannya

$$\text{yaitu, } DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \text{ . Contohnya jika } k=3, \text{ maka}$$

$$DBI = \frac{1}{3} \sum_{i=1}^3 (15.2643 + 10.0000 + 15.6525) = 0.1363 \text{ . Diperoleh nilai DBI}$$

yaitu 0.1363.

#### c) Nilai Validasi DBI dengan 4 Cluster

1. Untuk melakukan pengukuran DBI pada *K-Means*, langkah pertama yaitu membagi dua nilai *clustering*. Sebagai contoh jika nilai *cluster* yang diperoleh

5.84 pada *cluster* 1, maka nilai menjadi 5 dan 8. Hal ini dilakukan untuk memperoleh nilai **a** dan **b** pada saat proses penentuan nilai validasi *Davies Bouldin Index*. Berikut adalah tabel pembagian nilai *cluster*.

Tabel 4.91. Tabel Pembagian Nilai *Cluster*

C1		C2		C3		C4	
5	9	4	1	6	4	8	1
2	0	3	1	1	7	2	8
2	0	0	0	3	0	4	4
...	...	...	...	...	...	...	...
3	0	4	3	2	8	3	0
1	7	2	6	2	8	3	8
2	0	3	4	1	7	2	0

- Langkah selanjutnya yaitu melakukan perhitungan jarak *intracluster* atau jarak antar objek di dalam *cluster* setiap objek ke objek lainnya. Metrik yang digunakan yaitu Euclidean. Sebagai contoh, pada Tabel 4.61 data ke-1 pada C1 yaitu 5 dan 9, maka akan dihitung jarak antar 5 dan 9 (jarak dirinya sendiri/jarak data ke-1), jarak data ke-2, dan jarak data ke-3 hingga selesai pada C1. Lalu data ke-2 pada C1 yaitu 2 dan 0, maka akan dihitung jarak antar 5 dan 9 pada data ke-1, jarak 2 dan 0 (jarak dirinya sendiri/jarak data ke-2), dan jarak data ke-3 hingga selesai pada C1. Hal ini juga berlaku pada C2, C3, dan C4. Contohnya, jika nilai data ke-1 pada C1 adalah 5 dan 9, maka jarak antar data yaitu  $d(x_1, c_1) = \sqrt{(5-5)^2 + (9-9)^2} = 0$ . Berikut adalah tabel hasil perhitungan jarak *intracluster*.

Tabel 4.92. Jarak *Intracluster K-Means*

Dari Data ke-i	Ke Data ke-i	Intracluster Within C1	Intracluster Within C2	Intracluster Within C3	Intracluster Within C4
1	1	0.0000	0.0000	0.0000	0.0000
1	2	9.4868	1.0000	5.8310	9.2195
1	3	9.4868	4.1231	5.0000	5.0000
...	...	...	...	...	...

Dari Data ke-i	Ke Data ke-i	Intracluster Within C1	Intracluster Within C2	Intracluster Within C3	Intracluster Within C4
10	8	1.0000	1.4142	1.4142	1.0000
10	9	7.0711	2.2361	1.4142	8.0623
10	10	0.0000	0.0000	0.0000	0.0000

3. Setelah melakukan perhitungan jarak *intracluster*, selanjutnya menghitung jarak maksimal masing-masing *cluster*. Jarak maksimal yang diperoleh pada C1, C2, C3, dan C4, yaitu 9.8489, 9.0554, 8.2462, dan 9.2195. Jumlah nilai dari C1, C2, C3, dan C4 yaitu 36.3700.

4. Langkah selanjutnya yaitu menghitung nilai total SSW. Adapun rumus SSW,

yaitu  $SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, c_i)$ . Contohnya, jika  $m$  (banyak data) = 100 maka

$$SSW = \frac{1}{100} (9.8489 + 9.0554 + 8.2462 + 9.2195) = 0.3637$$

5. Tahapan selanjutnya yaitu melakukan perhitungan SSB (*Sum of Square Between Cluster*) atau perhitungan jarak antar objek di luar *cluster* setiap objek ke objek lainnya. Metrik yang digunakan yaitu Euclidean. Sebagai contoh, pada Tabel 4.61 data ke-1 pada C1 yaitu 5 dan 9, maka akan dihitung jarak antar 4 dan 1 pada data ke-1, jarak data ke-2, dan jarak data ke-3 hingga selesai pada C2. Lalu data ke-2 pada C1 yaitu 2 dan 0, maka akan dihitung jarak antar 4 dan 1 pada data ke-1, 3 dan 1 pada data ke-2, dan jarak data ke-3 hingga selesai pada C2. Hal ini juga berlaku pada C2, C3, dan C4. Contohnya, jika nilai data ke-1 pada C1 adalah 5 dan 9 dan nilai data ke-1 pada C2 adalah 4 dan 1, maka jarak antar data yaitu

$$d(c_1, x_1; c_2, x_2) = \sqrt{(5-4)^2 + (9-1)^2} = 8.0623$$

Hasil perhitungan jarak SSB dapat dilihat pada Tabel 4.63.

6. Setelah melakukan perhitungan jarak *intercluster*, langkah selanjutnya yaitu menghitung jarak minimal masing-masing *cluster*. Jarak minimal masing-masing *cluster* disebut sebagai total SSB. Adapun rumus SSB, yaitu

$SSB_{i,j}=d(c_i, c_j)$ . Jarak minimal yang diperoleh pada C1, C2, C3, dan C4, yaitu 0.0000. Jumlah nilai minimum dari C1, C2, C3, dan C4 yaitu 0.0000.

7. Langkah selanjutnya yaitu menghitung  $R_{i,j}$  atau rasio. Adapun rumusnya, yaitu

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_i} . \text{ Nilai SSW yang diperoleh yaitu 0.3637 dan nilai SSB}$$

yang diperoleh yaitu 0.0000. Maka  $R_{i,j} = \frac{0.3637}{0.000} = \infty$  .

Tabel 4.93. Jarak *Intercluster K-Means*

Dari Data ke-i	Ke Data ke-i	Dist. C1 to C2	Dist. C1 to C3	Dist. C1 to C4	Dist. C2 to C1	Dist. C2 to C3	Dist. C2 to C4	Dist. C3 to C1	Dist. C3 to C2	Dist. C3 to C4	Dist. C4 to C1	Dist. C4 to C2	Dist. C4 to C3
1	1	8.0623	5.0990	8.5440	8.0623	3.6056	4.0000	5.0990	3.6056	3.6056	8.5440	4.0000	3.6056
1	2	8.2462	4.4721	3.1623	2.2361	6.7082	7.2801	5.6569	4.2426	5.6569	6.0828	5.0000	9.2195
1	3	10.2956	9.2195	5.0990	2.2361	1.4142	3.0000	5.6569	7.2111	2.0000	6.0828	8.0623	5.0990
1	4	4.0000	4.1231	8.9443	6.7082	7.6158	3.0000	5.8310	7.0711	5.8310	9.2195	10.6301	9.8995
1	5	5.6569	6.4031	4.4721	3.0000	4.2426	6.7082	5.8310	5.0990	5.8310	7.0000	8.0623	7.6158
1	6	2.2361	9.4868	7.6158	5.0990	2.2361	2.2361	3.6056	3.6056	4.4721	7.0711	7.2111	6.0828
1	7	4.4721	9.4868	9.2195	3.1623	2.2361	1.4142	6.4031	5.8310	5.0000	7.0711	9.2195	6.0828
1	8	6.0828	3.1623	9.2195	1.4142	7.2801	1.4142	5.0000	2.2361	5.0000	5.0990	4.4721	9.2195
...	...	...	...	...	...	...	...	...	...	...	...	...	...
10	3	2.0000	1.0000	4.4721	4.1231	4.0000	1.0000	7.0711	7.0711	4.2426	0.0000	2.0000	1.0000
10	4	9.0554	8.0623	1.4142	3.6056	4.4721	3.6056	0.0000	2.0000	6.0000	7.0711	9.0554	8.0623
10	5	5.0990	4.1231	7.0711	3.6056	2.0000	3.6056	6.0000	2.0000	0.0000	1.4142	5.0990	4.1231
10	6	7.2801	0.0000	2.0000	2.0000	4.1231	2.2361	2.2361	3.0000	5.0990	6.0828	7.2801	0.0000
10	7	7.0711	0.0000	1.0000	4.4721	4.1231	4.0000	7.0000	0.0000	7.2801	1.0000	7.0711	0.0000
10	8	3.6056	8.0000	1.0000	4.0000	4.1231	4.0000	7.2801	5.0000	7.2801	1.0000	3.6056	8.0000
10	9	6.0000	8.0000	8.0623	3.6056	4.1231	4.0000	0.0000	1.4142	2.2361	7.0711	6.0000	8.0000
10	10	4.1231	7.0711	0.0000	4.1231	3.6056	4.1231	7.0711	3.6056	7.0711	0.0000	4.1231	7.0711

8. Langkah terakhir yaitu menghitung nilai DBI. Adapun rumus perhitungannya

yaitu,  $DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j})$  . Contohnya jika  $k=3$ , maka

$$DBI = \frac{1}{3} \sum_1^3 (9.8489 + 9.0554 + 8.2462 + 9.2195) = 0.1212 \text{ . Diperoleh nilai}$$

DBI yaitu 0.1212.

#### 4. DBI pada *Fuzzy C-Means*

Nilai validasi *Davies Bouldin Index* (DBI) pada *Fuzzy C-Means* dapat diukur menggunakan aplikasi *spreadsheet*. Aplikasi *spreadsheet* yang digunakan yaitu LibreOffice Calc 6.1.5.2. Berikut adalah langkah-langkah perhitungan pengukuran nilai validasi *Davies Bouldin Index* (DBI) pada aplikasi *spreadsheet*.

##### a) Nilai Validasi DBI dengan 2 Cluster

1. Untuk melakukan pengukuran DBI pada *Fuzzy C-Means*, langkah pertama yaitu membagi dua nilai *clustering*. Sebagai contoh jika nilai *cluster* yang diperoleh 5.84 pada *cluster* 1, maka nilai menjadi 5 dan 8. Hal ini dilakukan untuk memperoleh nilai **a** dan **b** pada saat proses penentuan nilai validasi *Davies Bouldin Index*. Berikut adalah tabel pembagian nilai *cluster*.

Tabel 4.94. Tabel Pembagian Nilai *Cluster*

U1		U2	
0	5	0	4
0	4	0	5
0	6	0	3
...	...	...	...
0	4	0	5
0	4	0	5
0	4	0	5

2. Langkah selanjutnya yaitu melakukan perhitungan jarak *intracluster* atau jarak antar objek di dalam *cluster* setiap objek ke objek lainnya. Metrik yang digunakan yaitu Euclidean. Sebagai contoh, pada Tabel 4.64 data ke-1 pada U1



yaitu 0 dan 0, maka akan dihitung jarak antar 0 dan 0 (jarak dirinya sendiri/jarak data ke-1), jarak data ke-2, dan jarak data ke-3 hingga selesai pada U1. Lalu data ke-2 pada U1 yaitu 0 dan 4, maka akan dihitung jarak antar 0 dan 0 pada data ke-1, jarak 0 dan 4 (jarak dirinya sendiri/jarak data ke-2), dan jarak data ke-3 hingga selesai pada U1. Hal ini juga berlaku pada U2. Contohnya, jika nilai data ke-1 pada U1 adalah 8 dan 0, maka jarak antar data yaitu  $d(x_1, c_1) = \sqrt{(0-0)^2 + (0-0)^2} = 0$ . Berikut adalah tabel hasil perhitungan jarak *intracluster*.

Tabel 4.95. Jarak *Intracluster Fuzzy C-Means*

Dari Data ke-i	Ke Data ke-i	Intracluster Within U1	Intracluster Within U2
1	1	0.0000	0.0000
1	2	1.0000	1.0000
1	3	1.0000	1.0000
...	...	...	...
10	8	0.0000	0.0000
10	9	0.0000	0.0000
10	10	0.0000	0.0000

- Setelah melakukan perhitungan jarak *intracluster*, selanjutnya menghitung jarak maksimal masing-masing *cluster*. Jarak maksimal yang diperoleh pada U1 dan U2, yaitu 2 dan 2. Jumlah nilai dari U1 dan U2 yaitu 4.
- Langkah selanjutnya yaitu menghitung nilai total SSW. Adapun rumus SSW,

yaitu  $SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, c_i)$ . Contohnya, jika m (banyak data) = 100 maka

$$SSW = \frac{1}{100} (2+2) = 0.04$$

- Tahapan selanjutnya yaitu melakukan perhitungan SSB (*Sum of Square Between Cluster*) atau perhitungan jarak antar objek di luar *cluster* setiap objek ke objek lainnya. Metrik yang digunakan yaitu Euclidean. Sebagai contoh, pada Tabel 4.64 data ke-1 pada U1 yaitu 0 dan 0, maka akan dihitung jarak

antar 0 dan 4 pada data ke-1, jarak data ke-2, dan jarak data ke-3 hingga selesai pada U2. Lalu data ke-2 pada U1 yaitu 0 dan 4, maka akan dihitung jarak antar 0 dan 4 pada data ke-1, 0 dan 5 pada data ke-2, dan jarak data ke-3 hingga selesai pada U2. Hal ini juga berlaku pada U2. Contohnya, jika nilai data ke-1 pada U1 adalah 0 dan 4 dan nilai data ke-1 pada U2 adalah 0 dan 4, maka jarak antar data yaitu  $d(c_1, x_1; c_2, x_2) = \sqrt{(0-4)^2 + (0-4)^2} = 1$ . Berikut adalah tabel hasil perhitungan jarak SSB.

Tabel 4.96. Jarak *Intercluster Fuzzy C-Means*

Dari Data ke-i	Ke Data ke-i	Dist. U1 to U2	Dist. U2 to U1
1	1	1.0000	1.0000
1	2	0.0000	0.0000
1	3	2.0000	2.0000
...	...	...	...
10	8	1.0000	1.0000
10	9	1.0000	1.0000
10	10	1.0000	1.0000

- Setelah melakukan perhitungan jarak *intercluster*, langkah selanjutnya yaitu menghitung jarak minimal masing-masing *cluster*. Jarak minimal masing-masing *cluster* disebut sebagai total SSB. Adapun rumus SSB, yaitu  $SSB_{i,j} = d(c_i, c_j)$ . Jarak minimal yang diperoleh pada U1 dan U2, yaitu 0. Jumlah nilai minimum dari U1 dan U2 yaitu 0.
- Langkah selanjutnya yaitu menghitung  $R_{i,j}$  atau rasio. Adapun rumusnya, yaitu

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_i} . \text{ Nilai SSW yang diperoleh yaitu 0.3637 dan nilai SSB}$$

yang diperoleh yaitu 0.0000. Maka  $R_{i,j} = \frac{1}{0} = \infty$ .

8. Langkah terakhir yaitu menghitung nilai DBI. Adapun rumus perhitungannya

yaitu,  $DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j})$  . Contohnya jika  $k=3$ , maka

$$DBI = \frac{1}{3} \sum_1^3 (2+2) = 0.0133 \text{ . Diperoleh nilai DBI yaitu } 0.0133.$$

**b) Nilai Validasi DBI dengan 3 Cluster**

1. Untuk melakukan pengukuran DBI pada *Fuzzy C-Means*, langkah pertama yaitu membagi dua nilai *clustering*. Sebagai contoh jika nilai *cluster* yang diperoleh 5.84 pada *cluster* 1, maka nilai menjadi 5 dan 8. Hal ini dilakukan untuk memperoleh nilai **a** dan **b** pada saat proses penentuan nilai validasi *Davies Bouldin Index*. Berikut adalah tabel pembagian nilai *cluster*.

Tabel 4.97. Tabel Pembagian *Cluster*

U1		U2		U3	
0	3	0	3	0	3
0	2	0	3	0	3
0	4	0	3	0	2
...	...	...	...	...	...
0	2	0	3	0	3
0	2	0	3	0	3
0	2	0	3	0	4

2. Langkah selanjutnya yaitu melakukan perhitungan jarak *intracluster* atau jarak antar objek di dalam *cluster* setiap objek ke objek lainnya. Metrik yang digunakan yaitu Euclidean. Sebagai contoh, pada Tabel 4.67 data ke-1 pada U1 yaitu 0 dan 3, maka akan dihitung jarak antar 0 dan 3 (jarak dirinya sendiri/jarak data ke-1), jarak data ke-2, dan jarak data ke-3 hingga selesai pada U1. Lalu data ke-2 pada U1 yaitu 0 dan 2, maka akan dihitung jarak antar 0 dan 3 pada data ke-1, jarak 0 dan 2 (jarak dirinya sendiri/jarak data ke-2), dan jarak data ke-3 hingga selesai pada U1. Hal ini juga berlaku pada U2 dan U3. Contohnya, jika nilai data ke-1 pada U1 adalah 0 dan 3, maka jarak antar

data yaitu  $d(x_1, c_1) = \sqrt{(0-0)^2 + (3-3)^2} = 1$ . Berikut adalah tabel hasil perhitungan jarak *intracluster*.

Tabel 4.98. Jarak *Intracluster Fuzzy C-Means*

Dari Data ke-i	Ke Data ke-i	Intracluster Within U1	Intracluster Within U2	Intracluster Within U3
1	1	1.0000	0.0000	0.0000
1	2	0.0000	0.0000	0.0000
1	3	2.0000	0.0000	1.0000
...	...	...	...	...
10	8	2.0000	3.0000	3.0000
10	9	2.0000	3.0000	3.0000
10	10	2.0000	3.0000	4.0000

- Setelah melakukan perhitungan jarak *intracluster*, selanjutnya menghitung jarak maksimal masing-masing *cluster*. Jarak maksimal yang diperoleh pada U1, U2, dan U3, yaitu 3, 2, dan 5. Jumlah nilai dari U1, U2, dan U3 yaitu 11.
- Langkah selanjutnya yaitu menghitung nilai total SSW. Adapun rumus SSW,

yaitu  $SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, c_i)$ . Contohnya, jika  $m$  (banyak data) = 100 maka

$$SSW = \frac{1}{100} (3+3+5) = 0.11.$$

- Tahapan selanjutnya yaitu melakukan perhitungan SSB (*Sum of Square Between Cluster*) atau perhitungan jarak antar objek di luar *cluster* setiap objek ke objek lainnya. Metrik yang digunakan yaitu Euclidean. Sebagai contoh, pada Tabel 4.67 data ke-1 pada U1 yaitu 0 dan 3, maka akan dihitung jarak antar 0 dan 3 pada data ke-1, jarak data ke-2, dan jarak data ke-3 hingga selesai pada U2. Lalu data ke-2 pada U1 yaitu 0 dan 2, maka akan dihitung jarak antar 0 dan 3 pada data ke-1, 0 dan 3 pada data ke-2, dan jarak data ke-3 hingga selesai pada U2. Hal ini juga berlaku pada U2 dan U3. Contohnya, jika nilai data ke-1 pada U1 adalah 0 dan 3 dan nilai data ke-1 pada U2 adalah 0 dan 3,

maka jarak antar data yaitu  $d(c_1, x_1; c_2, x_2) = \sqrt{(0-0)^2 + (3-3)^2} = 0$ . Berikut adalah tabel hasil perhitungan jarak SSB.

Tabel 4.99. Jarak *Intercluster Fuzzy C-Means*

Dari Data ke-i	Ke Data ke-i	Dist. U1 to U2	Dist. U1 to U3	Dist. U2 to U1	Dist. U2 to U3	Dist. U3 to U1	Dist. U3 to U2
1	1	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000
1	2	0.0000	1.0000	1.0000	0.0000	1.0000	0.0000
1	3	0.0000	0.0000	1.0000	1.0000	1.0000	0.0000
...	...	...	...	...	...	...	...
10	8	3.0000	3.0000	2.0000	3.0000	2.0000	3.0000
10	9	3.0000	3.0000	2.0000	3.0000	2.0000	3.0000
10	10	3.0000	4.0000	2.0000	4.0000	2.0000	3.0000

- Setelah melakukan perhitungan jarak *intercluster*, langkah selanjutnya yaitu menghitung jarak minimal masing-masing *cluster*. Jarak minimal masing-masing *cluster* disebut sebagai total SSB. Adapun rumus SSB, yaitu  $SSB_{i,j} = d(c_i, c_j)$ . Jarak minimal yang diperoleh pada U1, U2, dan U3, yaitu 0. Jumlah nilai minimum dari U1, U2, dan U3 yaitu 0.
- Langkah selanjutnya yaitu menghitung  $R_{i,j}$  atau rasio. Adapun rumusnya, yaitu

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_i} \text{ . Nilai SSW yang diperoleh yaitu 0.4091 dan nilai SSB}$$

yang diperoleh yaitu 0.0000. Maka  $R_{i,j} = \frac{0.11}{0} = \infty$  .

