

**PERBANDINGAN METODE *FUZZY C MEANS*,  
*FUZZY POSSIBILISTICS C MEANS* DAN  
*POSSIBILISTICS FUZZY C MEANS* PADA *NOISY DATA***  
(Studi Kasus: Fasilitas Tempat Tinggal Setiap Propinsi Tahun 2017)

**TUGAS AKHIR**



**Khusnul Hajar Nuansari**

**14 611 217**

**JURUSAN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS ISLAM INDONESIA  
YOGYAKARTA  
2018**

**PERBANDINGAN METODE *FUZZY C MEANS*,  
*FUZZY POSSIBILISTICS C MEANS* DAN  
*POSSIBILISTICS FUZZY C MEANS* PADA *NOISY DATA***  
(Studi Kasus: Fasilitas Tempat Tinggal Setiap Propinsi Tahun 2017)

**TUGAS AKHIR**

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana  
Jurusan Statistika



**Khusnul Hajar Nuansari**

**14 611 217**

**JURUSAN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS ISLAM INDONESIA  
YOGYAKARTA  
2018**

## HALAMAN PERSETUJUAN PEMBIMBING

### TUGAS AKHIR

Judul : Perbandingan Metode *Fuzzy C Means*, *Fuzzy Possibilistics C Means* dan *Possibilistics Fuzzy C Means* Pada *Noisy Data*

Nama : Khusnul Hajar Nuansari

No Mahasiswa : 14 611 217

TUGAS AKHIR INI TELAH DIPERIKSA DAN DISETUJUI UNTUK  
DIUJIKAN

Yogyakarta, 2 Mei 2018

Pembimbing



(Kariyam, S.Si., M.Si)

## HALAMAN PENGESAHAN

### TUGAS AKHIR

#### PERBANDINGAN METODE *FUZZY C MEANS*, *FUZZY POSSIBILISTICS C MEANS* DAN *POSSIBILISTICS FUZZY C MEANS* PADA *NOISY DATA*

Nama Mahasiswa : Khusnul Hajar Nuansari

No Mahasiswa : 14611217

TUGAS AKHIR INI UNTUK DIUJIKAN  
PADA TANGGAL 5 JUNI 2018

Nama Penguji

Tanda Tangan

1. Dr. Kartiko, M.Si

2. M. Hasan Sidiq K, M.Sc

3. Kariyam, S.Si., M.Si

Mengetahui,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

(Drs. Allywar, M.Sc., Ph. D.)



## KATA PENGANTAR



*Assalamu'alaikum Wr.Wb*

*Alhamdulillah* "aalamiin, puji syukur kehadiran Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya selama melaksanakan Tugas Akhir sehingga dapat terselesaikan. Shalawat serta salam tercurah kepada Nabi Muhammad SAW beserta keluarga dan para pengikut-pengikutnya. Tugas akhir yang berjudul "Perbandingan Metode *Fuzzy C Means*, *Fuzzy Possibilistics C Means* dan *Possibilistics Fuzzy C Means* Pada *Noisy Data* (Studi Kasus: Fasilitas Tempat Tinggal Setiap Propinsi Tahun 2017)" ini sebagai salah satu syarat untuk memperoleh gelar sarjana Jurusan Statistika di Universitas Islam Indonesia. Selama menulis Tugas Akhir, penulis telah banyak mendapat bimbingan dan bantuan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis bermaksud menyampaikan ucapan terima kasih kepada:

1. Bapak Drs. Allwar, M.Sc.,Ph.D. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam beserta jajarannya.
2. Bapak Dr. RB Fajriya Hakim, S.Si., M.Si. selaku Ketua Prodi Statistika yang telah banyak memberikan do'a, dukungan, dan masukan kepada penulis.
3. Ibu Kariyam, S.Si., M.Si selaku Dosen Pembimbing penulis yang selalu bersedia meluangkan waktunya untuk membimbing, memberikan arahan dan saran dalam penyusunan Tugas Akhir.
4. Dosen-dosen Statistika Universitas Islam Indonesia yang telah mendidik dan memberikan ilmunya kepada penulis.
5. Kedua orang tua tercinta, Bapak Joko Wiyono dan Ibu Suryati yang selalu memberikan semangat, doa, dan dukungan.
6. Kedua adik penulis, Faidhatul Achsani Ima Rahma Wati dan Fahlul Rozzaqi Aqsanu Robbiy yang selalu mendukung dan memberi semangat.

7. Teman-teman satu bimbingan tugas akhir (bimbingan Ibu Kariyam), Zarina, Eta, Evi, Winda, Wilda, Andre, Roro, Mia, Indri, Dian, Rina, Nur, Avit, Lia, Antisa dan Nayla yang selalu berbagi ilmu, berbagi cerita, dan pengalaman.
8. Sahabat dan teman seperjuangan, Dwi, Rabi, Nisa, Dhea, Ditia, Maulida, Reny, dan Ayu yang selalu berbagi ilmu dan pengalaman.
9. Sahabat Statistika 2014, sahabat KKN Unit 117, keluarga besar Ikatan Keluarga Statistika (IKS), keluarga besar Pondok Pesantren Universitas Islam Indonesia (PP UII) angkatan 2015, keluarga besar Jamaah Al-Ghuroba yang sudah banyak memberikan semangat dan dukungan selama penulisan tugas akhir ini.
10. Semua pihak yang telah membantu yang tidak bisa penulis sebutkan satu persatu. Semoga Allah SWT selalu memberikan rahmat dan anugerah-Nya kepada mereka semua.

Penulis menyadari bahwa tugas akhir ini masih jauh dari sempurna. Oleh karena itu segala kritik dan saran yang bersifat membangun selalu penulis harapkan. Semoga tugas akhir ini dapat bermanfaat bagi penulis khususnya dan bagi semua yang membutuhkan. Akhir kata, semoga Allah SWT selalu melimpahkan rahmat serta hidayah-Nya kepada kita semua. Amiin

***Wassalamu'alaikum Warahmatullaahi Wabarakaatuh.***

Yogyakarta, 2 Mei 2018

Penulis

## DAFTAR ISI

<b>HALAMAN JUDUL</b> .....	i
<b>HALAMAN PERSETUJUAN PEMBIMBING TUGAS AKHIR</b> .....	ii
<b>HALAMAN PENGESAHAN</b> .....	iii
<b>KATA PENGANTAR</b> .....	iv
<b>DAFTAR ISI</b> .....	vi
<b>DAFTAR TABEL</b> .....	viii
<b>DAFTAR GAMBAR</b> .....	ix
<b>DAFTAR LAMPIRAN</b> .....	x
<b>PERNYATAAN</b> .....	xi
<b>INTISARI</b> .....	xii
<b>ABSTRACT</b> .....	xiii
<b>BAB I PENDAHULUAN</b> .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	2
1.3 Batasan Masalah .....	2
1.4 Jenis Penelitian dan Metode Analisis .....	3
1.5 Tujuan Penelitian .....	3
1.6 Manfaat Penelitian .....	3
<b>BAB II TINJAUAN PUSTAKA</b> .....	4
<b>BAB III LANDASAN TEORI</b> .....	13
3.1 Data Mining .....	13
3.2 <i>Clustering</i> .....	14
3.3 <i>Fuzzy C Means</i> .....	15
3.4 <i>Fuzzy Possibilistics C Means</i> .....	17
3.5 <i>Possibilistics Fuzzy C Means</i> .....	18
3.6 <i>Noisy Data</i> .....	20
3.7 <i>Validitas Cluster</i> .....	21
3.8 <i>Kruskall Wallis</i> .....	22
3.9 Sumber Air Minum Layak .....	23