- Langkah terakhir yaitu menghitung nilai DBI. Adapun rumus perhitungannya

yaitu,  $DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j})$  . Contohnya jika  $k=3$ , maka

$$DBI = \frac{1}{3} \sum_{i=1}^3 (3+3+5) = 0.0367 \text{ . Diperoleh nilai DBI yaitu 0.0367.}$$

**c) Nilai Validasi DBI dengan 4 Cluster**

1. Untuk melakukan pengukuran DBI pada *Fuzzy C-Means*, langkah pertama yaitu membagi dua nilai *clustering*. Sebagai contoh jika nilai *cluster* yang diperoleh 5.84 pada *cluster* 1, maka nilai menjadi 5 dan 8. Hal ini dilakukan untuk memperoleh nilai **a** dan **b** pada saat proses penentuan nilai validasi *Davies Bouldin Index*. Berikut adalah tabel pembagian nilai *cluster*.

Tabel 4.100. Tabel Pembagian *Cluster*

U1		U2		U3		U4	
0	3	0	3	0	3	0	0
0	2	0	3	0	3	0	0
0	4	0	2	0	2	0	0
...	...	...	...	...	...	...	...
0	2	0	3	0	3	0	0
0	3	0	3	0	3	0	0
0	2	0	3	0	3	0	0

2. Langkah selanjutnya yaitu melakukan perhitungan jarak *intracluster* atau jarak antar objek di dalam *cluster* setiap objek ke objek lainnya. Metrik yang digunakan yaitu Euclidean. Sebagai contoh, pada Tabel 4.70 data ke-1 pada U1 yaitu 0 dan 3, maka akan dihitung jarak antar 0 dan 3 (jarak dirinya sendiri/jarak data ke-1), jarak data ke-2, dan jarak data ke-3 hingga selesai pada U1. Lalu data ke-2 pada U1 yaitu 0 dan 2, maka akan dihitung jarak antar 0 dan 3 pada data ke-1, jarak 0 dan 2 (jarak dirinya sendiri/jarak data ke-2), dan jarak data ke-3 hingga selesai pada U1. Hal ini juga berlaku pada U2, U3, dan U4. Contohnya, jika nilai data ke-1 pada U1 adalah 0 dan 3, maka jarak antar data yaitu  $d(x_1, c_1) = \sqrt{(0-0)^2 + (3-3)^2} = 0$ . Berikut adalah tabel hasil perhitungan jarak *intracluster*.

Tabel 4.101. Jarak *Intracluster Fuzzy C-Means*

Dari Data ke-i	Ke Data ke-i	Intracluster Within U1	Intracluster Within U2	Intracluster Within U3	Intracluster Within U4
1	1	0.0000	0.0000	0.0000	0.0000

Dari Data ke-i	Ke Data ke-i	Intracluster Within U1	Intracluster Within U2	Intracluster Within U3	Intracluster Within U4
1	2	1.0000	0.0000	0.0000	0.0000
1	3	1.0000	1.0000	1.0000	0.0000
...	...	...	...	...	...
10	8	0.0000	0.0000	0.0000	0.0000
10	9	1.0000	0.0000	0.0000	0.0000
10	10	0.0000	0.0000	0.0000	0.0000

- Setelah melakukan perhitungan jarak *intracluster*, selanjutnya menghitung jarak maksimal masing-masing *cluster*. Jarak maksimal yang diperoleh pada U1, U2, U3, dan U4, yaitu 2, 1, 2, dan 4. Jumlah nilai dari U1, U2, U3, dan U4 yaitu 9.
- Langkah selanjutnya yaitu menghitung nilai total SSW. Adapun rumus SSW,

yaitu  $SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, c_i)$ . Contohnya, jika  $m$  (banyak data) = 100 maka

$$SSW = \frac{1}{100} (2 + 1 + 2 + 4) = 0.09$$

- Tahapan selanjutnya yaitu melakukan perhitungan SSB (*Sum of Square Between Cluster*) atau perhitungan jarak antar objek di luar *cluster* setiap objek ke objek lainnya. Metrik yang digunakan yaitu Euclidean. Sebagai contoh, pada Tabel 4.70 data ke-1 pada U1 yaitu 0 dan 3, maka akan dihitung jarak antar 0 dan 3 pada data ke-1, jarak data ke-2, dan jarak data ke-3 hingga selesai pada U2. Lalu data ke-2 pada C1 yaitu 0 dan 2, maka akan dihitung jarak antar 0 dan 3 pada data ke-1, 0 dan 3 pada data ke-2, dan jarak data ke-3 hingga selesai pada U2. Hal ini juga berlaku pada C2, C3, dan C4. Contohnya, jika nilai data ke-1 pada U1 adalah 0 dan 3 dan nilai data ke-1 pada U2 adalah 0 dan 3, maka jarak antar data yaitu  $d(c_1, x_1; c_2, x_2) = \sqrt{(0-0)^2 + (3-3)^2} = 0$ . Hasil perhitungan jarak SSB dapat dilihat pada Tabel 4.72.
- Setelah melakukan perhitungan jarak *intercluster*, langkah selanjutnya yaitu menghitung jarak minimal masing-masing *cluster*. Jarak minimal masing-

masing *cluster* disebut sebagai total SSB. Adapun rumus SSB, yaitu  $SSB_{i,j} = d(c_i, c_j)$ . Jarak minimal yang diperoleh pada U1, U2, U3, dan U4, yaitu 0.0000. Jumlah nilai minimum dari U1, U2, U3, dan U4 yaitu 0.0000.

7. Langkah selanjutnya yaitu menghitung  $R_{i,j}$  atau rasio. Adapun rumusnya, yaitu

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_i} . \text{ Nilai SSW yang diperoleh yaitu 0.09 dan nilai SSB}$$

yang diperoleh yaitu 0.0000. Maka  $R_{i,j} = \frac{0.09}{0} = \infty$  .

8. Langkah terakhir yaitu menghitung nilai DBI. Adapun rumus perhitungannya

yaitu,  $DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j})$  . Contohnya jika  $k=3$ , maka

$$DBI = \frac{1}{3} \sum_1^3 (2+1+2+4) = 0.03 . \text{ Diperoleh nilai DBI yaitu 0.03.}$$



Tabel 4.102. Jarak *Intercluster Fuzzy C-Means*

Dari Data ke-i	Ke Data ke-i	Dist. U1 to U2	Dist. U1 to U3	Dist. U1 to U4	Dist. U2 to U1	Dist. U2 to U3	Dist. U2 to U4	Dist. U3 to U1	Dist. U3 to U2	Dist. U3 to U4	Dist. U4 to U1	Dist. U4 to U2	Dist. U4 to U3
1	1	0.0000	0.0000	3.0000	0.0000	0.0000	3.0000	0.0000	0.0000	3.0000	3.0000	3.0000	3.0000
1	2	0.0000	0.0000	3.0000	1.0000	0.0000	3.0000	1.0000	0.0000	3.0000	2.0000	3.0000	3.0000
1	3	1.0000	1.0000	3.0000	1.0000	1.0000	3.0000	1.0000	1.0000	3.0000	4.0000	2.0000	2.0000
1	4	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000	3.0000	3.0000	3.0000
1	5	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000	3.0000	3.0000	3.0000
1	6	0.0000	0.0000	3.0000	0.0000	0.0000	3.0000	0.0000	0.0000	3.0000	3.0000	3.0000	3.0000
1	7	0.0000	1.0000	3.0000	1.0000	1.0000	3.0000	1.0000	0.0000	3.0000	2.0000	3.0000	4.0000
1	8	0.0000	0.0000	3.0000	1.0000	0.0000	3.0000	1.0000	0.0000	3.0000	2.0000	3.0000	3.0000
...	...	...	...	...	...	...	...	...	...	...	...	...	...
10	3	0.0000	0.0000	2.0000	1.0000	1.0000	3.0000	1.0000	1.0000	3.0000	4.0000	2.0000	2.0000
10	4	1.0000	1.0000	2.0000	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000	3.0000	3.0000	3.0000
10	5	1.0000	1.0000	2.0000	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000	3.0000	3.0000	3.0000
10	6	1.0000	1.0000	2.0000	0.0000	0.0000	3.0000	0.0000	0.0000	3.0000	3.0000	3.0000	3.0000
10	7	1.0000	2.0000	2.0000	1.0000	1.0000	3.0000	1.0000	0.0000	3.0000	2.0000	3.0000	4.0000
10	8	1.0000	1.0000	2.0000	1.0000	0.0000	3.0000	1.0000	0.0000	3.0000	2.0000	3.0000	3.0000
10	9	1.0000	1.0000	2.0000	0.0000	0.0000	3.0000	0.0000	0.0000	3.0000	3.0000	3.0000	3.0000
10	10	1.0000	1.0000	2.0000	1.0000	0.0000	3.0000	1.0000	0.0000	3.0000	2.0000	3.0000	3.0000
10	3	0.0000	0.0000	2.0000	1.0000	1.0000	3.0000	1.0000	1.0000	3.0000	4.0000	2.0000	2.0000

## b. Pengukuran Nilai Validasi dengan RStudio

Nilai validasi *Silhouette Index* (SI) dan *Davies Bouldin Index* (DBI) pada *K-Means* dan *Fuzzy C-Means* dapat diukur menggunakan aplikasi RStudio. Versi aplikasi RStudio yang digunakan yaitu versi 1.1.456. Berikut adalah hasil perhitungan pengukuran nilai validasi *Silhouette Index* (SI) dan *Davies Bouldin Index* (DBI) pada aplikasi RStudio.

### 1. Pengukuran Nilai Validasi dengan 510 Data

Tabel 4.103. Pengukuran Nilai Validasi dengan 2 *Cluster*

Algoritma	Validasi DBI		Validasi SI	
	K Random	Nilai Validasi	K Random	Nilai Validasi
<i>K-Means</i>	2	0.7651	2	0.7713
<i>Fuzzy C-Means</i>	2	0.6332	2	0.7838

Tabel 4.104. Pengukuran Nilai Validasi dengan 3 *Cluster*

Algoritma	Validasi DBI		Validasi SI	
	K Random	Nilai Validasi	K Random	Nilai Validasi
<i>K-Means</i>	3	0.4893	3	0.7517
<i>Fuzzy C-Means</i>	3	0.3371	3	0.7547

Tabel 4.105. Pengukuran Nilai Validasi dengan 4 *Cluster*

Algoritma	Validasi DBI		Validasi SI	
	K Random	Nilai Validasi	K Random	Nilai Validasi
<i>K-Means</i>	4	0.3348	4	0.7332
<i>Fuzzy C-Means</i>	4	0.6175	4	0.6219

Pada pengukuran validasi menggunakan 2 *cluster*, *K-Means* memperoleh hasil DBI = 0.7651 dan SI = 0.7771. Sedangkan pada *Fuzzy C-Means* memperoleh hasil DBI = 0.6332 dan SI = 0.7838. Pada pengukuran validasi menggunakan 3 *cluster*, *K-Means* memperoleh hasil DBI = 0.4893 dan SI = 0.7517. Sedangkan pada *Fuzzy C-Means* memperoleh hasil DBI = 0.3371 dan SI = 0.7547. Pada pengukuran validasi menggunakan 4 *cluster*, *K-Means* memperoleh hasil DBI = 0.3348 dan SI = 0.7332. Sedangkan pada *Fuzzy C-Means* memperoleh hasil DBI = 0.6175 dan SI = 0.6219.

## 2. Pengukuran Nilai Validasi dengan 520 Data

Tabel 4.106. Pengukuran Nilai Validasi dengan 2 *Cluster*

Algoritma	Validasi DBI		Validasi SI	
	K Random	Nilai Validasi	K Random	Nilai Validasi
<i>K-Means</i>	2	0.7639	2	0.7779
<i>Fuzzy C-Means</i>	2	0.5428	2	0.7808

Tabel 4.107. Pengukuran Nilai Validasi dengan 3 *Cluster*

Algoritma	Validasi DBI		Validasi SI	
	K Random	Nilai Validasi	K Random	Nilai Validasi
<i>K-Means</i>	3	0.4878	3	0.7522
<i>Fuzzy C-Means</i>	3	0.4878	3	0.7522

Tabel 4.108. Pengukuran Nilai Validasi dengan 4 *Cluster*

Algoritma	Validasi DBI		Validasi SI	
	K Random	Nilai Validasi	K Random	Nilai Validasi
<i>K-Means</i>	4	0.3346	4	0.7343
<i>Fuzzy C-Means</i>	4	0.3346	4	0.7343

Pada pengukuran validasi menggunakan 2 *cluster*, *K-Means* memperoleh hasil DBI = 0.7639 dan SI = 0.7779. Sedangkan pada *Fuzzy C-Means* memperoleh hasil DBI = 0.5428 dan SI = 0.7808. Pada pengukuran validasi menggunakan 3 *cluster*, *K-Means* memperoleh hasil DBI = 0.4878 dan SI = 0.7522. Sedangkan pada *Fuzzy C-Means* memperoleh hasil DBI = 0.4878 dan SI = 0.7522. Pada pengukuran validasi menggunakan 4 *cluster*, *K-Means* memperoleh hasil DBI = 0.3346 dan SI = 0.7343. Sedangkan pada *Fuzzy C-Means* memperoleh hasil DBI = 0.3346 dan SI = 0.7343.

## 3. Pengukuran Nilai Validasi dengan 530 Data

Tabel 4.109. Pengukuran Nilai Validasi dengan 2 *Cluster*

Algoritma	Validasi DBI		Validasi SI	
	K Random	Nilai Validasi	K Random	Nilai Validasi
<i>K-Means</i>	2	0.7713	2	0.7738
<i>Fuzzy C-Means</i>	2	0.6324	2	0.7814

Tabel 4.110. Pengukuran Nilai Validasi dengan 3 *Cluster*

Algoritma	Validasi DBI		Validasi SI	
	K Random	Nilai Validasi	K Random	Nilai Validasi
<i>K-Means</i>	3	0.4966	3	0.7498
<i>Fuzzy C-Means</i>	3	0.3387	3	0.7537

Tabel 4.111. Pengukuran Nilai Validasi dengan 4 *Cluster*

Algoritma	Validasi DBI		Validasi SI	
	K Random	Nilai Validasi	K Random	Nilai Validasi
<i>K-Means</i>	4	0.3348	4	0.7288
<i>Fuzzy C-Means</i>	4	0.5981	4	0.6284

Pada pengukuran validasi menggunakan 2 *cluster*, *K-Means* memperoleh hasil DBI = 0.7713 dan SI = 0.7738. Sedangkan pada *Fuzzy C-Means* memperoleh hasil DBI = 0.6324 dan SI = 0.7814. Pada pengukuran validasi menggunakan 3 *cluster*, *K-Means* memperoleh hasil DBI = 0.4966 dan SI = 0.7498. Sedangkan pada *Fuzzy C-Means* memperoleh hasil DBI = 0.3387 dan SI = 0.7537. Pada pengukuran validasi menggunakan 4 *cluster*, *K-Means* memperoleh hasil DBI = 0.3348 dan SI = 0.7288. Sedangkan pada *Fuzzy C-Means* memperoleh hasil DBI = 0.5981 dan SI = 0.6284.

#### 4. Pengukuran Nilai Validasi dengan 540 Data

Tabel 4.112. Pengukuran Nilai Validasi dengan 2 *Cluster*

Algoritma	Validasi DBI		Validasi SI	
	K Random	Nilai Validasi	K Random	Nilai Validasi
<i>K-Means</i>	2	0.2232	2	0.7542
<i>Fuzzy C-Means</i>	2	0.4004	2	0.7860

Tabel 4.113. Pengukuran Nilai Validasi dengan 3 *Cluster*

Algoritma	Validasi DBI		Validasi SI	
	K Random	Nilai Validasi	K Random	Nilai Validasi
<i>K-Means</i>	3	0.4516	3	0.7528
<i>Fuzzy C-Means</i>	3	0.5299	3	0.7031

Tabel 4.114. Pengukuran Nilai Validasi dengan 4 Cluster

Algoritma	Validasi DBI		Validasi SI	
	K Random	Nilai Validasi	K Random	Nilai Validasi
<i>K-Means</i>	4	0.6396	4	0.7254
<i>Fuzzy C-Means</i>	4	0.7032	4	0.6156

Pada pengukuran validasi menggunakan 2 cluster, *K-Means* memperoleh hasil DBI = 0.2232 dan SI = 0.7542. Sedangkan pada *Fuzzy C-Means* memperoleh hasil DBI = 0.4004 dan SI = 0.7860. Pada pengukuran validasi menggunakan 3 cluster, *K-Means* memperoleh hasil DBI = 0.4516 dan SI = 0.7528. Sedangkan pada *Fuzzy C-Means* memperoleh hasil DBI = 0.5299 dan SI = 0.7031. Pada pengukuran validasi menggunakan 4 cluster, *K-Means* memperoleh hasil DBI = 0.3696 dan SI = 0.7254. Sedangkan pada *Fuzzy C-Means* memperoleh hasil DBI = 0.7032 dan SI = 0.6156.

### 5. Pengukuran Nilai Validasi dengan 550 Data

Tabel 4.115. Pengukuran Nilai Validasi dengan 2 Cluster

Algoritma	Validasi DBI		Validasi SI	
	K Random	Nilai Validasi	K Random	Nilai Validasi
<i>K-Means</i>	2	0.5192	2	0.7713
<i>Fuzzy C-Means</i>	2	0.5192	2	0.7713

Tabel 4.116. Pengukuran Nilai Validasi dengan 3 Cluster

Algoritma	Validasi DBI		Validasi SI	
	K Random	Nilai Validasi	K Random	Nilai Validasi
<i>K-Means</i>	3	0.4571	3	0.7486
<i>Fuzzy C-Means</i>	3	0.5853	3	0.6987

Tabel 4.117. Pengukuran Nilai Validasi dengan 4 Cluster

Algoritma	Validasi DBI		Validasi SI	
	K Random	Nilai Validasi	K Random	Nilai Validasi
<i>K-Means</i>	4	0.7618	4	0.5694
<i>Fuzzy C-Means</i>	4	0.5005	4	0.6826

Pada pengukuran validasi menggunakan 2 *cluster*, *K-Means* memperoleh hasil DBI = 0.5192 dan SI = 0.7713. Sedangkan pada *Fuzzy C-Means* memperoleh hasil DBI = 0.5192 dan SI = 0.7713. Pada pengukuran validasi menggunakan 3 *cluster*, *K-Means* memperoleh hasil DBI = 0.4571 dan SI = 0.7486. Sedangkan pada *Fuzzy C-Means* memperoleh hasil DBI = 0.5853 dan SI = 0.6987. Pada pengukuran validasi menggunakan 4 *cluster*, *K-Means* memperoleh hasil DBI = 0.7618 dan SI = 0.5694. Sedangkan pada *Fuzzy C-Means* memperoleh hasil DBI = 0.5005 dan SI = 0.6826.

#### 4.1.5 Perbandingan Hasil Cluster

Setelah mengukur nilai validasi antara SI dan DBI berdasarkan jumlah data, selanjutnya yaitu membandingkan hasil *cluster* yang lebih baik antara *K-Means* dan *Fuzzy C-Means* untuk mengetahui algoritma yang lebih baik antara *K-Means* dan *Fuzzy C-Means*. Untuk membandingkan hasil *cluster* yang lebih baik antara *K-Means* dan *Fuzzy C-Means*, maka akan dibandingkan nilai validasi pada setiap *cluster*. Pada bagian ini akan dilakukan pengukuran dengan 550 data sampel yang diperoleh menggunakan rumus *Cochran*. Berikut adalah tabel perbandingan hasil *cluster* antara *K-Means* dan *Fuzzy C-Means*

Tabel 4.118. Tabel Perbandingan Hasil Cluster

No	Jumlah Cluster	Algoritma	Validasi Davies Bouldin Index	Validasi Silhouette Index	Algoritma Yang Lebih Baik
1	2	<i>K-Means</i>	0.5192	0.7713	<i>K-Means dan Fuzzy C-Means</i>
	2	<i>Fuzzy C-Means</i>	0.5192	0.7713	
2	3	<b><i>K-Means</i></b>	<b>0.4571</b>	<b>0.7486</b>	<i>K-Means</i>
	3	<i>Fuzzy C-Means</i>	0.5853	0.6987	
3	4	<i>K-Means</i>	0.7168	0.5694	<i>Fuzzy C-Means</i>
	4	<b><i>Fuzzy C-Means</i></b>	<b>0.5005</b>	<b>0.6826</b>	

Berdasarkan tabel diatas, pada percobaan dengan 2 *cluster*, nilai validasi DBI dan SI pada *K-Means* dan *Fuzzy C-Means* adalah sama. Maka algoritma *cluster* terbaik adalah *K-Means* dan *Fuzzy C-Means*. Pada percobaan dengan 3 *cluster*, nilai validasi DBI dan SI pada *K-Means* adalah 0.4571 dan 0.7486. Sedangkan pada *Fuzzy C-Means* nilai DBI dan SI adalah 0.5853 dan 0.6987. Maka algoritma *cluster* yang lebih baik adalah *K-Means* karena nilai validasi DBI lebih mendekati 0 dan nilai validasi SI lebih mendekati 1. Disamping itu nilai validasi SI termasuk kategori *strong structure*. Pada percobaan 4 *cluster*, nilai validasi DBI dan SI pada *K-Means* adalah 0.7168 dan 0.5694. Sedangkan pada *Fuzzy C-Means* nilai

validasi DBI dan SI adalah 0.5005 dan 0.6826. Maka algoritma yang lebih baik adalah *Fuzzy C-Means* karena nilai validasi DBI lebih mendekati 0 dan nilai validasi SI lebih mendekati 1. Disamping itu nilai validasi SI termasuk kategori *strong structure*.

Selanjutnya yaitu membandingkan nilai validasi DBI dan SI pada 3 *cluster* dan 4 *cluster* pada masing-masing algoritma. Pada percobaan dengan 3 *cluster* dengan algoritma *cluster* yang lebih baik yaitu *K-Means*, nilai DBI dan SI yang diperoleh yaitu 0.4571 dan 0.7486. Sedangkan pada percobaan dengan 4 *cluster* dengan algoritma *cluster* yang lebih baik yaitu *Fuzzy C-Means*, nilai DBI dan SI yang diperoleh yaitu 0.5005 dan 0.6826. Pada percobaan dengan 3 *cluster*, nilai validasi DBI lebih mendekati 0 dan lebih baik dibandingkan dengan penelitian terkait sedangkan nilai validasi SI termasuk pada kategori *strong structure* menurut Kaufman dan Rousseuw. Sedangkan pada percobaan dengan 4 *cluster* nilai validasi DBI lebih mendekati 0 dan lebih baik dibandingkan dengan penelitian terkait sedangkan nilai validasi SI termasuk pada kategori *medium structure* menurut Kaufman dan Rousseuw.

Dengan demikian nilai validasi pada percobaan dengan 3 *cluster* lebih baik baik dibandingkan dengan 4 *cluster*. Sehingga dapat disimpulkan bahwa percobaan menggunakan 3 *cluster* dengan algoritma *K-Means*, DBI = 0.4571, SI = 0.7486, dan 550 data sampel merupakan hasil *cluster* yang lebih baik.

## 4.2 Analisis

Pada tahap ini, penulis akan membandingkan penelitian yang telah dilakukan dengan penelitian terkait. Penelitian yang dibahas oleh Risma Rustiyan dan Mustakim dengan judul, “Penerapan Algoritma *Fuzzy C-Means* untuk Analisis Permasalahan Simpanan Wajib Anggota Koperasi”, dengan menerapkan *Fuzzy C-Means* diperoleh hasil pengujian nilai validitas PC sebesar 0.323732 dengan kualitas klaster yang belum optimal. Sedangkan pada penelitian yang telah dilakukan oleh peneliti dengan judul, “Komparasi Validasi *K-Means* dan *Fuzzy C-Means* dalam Menentukan *Cluster* Terbaik Menggunakan *Silhouette Index* dan *Davies Bouldin Index*”, diperoleh hasil pengujian berupa algoritma *cluster* yang lebih baik yaitu *K-Means* dan mendapatkan kualitas *cluster* yang optimal dengan nilai validasi DBI = 0.4571 dan SI = 0.7486. Sehingga dapat disimpulkan bahwa nilai DBI dan SI pada penelitian terkait lebih rendah dibandingkan dengan penelitian yang telah dilakukan oleh peneliti.

Penelitian yang dilakukan oleh Ahmad Saiful dan Joko Lianto Buliali dengan judul, “Implementasi *Particle Swarm Optimization* pada *K-Means* untuk *Clustering Data Automatic Dependent Surveillance Broadcast*”, dengan menerapkan kombinasi algoritma *K-Means* dan

*Particle Swarm Optimization* diperoleh hasil pengujian nilai validitas DBI = 0.779 dan kualitas klaster yang optimal. Sedangkan pada penelitian yang telah dilakukan oleh peneliti dengan judul, “Komparasi Validasi *K-Means* dan *Fuzzy C-Means* dalam Menentukan *Cluster* Terbaik Menggunakan *Silhouette Index* dan *Davies Bouldin Index*”, diperoleh hasil pengujian berupa algoritma *cluster* yang lebih baik yaitu *K-Means* dan mendapatkan kualitas *cluster* yang optimal dengan nilai validasi DBI = 0.4571 dan SI = 0.7486. Sehingga dapat disimpulkan bahwa nilai DBI dan SI pada penelitian terkait lebih rendah dibandingkan dengan penelitian yang telah dilakukan oleh peneliti.

Penelitian yang dilakukan oleh Dila Fitriani Azuri, dkk dengan judul, “Pengelompokan Kabupaten/Kota di Pulau Jawa Berdasarkan Pembangunan Manusia Berbasis Gender Menggunakan *Bisecting K-Means*”, dengan menerapkan *Bisecting K-Means* diperoleh hasil pengujian nilai validitas SI pada klaster laki-laki yaitu 0.3 dan nilai validitas SI pada klaster perempuan yaitu 0.2 dengan klaster yang terbentuk masih tergolong lemah. Sedangkan pada penelitian yang telah dilakukan oleh peneliti dengan judul, “Komparasi Validasi *K-Means* dan *Fuzzy C-Means* dalam Menentukan *Cluster* Terbaik Menggunakan *Silhouette Index* dan *Davies Bouldin Index*”, diperoleh hasil pengujian berupa algoritma *cluster* yang lebih baik yaitu *K-Means* dan mendapatkan kualitas *cluster* yang optimal dengan nilai validasi DBI = 0.4571 dan SI = 0.7486. Sehingga dapat disimpulkan bahwa nilai DBI dan SI pada penelitian terkait lebih rendah dibandingkan dengan penelitian yang telah dilakukan oleh peneliti.

Penelitian yang dilakukan oleh Christina Deni Rumiarti dan Indra Budi dengan judul, “Segmentasi Pelanggan pada *Customer Relationship Management* di Perusahaan Ritel: Studi Kasus PT Gramedia Asri Media.”, dengan menerapkan *K-Means* diperoleh hasil pengujian nilai *average silhouette* pada dua *cluster* adalah 0.42. Nilai *Calinski Harabasz index* pada dua *cluster* adalah 1.31. Sedangkan pada penelitian yang telah dilakukan oleh peneliti dengan judul, “Komparasi Validasi *K-Means* dan *Fuzzy C-Means* dalam Menentukan *Cluster* Terbaik Menggunakan *Silhouette Index* dan *Davies Bouldin Index*”, diperoleh hasil pengujian berupa algoritma *cluster* yang lebih baik yaitu *K-Means* dan mendapatkan kualitas *cluster* yang optimal dengan nilai validasi DBI = 0.4571 dan SI = 0.7486. Sehingga dapat disimpulkan bahwa nilai DBI dan SI pada penelitian terkait lebih rendah dibandingkan dengan penelitian yang telah dilakukan oleh peneliti.



Penelitian yang dilakukan oleh Aditya Ramadhan, dkk dengan judul, "Perbandingan *K-Means* dan *Fuzzy C-Means* untuk Pengelompokkan *Data User Knowledge Modelling*", dengan menerapkan *K-Means* dan *Fuzzy C-Means*, diperoleh hasil penelitian yaitu nilai SI pada algoritma *K-Means* bernilai 0.1866, sedangkan nilai PCI pada algoritma *Fuzzy C-Means* bernilai 0.2854. Sedangkan pada penelitian yang telah dilakukan oleh peneliti dengan judul, "Komparasi Validasi *K-Means* dan *Fuzzy C-Means* dalam Menentukan *Cluster* Terbaik Menggunakan *Silhouette Index* dan *Davies Bouldin Index*", diperoleh hasil pengujian berupa algoritma *cluster* yang lebih baik yaitu *K-Means* dan mendapatkan kualitas *cluster* yang optimal dengan nilai validasi DBI = 0.4571 dan SI = 0.7486. Sehingga dapat disimpulkan bahwa nilai DBI dan SI pada penelitian terkait lebih rendah dibandingkan dengan penelitian yang telah dilakukan oleh peneliti.

Dari penelitian yang telah dilakukan dan dibandingkan dengan penelitian terkait, maka dapat disimpulkan bahwa algoritma *cluster* yang lebih baik yaitu algoritma *K-Means* dengan 3 *cluster*, memperoleh hasil DBI = 0.4571 (nilai DBI semakin baik jika mendekati 0), SI = 0.7486 (nilai SI semakin baik jika mendekati 1), jumlah anggota *cluster* 1 = 32 daerah kabupaten/kota, jumlah anggota *cluster* 2 = 161 daerah kabupaten/kota, dan jumlah anggota *cluster* 3 = 357 daerah kabupaten/kota. Sehingga kualitas *cluster* yang dihasilkan optimal dan nilai validasi yang cukup tinggi. Disamping itu, *K-Means* lebih baik dibandingkan dengan *Fuzzy C-Means* karena *K-Means* mempunyai kemampuan mengelompokkan data dalam jumlah yang cukup besar dengan waktu komputasi yang relatif lebih cepat dan efisien.

Dari penelitian yang telah dilakukan, diperoleh nilai validasi yang tinggi karena menggunakan pustaka (*library*) NbClust dengan metode *complete* pada *K-Means* dan metode *average* pada *Fuzzy C-Means*.

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Pada penelitian ini dapat disimpulkan beberapa hal diantaranya adalah sebagai berikut.

1. Algoritma *cluster* yang lebih baik antara *K-Means* dan *Fuzzy C-Means*, yaitu algoritma *K-Means*.
2. Nilai validasi yang diperoleh yaitu  $DBI = 0.4571$  dan  $SI = 0.7486$ . Jumlah *cluster* optimal yang diperoleh yaitu 3 *cluster*.
3. Jumlah setiap data pada setiap *cluster* pada algoritma *K-Means*, yaitu *cluster* 1 berjumlah 32 daerah kabupaten/kota, *cluster* 2 berjumlah 161 daerah kabupaten/kota, dan *cluster* 3 berjumlah 357 daerah kabupaten/kota.

#### 5.2 Saran

Berdasarkan hasil penelitian ini, terdapat beberapa saran yang diberikan untuk penelitian selanjutnya, yaitu:

1. Pada penelitian ini menggunakan bahasa pemrograman R dan *class fanny* untuk mengelompokkan data pada algoritma *Fuzzy C-Means*. Sehingga selisih nilai fungsi objektif terhadap nilai epsilon (*error*) yang dihasilkan masih lebih besar dan dapat berpengaruh terhadap kualitas *cluster* yang dihasilkan dan nilai validasi DBI dan SI.
2. Penelitian yang dilakukan adalah perbandingan antara *exclusive clustering* dan *overlapping clustering*, sehingga penelitian ini dapat dikembangkan dengan metode *clustering* yang lain, misalnya dengan menggunakan *exclusive clustering* dan *hierarchical clustering*, atau *overlapping clustering* dan *hierarchical clustering*.
3. Penentuan nilai K dapat dikembangkan dengan metode yang lain, misalnya yaitu *Gap Method* atau *Elbow Method*.
4. Penelitian selanjutnya dapat dikembangkan dengan menggunakan perbandingan akurasi berdasarkan nilai persentase antara *K-Means* dan *Fuzzy C-Means* dengan mengoptimasi algoritma *K-Means* dengan algoritma Genetika agar mendapatkan hasil optimasi yang lebih baik dibandingkan hanya dengan metode *K-Means* biasa.

5. Indeks validasi yang digunakan dapat dikembangkan menggunakan indeks validasi yang lain.

## DAFTAR PUSTAKA

- [1] B. P. Statistik, “Statistik Indonesia 2018,” Badan Pusat Statistik Indonesia, 2018.
- [2] Katadata, “Berapa Jumlah Penduduk Indonesia?,” <https://databoks.katadata.co.id/>, 2018. [Online]. Available: <https://databoks.katadata.co.id/datapublish/2018/01/12/berapa-jumlah-penduduk-indonesia>. [Accessed: 27-Feb-2019].
- [3] N. Atthina and L. Iswari, “Klasterisasi Data Kesehatan Penduduk untuk Menentukan Rentang Derajat Kesehatan Daerah dengan Metode K-Means,” *Semin. Nas. Apl. Teknol. Infromasi*, vol. 1, p. B-52-B-59, 2014.
- [4] R. Rustiyan and Mustakim, “Penerapan Algoritma Fuzzy C Means untuk Analisis Permasalahan Simpanan Wajib Anggota Koperasi,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 2, p. 171, 2018.
- [5] D. T. Larose, *Data Mining Methods and Models*. 2006.
- [6] A. Ramadhan, Z. Efendi, and Mustakim, “Perbandingan K-Means dan Fuzzy C-Means untuk Pengelompokan Data User Knowledge Modeling,” *Semin. Nas. Teknol. Informasi, Komun. dan Ind.* 9, pp. 18–19, 2017.
- [7] J. Saragih, O. S. Sitompul, and Z. Situmorang, *Analysis of Intergrade Variables in the Fuzzy C-Means and Improved Algorithm Cat Swarm Optimization(FCM-ISO) in Search Segmentation*, vol. 930, no. 1. 2017.
- [8] D. F. Azuri, Zulhanif, and R. S. Pontoh, “Pengelompokan Kabupaten/Kota Di Pulau Jawa Berdasarkan Pembangunan Manusia Berbasis Gender Menggunakan Bisecting K-Means,” in *Peran Penelitian Ilmu Dasar dalam Menunjang Pembangunan Berkelanjutan*, 2016, pp. 78–83.
- [9] A. Saiful and J. L. Buliali, “Implementasi Particle Swarm Optimization pada K-Means untuk Clustering Data Automatic Dependent Surveillance-Broadcast,” *J. Eksplora Inform.*, vol. 8, no. 1, pp. 30–35, 2018.
- [10] C. D. Rumiarti and I. Budi, “Segmentasi Pelanggan Pada Customer Relationship Management di Perusahaan Ritel: Studi Kasus PT Gramedia Asri Media,” *J. Sist. Inf.*, vol. 13, no. 1, pp. 1–10, 2017.
- [11] D. K. Kabupaten Aceh Singkil, “Profil Kesehatan Kabupaten Aceh Singkil,” Aceh, 2017.
- [12] Katadata, “2010-2035, Angka Kelahiran Turun Sedangkan Angka Kematian,” [databoks.katadata.co.id](https://databoks.katadata.co.id/), 2018. [Online]. Available: <https://databoks.katadata.co.id/datapublish/2018/03/20/2010-2035-angka-kelahiran-turun-sedangkan-angka-kematian-naik>. [Accessed: 28-Mar-2019].
- [13] L. N. Rani, “Klasifikasi Nasabah Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit,” *J. KomTekInfo Fak. Ilmu Komput.*, vol. 2, no. 2, pp. 33–38, 2015.
- [14] I. H. Witten and E. Frank, *Data Mining - Practical Machine Learning Tools and Techniques*, 2nd ed. 2005.

- [15] I. H. Witten, E. Frank, and A. M. Hall, *Data Mining - Practical Machine Learning Tools and Techniques*, 3rd ed. 2011.
- [16] E. Turban, J. E. Aronson, and T. Liang, *Decision Support Systems and Intelligent Systems*. 2005.
- [17] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Technics*, 3rd ed. 2000.
- [18] J. Wu, *Advances in K-means Clustering*. 2012.
- [19] P. K. Agarwal, *Clustering and Classification*. 2003.
- [20] S. Khanmohammadi, N. Adibeig, and S. Shanehbandy, "An improved overlapping k-means clustering method for medical applications," *Expert Syst. Appl.*, vol. 67, pp. 12–18, 2017.
- [21] D. J. Bora and A. K. Gupta, "A Comparative Study Between Fuzzy Clustering Algorithm and Hard Clustering Algorithm," *Int. J. Comput. Trends Technol.*, vol. 10, no. 2, pp. 108–113, 2014.
- [22] B. Panda, S. Sahoo, and S. K. Patnaik, "A Comparative Study of Hard and Soft Clustering Using Swarm Optimization," *Int. J. Sci. Eng. Res.*, vol. 4, no. 10, p. 787, 2013.
- [23] D. Atmajaya, "K-Means Algorithm - Dedy Atmajaya," 2016. [Online]. Available: <http://lecturer.fikom.umi.ac.id/dedyatmajaya/k-means-algorithm/>. [Accessed: 08-May-2018].
- [24] N. I. Selviana and Mustakim, "Analisis Perbandingan K-Means dan Fuzzy C-Means untuk Pemetaan Motivasi Balajar Mahasiswa," *Semin. Nas. Teknol. Informasi, Komun. dan Ind.* 8, pp. 95–105, 2016.
- [25] T. Alfina, B. Santosa, and A. R. Barakbah, "Analisa Perbandingan Metode Hierarchical Clustering, K-Means dan Gabungan Keduanya dalam Cluster Data (Studi Kasus: Problem Kerja Praktek Jurusan Teknik Industri ITS)," *J. Tek. POMITS*, vol. 1, no. 1, pp. 1–5, 2012.
- [26] X. Du and S. Dua, *Data Mining and Machine Learning in Cybersecurity*. 2011.
- [27] Amirullah, "Populasi dan Sampel (Pemahaman, Jenis dan Teknik)," in *Metode Penelitian Manajemen*, Malang: Bayumedia Publishing Malang, 2015, pp. 67–68.
- [28] J. E. Barlett, J. W. Kotrlik, and C. C. Higgins, "Organizational Research: Determining Sample Size in Survey Research," *Inf. Technol. Learn. Perform. J.*, vol. 19, no. 1, p. 46, 2001.
- [29] B. D. Satoto, A. Muhammad, and B. K. Khotimah, "Pengelompokan Tingkat Kesehatan Masyarakat Menggunakan Shelf Organizing Maps Dengan Cluster Validation," in *Seminar Nasional Aplikasi Teknologi Informasi (SNATI) 2015*, 2015, pp. 1–6.
- [30] P. U. Gio and A. R. Effendie, *Belajar Bahasa Pemrograman R*. Medan: USU Press, 2017.
- [31] H. Wickham, "ggplot2 - Elegant Graphics for Data Analysis (2nd Edition)," *J. Stat. Softw.*, vol. 77, no. Book Review 2, pp. 3–5, 2017.