3.10 Sanitasi Layak.....	24
3.10 Penerangan Dari Listrik.....	25
<b>BAB IV METODOLOGI PENELITIAN .....</b>	<b>26</b>
4.1 Studi Kasus .....	26
4.2 Metode Penelitian .....	26
<b>BAB V PEMBAHASAN .....</b>	<b>28</b>
5.1 Statistka Deskriptif .....	28
5.2 <i>Noisy Data</i> .....	29
5.3 Perbandingan <i>Fuzzy C Means</i> , <i>Fuzzy Possibilistics C Means</i> dan <i>Possibilistics Fuzzy Cmeans</i> .....	29
5.4 Penentuan Jumlah <i>Cluster</i> Optimal .....	32
5.5 Profilisasi <i>Fuzzy Possibilistics C Means</i> Terhadap <i>Hasil Cluster</i> .....	32
3.6 Uji <i>Kruskall Wallis</i> .....	34
<b>BAB VI PENUTUP .....</b>	<b>38</b>
5.1 Kesimpulan.....	38
5.2 Saran .....	38
<b>DAFTAR PUSTAKA .....</b>	<b>39</b>
<b>LAMPIRAN.....</b>	<b>42</b>



## DAFTAR TABEL

<b>Tabel 2.1</b>	Rangkuman Penelitian Terdahulu .....	7
<b>Tabel 3.1</b>	Penjelasan Variabel .....	26
<b>Tabel 5.1</b>	Statistika Deskriptif .....	28
<b>Tabel 5.2</b>	Perbandingan Waktu Komputasi .....	30
<b>Tabel 5.3</b>	Perbandingan Jumlah Iterasi dan Fungsi Objektif Minimum .....	45
<b>Tabel 5.4</b>	Penentuan Metode Terbaik .....	31
<b>Tabel 5.5</b>	Penentuan Jumlah <i>Cluster</i> .....	32
<b>Tabel 5.6</b>	Profilisasi <i>Fuzzy Possibilistics C Means</i> Terhadap Hasi <i>Cluster</i> .....	32
<b>Tabel 5.7</b>	Hasil <i>Cluster</i> .....	33
<b>Tabel 5.8</b>	Uji <i>Kruskall Wallis</i> .....	34
<b>Tabel 5.9</b>	Selisih <i>Mean Ranks</i> .....	35
<b>Tabel 5.10</b>	Uji Perbandingan Ganda .....	36

## DAFTAR GAMBAR

<b>Gambar 4.1</b> Diagram Alir.....	27
<b>Gambar 5.1</b> <i>Boxplot</i> .....	29

## DAFTAR LAMPIRAN

<b>Lampiran 1</b> Data .....	43
<b>Lampiran 2</b> Statistika Deskriptif .....	45
<b>Lampiran 3</b> Hasil Analisis FCM, FPCM dan PFCM .....	46
<b>Lampiran 4</b> Uji <i>Kruskall Wallis</i> .....	71

## PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang sebelumnya pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 8 Juni 2018



Penulis

# **PERBANDINGAN METODE *FUZZY C MEANS*, *FUZZY POSSIBILISTICS C MEANS* DAN *POSSIBILISTICS FUZZY C MEANS* PADA *NOISY DATA***

(Studi Kasus: Fasilitas Tempat Tinggal Setiap Propinsi Tahun 2017)

Khusnul Hajar Nuansari

Program Studi Statistika Fakultas MIPA

Universitas Islam Indonesia

## **INTISARI**

Di era digital ini, data telah menjadi bagian yang sangat penting dalam peradaban manusia. Melimpahnya data saat ini akan menghasilkan suatu informasi yang mudah dimengerti, bermakna dan bermanfaat apabila dilakukan pengolahan terhadap data tersebut, termasuk didalamnya adalah pengolahan data error (*noisy data*). Salah satu kasus *noisy data* dalam bentuk outlier yang perlu diolah adalah data fasilitas tempat tinggal setiap propinsi tahun 2017. Data tersebut berasal dari website BPS. Dalam upaya mengatasi kasus *noisy data* yang berupa outlier pada data fasilitas tempat tinggal setiap propinsi tahun 2017, maka dapat dilakukan dengan menggunakan metode *clustering*. Metode *clustering* yang digunakan dalam penelitian ini adalah metode *Fuzzy C Means*, *Fuzzy Possibilistics C Means* dan *Possibilistics Fuzzy C Means*. Ketiga metode tersebut dibandingkan keakuratannya dalam mengelompokkan *noisy data* berdasarkan indeks *Modified Partition Coefficient* (MPC). Berdasarkan indeks *Modified Partition Coefficient* (MPC) didapatkan hasil bahwa metode *Fuzzy Possibilistics C Means* (FPCM) merupakan metode terbaik dibandingkan dengan metode *Fuzzy Possibilistics C Means* (FPCM) dan *Possibilistics Fuzzy C Means* (PFCM). Jumlah cluster yang optimal pada metode *Fuzzy Possibilistics C Means* (FPCM) ialah 5 cluster. Dari kelima cluster tersebut, cluster 4 merupakan cluster dengan fasilitas tempat tinggal paling layak dibandingkan dengan yang lainnya.

**Kata Kunci :** *Fuzzy C Means*, *Fuzzy Possibilistics C Means*, *Possibilistics Fuzzy C Means*, *Noisy Data*, *Modified Partition Coefficient*

**COMPARISON OF FUZZY C MEANS METHODS, FUZZY  
POSSIBILISTICS C MEANS AND POSSIBILISTICS FUZZY C MEANS  
FOR NOISY DATA**

**(Case Study: Residential Facilities of Each Province in 2017)**

Khusnul Hajar Nuansari

Statistics Department, Faculty of Mathematics and Natural Sciences  
Islamic University of Indonesia

**ABSTRACT**

In this digital age, data has become a very important part of human civilization. The abundance of current data will result in an information that is easy to understand, meaningful and useful if done processing of the data, including in it is data processing error (noisy data). One case of noisy data in the form of outliers that need to be processed is data of residence facilities every province in 2017. The data comes from the BPS website. In an effort to overcome the case of noisy data in the form of outlier data on residential facilities of each province in 2017, it can be done by using clustering method. The clustering method used in this research is Fuzzy C Means method, Fuzzy Possibilistics C Means and Possibilistics Fuzzy C Means. The three methods compare their accuracy in classifying the noisy data based on the Modified Partition Coefficient (MPC) index. Based on the Modified Partition Coefficient (MPC) index, it is found that Fuzzy Possibilistics C Means (FPCM) method is the best method compared to Fuzzy Possibilistics C Means (FPCM) and Possibilistics Fuzzy C Means (PFCM) method. The optimal number of clusters in the Fuzzy Possibilistics C Means (FPCM) method is 5 clusters. Of the five clusters, cluster 4 is a cluster with the most decent living facilities compared to others.