- [32] M. Charrad, G. Nadia, V. Boiteau, and A. Niknafs, “NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set,” *J. Stat. Softw.*, vol. 61, no. 6, pp. 1–36, 2014.
- [33] T. Thinsungneon, N. Kaoungku, and P. Durongdumronchai, “The Clustering Validity with Silhouette and Sum of Squared Errors,” in *Proceedings of the 3rd International Conference on Industrial Application Engineering 2015*, 2015, pp. 44–51.
- [34] F. Irhamni, F. Damayanti, B. Khusnul, and A. Miftachul, “Optimalisasi Pengelompokan Kecamatan Berdasarkan Indikator Pendidikan Menggunakan Metode Clustering dan Davies Bouldin Index,” *Semin. Nas. dan Teknol. UMJ*, no. 12 November, pp. 1–6, 2014.
- [35] N. V. M. Waworuntu and M. F. Amin, “Penerapan Metode K-Means Pemetaan Calon Penerima Jamkesda,” *Kumpul. J. Ilmu Komput.*, vol. 05, no. 02, pp. 195–200, 2018.

## LAMPIRAN

### Lampiran 1. *Data Set* Penelitian

Lampiran 1. *Data Set* Kesehatan Kabupaten/Kota Tahun 2017

No.	Puskesmas	Kematian Neonatal (jiwa)	Kematian Bayi (jiwa)	Kematian Anak Balita (jiwa)	Kematian Balita (jiwa)	Kematian Ibu (jiwa)
1	Singkil	2	5	0	5	4
2	Pulau Banyak	0	1	1	2	1
3	Singkil Utara	2	3	0	3	1
4	Gunung Meriah	9	10	1	11	3
5	Simpang Kanan	8	11	1	12	0
6	Danau Paris	0	0	0	0	2
7	Suro	2	2	0	2	0
8	Singkohor	0	0	2	2	0
9	Kuta Baharu	0	2	1	3	0
10	Kuala Baru	1	2	0	0	0
11	Pulau Banyak Barat	0	0	0	0	0
12	Blangkejeren	1	2	0	2	0
13	Labuhan Haji	1	2	0	2	1
14	Peulumat	0	1	0	1	0
15	Meukek	1	0	1	1	1
16	Drien Jalo	0	0	0	0	0
17	Sawang	5	7	0	7	0
18	Samadua	1	3	0	3	0
19	Tapaktuan	1	1	1	2	1
20	Lhok Bengkuang	0	0	0	0	1
21	Ladang Tuha	0	0	1	1	0
22	Uj. Padang Rasian	0	0	0	0	0
23	Kluet Utara	1	2	0	2	0
24	Kampong Raya	0	0	0	0	0
25	Kuala Bâ€™TMu	3	4	1	5	0
26	Manggamat	1	1	0	1	0
27	Kluet Timur	2	2	0	2	1
28	Durian Kawan	0	0	0	0	0

29	Kluet Selatan	1	2	1	3	0
30	Bakongan	0	0	0	0	0
31	Bukit Gadeng	0	0	0	0	0
32	Seubadeh	2	2	0	2	1
33	Trumon	3	4	0	4	1
34	Ladang Rimba	2	2	0	2	0
35	Krueng Luas	0	0	0	0	0
36	Linge	0	2	0	2	0
37	Atu Lintang	2	2	0	2	0
38	Jagong Jeget	1	1	0	1	0
39	Bintang	1	2	0	2	0
40	Lut Tawar	5	5	0	5	1
41	Kebayakan	4	5	0	5	0
42	Pegasing	5	5	0	5	0
43	Bies	3	3	0	3	1
44	Bebesen	15	19	0	19	1
45	Kute Panang	1	1	0	1	0
46	Silih Nara	4	4	0	4	0
47	Ketol	2	3	0	3	0
48	Celala	1	1	0	1	0
49	Rusip Antara	2	5	0	5	0
50	Johan Pahlawan	5	7	0	7	1
51	Suak Ribe	1	1	0	1	0
52	Meureubo	6	7	0	7	5
53	Peureumeu	7	13	1	14	0
54	Cot Seumeureng	1	1	0	1	0
55	Kuta Padang Layung	0	2	0	2	0
56	Kuala Bhee	2	3	1	4	2
57	Tangkeh	1	1	1	2	0
58	Drien Rampak	0	2	0	2	0
59	Pasi Mali	2	3	0	3	0
60	Meutulang	2	3	0	3	1
61	Pante Ceureumen	2	3	0	3	0
62	Kajeng	2	5	0	5	0
63	Geumpang	1	1	0	1	0
64	Menee	0	0	1	1	0



65	Glp.Tiga	1	1	0	1	1
66	Teupin Raya	0	0	0	0	3
67	Glp.Bar	4	6	1	7	0
68	Mutiara Timur	2	2	0	2	1
69	Ujong Rimba	2	4	1	5	1
70	Mutiara Barat	3	4	1	5	0
71	Tiro	3	4	1	5	0
72	Tangse	6	7	0	7	2
73	Keumala	1	3	0	3	0
74	Titeu	0	0	0	0	0
75	Sakti	5	10	1	11	2
76	Mila	0	0	0	0	0
77	Padang Tiji	10	11	0	11	0
78	Delima	4	6	1	7	0
79	Reubee	1	1	0	1	0
80	Grong-Grong	6	6	1	7	0
81	Indrajaya	5	9	1	10	0
82	Peukan Baro	3	3	3	6	1
83	Kb.Tanjong	3	4	0	4	0
84	Simpang Tiga	2	4	0	4	1
85	Kota Sigli	9	9	0	9	1
86	Pidie	11	14	4	18	2
87	Batee	2	2	0	2	0
88	Muara Tiga	5	5	0	5	0
89	Samalanga	5	6	0	6	1
90	Simpang Mamplam	6	6	0	6	1
91	Pandrah	2	2	0	2	0
92	Jeunieb	11	13	0	13	0
93	Peulimbang	7	7	1	8	0
94	Peudada	3	3	1	4	0
95	Jeumpa	6	12	1	13	1
96	Kuala	5	5	0	5	3
97	Juli	8	11	0	11	0
98	Juli 2	2	3	0	3	0
99	Kota Juang	10	12	0	12	0
100	Jangka	1	1	0	1	1

101	Peusangan Selatan	3	4	0	4	1
102	Peusangan Siblah	3	5	0	5	1
103	Peusangan	4	5	0	5	1
104	Kuta Blang	5	5	0	5	0
105	Makmur	0	2	1	3	0
106	Gandapura	4	6	2	8	0
107	Blangpidie	3	2	0	0	2
108	Tangan-Tangan	3	4	2	0	0
109	Bineh Krueng	0	3	0	0	0
110	Manggeng	3	3	0	0	0
111	Susoh	4	5	0	0	0
112	Sangkalan	1	1	0	0	0
113	Kuala Batee	2	0	0	0	0
114	Alue Pisang	0	0	0	0	1
115	Babahrot	0	0	0	0	0
116	Ie Mirah	2	1	0	0	0
117	Lembah Sabil	1	2	0	0	0
118	Lhang	0	1	0	0	0
119	A.S. Pinang	1	1	0	0	1
120	Kuta Panjang	1	1	0	1	0
121	Blang Jerango	0	1	0	1	0
122	Blangkejeren	0	0	0	0	0
123	Gumpang	0	0	0	0	0
124	Badak	1	3	0	0	0
125	Cinta Maju	0	0	0	0	0
126	Pinturime	0	0	0	0	0
127	Pining	0	0	0	0	1
128	Rikit Gaib	2	4	0	4	0
129	Kenyaran	0	0	0	0	0
130	Terangun	1	3	0	0	0
131	Rerebe	1	3	0	0	0
132	Minyak Payed	2	2	0	2	1
133	Bendahara	2	3	0	3	0
134	Sungai Iyu	1	1	0	1	0
135	Seruway	10	13	0	13	1
136	Karang Baru	13	13	0	13	0

137	Kualasimpang	3	4	0	4	1
138	Kejuruan Muda	6	7	0	7	0
139	Rantau	4	6	0	6	1
140	Sapta Jaya	7	8	0	8	0
141	Tamiang Hulu	3	5	0	5	1
142	Sekerak	2	2	0	2	0
143	Banda Mulia	1	1	0	1	0
144	Tenggulun	3	5	0	5	0
145	Bandar Pusaka	4	5	0	5	0
146	Beutong Ateuh	0	0	0	0	0
147	Beutong	2	4	0	4	0
148	Uteun Pulo	2	4	2	6	1
149	Jeuram	5	7	0	7	0
150	Cot Kuta	2	2	0	2	1
151	Ujong Patihah	0	1	0	1	1
152	Padang Panyang	0	0	0	0	0
153	Padang Rubek	1	3	0	3	0
154	Simpang Jaya	1	1	0	1	1
155	Kuala Tadu	3	3	0	3	0
156	Alue Bilie	2	2	0	2	1
157	Suka Mulia	2	2	0	2	0
158	Alue Rambot	0	0	0	0	0
159	Lueng Kebeu Jagat	3	4	0	4	0
160	Lamno	2	3	0	3	0
161	Indra Jaya	1	1	1	2	3
162	Lhok Kruet	2	2	0	2	0
163	Patek	4	6	0	6	0
164	Lageun	3	3	0	3	0
165	Calang	1	3	0	3	0
166	Krueng Sabee	3	3	0	3	0
167	Panga	5	5	0	5	0
168	Teunom	0	0	0	0	1
169	Pasie Raya	4	5	0	5	0
170	Belang Rakal	1	0	0	0	0
171	Singah Mulo	1	1	0	1	0
172	Ronga Ronga	1	1	2	3	1

173	Lampahan	2	4	0	4	1
174	Pante Raya	4	5	1	6	0
175	Simpang Teritit	4	5	1	6	0
176	Simpang Tiga	9	10	0	10	2
177	Bandar	6	8	0	8	1
178	Bener Kalipah	2	2	1	3	1
179	Buntul Kemumu	1	2	0	2	0
180	Ramung	0	1	0	1	0
181	Samar Kilang	0	0	1	1	0
182	Mesidah	3	3	0	3	1
183	Bandar Baru	3	0	0	0	0
184	Cubo	0	2	0	1	1
185	Panteraja	0	1	0	0	0
186	Trienggadeng	0	1	0	0	0
187	Meureudu	4	2	0	1	0
188	Meurah Dua	4	0	0	0	2
189	Ulim	2	1	0	0	0
190	Jangka Buya	2	0	0	0	0
191	Bandar Dua	2	0	0	1	0
192	Kuta Krueng	0	0	0	0	0
193	Blangkuta	2	3	0	2	0
194	Meuraxa	1	2	0	2	0
195	Jaya Baru	3	3	0	3	1
196	Banda Raya	0	0	0	0	0
197	Baiturrahman	1	2	0	2	1
198	Batoh	0	0	0	0	0
199	Kuta Alam	0	0	0	0	0
200	Lampulo	0	0	0	0	0
201	Lampaseh	0	0	0	0	0
202	Kopelma Darussalam	2	2	0	2	0
203	Jeulingke	0	0	0	0	0
204	Ulee Kareng	2	3	0	3	0
205	UPTD PKM Langsa Barat	7	9	1	10	1
206	UPTD PKM Langsa Baro	1	1	1	2	0
207	UPTD PKM Langsa Lama	2	4	5	9	0
208	UPTD PKM Langsa Kota	3	5	0	5	1

209	UPTD PKM Langsa Timur	8	0	0	10	3
210	Banda Sakti	5	6	0	6	3
211	Mongeudong	2	2	0	2	1
212	Muara Dua	3	3	0	3	0
213	Muara Satu	3	3	0	3	0
214	Blang Mangat	2	2	0	2	2
215	Blang Cut	1	1	0	1	0
216	Hiliduho	1	2	2	0	0
217	Botombawo	0	0	0	0	1
218	Botomuzoi	1	0	0	0	0
219	Hiliweto Gido	7	2	0	0	2
220	Somolomolo	0	0	0	0	0
221	Maâ€™u	2	1	0	0	2
222	Idanogawo	0	0	0	0	0
223	Ulugawo	0	0	0	0	3
224	Bawolato	1	1	0	0	0
225	Sogaâ€™adu	0	0	0	0	0
226	Pulo Pakkat	0	0	1	1	1
227	Sibabangun	0	0	0	0	0
228	Pinangsori	1	0	0	1	1
229	Hutabalang	0	0	0	0	0
230	Pandan	0	0	0	0	0
231	Kalangan	1	0	0	1	0
232	Tukka	0	0	0	0	0
233	Sarudik	2	0	0	2	0
234	Poriaha	2	0	0	2	2
235	Aek Raisan	2	0	0	2	0
236	Kolang	3	2	1	6	1
237	Sorkam	3	0	0	3	0
238	Gonting Mahe	0	0	0	0	0
239	Sipea-Pea	3	0	0	3	1
240	Pasaributobing	0	0	0	0	0
241	Siantar Ca	0	0	0	0	0
242	Barus	0	0	0	0	0
243	Andam Dewi	0	0	0	0	0
244	Sirandorung	1	0	0	1	2

245	Manduamas	0	0	1	1	0
246	Saragih	0	0	0	0	0
247	Lumut	0	0	0	0	0
248	Barus Utara	0	0	0	0	0
249	Parmonangan	0	1	2	3	0
250	Aek Raja	2	2	0	2	0
251	Parsingkaman	3	3	0	3	1
252	Sitadatada	3	3	1	4	0
253	Situmeanghabinsaran	3	6	0	6	0
254	Hutabaginda	3	5	0	5	1
255	Siatas Barita	3	3	0	3	0
256	Onan Hasang	3	4	0	4	0
257	Sarulla	3	3	0	3	0
258	Janji Angkola	0	0	1	1	0
259	Simangbumban	0	0	0	0	1
260	Lumban Sinaga	1	1	0	1	1
261	Pangaribuan	3	4	0	4	0
262	Garoga	1	1	1	2	0
263	Sipahutar	2	4	0	4	2
264	Siborongborong	5	5	0	5	0
265	Silangit	0	0	0	0	0
266	Butar	2	3	0	3	2
267	Muara	1	3	0	3	0
268	Suka Makmur	1	1	0	1	0
269	Janji	5	0	0	0	0
270	Rantauprapat	3	11	0	11	2
271	Perlayuan	2	2	0	2	1
272	Sigambal	7	7	1	8	2
273	Perbaungan	3	3	0	3	1
274	Lingga Tiga	2	2	0	2	3
275	Gunung Selamat	2	2	0	2	0
276	Pangkatan	3	4	0	4	0
277	Negeri Lama	9	6	0	6	2
278	Tanjung Haloban	2	3	0	3	0
279	Teluk Sentosa	7	4	0	4	0
280	Labuhan Bilik	4	2	0	2	1

281	Sei Berombang	5	3	0	3	2
282	Sei Penggantungan	0	0	0	0	1
283	Bp. Mandoge	5	5	0	5	1
284	Aek Songsongan	1	1	0	1	0
285	Pulau Rakyat	1	1	0	1	0
286	Aek Loba	4	4	0	4	0
287	Aek Ledong	0	0	0	0	0
288	Sei Kepyang	0	0	0	0	1
289	Sei Kepyang Barat	0	0	0	0	0
290	Sei Apung	0	0	0	0	0
291	Bagan Asahan	1	1	0	1	0
292	Simpang Empat	0	0	0	0	1
293	Air Batu	3	3	0	3	0
294	Hessa Air Genting	4	4	0	4	2
295	Sei Dadap	0	0	0	0	1
296	Prapat Janji	1	1	0	1	0
297	Tinggi Raja	0	0	0	0	0
298	Setia Janji	2	2	0	2	0
299	Meranti	1	1	0	1	0
300	Rawang Pasar IV	0	0	0	0	1
301	Binjai Serbangun	0	0	0	0	1
302	Sidodadi	0	0	0	0	2
303	Rs. Ibu Kartini	0	0	0	0	0
304	RSUD HAMS	0	0	0	0	0
305	Gambir Baru	0	0	0	0	1
306	Mutiara	5	5	1	6	0
307	Saribu Dolok	0	0	0	0	0
308	Pamatang Silimahuta	0	0	0	0	0
309	Tiga Runggu	2	0	0	0	0
310	Haranggaol	2	1	0	1	0
311	Sipintu Angin	4	0	0	0	1
312	Sarimatondang	0	0	0	0	1
313	Pamatang Sidamanik	0	0	0	0	0
314	Parapat	4	0	0	0	0
315	Tanah Jawa	3	0	0	0	0
316	Marubun Jaya	0	0	0	0	0

317	Hatonduhan	1	0	1	1	0
318	Buttu Turunan	1	0	0	0	0
319	Tiga Dolok	3	1	0	1	0
320	Tiga Balata	0	0	0	0	0
321	Panei Tongah	0	0	0	0	0
322	Panombean Pane	1	0	0	0	0
323	Pamatang Raya	2	0	0	0	1
324	Saran Padang	0	0	0	0	0
325	Cingkes	0	0	0	0	0
326	Nagori Dolok	1	0	0	0	0
327	Sinasih	1	0	0	0	0
328	Sindar Raya	0	0	0	0	0
329	Bah Tonang	0	0	0	0	0
330	Tapian Dolok	3	0	0	0	0
331	Serbelawan	1	1	0	1	0
332	Silou Malaha	2	1	0	1	0
333	Batu Anam	3	0	0	0	0
334	Rambung Merah	1	0	0	0	0
335	Simpang Bah Jambi	0	0	0	0	0
336	Bandar Siantar	1	0	0	0	0
337	Gunung Maligas	2	0	0	0	1
338	Raja Maligas	0	0	0	0	0
339	Huta Bayu	3	0	0	2	1
340	Jawa Maraja Bah Jambi	0	0	0	0	0
341	Pematang Bandar	2	0	0	0	1
342	Kerasaan	1	0	0	0	1
343	Bandar Huluan	0	1	0	1	0
344	Perdagangan	3	0	0	0	1
345	Marihat Bandar	3	0	0	0	0
346	Bandar Masilam	1	0	0	0	0
347	Bandar Tinggi	0	0	0	0	0
348	Bosar Maligas	0	1	0	1	0
349	Parbutaran	0	0	0	0	0
350	Ujung Padang	3	1	0	1	0
351	Sayur Matinggi	2	0	0	0	0
352	Bah Bolon	0	0	0	0	0



353	Kabanjahe	2	2	0	2	0
354	Berastagi	0	0	0	0	0
355	Korpri	0	0	0	0	0
356	Tigapanah	1	2	0	2	0
357	Singa	0	0	0	0	0
358	Dolat Rayat	0	0	0	0	0
359	Merdeka	1	1	0	1	0
360	Merek	2	2	0	2	0
361	Barusjahe	0	0	0	0	1
362	Simpang Empat	1	1	0	1	0
363	Naman Teran	0	0	0	0	0
364	Tiganderket	2	2	0	2	0
365	Payung	0	0	1	1	1
366	Munthe	2	2	0	2	2
367	Juhar	1	1	0	1	1
368	Tigabinanga	0	0	0	0	0
369	Kutabuluh	0	0	0	0	1
370	Laubaleng	1	1	1	2	0
371	Mardinding	0	0	0	0	0
372	Bahorok	3	3	0	3	0
373	Bukit Lawang	0	0	0	0	0
374	Serapit	0	0	0	0	0
375	Tanjung Langkat	1	1	1	2	0
376	Marike	1	1	0	1	0
377	Namu Ukur	2	2	0	2	1
378	Namu Trasi	0	0	1	2	0
379	Kuala	0	0	0	0	0
380	Selesai	1	1	0	1	1
381	Sambirejo	9	9	0	9	0
382	Stabat	0	0	0	0	0
383	Karang Rejo	0	0	0	0	0
384	Stabat Lama	4	4	0	4	0
385	Hinai Kiri	0	0	0	0	2
386	Desa Teluk	1	1	0	1	0
387	Secanggang	1	1	0	1	1
388	Tanjung Beringin	2	2	0	2	1

389	Tanjung Selamat	0	0	0	0	1
390	Sei Bamban	0	0	0	0	0
391	Sawit Seberang	2	2	0	2	0
392	Pantai Cermin	4	4	0	3	4
393	Gebang	2	2	0	2	0
394	Securai	1	1	0	1	1
395	Pangkalan Brandan	0	0	0	0	0
396	Desa Lama	1	1	0	1	1
397	Tangkahan Durian	0	0	0	0	0
398	Pangkalan Susu	0	0	0	0	0
399	Beras Basah	0	0	0	0	0
400	Besitang	3	5	0	1	0
401	Pematang Jaya	1	1	0	1	0
402	Telukdalam	4	4	0	4	1
403	Onolalu	3	3	0	3	0
404	Luahagundre Maniamolo	1	1	0	1	0
405	Bawomataluo	1	1	0	1	0
406	Hilisataro	0	0	0	0	0
407	Hilisimaetano	1	1	0	1	0
408	Hilizalootano	1	1	0	1	0
409	Amandraya	4	4	0	4	0
410	Ulususua	1	1	0	1	0
411	Aramo	0	0	0	0	0
412	Lolowau	4	4	0	4	1
413	O'o'u	3	3	0	3	0
414	Onohazumba	2	2	0	2	0
415	Hilisalawaaha	2	2	0	2	0
416	Hilimegai	0	0	0	0	0
417	Huruna	0	0	0	0	0
418	Lolomatua	2	2	0	2	0
419	Ulunoyo	1	1	0	1	0
420	Lahusa	3	3	0	3	1
421	Siduaori	0	0	0	0	0
422	Somambawa	2	2	0	2	0
423	Gomo	2	2	0	2	0
424	Boronadu	0	0	0	0	0

425	Idanotae	0	0	0	0	0
426	Uluidanotae	0	0	0	0	0
427	Susua	0	0	0	0	0
428	Mazo	0	0	0	0	1
429	Umbunasi	0	0	0	0	0
430	Pulau Tello	0	0	0	0	0
431	P.P. Batu Barat	0	0	0	0	0
432	P.P. Batu Utara	0	0	0	0	0
433	Simuk	1	1	0	1	0
434	Hibala	0	0	0	0	0
435	Tanah Masa	0	0	0	0	0
436	Labuhan Hiu	0	0	0	0	0
437	Hilianombasela	1	1	0	1	0
438	Parlilitan	1	1	0	1	0
439	Hutagalung	1	1	0	1	0
440	Hutapaung	3	5	0	5	0
441	Baktiraja	1	1	0	1	0
442	Paranginan	4	5	0	5	0
443	Sigompul	3	3	0	3	0
444	Matiti	6	6	0	6	1
445	Saitnihuta	5	6	0	6	0
446	Bonandolok	0	0	0	0	0
447	Onanganjang	4	4	0	4	0
448	Pakkat	3	5	0	5	3
449	Tarabintang	3	3	0	3	4
450	Salak	1	0	0	0	0
451	Sukaramai	2	0	0	0	0
452	Sibande	0	0	0	0	0
453	Tinada	0	0	0	0	0
454	Kecupak	1	0	2	2	0
455	Singgabur	0	1	0	1	0
456	Siempat Rube	3	0	0	0	0
457	Sibagindar	0	0	0	0	0
458	Perbaungan	7	1	0	8	4
459	Melati	1	1	0	2	1
460	Pantai Cermin	8	1	1	10	1

461	Pegajahan	0	0	0	0	1
462	Sialang Buah	1	4	0	5	0
463	Sei Rampah	2	1	0	3	2
464	PKL Budiman	3	0	0	3	0
465	T. Beringin	10	2	0	12	0
466	Desa Pon	1	0	0	1	0
467	Paya Lombang	6	0	0	6	0
468	N. Kesiangan	0	1	0	1	0
469	T. Syahbandar	3	1	0	1	0
470	Bdr. Kalipah	0	0	0	0	0
471	Dlk. Merawan	0	0	0	0	0
472	Dolok Masihul	1	5	0	6	0
473	Sipispis	0	1	0	1	0
474	Kuala Bali	3	2	1	6	0
475	Bintang Bayu	0	0	0	0	0
476	Kotarih	2	0	0	2	0
477	Silinda	2	2	0	4	0
478	Kotapinang	3	2	1	3	0
479	Batu Ajo	0	1	0	1	0
480	Sisumut	1	2	0	2	0
481	Mampang	1	1	0	1	0
482	Tanjung Medan	3	3	0	3	0
483	Pekan Tolan	0	0	0	0	0
484	Teluk Panji	3	2	0	2	2
485	Hutagondang	1	1	0	1	0
486	Langga Payung	2	2	0	2	0
487	Aek Goti	1	1	0	1	0
488	Ulumahuam	0	1	0	1	0
489	Aek batu	2	1	0	1	0
490	Aek Raso	0	1	0	1	0
491	Beringin Jaya	0	1	0	1	0
492	Bunut	3	1	1	2	0
493	Cikampak	3	2	0	2	0
494	Rasau	1	1	0	1	0
495	Pintu Angin	1	0	0	0	0
496	Sambas	3	0	0	0	0

497	Pelabuhan Sambas	3	0	0	0	0
498	Aek Habil	0	0	0	0	2
499	Aek Parombunan	1	0	0	0	0
500	Datuk Bandar	0	0	0	0	0
501	Semula Jadi	0	0	0	0	2
502	Mu Damanik	0	2	0	2	1
503	RSUD dr.T. Mansyur	17	21	3	24	1
504	Kabung Baru	0	0	0	0	0
505	Kampung Persatuan	0	0	0	0	0
506	Seitualang Raso	0	2	0	2	0
507	Teluk Nibung	0	0	0	0	1
508	Siporipori	0	0	0	0	0
509	Rantau Laban	2	4	2	6	0
510	Tanjung Marulak	0	1	0	1	2
511	Sri Padang	1	2	0	2	0
512	Satria	4	3	1	4	0
513	Rambung	1	1	0	2	1
514	Pasar Gambir	0	0	0	0	1
515	Pabatu	2	4	0	4	0
516	Teluk Karang	2	2	0	2	1
517	Berohol	3	4	1	4	2
518	Binjai Estate	3	3	0	3	1
519	Rambung	1	1	0	1	1
520	Binjai Kota	2	2	0	2	0
521	Tanah Tinggi	5	5	0	5	2
522	Kebun Lada	3	3	0	3	0
523	Jati Makmur	1	1	0	1	2
524	Bandar Senembah	0	0	0	0	0
525	H,A,H. Hasan	2	2	0	2	0
526	Pijorkoling	3	5	1	6	2
527	Labuhan Rasoki	0	2	0	2	0
528	Padangmatinggi	10	13	1	14	2
529	Sidangkal	11	11	0	11	0
530	Batunadua	7	7	0	7	1
531	Sadabuan	17	21	2	23	2
532	Hutaimbaru	6	7	0	7	1

533	Pokenijor	2	2	0	2	0
534	Pintu Langit	1	1	0	1	0
535	UPTD Puskesmas Kecamatan Gunungsitoli	12	3	0	15	0
536	UPTD Puskesmas Kecamatan Gunungsitoli Utara	4	0	0	4	0
537	UPTD Puskesmas Kecamatan Gunungsitoli Selatan	5	2	0	7	0
538	UPTD Puskesmas Kecamatan Gunungsitoli Barat	3	0	0	3	0
539	UPTD Puskesmas Kecamatan Gunungsitoli Idanoi	5	2	0	7	1
540	UPTD Puskesmas Kecamatan Gunungsitoli Alo'oa	0	2	0	2	0
541	Sungai Lansek	2	4	0	4	0
542	Kamang	5	5	1	6	0
543	Air Amo	3	8	0	8	1
544	Tanjung Gadang	7	12	1	13	1
545	Sijunjung	3	9	1	10	2
546	Gambok	4	5	1	6	0
547	Lubuk Tarok	2	8	0	8	0
548	Muaro Bodi	2	4	0	4	0
549	Padang Sibusuk	1	3	0	3	0
550	Tanjung Ampalu	5	10	0	10	1

## Lampiran 2. Kode Sumber Algoritma *K-Means*

### Lampiran 2. *Source Code* Algoritma *K-Means*

#### ***Source Code K-Means***

```
#K-Means 550 data

#impor dataset (1)
impor.dataset5 <- read.csv(file.choose(), header = TRUE)
View(impor.dataset5)

#menghapus nilai yang hilang pada data dan menghapus kolom berisi
string (2)
```

```

ds5<-impor.dataset5
df5<-na.omit(ds5)
x5<-df5[-1]
View(x5)

#standarisasi data (3)
sd5<-scale(x5)
View(sd5)

#proses k-means dengan k=3 (4)
kmeans.proc5 <- kmeans(sd5, centers = 3)
kmeans.proc5
kmeans.proc5$iter
#-kmeans = fungsi
#-sd = data yang akan diproses dengan k-means
#-centers = jumlah cluster (k)

#menambahkan kolom cluster ke data asli
dd5<-cbind(impor.dataset5, cluster = kmeans.proc5$cluster)
View(dd5)

#plot/visualisasi (5)
library(factoextra)
library(ggplot2)
fviz_cluster(kmeans.proc5, data = sd5)

#validasi dengan DBI (Davies-Bouldin Index) (6)
library(NbClust)
DBIKMeans5<-NbClust(sd5, distance = "euclidean",
                    min.nc = 3, max.nc = 3,
                    index = "db", method = "complete")
DBIKMeans5$Best.nc #rekomendasi jumlah kluster terbaik dan nilai
validasinya
DBIKMeans5$All.index #banyaknya jumlah kluster dan nilai
validasinya masing-masing

#validasi dengan SI (Silhouette Index) (6)
library(NbClust)
SIKMeans5<-NbClust(sd5, distance = "euclidean",
                   min.nc = 3, max.nc = 3,
                   index = "silhouette", method = "complete")
SIKMeans5$Best.nc #rekomendasi jumlah kluster terbaik dan nilai
validasinya
SIKMeans5$All.index #banyaknya jumlah kluster dan nilai validasinya
masing-masing

```

### Lampiran 3. Kode Sumber Algoritma *Fuzzy C-Means*

#### Lampiran 3. *Source Code* Algoritma *Fuzzy C-Means*

***Source Code Fuzzy C-Means***

```

#Fuzzy C-Means 550 data

#impor dataset (1)
impor.dataset5 <- read.csv(file.choose(), header = TRUE)
View(impor.dataset5)

#menghapus nilai yang hilang pada data (2)
ds5<-impor.dataset5
df5<-na.omit(ds5)
x5<-df5[-1]

#standarisasi data (3)
sd5<-scale(x5)
View(sd5)

#proses fuzzy c-means dengan k=5 (4)
library(cluster)
proses.fcm5 <- fanny(sd5, 3, metric = "euclidean", stand = FALSE)
proses.fcm5

#menambahkan kolom cluster ke data asli
dd5<-cbind(impor.dataset5, cluster = proses.fcm5$cluster)
View(dd5)

#plot/visualisasi (5)
library(factoextra)
library(ggplot2)
fviz_cluster(proses.fcm5, data = sd5)

#validasi dengan DBI (Davies-Bouldin Index) (6)
library(NbClust)
DBIFCM5<-NbClust(sd5, distance = "euclidean",
                 min.nc = 3, max.nc = 3,
                 index = "db", method = "average")
DBIFCM5$Best.nc #rekomendasi jumlah kluster terbaik dan nilai
validasinya
DBIFCM5$All.index #banyaknya jumlah kluster dan nilai validasinya
masing-masing

#validasi dengan SI (Silhouette Index) (6)
library(NbClust)
SIFCM5<-NbClust(sd5, distance = "euclidean",
                 min.nc = 3, max.nc = 3,
                 index = "silhouette", method = "average")
SIFCM5$Best.nc #rekomendasi jumlah kluster terbaik dan nilai
validasinya
SIFCM5$All.index #banyaknya jumlah kluster dan nilai validasinya
masing-masing

```



#### Lampiran 4. Pembagian Anggota *Cluster K-Means* pada Data Kesehatan Daerah

Lampiran 4. Pembagian Anggota *Cluster K-Means* pada Data Kesehatan Daerah

No.	Puskesmas	Kematian Neonatal (jiwa)	Kematian Bayi (jiwa)	Kematian Anak Balita (jiwa)	Kematian Balita (jiwa)	Kematian Ibu (jiwa)	Anggota Cluster ke-i
1	Singkil	2	5	0	5	4	2
2	Pulau Banyak	0	1	1	2	1	3
3	Singkil Utara	2	3	0	3	1	2
4	Gunung Meriah	9	10	1	11	3	1
5	Simpang Kanan	8	11	1	12	0	1
6	Danau Paris	0	0	0	0	2	3
7	Suro	2	2	0	2	0	3
8	Singkohor	0	0	2	2	0	3
9	Kuta Baharu	0	2	1	3	0	3
10	Kuala Baru	1	2	0	0	0	3
11	Pulau Banyak Barat	0	0	0	0	0	3
12	Blangkejeren	1	2	0	2	0	3
13	Labuhan Haji	1	2	0	2	1	3
14	Peulumat	0	1	0	1	0	3
15	Meukek	1	0	1	1	1	3
16	Drien Jalo	0	0	0	0	0	3
17	Sawang	5	7	0	7	0	2
18	Samadua	1	3	0	3	0	3
19	Tapaktuan	1	1	1	2	1	3
20	Lhok Bengkuang	0	0	0	0	1	3
21	Ladang Tuha	0	0	1	1	0	3
22	Uj. Padang Rasian	0	0	0	0	0	3
23	Kluet Utara	1	2	0	2	0	3
24	Kampung Raya	0	0	0	0	0	3
25	Kuala Ba <sup>TM</sup> u	3	4	1	5	0	2
26	Manggamat	1	1	0	1	0	3
27	Kluet Timur	2	2	0	2	1	2
28	Durian Kawan	0	0	0	0	0	3

<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
29	Kluet Selatan	1	2	1	3	0	3
30	Bakongan	0	0	0	0	0	3
31	Bukit Gadeng	0	0	0	0	0	3
32	Seubadeh	2	2	0	2	1	2
33	Trumon	3	4	0	4	1	2
34	Ladang Rimba	2	2	0	2	0	3
35	Krueng Luas	0	0	0	0	0	3
36	Linge	0	2	0	2	0	3
37	Atu Lintang	2	2	0	2	0	3
38	Jagong Jeget	1	1	0	1	0	3
39	Bintang	1	2	0	2	0	3
40	Lut Tawar	5	5	0	5	1	2
41	Kebayakan	4	5	0	5	0	2
42	Pegasing	5	5	0	5	0	2
43	Bies	3	3	0	3	1	2
44	Bebesen	15	19	0	19	1	1
45	Kute Panang	1	1	0	1	0	3
46	Silih Nara	4	4	0	4	0	2
47	Ketol	2	3	0	3	0	3
48	Celala	1	1	0	1	0	3
49	Rusip Antara	2	5	0	5	0	2
50	Johan Pahlawan	5	7	0	7	1	2
51	Suak Ribe	1	1	0	1	0	3
52	Meureubo	6	7	0	7	5	2
53	Peureumeu	7	13	1	14	0	1
54	Cot Seumeureng	1	1	0	1	0	3
55	Kuta Padang Layung	0	2	0	2	0	3
56	Kuala Bhee	2	3	1	4	2	2
57	Tangkeh	1	1	1	2	0	3
58	Drien Rampak	0	2	0	2	0	3

<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
59	Pasi Mali	2	3	0	3	0	<b>3</b>
60	Meutulang	2	3	0	3	1	<b>2</b>
61	Pante Ceureumen	2	3	0	3	0	<b>3</b>
62	Kajeng	2	5	0	5	0	<b>2</b>
63	Geumpang	1	1	0	1	0	<b>3</b>
64	Menee	0	0	1	1	0	<b>3</b>
65	Glp.Tiga	1	1	0	1	1	<b>3</b>
66	Teupin Raya	0	0	0	0	3	<b>2</b>
67	Glp.Baroe	4	6	1	7	0	<b>2</b>
68	Mutiara Timur	2	2	0	2	1	<b>2</b>
69	Ujong Rimba	2	4	1	5	1	<b>2</b>
70	Mutiara Barat	3	4	1	5	0	<b>2</b>
71	Tiro	3	4	1	5	0	<b>2</b>
72	Tangse	6	7	0	7	2	<b>2</b>
73	Keumala	1	3	0	3	0	<b>3</b>
74	Titeu	0	0	0	0	0	<b>3</b>
75	Sakti	5	10	1	11	2	<b>1</b>
76	Mila	0	0	0	0	0	<b>3</b>
77	Padang Tiji	10	11	0	11	0	<b>1</b>
78	Delima	4	6	1	7	0	<b>2</b>
79	Reubee	1	1	0	1	0	<b>3</b>
80	Grong-Grong	6	6	1	7	0	<b>2</b>
81	Indrajaya	5	9	1	10	0	<b>1</b>
82	Peukan Baro	3	3	3	6	1	<b>2</b>
83	Kb.Tanjong	3	4	0	4	0	<b>2</b>
84	Simpang Tiga	2	4	0	4	1	<b>2</b>
85	Kota Sigli	9	9	0	9	1	<b>1</b>
86	Pidie	11	14	4	18	2	<b>1</b>
87	Batee	2	2	0	2	0	<b>3</b>
88	Muara Tiga	5	5	0	5	0	<b>2</b>
89	Samalanga	5	6	0	6	1	<b>2</b>

<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
90	Simpang Mamplam	6	6	0	6	1	2
91	Pandrah	2	2	0	2	0	3
92	Jeunieb	11	13	0	13	0	1
93	Peulimbang	7	7	1	8	0	1
94	Peudada	3	3	1	4	0	2
95	Jeumpa	6	12	1	13	1	1
96	Kuala	5	5	0	5	3	2
97	Juli	8	11	0	11	0	1
98	Juli 2	2	3	0	3	0	3
99	Kota Juang	10	12	0	12	0	1
100	Jangka	1	1	0	1	1	3
101	Peusangan Selatan	3	4	0	4	1	2
102	Peusangan Siblah	3	5	0	5	1	2
103	Peusangan	4	5	0	5	1	2
104	Kuta Blang	5	5	0	5	0	2
105	Makmur	0	2	1	3	0	3
106	Gandapura	4	6	2	8	0	1
107	Blangpidie	3	2	0	0	2	2
108	Tangan-Tangan	3	4	2	0	0	2
109	Bineh Krueng	0	3	0	0	0	3
110	Manggeng	3	3	0	0	0	3
111	Susoh	4	5	0	0	0	2
112	Sangkalan	1	1	0	0	0	3
113	Kuala Batee	2	0	0	0	0	3
114	Alue Pisang	0	0	0	0	1	3
115	Babahrot	0	0	0	0	0	3
116	Ie Mirah	2	1	0	0	0	3
117	Lembah Sabil	1	2	0	0	0	3
118	Lhang	0	1	0	0	0	3
119	A.S. Pinang	1	1	0	0	1	3
120	Kuta Panjang	1	1	0	1	0	3

<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
121	Blang Jerango	0	1	0	1	0	3
122	Blangkejeren	0	0	0	0	0	3
123	Gumpang	0	0	0	0	0	3
124	Badak	1	3	0	0	0	3
125	Cinta Maju	0	0	0	0	0	3
126	Pinturime	0	0	0	0	0	3
127	Pining	0	0	0	0	1	3
128	Rikit Gaib	2	4	0	4	0	2
129	Kenyaran	0	0	0	0	0	3
130	Terangun	1	3	0	0	0	3
131	Rerebe	1	3	0	0	0	3
132	Minyak Payed	2	2	0	2	1	2
133	Bendahara	2	3	0	3	0	3
134	Sungai Iyu	1	1	0	1	0	3
135	Seruway	10	13	0	13	1	1
136	Karang Baru	13	13	0	13	0	1
137	Kualasimpang	3	4	0	4	1	2
138	Kejuruan Muda	6	7	0	7	0	2
139	Rantau	4	6	0	6	1	2
140	Sapta Jaya	7	8	0	8	0	2
141	Tamiang Hulu	3	5	0	5	1	2
142	Sekerak	2	2	0	2	0	3
143	Banda Mulia	1	1	0	1	0	3
144	Tenggulun	3	5	0	5	0	2
145	Bandar Pusaka	4	5	0	5	0	2
146	Beutong Ateuh	0	0	0	0	0	3
147	Beutong	2	4	0	4	0	2
148	Uteun Pulo	2	4	2	6	1	2
149	Jeuram	5	7	0	7	0	2
150	Cot Kuta	2	2	0	2	1	2
151	Ujong Patihah	0	1	0	1	1	3