**Keyword:** Fuzzy C Means, Fuzzy Possibilistics C Means, Possibilistics Fuzzy C Means, Noisy Data, Modified Partition Coefficient

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Di era digital ini, data telah menjadi bagian yang sangat penting dalam peradaban manusia. Semua orang menghasilkan data bahkan untuk hal kecil sekalipun. Data dibuat dan berada dimana-mana, di berbagai sektor dan bidang kehidupan. Kumpulan dari berbagai macam data tersebut biasanya disimpan dalam suatu tempat yang namanya *database*. *Database* berfungsi sebagai tempat untuk menyimpan dan mengelompokkan data secara sistematis dan terorganisir sehingga lebih memudahkan dalam mengakses data-data, *maintenance* data, serta pengelolaan data-data yang berada dalam *database* tersebut.

Melimpahnya data saat ini akan menghasilkan suatu informasi yang mudah dimengerti, bermakna dan bermanfaat apabila dilakukan pengolahan terhadap data tersebut. Namun begitu, sering kita jumpai bahwa data yang digunakan mengandung *error (noisy)* seperti data yang memiliki tipe salah (data string di atribut numerik), data dengan nilai yang sangat berbeda dengan nilai data lainnya, data yang salah diinputkan, data *missing*, data *outlier* dan lain sebagainya yang mana akan berakibat pada hasil keputusan. Di sisi lain, terkadang data tersebut juga mengandung makna yang berarti. Salah satu kasus *noisy* data dalam bentuk *outlier* yang perlu diolah adalah data fasilitas tempat tinggal pada indikator perumahan.

Perumahan merupakan salah satu kebutuhan pokok manusia. Fungsi rumah sebagai tempat tinggal selayaknya memiliki bentuk fisik dan fasilitas yang memadai. Dengan demikian, rumah sebagai sarana pembinaan keluarga dapat memperoleh hasil yang maksimal yaitu tercapainya peningkatan kualitas sumber daya manusia. Salah satu indikator perumahan dan permukiman adalah fasilitas tempat tinggal seperti sumber air minum, sumber penerangan dan akses terhadap sanitasi. Untuk mengatasi kasus *noisy* data yang berupa data *outlier* pada data fasilitas tempat tinggal maka dapat dilakukan dengan menggunakan *clustering*.

*Clustering* merupakan merupakan proses pembagian data ke dalam suatu kelompok berdasarkan tingkat kesamaannya sehingga data-data yang memiliki kemiripan karakteristik akan berkumpul dalam satu kelompok. Sementara data-data yang memiliki perbedaan karakteristik, akan berkumpul dalam kelompok yang berbeda (Siyamto, 2017). Dalam kaitannya dengan noisy data khususnya *outlier*, metode clustering dapat digunakan sebagai *outlier detection*.

Salah satu jenis metode *clustering* ialah metode *partition clustering*. *Partition clustering* merupakan suatu metode penyusunan partisi dari objek ke dalam suatu *cluster*. Dengan kata lain, jumlah *cluster* dalam hal ini ditentukan terlebih dahulu.

Pada penelitian ini, analisis difokuskan pada metode *partition clustering* yakni *Fuzzy*. *Fuzzy clustering* merupakan suatu metode pengelompokkan dimana suatu data dapat berada pada lebih dari satu *cluster*. Metode fuzzy yang digunakan untuk membandingkan pengelompokkan *noisy* data ialah *Fuzzy C Means* (FCM), *Fuzzy Possibilistics C Means* (FPCM) dan *Possibilistics Fuzzy C Means* (PFCM). Ketiga metode tersebut akan dibandingkan keakurasiannya berdasarkan indeks validitas *Modified Partition Coefficient* (MPC) sehingga akan didapatkan metode yang tepat dalam pengelompokkan *noisy* data.

## 1.2 Rumusan Masalah

Berdasarkan penjelasan pada latar belakang, maka permasalahan yang dapat diidentifikasi dalam penelitian ini antara lain adalah:

1. Bagaimana hasil perbandingan indeks validitas *cluster* pada metode *Fuzzy C Means*, *Fuzzy Possibilistic C Means* dan *Possibilistic Fuzzy C Means* pada *noisy* data?
2. Bagaimana hasil profilisasi pada pengelompokkan *noisy* data?

## 1.3 Batasan Masalah

Batasan masalah dalam penelitian ini adalah:

1. Data yang digunakan dalam penelitian ini adalah *noisy* data yang berupa data *outlier*.



2. Sumber data yang digunakan dalam penelitian ini adalah data sekunder yang didapatkan dari *website* BPS.
3. Data yang digunakan adalah data fasilitas tempat tinggal setiap propinsi dengan variabel sumber air minum layak setiap, sumber penerangan dari listrik serta akses sanitasi layak.
4. Metode yang digunakan dalam penelitian ini adalah metode *Fuzzy C Means* (FCM), *Fuzzy Possibilistic C Means* (FPCM) dan *Possibilistic Fuzzy C Means* (PFCM).
5. Kinerja yang dilihat untuk *Fuzzy C Means*, *Fuzzy Possibilistic C Means* dan *Possibilistic Fuzzy C Means* adalah keakuratannya berdasarkan nilai indeks validitas *Modified Partition Coefficient* (MPC).
6. Data diolah dengan menggunakan bantuan *software* R, SPSS dan *Microsoft Excel*.

#### **1.4 Jenis Penelitian dan Metode Analisis**

Jenis penelitian pada tugas akhir ini adalah penelitian aplikatif. Metode analisis yang digunakan adalah membandingkan kinerja metode *Fuzzy C Means*, *Fuzzy Possibilistic C Means* dan *Possibilistic Fuzzy C Means* pada *noisy* data berdasarkan keakurasiannya ketika meng*cluster*.

#### **1.5 Tujuan Penelitian**

Tujuan penelitian tugas akhir ini adalah:

1. Mengetahui kinerja *Fuzzy C Means*, *Fuzzy Possibilistic C Means* dan *Possibilistic Fuzzy C Means* pada *noisy* data dari segi keakuratannya.
2. Mengetahui hasil profilisasi pada pengelompokkan *noisy* data

#### **1.6 Manfaat Penelitian**

Manfaat dari penelitian ini adalah mempermudah penyelesaian kasus untuk pengelompokkan *noisy* data dengan menggunakan metode *clustering*. Selain itu bisa dijadikan dasar dalam pengambilan keputusan dan kebijakan berdasarkan studi kasus yang digunakan.

## BAB II

### TINJAUAN PUSTAKA

Pada penelitian-penelitian sebelumnya, penelitian terkait metode *Fuzzy C Means* (FCM), *Fuzzy Possibilistic C Means* (FPCM) dan *Possibilistic Fuzzy C Means* (PFCM) telah banyak dilakukan. Diantara penelitian-penelitian terdahulu adalah penelitian yang dilakukan oleh Hyeong-Seog Kim (2011) dengan judul “*Pattern Classification of Typhoon Tracks Using the Fuzzy C Means Clustering Method*” didapatkan hasil bahwa metode *Fuzzy C Means* menghasilkan hasil yang baik untuk *dataset* seperti track TC yang terlalu kompleks.