<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
152	Padang Panyang	0	0	0	0	0	3
153	Padang Rubek	1	3	0	3	0	3
154	Simpang Jaya	1	1	0	1	1	3
155	Kuala Tadu	3	3	0	3	0	2
156	Alue Bilie	2	2	0	2	1	2
157	Suka Mulia	2	2	0	2	0	3
158	Alue Rambot	0	0	0	0	0	3
159	Lueng Kebeu Jagat	3	4	0	4	0	2
160	Lamno	2	3	0	3	0	3
161	Indra Jaya	1	1	1	2	3	2
162	Lhok Kruet	2	2	0	2	0	3
163	Patek	4	6	0	6	0	2
164	Lageun	3	3	0	3	0	2
165	Calang	1	3	0	3	0	3
166	Krueng Sabee	3	3	0	3	0	2
167	Panga	5	5	0	5	0	2
168	Teunom	0	0	0	0	1	3
169	Pasie Raya	4	5	0	5	0	2
170	Belang Rakal	1	0	0	0	0	3
171	Singah Mulo	1	1	0	1	0	3
172	Ronga Ronga	1	1	2	3	1	2
173	Lampahan	2	4	0	4	1	2
174	Pante Raya	4	5	1	6	0	2
175	Simpang Teritit	4	5	1	6	0	2
176	Simpang Tiga	9	10	0	10	2	1
177	Bandar	6	8	0	8	1	2
178	Bener Kalipah	2	2	1	3	1	2
179	Buntul Kemumu	1	2	0	2	0	3
180	Ramung	0	1	0	1	0	3
181	Samar Kilang	0	0	1	1	0	3
182	Mesidah	3	3	0	3	1	2

<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
183	Bandar Baru	3	0	0	0	0	<b>3</b>
184	Cubo	0	2	0	1	1	<b>3</b>
185	Panteraja	0	1	0	0	0	<b>3</b>
186	Trienggadeng	0	1	0	0	0	<b>3</b>
187	Meureudu	4	2	0	1	0	<b>3</b>
188	Meurah Dua	4	0	0	0	2	<b>2</b>
189	Ulim	2	1	0	0	0	<b>3</b>
190	Jangka Buya	2	0	0	0	0	<b>3</b>
191	Bandar Dua	2	0	0	1	0	<b>3</b>
192	Kuta Krueng	0	0	0	0	0	<b>3</b>
193	Blangkuta	2	3	0	2	0	<b>3</b>
194	Meuraxa	1	2	0	2	0	<b>3</b>
195	Jaya Baru	3	3	0	3	1	<b>2</b>
196	Banda Raya	0	0	0	0	0	<b>3</b>
197	Baiturrahman	1	2	0	2	1	<b>3</b>
198	Batoh	0	0	0	0	0	<b>3</b>
199	Kuta Alam	0	0	0	0	0	<b>3</b>
200	Lampulo	0	0	0	0	0	<b>3</b>
201	Lampaseh	0	0	0	0	0	<b>3</b>
202	Kopelma Darussalam	2	2	0	2	0	<b>3</b>
203	Jeulingke	0	0	0	0	0	<b>3</b>
204	Ulee Kareng	2	3	0	3	0	<b>3</b>
205	UPTD PKM Langsa Barat	7	9	1	10	1	<b>1</b>
206	UPTD PKM Langsa Baro	1	1	1	2	0	<b>3</b>
207	UPTD PKM Langsa Lama	2	4	5	9	0	<b>1</b>
208	UPTD PKM Langsa Kota	3	5	0	5	1	<b>2</b>
209	UPTD PKM Langsa Timur	8	0	0	10	3	<b>2</b>

<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
210	Banda Sakti	5	6	0	6	3	2
211	Mongeudong	2	2	0	2	1	2
212	Muara Dua	3	3	0	3	0	2
213	Muara Satu	3	3	0	3	0	2
214	Blang Mangat	2	2	0	2	2	2
215	Blang Cut	1	1	0	1	0	3
216	Hiliduho	1	2	2	0	0	3
217	Botombawo	0	0	0	0	1	3
218	Botomuzoi	1	0	0	0	0	3
219	Hiliweto Gido	7	2	0	0	2	2
220	Somolomolo	0	0	0	0	0	3
221	Maâ€™u	2	1	0	0	2	2
222	Idanogawo	0	0	0	0	0	3
223	Ulugawo	0	0	0	0	3	2
224	Bawolato	1	1	0	0	0	3
225	Sogaeâ€™adu	0	0	0	0	0	3
226	Pulo Pakkat	0	0	1	1	1	3
227	Sibabangun	0	0	0	0	0	3
228	Pinangsori	1	0	0	1	1	3
229	Hutabalang	0	0	0	0	0	3
230	Pandan	0	0	0	0	0	3
231	Kalangan	1	0	0	1	0	3
232	Tukka	0	0	0	0	0	3
233	Sarudik	2	0	0	2	0	3
234	Poriaha	2	0	0	2	2	2
235	Aek Raisan	2	0	0	2	0	3
236	Kolang	3	2	1	6	1	2
237	Sorkam	3	0	0	3	0	3
238	Gonting Mahe	0	0	0	0	0	3
239	Sipea-Pea	3	0	0	3	1	2
240	Pasaributobing	0	0	0	0	0	3



<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
241	Siantar Ca	0	0	0	0	0	3
242	Barus	0	0	0	0	0	3
243	Andam Dewi	0	0	0	0	0	3
244	Sirandorung	1	0	0	1	2	3
245	Manduamas	0	0	1	1	0	3
246	Saragih	0	0	0	0	0	3
247	Lumut	0	0	0	0	0	3
248	Barus Utara	0	0	0	0	0	3
249	Parmonangan	0	1	2	3	0	3
250	Aek Raja	2	2	0	2	0	3
251	Parsingkaman	3	3	0	3	1	2
252	Sitadatada	3	3	1	4	0	2
253	Situmeanghabinsara n	3	6	0	6	0	2
254	Hutabaginda	3	5	0	5	1	2
255	Siatar Barita	3	3	0	3	0	2
256	Onan Hasang	3	4	0	4	0	2
257	Sarulla	3	3	0	3	0	2
258	Janji Angkola	0	0	1	1	0	3
259	Simangbumban	0	0	0	0	1	3
260	Lumban Sinaga	1	1	0	1	1	3
261	Pangaribuan	3	4	0	4	0	2
262	Garoga	1	1	1	2	0	3
263	Sipahutar	2	4	0	4	2	2
264	Siborongborong	5	5	0	5	0	2
265	Silangit	0	0	0	0	0	3
266	Butar	2	3	0	3	2	2
267	Muara	1	3	0	3	0	3
268	Suka Makmur	1	1	0	1	0	3
269	Janji	5	0	0	0	0	3
270	Rantauprapat	3	11	0	11	2	1

<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
271	Perlayuan	2	2	0	2	1	<b>2</b>
272	Sigambal	7	7	1	8	2	<b>1</b>
273	Perbaungan	3	3	0	3	1	<b>2</b>
274	Lingga Tiga	2	2	0	2	3	<b>2</b>
275	Gunung Selamat	2	2	0	2	0	<b>3</b>
276	Pangkatan	3	4	0	4	0	<b>2</b>
277	Negeri Lama	9	6	0	6	2	<b>2</b>
278	Tanjung Haloban	2	3	0	3	0	<b>3</b>
279	Teluk Sentosa	7	4	0	4	0	<b>2</b>
280	Labuhan Bilik	4	2	0	2	1	<b>2</b>
281	Sei Berombang	5	3	0	3	2	<b>2</b>
282	Sei Penggantungan	0	0	0	0	1	<b>3</b>
283	Bp. Mandoge	5	5	0	5	1	<b>2</b>
284	Aek Songsongan	1	1	0	1	0	<b>3</b>
285	Pulau Rakyat	1	1	0	1	0	<b>3</b>
286	Aek Loba	4	4	0	4	0	<b>2</b>
287	Aek Ledong	0	0	0	0	0	<b>3</b>
288	Sei Kepayang	0	0	0	0	1	<b>3</b>
289	Sei Kepayang Barat	0	0	0	0	0	<b>3</b>
290	Sei Apung	0	0	0	0	0	<b>3</b>
291	Bagan Asahan	1	1	0	1	0	<b>3</b>
292	Simpang Empat	0	0	0	0	1	<b>3</b>
293	Air Batu	3	3	0	3	0	<b>2</b>
294	Hessa Air Genting	4	4	0	4	2	<b>2</b>
295	Sei Dadap	0	0	0	0	1	<b>3</b>
296	Prapat Janji	1	1	0	1	0	<b>3</b>
297	Tinggi Raja	0	0	0	0	0	<b>3</b>
298	Setia Janji	2	2	0	2	0	<b>3</b>
299	Meranti	1	1	0	1	0	<b>3</b>
300	Rawang Pasar IV	0	0	0	0	1	<b>3</b>
301	Binjai Serbangun	0	0	0	0	1	<b>3</b>

<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
302	Sidodadi	0	0	0	0	2	3
303	Rs. Ibu Kartini	0	0	0	0	0	3
304	RSUD HAMS	0	0	0	0	0	3
305	Gambir Baru	0	0	0	0	1	3
306	Mutiara	5	5	1	6	0	2
307	Saribu Dolok	0	0	0	0	0	3
308	Pamatang Silimahuta	0	0	0	0	0	3
309	Tiga Runggu	2	0	0	0	0	3
310	Haranggaol	2	1	0	1	0	3
311	Sipintu Angin	4	0	0	0	1	3
312	Sarimatondang	0	0	0	0	1	3
313	Pamatang Sidamanik	0	0	0	0	0	3
314	Parapat	4	0	0	0	0	3
315	Tanah Jawa	3	0	0	0	0	3
316	Marubun Jaya	0	0	0	0	0	3
317	Hatonduhan	1	0	1	1	0	3
318	Buttu Turunan	1	0	0	0	0	3
319	Tiga Dolok	3	1	0	1	0	3
320	Tiga Balata	0	0	0	0	0	3
321	Panei Tongah	0	0	0	0	0	3
322	Panombean Pane	1	0	0	0	0	3
323	Pamatang Raya	2	0	0	0	1	3
324	Saran Padang	0	0	0	0	0	3
325	Cingkes	0	0	0	0	0	3
326	Nagori Dolok	1	0	0	0	0	3
327	Sinasih	1	0	0	0	0	3
328	Sindar Raya	0	0	0	0	0	3
329	Bah Tonang	0	0	0	0	0	3
330	Tapian Dolok	3	0	0	0	0	3

<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
331	Serbelawan	1	1	0	1	0	3
332	Silou Malaha	2	1	0	1	0	3
333	Batu Anam	3	0	0	0	0	3
334	Rambung Merah	1	0	0	0	0	3
335	Simpang Bah Jambi	0	0	0	0	0	3
336	Bandar Siantar	1	0	0	0	0	3
337	Gunung Maligas	2	0	0	0	1	3
338	Raja Maligas	0	0	0	0	0	3
339	Huta Bayu	3	0	0	2	1	3
340	Jawa Maraja Bah Jambi	0	0	0	0	0	3
341	Pematang Bandar	2	0	0	0	1	3
342	Kerasaan	1	0	0	0	1	3
343	Bandar Huluan	0	1	0	1	0	3
344	Perdagangan	3	0	0	0	1	3
345	Marihat Bandar	3	0	0	0	0	3
346	Bandar Masilam	1	0	0	0	0	3
347	Bandar Tinggi	0	0	0	0	0	3
348	Bosar Maligas	0	1	0	1	0	3
349	Parbutaran	0	0	0	0	0	3
350	Ujung Padang	3	1	0	1	0	3
351	Sayur Matinggi	2	0	0	0	0	3
352	Bah Bolon	0	0	0	0	0	3
353	Kabanjahe	2	2	0	2	0	3
354	Berastagi	0	0	0	0	0	3
355	Korpri	0	0	0	0	0	3
356	Tigapanah	1	2	0	2	0	3
357	Singa	0	0	0	0	0	3
358	Dolat Rayat	0	0	0	0	0	3
359	Merdeka	1	1	0	1	0	3
360	Merek	2	2	0	2	0	3

<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
361	Barusjahe	0	0	0	0	1	3
362	Simpang Empat	1	1	0	1	0	3
363	Naman Teran	0	0	0	0	0	3
364	Tiganderket	2	2	0	2	0	3
365	Payung	0	0	1	1	1	3
366	Munthe	2	2	0	2	2	2
367	Juhar	1	1	0	1	1	3
368	Tigabinanga	0	0	0	0	0	3
369	Kutabuluh	0	0	0	0	1	3
370	Laubaleng	1	1	1	2	0	3
371	Mardinding	0	0	0	0	0	3
372	Bahorok	3	3	0	3	0	2
373	Bukit Lawang	0	0	0	0	0	3
374	Serapit	0	0	0	0	0	3
375	Tanjung Langkat	1	1	1	2	0	3
376	Marike	1	1	0	1	0	3
377	Namu Ukur	2	2	0	2	1	2
378	Namu Trasi	0	0	1	2	0	3
379	Kuala	0	0	0	0	0	3
380	Selesai	1	1	0	1	1	3
381	Sambirejo	9	9	0	9	0	1
382	Stabat	0	0	0	0	0	3
383	Karang Rejo	0	0	0	0	0	3
384	Stabat Lama	4	4	0	4	0	2
385	Hinai Kiri	0	0	0	0	2	3
386	Desa Teluk	1	1	0	1	0	3
387	Secanggang	1	1	0	1	1	3
388	Tanjung Beringin	2	2	0	2	1	2
389	Tanjung Selamat	0	0	0	0	1	3
390	Sei Bamban	0	0	0	0	0	3
391	Sawit Seberang	2	2	0	2	0	3

<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
392	Pantai Cermin	4	4	0	3	4	<b>2</b>
393	Gebang	2	2	0	2	0	<b>3</b>
394	Securai	1	1	0	1	1	<b>3</b>
395	Pangkalan Brandan	0	0	0	0	0	<b>3</b>
396	Desa Lama	1	1	0	1	1	<b>3</b>
397	Tangkahan Durian	0	0	0	0	0	<b>3</b>
398	Pangkalan Susu	0	0	0	0	0	<b>3</b>
399	Beras Basah	0	0	0	0	0	<b>3</b>
400	Besitang	3	5	0	1	0	<b>2</b>
401	Pematang Jaya	1	1	0	1	0	<b>3</b>
402	Telukdalam	4	4	0	4	1	<b>2</b>
403	Onolalu	3	3	0	3	0	<b>2</b>
404	Luahagundre Maniamolo	1	1	0	1	0	<b>3</b>
405	Bawomataluo	1	1	0	1	0	<b>3</b>
406	Hilisataro	0	0	0	0	0	<b>3</b>
407	Hilisimaetano	1	1	0	1	0	<b>3</b>
408	Hilizalootano	1	1	0	1	0	<b>3</b>
409	Amandraya	4	4	0	4	0	<b>2</b>
410	Ulususua	1	1	0	1	0	<b>3</b>
411	Aramo	0	0	0	0	0	<b>3</b>
412	Lolowau	4	4	0	4	1	<b>2</b>
413	O'o'u	3	3	0	3	0	<b>2</b>
414	Onohazumba	2	2	0	2	0	<b>3</b>
415	Hilisalawahe	2	2	0	2	0	<b>3</b>
416	Hilimegai	0	0	0	0	0	<b>3</b>
417	Huruna	0	0	0	0	0	<b>3</b>
418	Lolomatua	2	2	0	2	0	<b>3</b>
419	Ulunoyo	1	1	0	1	0	<b>3</b>
420	Lahusa	3	3	0	3	1	<b>2</b>
421	Siduaori	0	0	0	0	0	<b>3</b>

<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
422	Somambawa	2	2	0	2	0	3
423	Gomo	2	2	0	2	0	3
424	Boronadu	0	0	0	0	0	3
425	Idanotae	0	0	0	0	0	3
426	Uluidanotae	0	0	0	0	0	3
427	Susua	0	0	0	0	0	3
428	Mazo	0	0	0	0	1	3
429	Umbunasi	0	0	0	0	0	3
430	Pulau Tello	0	0	0	0	0	3
431	P.P. Batu Barat	0	0	0	0	0	3
432	P.P. Batu Utara	0	0	0	0	0	3
433	Simuk	1	1	0	1	0	3
434	Hibala	0	0	0	0	0	3
435	Tanah Masa	0	0	0	0	0	3
436	Labuhan Hiu	0	0	0	0	0	3
437	Hilianombasela	1	1	0	1	0	3
438	Parlilitan	1	1	0	1	0	3
439	Hutagalung	1	1	0	1	0	3
440	Hutapaung	3	5	0	5	0	2
441	Baktiraja	1	1	0	1	0	3
442	Paranginan	4	5	0	5	0	2
443	Sigompul	3	3	0	3	0	2
444	Matiti	6	6	0	6	1	2
445	Saitnihuta	5	6	0	6	0	2
446	Bonandolok	0	0	0	0	0	3
447	Onanganjang	4	4	0	4	0	2
448	Pakkat	3	5	0	5	3	2
449	Tarabintang	3	3	0	3	4	2
450	Salak	1	0	0	0	0	3
451	Sukaramai	2	0	0	0	0	3
452	Sibande	0	0	0	0	0	3

<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
453	Tinada	0	0	0	0	0	3
454	Kecupak	1	0	2	2	0	3
455	Singgabur	0	1	0	1	0	3
456	Siempat Rube	3	0	0	0	0	3
457	Sibagindar	0	0	0	0	0	3
458	Perbaungan	7	1	0	8	4	2
459	Melati	1	1	0	2	1	3
460	Pantai Cermin	8	1	1	10	1	2
461	Pegajahan	0	0	0	0	1	3
462	Sialang Buah	1	4	0	5	0	2
463	Sei Rampah	2	1	0	3	2	2
464	PKL Budiman	3	0	0	3	0	3
465	T. Beringin	10	2	0	12	0	1
466	Desa Pon	1	0	0	1	0	3
467	Paya Lombang	6	0	0	6	0	2
468	N. Kesiangan	0	1	0	1	0	3
469	T. Syahbandar	3	1	0	1	0	3
470	Bdr. Kalipah	0	0	0	0	0	3
471	Dlk. Merawan	0	0	0	0	0	3
472	Dolok Masihul	1	5	0	6	0	2
473	Sipispis	0	1	0	1	0	3
474	Kuala Bali	3	2	1	6	0	3
475	Bintang Bayu	0	0	0	0	0	3
476	Kotarih	2	0	0	2	0	3
477	Silinda	2	2	0	4	0	3
478	Kotapinang	3	2	1	3	0	2
479	Batu Ajo	0	1	0	1	0	3
480	Sisumut	1	2	0	2	0	3
481	Mampang	1	1	0	1	0	3
482	Tanjung Medan	3	3	0	3	0	2
483	Pekan Tolan	0	0	0	0	0	3



<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
484	Teluk Panji	3	2	0	2	2	<b>2</b>
485	Hutagondang	1	1	0	1	0	<b>3</b>
486	Langga Payung	2	2	0	2	0	<b>3</b>
487	Aek Goti	1	1	0	1	0	<b>3</b>
488	Ulumahuam	0	1	0	1	0	<b>3</b>
489	Aek batu	2	1	0	1	0	<b>3</b>
490	Aek Raso	0	1	0	1	0	<b>3</b>
491	Beringin Jaya	0	1	0	1	0	<b>3</b>
492	Bunut	3	1	1	2	0	<b>3</b>
493	Cikampak	3	2	0	2	0	<b>3</b>
494	Rasau	1	1	0	1	0	<b>3</b>
495	Pintu Angin	1	0	0	0	0	<b>3</b>
496	Sambas	3	0	0	0	0	<b>3</b>
497	Pelabuhan Sambas	3	0	0	0	0	<b>3</b>
498	Aek Habil	0	0	0	0	2	<b>3</b>
499	Aek Parombunan	1	0	0	0	0	<b>3</b>
500	Datuk Bandar	0	0	0	0	0	<b>3</b>
501	Semula Jadi	0	0	0	0	2	<b>3</b>
502	Mu Damanik	0	2	0	2	1	<b>3</b>
503	RSUD dr.T. Mansyur	17	21	3	24	1	<b>1</b>
504	Kambung Baru	0	0	0	0	0	<b>3</b>
505	Kampung Persatuan	0	0	0	0	0	<b>3</b>
506	Seitualang Raso	0	2	0	2	0	<b>3</b>
507	Teluk Nibung	0	0	0	0	1	<b>3</b>
508	Siporipori	0	0	0	0	0	<b>3</b>
509	Rantau Laban	2	4	2	6	0	<b>2</b>
510	Tanjung Marulak	0	1	0	1	2	<b>3</b>
511	Sri Padang	1	2	0	2	0	<b>3</b>
512	Satria	4	3	1	4	0	<b>2</b>
513	Rambung	1	1	0	2	1	<b>3</b>

<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
514	Pasar Gambir	0	0	0	0	1	<b>3</b>
515	Pabatu	2	4	0	4	0	<b>2</b>
516	Teluk Karang	2	2	0	2	1	<b>2</b>
517	Berohol	3	4	1	4	2	<b>2</b>
518	Binjai Estate	3	3	0	3	1	<b>2</b>
519	Rambung	1	1	0	1	1	<b>3</b>
520	Binjai Kota	2	2	0	2	0	<b>3</b>
521	Tanah Tinggi	5	5	0	5	2	<b>2</b>
522	Kebun Lada	3	3	0	3	0	<b>2</b>
523	Jati Makmur	1	1	0	1	2	<b>2</b>
524	Bandar Senembah	0	0	0	0	0	<b>3</b>
525	H,A,H. Hasan	2	2	0	2	0	<b>3</b>
526	Pijorkoling	3	5	1	6	2	<b>2</b>
527	Labuhan Rasoki	0	2	0	2	0	<b>3</b>
528	Padangmatinggi	10	13	1	14	2	<b>1</b>
529	Sidangkal	11	11	0	11	0	<b>1</b>
530	Batunadua	7	7	0	7	1	<b>2</b>
531	Sadabuan	17	21	2	23	2	<b>1</b>
532	Hutaimbaru	6	7	0	7	1	<b>2</b>
533	Pokenijor	2	2	0	2	0	<b>3</b>
534	Pintu Langit	1	1	0	1	0	<b>3</b>
535	UPTD Puskesmas Kecamatan Gunungsitoli	12	3	0	15	0	<b>1</b>
536	UPTD Puskesmas Kecamatan Gunungsitoli Utara	4	0	0	4	0	<b>3</b>
537	UPTD Puskesmas Kecamatan Gunungsitoli Selatan	5	2	0	7	0	<b>2</b>

<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
539	UPTD Puskesmas Kecamatan Gunungsitoli Idanoi	5	2	0	7	1	<b>2</b>
540	UPTD Puskesmas Kecamatan Gunungsitoli Alo'oa	0	2	0	2	0	<b>3</b>
541	Sungai Lansek	2	4	0	4	0	<b>2</b>
542	Kamang	5	5	1	6	0	<b>2</b>
543	Air Amo	3	8	0	8	1	<b>2</b>
544	Tanjung Gadang	7	12	1	13	1	<b>1</b>
545	Sijunjung	3	9	1	10	2	<b>1</b>
546	Gambok	4	5	1	6	0	<b>2</b>
547	Lubuk Tarok	2	8	0	8	0	<b>2</b>
548	Muaro Bodi	2	4	0	4	0	<b>2</b>
549	Padang Sibusuk	1	3	0	3	0	<b>3</b>
550	Tanjung Ampalu	5	10	0	10	1	<b>1</b>
<b>Jumlah Anggota Cluster 1</b>							<b>32</b>
<b>Jumlah Anggota Cluster 2</b>							<b>161</b>
<b>Jumlah Anggota Cluster 3</b>							<b>357</b>
<b>Jumlah Anggota Seluruh Cluster</b>							<b>550</b>

## Lampiran 5. Pembagian Anggota *Cluster Fuzzy C-Means* pada Data Kesehatan Daerah

Lampiran 5. Pembagian Anggota *Cluster Fuzzy C-Means* pada Data Kesehatan Daerah

No.	Puskesmas	Kematian Neonatal (jiwa)	Kematian Bayi (jiwa)	Kematian Anak Balita (jiwa)	Kematian Balita (jiwa)	Kematian Ibu (jiwa)	Anggota Cluster ke-i
1	Singkil	2	5	0	5	4	1
2	Pulau Banyak	0	1	1	2	1	2
3	Singkil Utara	2	3	0	3	1	2
4	Gunung Meriah	9	10	1	11	3	1
5	Simpang Kanan	8	11	1	12	0	1
6	Danau Paris	0	0	0	0	2	2
7	Suro	2	2	0	2	0	2
8	Singkohor	0	0	2	2	0	1
9	Kuta Baharu	0	2	1	3	0	2
10	Kuala Baru	1	2	0	0	0	3
11	Pulau Banyak Barat	0	0	0	0	0	3
12	Blangkejeren	1	2	0	2	0	3
13	Labuhan Haji	1	2	0	2	1	2
14	Peulumat	0	1	0	1	0	3
15	Meukek	1	0	1	1	1	2
16	Drien Jalo	0	0	0	0	0	3
17	Sawang	5	7	0	7	0	1
18	Samadua	1	3	0	3	0	2
19	Tapaktuan	1	1	1	2	1	1
20	Lhok Bengkuang	0	0	0	0	1	3
21	Ladang Tuha	0	0	1	1	0	2
22	Uj. Padang Rasian	0	0	0	0	0	3
23	Kluet Utara	1	2	0	2	0	3
24	Kampung Raya	0	0	0	0	0	3
25	Kuala Ba <sup>TM</sup> u	3	4	1	5	0	1
26	Manggamat	1	1	0	1	0	3
27	Kluet Timur	2	2	0	2	1	2
28	Durian Kawan	0	0	0	0	0	3

<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
29	Kluet Selatan	1	2	1	3	0	<b>2</b>
30	Bakongan	0	0	0	0	0	<b>3</b>
31	Bukit Gadeng	0	0	0	0	0	<b>3</b>
32	Seubadeh	2	2	0	2	1	<b>2</b>
33	Trumon	3	4	0	4	1	<b>1</b>
34	Ladang Rimba	2	2	0	2	0	<b>2</b>
35	Krueng Luas	0	0	0	0	0	<b>3</b>
36	Linge	0	2	0	2	0	<b>3</b>
37	Atu Lintang	2	2	0	2	0	<b>2</b>
38	Jagong Jeget	1	1	0	1	0	<b>3</b>
39	Bintang	1	2	0	2	0	<b>3</b>
40	Lut Tawar	5	5	0	5	1	<b>1</b>
41	Kebayakan	4	5	0	5	0	<b>1</b>
42	Pegasing	5	5	0	5	0	<b>1</b>
43	Bies	3	3	0	3	1	<b>1</b>
44	Bebesen	15	19	0	19	1	<b>1</b>
45	Kute Panang	1	1	0	1	0	<b>3</b>
46	Silih Nara	4	4	0	4	0	<b>1</b>
47	Ketol	2	3	0	3	0	<b>2</b>
48	Celala	1	1	0	1	0	<b>3</b>
49	Rusip Antara	2	5	0	5	0	<b>1</b>
50	Johan Pahlawan	5	7	0	7	1	<b>1</b>
51	Suak Ribe	1	1	0	1	0	<b>3</b>
52	Meureubo	6	7	0	7	5	<b>1</b>
53	Peureumeu	7	13	1	14	0	<b>1</b>
54	Cot Seumeureng	1	1	0	1	0	<b>3</b>
55	Kuta Padang Layung	0	2	0	2	0	<b>3</b>
56	Kuala Bhee	2	3	1	4	2	<b>1</b>
57	Tangkeh	1	1	1	2	0	<b>2</b>
58	Drien Rampak	0	2	0	2	0	<b>3</b>

<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
59	Pasi Mali	2	3	0	3	0	<b>2</b>
60	Meutulang	2	3	0	3	1	<b>2</b>
61	Pante Ceureumen	2	3	0	3	0	<b>2</b>
62	Kajeng	2	5	0	5	0	<b>1</b>
63	Geumpang	1	1	0	1	0	<b>3</b>
64	Menee	0	0	1	1	0	<b>2</b>
65	Glp.Tiga	1	1	0	1	1	<b>2</b>
66	Teupin Raya	0	0	0	0	3	<b>1</b>
67	Glp.Baro	4	6	1	7	0	<b>1</b>
68	Mutiara Timur	2	2	0	2	1	<b>2</b>
69	Ujong Rimba	2	4	1	5	1	<b>1</b>
70	Mutiara Barat	3	4	1	5	0	<b>1</b>
71	Tiro	3	4	1	5	0	<b>1</b>
72	Tangse	6	7	0	7	2	<b>1</b>
73	Keumala	1	3	0	3	0	<b>2</b>
74	Titeu	0	0	0	0	0	<b>3</b>
75	Sakti	5	10	1	11	2	<b>1</b>
76	Mila	0	0	0	0	0	<b>3</b>
77	Padang Tiji	10	11	0	11	0	<b>1</b>
78	Delima	4	6	1	7	0	<b>1</b>
79	Reubee	1	1	0	1	0	<b>3</b>
80	Grong-Grong	6	6	1	7	0	<b>1</b>
81	Indrajaya	5	9	1	10	0	<b>1</b>
82	Peukan Baro	3	3	3	6	1	<b>1</b>
83	Kb.Tanjong	3	4	0	4	0	<b>2</b>
84	Simpang Tiga	2	4	0	4	1	<b>1</b>
85	Kota Sigli	9	9	0	9	1	<b>1</b>
86	Pidie	11	14	4	18	2	<b>1</b>
87	Batee	2	2	0	2	0	<b>2</b>
88	Muara Tiga	5	5	0	5	0	<b>1</b>
89	Samalanga	5	6	0	6	1	<b>1</b>

<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
90	Simpang Mamplam	6	6	0	6	1	<b>1</b>
91	Pandrah	2	2	0	2	0	<b>2</b>
92	Jeunieb	11	13	0	13	0	<b>1</b>
93	Peulimbang	7	7	1	8	0	<b>1</b>
94	Peudada	3	3	1	4	0	<b>1</b>
95	Jeumpa	6	12	1	13	1	<b>1</b>
96	Kuala	5	5	0	5	3	<b>1</b>
97	Juli	8	11	0	11	0	<b>1</b>
98	Juli 2	2	3	0	3	0	<b>2</b>
99	Kota Juang	10	12	0	12	0	<b>1</b>
100	Jangka	1	1	0	1	1	<b>2</b>
101	Peusangan Selatan	3	4	0	4	1	<b>1</b>
102	Peusangan Siblih	3	5	0	5	1	<b>1</b>
103	Peusangan	4	5	0	5	1	<b>1</b>
104	Kuta Blang	5	5	0	5	0	<b>1</b>
105	Makmur	0	2	1	3	0	<b>2</b>
106	Gandapura	4	6	2	8	0	<b>1</b>
107	Blangpidie	3	2	0	0	2	<b>1</b>
108	Tangan-Tangan	3	4	2	0	0	<b>1</b>
109	Bineh Krueng	0	3	0	0	0	<b>3</b>
110	Manggeng	3	3	0	0	0	<b>2</b>
111	Susoh	4	5	0	0	0	<b>2</b>
112	Sangkalan	1	1	0	0	0	<b>3</b>
113	Kuala Batee	2	0	0	0	0	<b>3</b>
114	Alue Pisang	0	0	0	0	1	<b>3</b>
115	Babahrot	0	0	0	0	0	<b>3</b>
116	Ie Mirah	2	1	0	0	0	<b>3</b>
117	Lembah Sabil	1	2	0	0	0	<b>3</b>
118	Lhang	0	1	0	0	0	<b>3</b>
119	A.S. Pinang	1	1	0	0	1	<b>2</b>
120	Kuta Panjang	1	1	0	1	0	<b>3</b>

<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
121	Blang Jerango	0	1	0	1	0	3
122	Blangkejeren	0	0	0	0	0	3
123	Gumpang	0	0	0	0	0	3
124	Badak	1	3	0	0	0	3
125	Cinta Maju	0	0	0	0	0	3
126	Pinturime	0	0	0	0	0	3
127	Pining	0	0	0	0	1	3
128	Rikit Gaib	2	4	0	4	0	2
129	Keneyaran	0	0	0	0	0	3
130	Terangun	1	3	0	0	0	3
131	Rerebe	1	3	0	0	0	3
132	Minyak Payed	2	2	0	2	1	2
133	Bendahara	2	3	0	3	0	2
134	Sungai Iyu	1	1	0	1	0	3
135	Seruway	10	13	0	13	1	1
136	Karang Baru	13	13	0	13	0	1
137	Kualasimpang	3	4	0	4	1	1
138	Kejuruhan Muda	6	7	0	7	0	1
139	Rantau	4	6	0	6	1	1
140	Sapta Jaya	7	8	0	8	0	1
141	Tamiang Hulu	3	5	0	5	1	1
142	Sekerak	2	2	0	2	0	2
143	Banda Mulia	1	1	0	1	0	3
144	Tenggulun	3	5	0	5	0	1
145	Bandar Pusaka	4	5	0	5	0	1
146	Beutong Ateuh	0	0	0	0	0	3
147	Beutong	2	4	0	4	0	2
148	Uteun Pulo	2	4	2	6	1	1
149	Jeuram	5	7	0	7	0	1
150	Cot Kuta	2	2	0	2	1	2
151	Ujong Patihah	0	1	0	1	1	2



<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
152	Padang Panyang	0	0	0	0	0	<b>3</b>
153	Padang Rubek	1	3	0	3	0	<b>2</b>
154	Simpang Jaya	1	1	0	1	1	<b>2</b>
155	Kuala Tadu	3	3	0	3	0	<b>2</b>
156	Alue Bilie	2	2	0	2	1	<b>2</b>
157	Suka Mulia	2	2	0	2	0	<b>2</b>
158	Alue Rambot	0	0	0	0	0	<b>3</b>
159	Lueng Kebeu Jagat	3	4	0	4	0	<b>2</b>
160	Lamno	2	3	0	3	0	<b>2</b>
161	Indra Jaya	1	1	1	2	3	<b>1</b>
162	Lhok Kruet	2	2	0	2	0	<b>2</b>
163	Patek	4	6	0	6	0	<b>1</b>
164	Lageun	3	3	0	3	0	<b>2</b>
165	Calang	1	3	0	3	0	<b>2</b>
166	Krueng Sabee	3	3	0	3	0	<b>2</b>
167	Panga	5	5	0	5	0	<b>1</b>
168	Teunom	0	0	0	0	1	<b>3</b>
169	Pasie Raya	4	5	0	5	0	<b>1</b>
170	Belang Rakal	1	0	0	0	0	<b>3</b>
171	Singah Mulo	1	1	0	1	0	<b>3</b>
172	Ronga Ronga	1	1	2	3	1	<b>1</b>
173	Lampahan	2	4	0	4	1	<b>1</b>
174	Pante Raya	4	5	1	6	0	<b>1</b>
175	Simpang Teritit	4	5	1	6	0	<b>1</b>
176	Simpang Tiga	9	10	0	10	2	<b>1</b>
177	Bandar	6	8	0	8	1	<b>1</b>
178	Bener Kalipah	2	2	1	3	1	<b>1</b>
179	Buntul Kemumu	1	2	0	2	0	<b>3</b>
180	Ramung	0	1	0	1	0	<b>3</b>
181	Samar Kilang	0	0	1	1	0	<b>2</b>
182	Mesidah	3	3	0	3	1	<b>1</b>

<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
183	Bandar Baru	3	0	0	0	0	<b>3</b>
184	Cubo	0	2	0	1	1	<b>2</b>
185	Panteraja	0	1	0	0	0	<b>3</b>
186	Trienggadeng	0	1	0	0	0	<b>3</b>
187	Meureudu	4	2	0	1	0	<b>2</b>
188	Meurah Dua	4	0	0	0	2	<b>1</b>
189	Ulim	2	1	0	0	0	<b>3</b>
190	Jangka Buya	2	0	0	0	0	<b>3</b>
191	Bandar Dua	2	0	0	1	0	<b>3</b>
192	Kuta Krueng	0	0	0	0	0	<b>3</b>
193	Blangkuta	2	3	0	2	0	<b>2</b>
194	Meuraxa	1	2	0	2	0	<b>3</b>
195	Jaya Baru	3	3	0	3	1	<b>1</b>
196	Banda Raya	0	0	0	0	0	<b>3</b>
197	Baiturrahman	1	2	0	2	1	<b>2</b>
198	Batoh	0	0	0	0	0	<b>3</b>
199	Kuta Alam	0	0	0	0	0	<b>3</b>
200	Lampulo	0	0	0	0	0	<b>3</b>
201	Lampaseh	0	0	0	0	0	<b>3</b>
202	Kopelma Darussalam	2	2	0	2	0	<b>2</b>
203	Jeulingke	0	0	0	0	0	<b>3</b>
204	Ulee Kareng	2	3	0	3	0	<b>2</b>
205	UPTD PKM Langsa Barat	7	9	1	10	1	<b>1</b>
206	UPTD PKM Langsa Baro	1	1	1	2	0	<b>2</b>
207	UPTD PKM Langsa Lama	2	4	5	9	0	<b>1</b>
208	UPTD PKM Langsa Kota	3	5	0	5	1	<b>1</b>
209	UPTD PKM Langsa Timur	8	0	0	10	3	<b>1</b>

<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
210	Banda Sakti	5	6	0	6	3	<b>1</b>
211	Mongeudong	2	2	0	2	1	<b>2</b>
212	Muara Dua	3	3	0	3	0	<b>2</b>
213	Muara Satu	3	3	0	3	0	<b>2</b>
214	Blang Mangat	2	2	0	2	2	<b>1</b>
215	Blang Cut	1	1	0	1	0	<b>3</b>
216	Hiliduh	1	2	2	0	0	<b>1</b>
217	Botombawo	0	0	0	0	1	<b>3</b>
218	Botomuzoi	1	0	0	0	0	<b>3</b>
219	Hiliweto Gido	7	2	0	0	2	<b>1</b>
220	Somolomolo	0	0	0	0	0	<b>3</b>
221	Maâ€™u	2	1	0	0	2	<b>1</b>
222	Idanogawo	0	0	0	0	0	<b>3</b>
223	Ulugawo	0	0	0	0	3	<b>1</b>
224	Bawolato	1	1	0	0	0	<b>3</b>
225	Sogaâ€™adu	0	0	0	0	0	<b>3</b>
226	Pulo Pakkat	0	0	1	1	1	<b>2</b>
227	Sibabangun	0	0	0	0	0	<b>3</b>
228	Pinangsori	1	0	0	1	1	<b>2</b>
229	Hutabalang	0	0	0	0	0	<b>3</b>
230	Pandan	0	0	0	0	0	<b>3</b>
231	Kalangan	1	0	0	1	0	<b>3</b>
232	Tukka	0	0	0	0	0	<b>3</b>
233	Sarudik	2	0	0	2	0	<b>3</b>
234	Poriaha	2	0	0	2	2	<b>1</b>
235	Aek Raisan	2	0	0	2	0	<b>3</b>
236	Kolang	3	2	1	6	1	<b>1</b>
237	Sorkam	3	0	0	3	0	<b>2</b>
238	Gonting Mahe	0	0	0	0	0	<b>3</b>
239	Sipea-Pea	3	0	0	3	1	<b>2</b>
240	Pasaributobing	0	0	0	0	0	<b>3</b>