Selanjutnya penelitian yang dilakukan oleh Christian Correa, Constantino Valero, Pilar Barreiro, Maria P. Diago, dan Javier Tardaguila (2014) dengan judul “*A Comparison of Fuzzy Clustering Algorithms Applied to Feature Extraction on Vineyard*” didapatkan hasil bahwa metode *Fuzzy C Means* (FCM) merupakan metode terbaik berdasarkan kecepatan dan performanya dibandingkan dengan *Possibilistic C-Means* (PCM), *Fuzzy Possibilistic CMeans* (FPCM), *Robust Fuzzy Possibilistic C-Means* (RFCM) dan *Fuzzy C-Means with Gustafson-Kessel algorithm* (FCM-GK).

Penelitian yang berjudul “*Kernel-Based Fuzzy and Possibilistic C Means Clustering*” oleh Dao-Qiang Zhang and Song-Can Chen (2003) didapatkan hasil bahwa metode *Kernel Fuzzy C Means* (KFCM) dan *Kernel Possibilistic C Means* (KPCM) memiliki ketahanan yang jauh lebih baik daripada *Fuzzy C Means* (FCM) dan *Possibilistic C Means* (PCM) ketika data *outlier*. Untuk yang sedikit data *outliernya*, metode *Fuzzy C Means* (FCM) dan *Kernel Fuzzy C Means* (KFCM) mempunyai hasil yang mewakili *dataset*.

Pada penelitian yang dilakukan oleh Hesam Izakian dan Ajith Abraham (2011) dengan judul “*Fuzzy C Means and Fuzzy Swarm for Fuzzy Clustering Problem*” didapatkan hasil bahwa penggabungan antara *Fuzzy C-Means* (FCM) dan *Fuzzy Particle Swarm Optimization* (FPSO) lebih efisien dan berkualitas dibandingkan dengan *Fuzzy C-Means* (FCM) dan *Fuzzy Particle Swarm*

*Optimization* (FPSO). Namun ketika jumlah data ditingkatkan, metode *Fuzzy C-Means* (FCM) lebih unggul dibandingkan metode *Fuzzy Particle Swarm Optimization* (FPSO).

Selanjutnya, penelitian dengan judul “*Image Segmentation Using PSO and PCM with Mahalanobis Distance*” yang dilakukan oleh Yong Zhang, Dan Huang, Min Ji dan Fuding Xie (2011) didapatkan hasil bahwa gambar dengan jarak mahalanobis dalam algoritma *Possibilistic C-Means* (PCM) yang kemudian dikombinasikan dengan pengoptimalan pusat pengelompokan awal dengan metode optimasi partikel swarm lebih baik daripada metode *Fuzzy C-Means* (FCM).

Pada penelitian yang dilakukan oleh S. Askari, N. Montazein, M. H. Fazal Zarandi dan E. Hakim (2017) dengan judul “*Generalized Entropy Based Possibilistic Fuzzy C-Means For Clustering Noisy Data and Its Convergence Proof*” didapatkan hasil bahwa nilai *error* pada metode *Generalized Entropy Based Possibilistic Fuzzy C Means* (GEPFCM) lebih dari 80% lebih kecil dari pada metode *Possibilistics Fuzzy C-Means* (PFCM).

Penelitian dengan judul “*Granular Fuzzy Possibilistic C Means Clustering Approach to DNA Microarray Problem*” oleh Hung Quoc Truong, Long Thanh Ngo dan Witold Pedrycz (2017) didapatkan hasil bahwa metode *Granular Fuzzy Possibilistic C-Means Clustering* (GrFPCM) lebih baik dibandingkan dengan metode *K Means*, *Fuzzy C Means* (FCM), dan *Fuzzy Possibilistic C Means* (FPCM)

Pada penelitian yang dilakukan oleh Cenk Budayan, Irem Dikmen, dan M. Talat Birgonul (2009) dengan judul “*Comparing the Performance of Traditional Cluster Analysis, Self-Organizing Maps and Fuzzy C-Means Method for Strategic Grouping*” didapatkan hasil bahwa metode *Self-Organizing Maps* (SOM) and *Fuzzy C Means* (FCM) lebih baik dalam menentukan tipologi kelompok sehingga dapat memberikan informasi yang lebih baik.

Selanjutnya berdasarkan penelitian yang dilakukan oleh Sueli A. Mingoti dan Joab O.Lima (2006) dengan judul “*Comparing SOM Neural Network with Fuzzy C-Means, K-Means and Traditional Hierarchical Clustering Algorithms*”

didapatkan hasil bahwa *Fuzzy C-Means* memiliki kinerja yang lebih baik dalam semua kasus bahkan di hadapan *outlier*.

Berdasarkan penelitian yang dilakukan oleh Ahmad Mauliyadi M, Hizir Sofyan dan Muhammad Subianto (2013) dengan judul “Perbandingan Metode *Fuzzy C-Means (FCM)* dan *Fuzzy Gustafson Kessel (FGK)* Menggunakan Data Citra Satelit Quickbird” didapatkan hasil bahwa Metode *Fuzzy Gustafson Kessel (FGK)* lebih akurat dibandingkan dengan metode *Fuzzy C-Means (FCM)* dengan akurasi sebesar 88%.

Pada penelitian yang dilakukan oleh Desy Rahmawati Ningrat, Di Asih I Maruddani dan Triastuti Wuryandari (2016) dengan judul “Analisis *Cluster* Dengan Algoritma *K-Means* dan *Fuzzy C Means Clustering* Untuk Pengelompokkan Data Obligasi Korporasi” didapatkan hasil bahwa berdasarkan kualitas ketepatan pengelompokkan rasio simpangan baku dalam *cluster* dan antar *cluster*, metode *K-Means* lebih tepat dibandingkan metode *Fuzzy C Means (FCM)*.

Selanjutnya berdasarkan penelitian dengan judul “Inisialisasi Pusat *Cluster* Menggunakan *Artificial Bee Colony* Pada Algoritma *Possibilistics Fuzzy C Means* Untuk Segmentasi Citra” oleh Amalia Nurani Basyrarah, Chastine Fatichah dan Darlis Heru Murti (2016) didapatkan hasil bahwa hasil segmentasi pada metode *Artificial Bee Colony* pada Algoritma *Possibilistics Fuzzy C Means (ABC-PFCM)* lebih baik daripada *Possibilistics Fuzzy C Means PFCM* berdasarkan nilai *Structural Similarity Index (SSIM)* yakni sebesar 0.9843.

Pada penelitian yang dilakukan oleh Baiq Nurul Haqiqi dan Robert Kurniawan (2015) dengan judul “Analisis Perbandingan Metode *Fuzzy C Means* dan *Subtractive Fuzzy C Means*” didapatkan hasil bahwa pada data sebanyak 20 metode *Subtractive Fuzzy C Means (SFCM)* memberikan hasil yang lebih baik daripada metode *Fuzzy C Means (FCM)*. Sementara pada data sebanyak 100, metode *Fuzzy C Means (FCM)* memberikan hasil yang lebih baik. Secara keseluruhan metode *Subtractive Fuzzy C Means (SFCM)* memberikan hasil yang lebih baik.

Selanjutnya pada tesis yang dilakukan oleh Arif Setiawan (2009) dengan judul “Klustering Skala Industri di Kabupaten Kudus Menggunakan *Fuzzy Possibilistics C Means* Standar” didapatkan hasil bahwa dari pengujian sistem yang dilakukan dengan metode *Fuzzy Possibilistics C Means* (FPCM) diperoleh kluster ideal dan setelah diimplementasikan ke data asli terbentuk kelompok industri berdasar skala yaitu besar, menengah dan kecil.

Selanjutnya pada penelitian yang dilakukan oleh S. Askari, N. Montazein, dan M. H. Fazel Zarandi (2017) dengan judul “*Generalized Possibilistic Fuzzy C-Means with Novel Cluster Validity Indices for Clustering Noisy Data*” didapatkan hasil bahwa Metode *Generalized Possibilistic Fuzzy C Means* (GPFM) lebih akurat dibandingkan dengan *Fuzzy C Means* (FCM), *Possibilistics C Means* (PCM), dan *Possibilistics Fuzzy C-Means* (PFCM) pada *noisy data*. Rangkuman penelitian terdahulu tersebut dapat disajikan dalam **Tabel 1**.

**Tabel 1** Rangkuman Penelitian Terdahulu

No	Nama/ Tahun	Metode	Judul Penelitian/Jur nal	Hasil Penelitian
1	Hyeong- Seog Kim (2011)	<i>Fuzzy C Means</i> (FCM)	<i>Pattern Classificaton of Typhoon Tracks Using the Fuzzy C Means Clustering Method/ Journal of Climate</i>	Metode FCM menghasilkan hasil klasifikasi yang lebih alami untuk dataset seperti track TC yang terlalu kompleks
2	Christian Correa, Constantin o Valero, Pilar Barreiro, Maria P. Diago, dan Javier Tardaguila (2014)	<i>Fuzzy C Means (FCM), Possibilistic C-Means (PCM), Fuzzy Possibilistic CMeans (FPCM), Robust Fuzzy Possibilistic C-Means</i>	<i>A Comparison of Fuzzy Clustering Algorithms Applied to Feature Extraction on Vineyard/ Research Gate</i>	Metode FCM merupakan metode terbaik antara kecepatan dan performa klasifikasi dibandingkan dengan keempat metode lainnya.

		(RFCM) dan <i>Fuzzy C-Means with Gustafson-Kessel algorithm</i> (FCM-GK)		
3	Dao-Qiang Zhang and Song-Can Chen (2003)	<i>Fuzzy C-Means</i> (FCM), <i>Possibilistic Fuzzy C Means</i> (PFCM), <i>Kernel Fuzzy C Means</i> (KFCM), dan <i>Kernel Possibilistic C Means</i> (KPCM)	<i>Kernel-Based Fuzzy And Possibilistic C-Means Clustering/ Journal Cite Seer</i>	KFCM dan KPCM memiliki ketahanan yang jauh lebih baik daripada FCM dan PCM ketika data <i>outlier</i>
4	Hesam Izakian dan Ajith Abraham (2011)	<i>Fuzzy C Means</i> (FCM), <i>Fuzzy Particle Swarm Optimization</i> (FPSO) dan gabungan FCM-FPSO	<i>Fuzzy C-Means And Fuzzy Swarm For Fuzzy Clustering Problem/ Journal Elsevier</i>	Hasil FCM-FPSO lebih efisien dan berkualitas dibandingkan dengan kedua metode lainnya. Namun ketika jumlah data ditingkatkan, metode FCM lebih unggul dibandingkan metode FPSO
5	Yong Zhang, Dan Huang, Min Ji dan Fuding Xie (2011)	<i>Fuzzy C Means</i> (FCM) dan Gabungan <i>Possibilistic C Means</i> (PCM) dan <i>Particle Swarm Optimization</i> (PSO)	<i>Image Segmentation Using PSO and PCM with Mahalanobis Distance/ Journal Elsevier</i>	Hasil gambar dengan jarak Mahalanobis dalam algoritma PCM yang kemudian dikombinasikan dengan pengoptimalan pusat pengelompokan

				awal dengan metode PSO lebih baik daripada metode FCM
6	S. Askari, N. Montazein, M. H. Fazel Zarandi dan E. Hakim (2017)	<i>Possibilistic Fuzzy C-Means (PFCM) dan Generalized Entropy Based Possibilistic Fuzzy C-Means (GEPFCM)</i>	<i>Generalized Entropy Based Possibilistic Fuzzy C-Means for Clustering Noisy Data and Its Convergence Proof/ Journal Elsevier</i>	Nilai error pada metode GEPFCM lebih dari 80% lebih kecil dari pada metode PFCM
7	Hung Quoc Truong, Long Thanh Ngo dan Witold Pedrycz (2017)	<i>K Means, Fuzzy C Means (FCM), Fuzzy Possibilistic C Means (FPCM) dan Granular Fuzzy Possibilistic C-Means (GrFPCM)</i>	<i>Granular Fuzzy Possibilistic CMeans Clustering Approach to DNA Microarray Problem/ Journal Elsevier</i>	Metode GrFPCM lebih baik dibandingkan ketiga metode lainnya.
8	Cenk Budayan, Irem Dikmen, dan M. Talat Birgonul (2009)	Metode tradisional, <i>Self-Organizing Maps (SOM)</i> dan <i>Fuzzy C-Means (FCM)</i>	<i>Comparing The Performance of Traditional Cluster Analysis, Self-Organizing Maps and Fuzzy C-Means Method For Strategic Grouping/ Journal Elsevier</i>	Metode SOM dan FCM lebih baik dalam menentukan tipologi kelompok sehingga dapat memberikan informasi yang lebih baik.
9	Sueli A.Mingoti Joab O.Lima (2006)	<i>SOM Neural Network, Fuzzy C Means, K Means</i> dan Metode	<i>Comparing SOM Neural Network with Fuzzy C Means, K Means and Traditional</i>	FCM memiliki kinerja yang lebih baik dalam semua kasus bahkan di hadapan <i>outlier</i>