<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
241	Siantar Ca	0	0	0	0	0	3
242	Barus	0	0	0	0	0	3
243	Andam Dewi	0	0	0	0	0	3
244	Sirandorung	1	0	0	1	2	2
245	Manduamas	0	0	1	1	0	2
246	Saragih	0	0	0	0	0	3
247	Lumut	0	0	0	0	0	3
248	Barus Utara	0	0	0	0	0	3
249	Parmonangan	0	1	2	3	0	1
250	Aek Raja	2	2	0	2	0	2
251	Parsingkaman	3	3	0	3	1	1
252	Sitadatada	3	3	1	4	0	1
253	Situmeanghabinsara n	3	6	0	6	0	1
254	Hutabaginda	3	5	0	5	1	1
255	Siatar Barita	3	3	0	3	0	2
256	Onan Hasang	3	4	0	4	0	2
257	Sarulla	3	3	0	3	0	2
258	Janji Angkola	0	0	1	1	0	2
259	Simangbumban	0	0	0	0	1	3
260	Lumban Sinaga	1	1	0	1	1	2
261	Pangaribuan	3	4	0	4	0	2
262	Garoga	1	1	1	2	0	2
263	Sipahutar	2	4	0	4	2	1
264	Siborongborong	5	5	0	5	0	1
265	Silangit	0	0	0	0	0	3
266	Butar	2	3	0	3	2	1
267	Muara	1	3	0	3	0	2
268	Suka Makmur	1	1	0	1	0	3
269	Janji	5	0	0	0	0	2
270	Rantauprapat	3	11	0	11	2	1

<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
271	Perlayuan	2	2	0	2	1	<b>2</b>
272	Sigambal	7	7	1	8	2	<b>1</b>
273	Perbaungan	3	3	0	3	1	<b>1</b>
274	Lingga Tiga	2	2	0	2	3	<b>1</b>
275	Gunung Selamat	2	2	0	2	0	<b>2</b>
276	Pangkatan	3	4	0	4	0	<b>2</b>
277	Negeri Lama	9	6	0	6	2	<b>1</b>
278	Tanjung Haloban	2	3	0	3	0	<b>2</b>
279	Teluk Sentosa	7	4	0	4	0	<b>1</b>
280	Labuhan Bilik	4	2	0	2	1	<b>1</b>
281	Sei Berombang	5	3	0	3	2	<b>1</b>
282	Sei Penggantungan	0	0	0	0	1	<b>3</b>
283	Bp. Mandoge	5	5	0	5	1	<b>1</b>
284	Aek Songsongan	1	1	0	1	0	<b>3</b>
285	Pulau Rakyat	1	1	0	1	0	<b>3</b>
286	Aek Loba	4	4	0	4	0	<b>1</b>
287	Aek Ledong	0	0	0	0	0	<b>3</b>
288	Sei Kepayang	0	0	0	0	1	<b>3</b>
289	Sei Kepayang Barat	0	0	0	0	0	<b>3</b>
290	Sei Apung	0	0	0	0	0	<b>3</b>
291	Bagan Asahan	1	1	0	1	0	<b>3</b>
292	Simpang Empat	0	0	0	0	1	<b>3</b>
293	Air Batu	3	3	0	3	0	<b>2</b>
294	Hessa Air Genting	4	4	0	4	2	<b>1</b>
295	Sei Dadap	0	0	0	0	1	<b>3</b>
296	Prapat Janji	1	1	0	1	0	<b>3</b>
297	Tinggi Raja	0	0	0	0	0	<b>3</b>
298	Setia Janji	2	2	0	2	0	<b>2</b>
299	Meranti	1	1	0	1	0	<b>3</b>
300	Rawang Pasar IV	0	0	0	0	1	<b>3</b>
301	Binjai Serbangun	0	0	0	0	1	<b>3</b>

<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
302	Sidodadi	0	0	0	0	2	<b>2</b>
303	Rs. Ibu Kartini	0	0	0	0	0	<b>3</b>
304	RSUD HAMS	0	0	0	0	0	<b>3</b>
305	Gambir Baru	0	0	0	0	1	<b>3</b>
306	Mutiara	5	5	1	6	0	<b>1</b>
307	Saribu Dolok	0	0	0	0	0	<b>3</b>
308	Pamatang Silimahuta	0	0	0	0	0	<b>3</b>
309	Tiga Runggu	2	0	0	0	0	<b>3</b>
310	Haranggaol	2	1	0	1	0	<b>3</b>
311	Sipintu Angin	4	0	0	0	1	<b>2</b>
312	Sarimatondang	0	0	0	0	1	<b>3</b>
313	Pamatang Sidamanik	0	0	0	0	0	<b>3</b>
314	Parapat	4	0	0	0	0	<b>2</b>
315	Tanah Jawa	3	0	0	0	0	<b>3</b>
316	Marubun Jaya	0	0	0	0	0	<b>3</b>
317	Hatonduhan	1	0	1	1	0	<b>2</b>
318	Buttu Turunan	1	0	0	0	0	<b>3</b>
319	Tiga Dolok	3	1	0	1	0	<b>2</b>
320	Tiga Balata	0	0	0	0	0	<b>3</b>
321	Panei Tongah	0	0	0	0	0	<b>3</b>
322	Panombean Pane	1	0	0	0	0	<b>3</b>
323	Pamatang Raya	2	0	0	0	1	<b>2</b>
324	Saran Padang	0	0	0	0	0	<b>3</b>
325	Cingkes	0	0	0	0	0	<b>3</b>
326	Nagori Dolok	1	0	0	0	0	<b>3</b>
327	Sinasih	1	0	0	0	0	<b>3</b>
328	Sindar Raya	0	0	0	0	0	<b>3</b>
329	Bah Tonang	0	0	0	0	0	<b>3</b>
330	Tapian Dolok	3	0	0	0	0	<b>3</b>

<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
331	Serbelawan	1	1	0	1	0	3
332	Silou Malaha	2	1	0	1	0	3
333	Batu Anam	3	0	0	0	0	3
334	Rambung Merah	1	0	0	0	0	3
335	Simpang Bah Jambi	0	0	0	0	0	3
336	Bandar Siantar	1	0	0	0	0	3
337	Gunung Maligas	2	0	0	0	1	2
338	Raja Maligas	0	0	0	0	0	3
339	Huta Bayu	3	0	0	2	1	2
340	Jawa Maraja Bah Jambi	0	0	0	0	0	3
341	Pematang Bandar	2	0	0	0	1	2
342	Kerasaan	1	0	0	0	1	3
343	Bandar Huluan	0	1	0	1	0	3
344	Perdagangan	3	0	0	0	1	2
345	Marihat Bandar	3	0	0	0	0	3
346	Bandar Masilam	1	0	0	0	0	3
347	Bandar Tinggi	0	0	0	0	0	3
348	Bosar Maligas	0	1	0	1	0	3
349	Parbutaran	0	0	0	0	0	3
350	Ujung Padang	3	1	0	1	0	2
351	Sayur Matinggi	2	0	0	0	0	3
352	Bah Bolon	0	0	0	0	0	3
353	Kabanjahe	2	2	0	2	0	2
354	Berastagi	0	0	0	0	0	3
355	Korpri	0	0	0	0	0	3
356	Tigapanah	1	2	0	2	0	3
357	Singa	0	0	0	0	0	3
358	Dolat Rayat	0	0	0	0	0	3
359	Merdeka	1	1	0	1	0	3
360	Merek	2	2	0	2	0	2

<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
361	Barusjahe	0	0	0	0	1	3
362	Simpang Empat	1	1	0	1	0	3
363	Naman Teran	0	0	0	0	0	3
364	Tiganderket	2	2	0	2	0	2
365	Payung	0	0	1	1	1	2
366	Munthe	2	2	0	2	2	1
367	Juhar	1	1	0	1	1	2
368	Tigabinanga	0	0	0	0	0	3
369	Kutabuluh	0	0	0	0	1	3
370	Laubaleng	1	1	1	2	0	2
371	Mardinding	0	0	0	0	0	3
372	Bahorok	3	3	0	3	0	2
373	Bukit Lawang	0	0	0	0	0	3
374	Serapit	0	0	0	0	0	3
375	Tanjung Langkat	1	1	1	2	0	2
376	Marike	1	1	0	1	0	3
377	Namu Ukur	2	2	0	2	1	2
378	Namu Trasi	0	0	1	2	0	2
379	Kuala	0	0	0	0	0	3
380	Selesai	1	1	0	1	1	2
381	Sambirejo	9	9	0	9	0	1
382	Stabat	0	0	0	0	0	3
383	Karang Rejo	0	0	0	0	0	3
384	Stabat Lama	4	4	0	4	0	1
385	Hinai Kiri	0	0	0	0	2	2
386	Desa Teluk	1	1	0	1	0	3
387	Secanggang	1	1	0	1	1	2
388	Tanjung Beringin	2	2	0	2	1	2
389	Tanjung Selamat	0	0	0	0	1	3
390	Sei Bamban	0	0	0	0	0	3
391	Sawit Seberang	2	2	0	2	0	2



<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
392	Pantai Cermin	4	4	0	3	4	<b>1</b>
393	Gebang	2	2	0	2	0	<b>2</b>
394	Securai	1	1	0	1	1	<b>2</b>
395	Pangkalan Brandan	0	0	0	0	0	<b>3</b>
396	Desa Lama	1	1	0	1	1	<b>2</b>
397	Tangkahan Durian	0	0	0	0	0	<b>3</b>
398	Pangkalan Susu	0	0	0	0	0	<b>3</b>
399	Beras Basah	0	0	0	0	0	<b>3</b>
400	Besitang	3	5	0	1	0	<b>2</b>
401	Pematang Jaya	1	1	0	1	0	<b>3</b>
402	Telukdalam	4	4	0	4	1	<b>1</b>
403	Onolalu	3	3	0	3	0	<b>2</b>
404	Luahagundre Maniamolo	1	1	0	1	0	<b>3</b>
405	Bawomataluo	1	1	0	1	0	<b>3</b>
406	Hilisataro	0	0	0	0	0	<b>3</b>
407	Hilisimaetano	1	1	0	1	0	<b>3</b>
408	Hilizalootano	1	1	0	1	0	<b>3</b>
409	Amandraya	4	4	0	4	0	<b>1</b>
410	Ulususua	1	1	0	1	0	<b>3</b>
411	Aramo	0	0	0	0	0	<b>3</b>
412	Lolowau	4	4	0	4	1	<b>1</b>
413	O'o'u	3	3	0	3	0	<b>2</b>
414	Onohazumba	2	2	0	2	0	<b>2</b>
415	Hilisalawahe	2	2	0	2	0	<b>2</b>
416	Hilimegai	0	0	0	0	0	<b>3</b>
417	Huruna	0	0	0	0	0	<b>3</b>
418	Lolomatua	2	2	0	2	0	<b>2</b>
419	Ulunoyo	1	1	0	1	0	<b>3</b>
420	Lahusa	3	3	0	3	1	<b>1</b>
421	Siduaori	0	0	0	0	0	<b>3</b>

<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
422	Somambawa	2	2	0	2	0	<b>2</b>
423	Gomo	2	2	0	2	0	<b>2</b>
424	Boronadu	0	0	0	0	0	<b>3</b>
425	Idanotae	0	0	0	0	0	<b>3</b>
426	Uluidanotae	0	0	0	0	0	<b>3</b>
427	Susua	0	0	0	0	0	<b>3</b>
428	Mazo	0	0	0	0	1	<b>3</b>
429	Umbunasi	0	0	0	0	0	<b>3</b>
430	Pulau Tello	0	0	0	0	0	<b>3</b>
431	P.P. Batu Barat	0	0	0	0	0	<b>3</b>
432	P.P. Batu Utara	0	0	0	0	0	<b>3</b>
433	Simuk	1	1	0	1	0	<b>3</b>
434	Hibala	0	0	0	0	0	<b>3</b>
435	Tanah Masa	0	0	0	0	0	<b>3</b>
436	Labuhan Hiu	0	0	0	0	0	<b>3</b>
437	Hilianombasela	1	1	0	1	0	<b>3</b>
438	Parlilitan	1	1	0	1	0	<b>3</b>
439	Hutagalung	1	1	0	1	0	<b>1</b>
440	Hutapaung	3	5	0	5	0	<b>3</b>
441	Baktiraja	1	1	0	1	0	<b>1</b>
442	Paranginan	4	5	0	5	0	<b>2</b>
443	Sigompul	3	3	0	3	0	<b>2</b>
444	Matiti	6	6	0	6	1	<b>1</b>
445	Saitnihuta	5	6	0	6	0	<b>1</b>
446	Bonandolok	0	0	0	0	0	<b>3</b>
447	Onanganjang	4	4	0	4	0	<b>1</b>
448	Pakkat	3	5	0	5	3	<b>1</b>
449	Tarabintang	3	3	0	3	4	<b>1</b>
450	Salak	1	0	0	0	0	<b>3</b>
451	Sukaramai	2	0	0	0	0	<b>3</b>
452	Sibande	0	0	0	0	0	<b>3</b>

<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
453	Tinada	0	0	0	0	0	<b>3</b>
454	Kecupak	1	0	2	2	0	<b>1</b>
455	Singgabur	0	1	0	1	0	<b>3</b>
456	Siempat Rube	3	0	0	0	0	<b>3</b>
457	Sibagindar	0	0	0	0	0	<b>3</b>
458	Perbaungan	7	1	0	8	4	<b>1</b>
459	Melati	1	1	0	2	1	<b>2</b>
460	Pantai Cermin	8	1	1	10	1	<b>1</b>
461	Pegajahan	0	0	0	0	1	<b>3</b>
462	Sialang Buah	1	4	0	5	0	<b>2</b>
463	Sei Rampah	2	1	0	3	2	<b>1</b>
464	PKL Budiman	3	0	0	3	0	<b>2</b>
465	T. Beringin	10	2	0	12	0	<b>1</b>
466	Desa Pon	1	0	0	1	0	<b>3</b>
467	Paya Lombang	6	0	0	6	0	<b>1</b>
468	N. Kesiangan	0	1	0	1	0	<b>3</b>
469	T. Syahbandar	3	1	0	1	0	<b>2</b>
470	Bdr. Kalipah	0	0	0	0	0	<b>3</b>
471	Dlk. Merawan	0	0	0	0	0	<b>3</b>
472	Dolok Masihul	1	5	0	6	0	<b>1</b>
473	Sipispis	0	1	0	1	0	<b>3</b>
474	Kuala Bali	3	2	1	6	0	<b>1</b>
475	Bintang Bayu	0	0	0	0	0	<b>3</b>
476	Kotarih	2	0	0	2	0	<b>3</b>
477	Silinda	2	2	0	4	0	<b>2</b>
478	Kotapinang	3	2	1	3	0	<b>1</b>
479	Batu Ajo	0	1	0	1	0	<b>3</b>
480	Sisumut	1	2	0	2	0	<b>3</b>
481	Mampang	1	1	0	1	0	<b>3</b>
482	Tanjung Medan	3	3	0	3	0	<b>2</b>
483	Pekan Tolan	0	0	0	0	0	<b>3</b>

<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
484	Teluk Panji	3	2	0	2	2	<b>1</b>
485	Hutagondang	1	1	0	1	0	<b>3</b>
486	Langga Payung	2	2	0	2	0	<b>2</b>
487	Aek Goti	1	1	0	1	0	<b>3</b>
488	Ulumahuam	0	1	0	1	0	<b>3</b>
489	Aek batu	2	1	0	1	0	<b>3</b>
490	Aek Raso	0	1	0	1	0	<b>3</b>
491	Beringin Jaya	0	1	0	1	0	<b>3</b>
492	Bunut	3	1	1	2	0	<b>2</b>
493	Cikampak	3	2	0	2	0	<b>2</b>
494	Rasau	1	1	0	1	0	<b>3</b>
495	Pintu Angin	1	0	0	0	0	<b>3</b>
496	Sambas	3	0	0	0	0	<b>3</b>
497	Pelabuhan Sambas	3	0	0	0	0	<b>3</b>
498	Aek Habil	0	0	0	0	2	<b>2</b>
499	Aek Parombunan	1	0	0	0	0	<b>3</b>
500	Datuk Bandar	0	0	0	0	0	<b>3</b>
501	Semula Jadi	0	0	0	0	2	<b>2</b>
502	Mu Damanik	0	2	0	2	1	<b>2</b>
503	RSUD dr.T. Mansyur	17	21	3	24	1	<b>1</b>
504	Kambung Baru	0	0	0	0	0	<b>3</b>
505	Kampung Persatuan	0	0	0	0	0	<b>3</b>
506	Seitualang Raso	0	2	0	2	0	<b>3</b>
507	Teluk Nibung	0	0	0	0	1	<b>3</b>
508	Siporipori	0	0	0	0	0	<b>3</b>
509	Rantau Laban	2	4	2	6	0	<b>1</b>
510	Tanjung Marulak	0	1	0	1	2	<b>2</b>
511	Sri Padang	1	2	0	2	0	<b>3</b>
512	Satria	4	3	1	4	0	<b>1</b>
513	Rambung	1	1	0	2	1	<b>2</b>

<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
514	Pasar Gambir	0	0	0	0	1	<b>3</b>
515	Pabatu	2	4	0	4	0	<b>2</b>
516	Teluk Karang	2	2	0	2	1	<b>2</b>
517	Berohol	3	4	1	4	2	<b>1</b>
518	Binjai Estate	3	3	0	3	1	<b>1</b>
519	Rambung	1	1	0	1	1	<b>2</b>
520	Binjai Kota	2	2	0	2	0	<b>2</b>
521	Tanah Tinggi	5	5	0	5	2	<b>1</b>
522	Kebun Lada	3	3	0	3	0	<b>2</b>
523	Jati Makmur	1	1	0	1	2	<b>1</b>
524	Bandar Senembah	0	0	0	0	0	<b>3</b>
525	H,A,H. Hasan	2	2	0	2	0	<b>2</b>
526	Pijorkoling	3	5	1	6	2	<b>1</b>
527	Labuhan Rasoki	0	2	0	2	0	<b>3</b>
528	Padangmatinggi	10	13	1	14	2	<b>1</b>
529	Sidangkal	11	11	0	11	0	<b>1</b>
530	Batunadua	7	7	0	7	1	<b>1</b>
531	Sadabuan	17	21	2	23	2	<b>1</b>
532	Hutaimbaru	6	7	0	7	1	<b>1</b>
533	Pokenijor	2	2	0	2	0	<b>2</b>
534	Pintu Langit	1	1	0	1	0	<b>3</b>
535	UPTD Puskesmas Kecamatan Gunungsitoli	12	3	0	15	0	<b>1</b>
536	UPTD Puskesmas Kecamatan Gunungsitoli Utara	4	0	0	4	0	<b>2</b>
537	UPTD Puskesmas Kecamatan Gunungsitoli Selatan	5	2	0	7	0	<b>1</b>

<b>No.</b>	<b>Puskesmas</b>	<b>Kematian Neonatal (jiwa)</b>	<b>Kematian Bayi (jiwa)</b>	<b>Kematian Anak Balita (jiwa)</b>	<b>Kematian Balita (jiwa)</b>	<b>Kematian Ibu (jiwa)</b>	<b>Anggota Cluster ke-i</b>
538	UPTD Puskesmas Kecamatan Gunungsitoli Barat	3	0	0	3	0	<b>2</b>
539	UPTD Puskesmas Kecamatan Gunungsitoli Idanoi	5	2	0	7	1	<b>1</b>
540	UPTD Puskesmas Kecamatan Gunungsitoli Alo'oa	0	2	0	2	0	<b>3</b>
541	Sungai Lansek	2	4	0	4	0	<b>2</b>
542	Kamang	5	5	1	6	0	<b>1</b>
543	Air Amo	3	8	0	8	1	<b>1</b>
544	Tanjung Gadang	7	12	1	13	1	<b>1</b>
545	Sijunjung	3	9	1	10	2	<b>1</b>
546	Gambok	4	5	1	6	0	<b>1</b>
547	Lubuk Tarok	2	8	0	8	0	<b>1</b>
548	Muaro Bodi	2	4	0	4	0	<b>2</b>
549	Padang Sibusuk	1	3	0	3	0	<b>2</b>
550	Tanjung Ampalu	5	10	0	10	1	<b>1</b>
<b>Jumlah Anggota Cluster 1</b>							<b>157</b>
<b>Jumlah Anggota Cluster 2</b>							<b>148</b>
<b>Jumlah Anggota Cluster 3</b>							<b>245</b>
<b>Jumlah Anggota Seluruh Cluster</b>							<b>550</b>