		Tradisional	<i>Hierarchical Clustering Algorithms/ Jurnal Elsevier</i>	
10	Ahmad Mauliyadi M, Hizir Sofyan dan Muhammad Subianto (2013)	Metode <i>Fuzzy C-Means</i> (FCM) dan <i>Fuzzy Gustafson Kessel</i> (FGK)	Perbandingan Metode <i>Fuzzy C-Means</i> (FCM) dan <i>Fuzzy Gustafson Kessel</i> (FGK) Menggunakan Data Citra Satelit Quickbird/ Reserch Gate	Metode <i>Fuzzy Gustafson Kessel</i> (FGK) lebih akurat dibandingkan dengan metode FCM dengkn akurasi sebesar 88%
11	Desy Rahmawati Ningrat, Di Asih I Maruddani dan Triastuti Wuryandari (2016)	K-Means dan <i>Fuzzy C Means</i> (FCM)	Analisis <i>Cluster</i> Dengan Algoritma <i>K-Means</i> dan <i>Fuzzy C Means Clustering</i> Untuk Pengelompokkan Data Obligasi Korporasi / Jurnal Gaussian	Berdasarkan kualitas ketepatan pengelompokkan rasio simpangan baku dalam cluster dan antar cluster, metode K-Means lebih tepat dibandingkan metode FCM
12	Amalia Nurani Basyrarah, Chastine Faticah, Darlis Heru Murti (2016)	<i>Artificial Bee Colony</i> pada Algoritma <i>Possibilistics Fuzzy C Means</i> (ABC PFCM) dan <i>Possibilistics Fuzzy C Means</i> (PFCM)	Inisialisasi Pusat <i>Cluster</i> Menggunakan <i>Artificial Bee Colony</i> Pada Algoritma <i>Possibilistics Fuzzy C Means</i> Untuk Segmentasi Citra / Journal <i>Inspiration</i>	Hail segmentasi pada metode ABC-PFCM lebih baik daripada PFCM berdasarkan nilai SSIMnya yakni sebesar 0.9843.
13	Baiq Nurul Haqiqi, Robert Kurniawan (2015)	Metode <i>Fuzzy C Means</i> (FCM) dan <i>Subtractive Fuzzy C Means</i> (SFCM)	Analisis Perbandingan Metode <i>Fuzzy C Means</i> dan <i>Subtractive Fuzzy C Means</i> /Researh Gate	Pada data sebanyak 20 metode SFCM memberikan hasil yang lebih baik daripada metode FCM. Sementara pada data sebanyak



				100, metode FCM memberikan hasil yang lebih baik. Secara keseluruhan metode SFCM memberikan hasil yang lebih baik.
14	Arif Setiawan, 2009	<i>Fuzzy Possibilistics C Means (FPCM)</i>	Klustering Skala Industri di Kabupaten Kudus Menggunakan <i>Fuzzy Possibilistics C Means</i> Standar/ Tesis S2 UGM Ilmu Komputer	Dalam penelitian ini disimpulkan bahwa dari pengujian sistem yang dilakukan diperoleh kluster ideal dan setelah diimplementasikan ke data asli terbentuk kelompok industri berdasar skala yaitu besar, menengah dan kecil.
15	S. Askari, N. Montazein, dan M. H. Fazel Zarandi (2017)	<i>Fuzzy C-Means (FCM), Possibilistics C-Means (PCM), Possibilistics Fuzzy C-Means (PFCM) dan Generalized Possibilistic Fuzzy C-Means (GPFCM)</i>	<i>Generalized Possibilistic Fuzzy C-Means with Novel Cluster Validity Indices for Clustering Noisy Data/ Jurnal Elsevier</i>	Metode GPFCM lebih akurat dibandingkan dengan ketiga metode lainnya pada data noisy

Penelitian ini memiliki persamaan dengan penelitian sebelumnya yakni sama-sama menggunakan metode Fuzzy C Means (FCM), Fuzzy Possibilistics C Means (FPCM) dan Possibilistics Fuzzy C Means (PFCM) dalam menyelesaikan kasus.

Perbedaan penelitian ini dengan penelitian sebelumnya adalah studi kasus penelitian. Pada penelitian ini, studi kasus yang digunakan adalah *noisy* data yang

berupa data *outlier*. Data yang digunakan adalah data fasilitas perumahan setiap provinsi di Indonesia pada tahun 2017. Pada penelitian ini juga ditambahkan analisis kruskall wallis.

## **BAB III**

### **LANDASAN TEORI**

#### **3.1 Data mining**

Data mining itu adalah proses pencarian pola-pola yang tersembunyi berupa pengetahuan yang tidak diketahui sebelumnya dari suatu sekumpulan data dimana data tersebut dapat berada di dalam database, data warehouse, atau media penyimpanan informasi yang lain. Data *mining* merupakan proses semi otomatis yang menggunakan teknik matematika, statistik, *machine learning*, dan kecerdasan buatan guna mengambil dan mengidentifikasi informasi yang didapatkan dari sekumpulan jumlah data yang besar (Ochtario & Karmila, 2013).

Data mining sering disebut sebagai *knowledge discovery in database* (KDD). KDD adalah suatu kegiatan pengumpulan dan pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam data yang berukuran besar (Santosa, 2007). Informasi yang didapatkan dari proses data mining ini nantinya bisa dipakai dalam pengambilan keputusan.

Secara umum, fungsi data mining terbagi menjadi enam yakni fungsi deskripsi, fungsi prediksi, fungsi klasifikasi, fungsi estimasi, fungsi pengelompokkan dan fungsi asosiasi (Larose, 2005). Fungsi deskripsi guna memberi gambaran secara umum terkait data. Fungsi prediksi guna memperkirakan nilai masa mendatang. Fungsi klasifikasi guna menemukan model atau fungsi yang menjelaskan atau membedakan konsep dengan tujuan agar dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Fungsi estimasi hampir sama dengan fungsi klasifikasi namun estimasi lebih ke arah numerik daripada kategorik. Fungsi pengelompokkan guna mengelompokkan data yang mempunyai karakteristik yang sama. Serta fungsi asosiasi guna mengidentifikasi hubungan antar item (Susanto & Suryadi, 2010).

Beberapa teknik yang termasuk dalam data mining yaitu *clustering*, *association rule mining*, *classification*, *genetic algorithm*, *neural network* dan lain-lain.

### 3.2 Clustering

*Clustering* merupakan proses pembagian data ke dalam suatu kelompok berdasarkan tingkat kesamaannya. Definisi lain menyebutkan bahwa *clustering* merupakan pekerjaan yang memisahkan data atau vektor ke dalam beberapa kelompok berdasarkan karakteristik masing-masing. Data-data yang memiliki kemiripan karakteristik akan berkumpul dalam satu kelompok. Sementara data-data yang memiliki perbedaan karakteristik, akan berkumpul dalam kelompok yang berbeda (Siyamto, 2017).

Analisis *cluster* bertujuan untuk mengidentifikasi sekelompok obyek yang mempunyai kemiripan karakteristik tertentu yang dapat dipisahkan dengan kelompok obyek lainnya. Oleh karenanya obyek yang berada dalam kelompok yang sama relatif lebih homogen dibandingkan obyek yang berada pada kelompok yang berbeda. Prinsip pada analisis *cluster* adalah memaksimalkan homogenitas (kesamaan) dalam satu kelompok serta memaksimalkan heterogenitas (ketidaksamaan) antar kelompok (Johnson & Wichern, 1998).

*Clustering* merupakan salah satu metode dalam data *mining* yang menggunakan teknik *unsupervised learning*. *Unsupervised learning* adalah suatu pendekatan yang tidak memiliki data latih atau dengan kata lain tidak ada fase *learning* pada data tersebut (Kusuma & Nazori, 2015).

Adapun data yang dapat digunakan untuk analisis *cluster* adalah data yang berupa data ordinal, interval dan rasio. Set data obyek juga harus mempunyai peubah dengan tipe yang sejenis sehingga tidak bercampur antara tipe yang satu dengan tipe yang lainnya.

Berdasarkan strukturnya, *clustering* terbagi menjadi dua yakni hirarki dan partisi. Pada pengelompokan berbasis hirarki, satu data tunggal dapat menjadi suatu *cluster*. Dua atau lebih *cluster* juga dapat bergabung menjadi sebuah *cluster* besar dan seterusnya (Prasetyo, 2014). Sementara pengelompokan berbasis partisi merupakan suatu metode penyusunan partisi tunggal dataset dari objek ke dalam satu set *cluster*, sehingga akan terbentuk objek yang lebih mirip satu sama lain dalam satu *cluster* dibandingkan dengan *cluster* lain. Atau dengan kata lain

tidak terdapat tumpang tindih antara satu *cluster* dengan *cluster* yang lain (Prasetyo, 2014).

Dalam *clustering* menggunakan konsep partisi terdapat tiga konsep yang dapat digunakan, yakni partisi klasik, partisi fuzzy dan partisi *possibilistic*. Pada partisi klasik, suatu data secara eksklusif menjadi anggota hanya pada satu *cluster* saja. Pada partisi fuzzy, nilai keanggotaan suatu data pada suatu *cluster* terletak pada interval  $[0,1]$ . Jumlah derajat keanggotaan setiap data pada semua *cluster* adalah 1. Sementara pada partisi *possibilistics*, jumlah nilai keanggotaan suatu data pada semua *cluster* tidak harus 1. Namun begitu, untuk menjamin suatu data menjadi anggota dari paling tidak satu *cluster* maka diharuskan nilai keanggotaannya lebih dari 0 (Kusumadewi dkk. 2006).

### 3.3 Fuzzy C Means

*Fuzzy C Means* (FCM) merupakan salah satu jenis *clustering* non hirarki yang termasuk dalam teknik fuzzy *clustering*. Teknik ini pertama kali diperkenalkan oleh Jim Bezdek pada tahun 1981. *Fuzzy C Means* (FCM) merupakan suatu teknik pengelompokan data dimana keberadaan tiap-tiap titik data dalam suatu kelompok ditentukan oleh derajat keanggotaan data tersebut (Kusumadewi dkk. 2006). Pada *Fuzzy C Means* (FCM), data dapat menjadi anggota dari semua kelompok yang terbentuk dengan derajat keanggotaan yang berada antara 0 hingga 1 (Luthfi, 2017).

Pengelompokan data dengan metode *Fuzzy C Means* (FCM) dimulai dengan tebakan awal tentang pusat *cluster*. Tiap titik data akan diberi suatu derajat keanggotaan terhadap *cluster* tebakan. Selanjutnya dilakukan iterasi guna memperbaharui pusat-pusat *cluster* dan derajat keanggotaan tiap titik data hingga sedekat mungkin dengan pusat-pusat *cluster* yang benar (Naba, 2009). Perulangan iterasi ini didasarkan pada minimasi fungsi obyektif (Ned & Roger, 2000). Hal tersebut menggambarkan jarak dari titik data yang diberikan ke pusat *cluster* yang terbobot oleh derajat keanggotaan titik data tersebut. Fungsi objektif yang dilambangkan dengan  $P_t$  pada metode *Fuzzy C Means* (FCM) adalah:

$$P_t = \sum_{i=1}^n \sum_{k=1}^c ([\sum_{j=1}^m (X_{ij} - V_{kj})^2] (\mu_{ik})^w) \quad (3.1)$$

Adapun langkah- langkah *Fuzzy C Means* (FCM) adalah sebagai berikut:

1. Memasukkan data  $(X_{ij})$  yang akan dicluster ke dalam sebuah matriks, dimana matriks berukuran  $n \times m$ , dengan  $n$  adalah jumlah data yang akan dicluster dan  $m$  adalah jumlah atribut atau variabel.

$$X_{ij} = \begin{bmatrix} X_{11} & \dots & X_{1m} \\ \vdots & \ddots & \vdots \\ X_{n1} & \dots & X_{nm} \end{bmatrix} \quad (3.2)$$

Dengan  $i = 1, 2, \dots, n; j = 1, 2, \dots, m$

2. Menentukan
  - a. Jumlah cluster =  $c (\geq 2)$ ;
  - b. Pangkat/pembobot =  $w (> 1)$ ;
  - c. Maksimum iterasi = MaksIter;
  - d. Kriteria penghentian =  $\xi$ ;
  - e. Fungsi Objektif awal =  $P_0 = 0$ ;
  - f. Iterasi awal =  $t = 1$ ;
3. Membangkitkan bilangan acak sebagai elemen matriks partisi awal  $U$  atau matriks kekhasan relatif . Matriks kekhasan relatif dilambangkan dengan  $\mu_{ik}$ . Berikut adalah bentuk matriks partisi awal:

$$\mu_{ik} = \begin{bmatrix} \mu_{11}(X_1) & \dots & \mu_{1c}(X_1) \\ \vdots & \ddots & \vdots \\ \mu_{n1}(X_n) & \dots & \mu_{nc}(X_n) \end{bmatrix}; 1 \leq i \leq n; 1 \leq k \leq c \quad (3.3)$$

Jumlah setiap nilai elemen kolom dalam satu baris adalah 1.

4. Menghitung pusat cluster. Pusat cluster dilambangkan dengan  $V_{kj}$ .

$$V_{kj} = \frac{\sum_{i=1}^n (\mu_{ik})^w X_{ij}}{\sum_{i=1}^n (\mu_{ik})^w}; 1 \leq i \leq n; 1 \leq j \leq m; 1 \leq k \leq c \quad (3.4)$$

5. Menghitung perubahan derajat keanggotaan setiap data pada setiap cluster/ memperbaiki matriks kekhasan relatif ( $\mu_{ik}$ ) dengan:

$$\mu_{ik} = \frac{[\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-\frac{1}{w-1}}}{\sum_{k=1}^c [\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-\frac{1}{w-1}}}; 1 \leq i \leq n; 1 \leq j \leq m; 1 \leq k \leq c \quad (3.5)$$

7. Mengecek kondisi berhenti:
  - Jika :  $(|P_t - P_{t-1}| < \xi)$  atau  $(t > \text{MaksIter})$  maka berhenti;
  - Jika tidak :  $t = t + 1$ , ulangi langkah 4

### 3.4 Fuzzy Possibilistics C Means

*Fuzzy Possibilistics C Means* (FPCM) merupakan suatu algoritma pengembangan dari *Fuzzy C Means* (FCM) dan *Possibilistics C Means* (PCM). Pada algoritma FCM, nilai  $\mu_{ik}$  dipengaruhi oleh  $X_{ij}$  dan semua pusat *cluster*. Sementara pada algoritma PCM, nilai  $t_{ik}$  dipengaruhi oleh  $X_{ij}$ , pusat cluster ke  $k$  serta  $\gamma_k$  (Kusumadewi dkk. 2006). Adapun fungsi objektif pada metode *Fuzzy Possibilistic C-Means* (FPCM) ialah:

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left( \left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right] (\mu_{ik}^w + t_{ik}^\eta) \right) \quad (3.6)$$

Langkah- langkah *Fuzzy Possibilistic C Means* (FPCM) adalah:

1. Memasukkan data ( $X_{ij}$ ) yang akan *dicluster* ke dalam sebuah matriks, dimana matriks berukuran  $n \times m$ , dengan  $n$  adalah jumlah data yang akan *dicluster* dan  $m$  adalah jumlah atribut atau variabel.

$$X_{ij} = \begin{bmatrix} X_{11} & \dots & X_{1m} \\ \vdots & \ddots & \vdots \\ X_{n1} & \dots & X_{nm} \end{bmatrix}; 1 \leq i \leq n; 1 \leq j \leq m \quad (3.7)$$

2. Menentukan
  - a. Jumlah *cluster* =  $c$  ( $\geq 2$ );
  - b. Pangkat/ pembobot =  $w$  ( $>1$ ) dan  $\eta$  ( $>1$ )
  - c. Maksimum iterasi = MaksIter;
  - d. Error yang diharapkan =  $\xi$ ;
  - e. Fungsi Objektif awal =  $P_0 = 0$ ;
  - f. Iterasi awal =  $t = 1$ ;
3. Memanggil hasil akhir yang berupa matriks kekhasan relatif ( $\mu_{ik}$ ) dan pusat *cluster* ( $V_{kj}$ ) pada algoritma FCM untuk menghitung matriks kekhasan absolut ( $t_{ik}$ ). Matriks kekhasan absolut dilambangkan dengan dengan persamaan sebagai berikut:

$$t_{ik} = \begin{bmatrix} t_{11} & \dots & t_{1c} \\ \vdots & \ddots & \vdots \\ t_{n1} & \dots & t_{nc} \end{bmatrix} \quad (3.8)$$

Dimana elemen matriksnya ialah:

$$t_{ik} = \frac{[\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-1/\eta-1}}{\sum_{i=1}^n [\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-1/\eta-1}} \quad (3.9)$$

Dengan  $i = 1, 2, \dots, n$ ;  $k=1, 2, \dots, c$ ;  $j=1, 2, \dots, m$

4. Memperbaiki pusat *cluster* ( $V_{kj}$ ) dengan perhitungan sebagai berikut:

$$V_{kj} = \frac{\sum_{i=1}^n (\mu_{ik}^w + t_{ik}^\eta) X_{ij}}{\sum_{i=1}^n (\mu_{ik}^w + t_{ik}^\eta)} \quad (3.10)$$

Dengan  $i = 1, 2, \dots, n$ ;  $k=1, 2, \dots, c$ ;  $j=1, 2, \dots, m$

5. Memperbaiki matriks kekhasan relatif ( $\mu_{ik}$ ) dengan perhitungan sebagai berikut:

$$\mu_{ik} = \frac{[\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-1/w-1}}{\sum_{i=1}^c [\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-1/w-1}} \quad (3.11)$$

Dengan  $i = 1, 2, \dots, n$ ;  $k=1, 2, \dots, c$ ;  $j=1, 2, \dots, m$

6. Memperbaiki matriks kekhasan absolut ( $t_{ik}$ ) dengan perhitungan sebagai berikut:

$$t_{ik} = \frac{[\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-1/\eta-1}}{\sum_{i=1}^n [\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-1/\eta-1}} \quad (3.12)$$

7. Mengecek kondisi berhenti:

- Jika :  $(|P_t - P_{t-1}| < \zeta)$  atau  $(t > \text{MaksIter})$  maka berhenti ;
- Jika tidak :  $t = t + 1$ , ulangi langkah 4

### 3.5 Possibilistic Fuzzy C-Means

*Possibilistic Fuzzy C Means* (PFCM) merupakan pengembangan dari *Fuzzy Possibilistic C Means* yang menggabungkan algoritma FCM dan algoritma PCM (Pal, Kal, & Bezdek, 1997). Dengan kata lain yakni algoritma clustering yang menggabungkan nilai keanggotaan *Fuzzy C-Means* (FCM) dan nilai kesesuaian *Possibilistic C-Means* (PCM). Pengembangan metode *Fuzzy Possibilistics C Means* tersebut dilakukan dengan memberikan derajat kepentingan yang berbeda antara matriks kekhasan relatif dan matriks kekhasan absolut (Kusumadewi, Hartati, Harjoko, & Wardoyo, 2006). Adapun fungsi obyektif ialah:



$$P_t = \sum_{i=1}^n \sum_{k=1}^c (\alpha \mu_{ik}^w + b t_{ik}^\eta) \left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right] + \sum_{k=1}^c \gamma_k \sum_{i=1}^n (1 - t_{ik})^\eta \quad (3.13)$$

Langkah- langkah Possibilistic Fuzzy C-Means (PFCM)

1. Memasukkan data  $(X_{ij})$  yang akan *dicluster* ke dalam sebuah matriks, dimana matriks berukuran  $n \times m$ , dengan  $n$  adalah jumlah data yang akan *dicluster* dan  $m$  adalah jumlah atribut atau variabel.

$$X_{ij} = \begin{bmatrix} X_{11} & \dots & X_{1m} \\ \vdots & \ddots & \vdots \\ X_{n1} & \dots & X_{nm} \end{bmatrix}; 1 \leq i \leq n; 1 \leq j \leq m \quad (3.14)$$

2. Menentukan
  - a. Jumlah *cluster* =  $c \geq 2$ ;
  - b. Pangkat/pembobot untuk FCM ( $w$ ) =  $w > 1$ ;
  - c. Pangkat/pembobot untuk PCM ( $\eta$ ) =  $\eta > 1$ ;
  - d. Maksimum iterasi;
  - e. Kriteria pengehtian=  $\zeta$ ;
  - f. Iterasi awal  $t = 1$ ;
  - g. Koefisien pembobot untuk kekhasan relatif =  $a (> 0)$
  - h. Koefisien pembobot untuk kekhasan absolut =  $b (> 0)$
  - i. Koefisien  $\gamma = K$  (biasanya  $K=1$ )
3. Memanggil hasil akhir yang berupa matriks kekhasan relatif  $(\mu_{ik})$  dan pusat *cluster*  $(V_{kj})$  pada algoritma FCM yang bertujuan untuk mencari nilai matriks kekhasan absolut  $(t_{ik})$ , adapun persamaannya adalah:

$$t_{ik} = \begin{bmatrix} t_{11} & \dots & t_{1c} \\ \vdots & \ddots & \vdots \\ t_{n1} & \dots & t_{nc} \end{bmatrix}; 1 \leq i \leq n; 1 \leq k \leq c \quad (3.15)$$

Dimana elemen matriksnya sebagai berikut:

$$t_{ik} = \left[ 1 + \left( \frac{b \cdot \left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]}{\gamma_k} \right)^{1/(\eta-1)} \right]^{-1};$$

$$\text{Dengan } 1 \leq i \leq n; 1 \leq k \leq c; 1 \leq j \leq m \quad (3.16)$$

Nilai  $\gamma_k$  dan  $d_{ik}$  dapat dicari dengan persamaan sebagai berikut:

$$\gamma_k = K \frac{\sum_{i=1}^n ((\mu_{ik})^w * \left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right])}{\sum_{i=1}^n (\mu_{ik})^w} \quad (3.17)$$

Dengan  $1 \leq i \leq n; 1 \leq k \leq c; 1 \leq j \leq m$

4. Memperbaiki pusat *cluster* ( $V_{kj}$ ) dengan rumus sebagai berikut:

$$V_{kj} = \frac{\sum_{i=1}^n (\alpha \mu_{ik}^w + b t_{ik}^\eta) X_{ij}}{\sum_{i=1}^n (\alpha \mu_{ik}^w + b t_{ik}^\eta)}; 1 \leq i \leq n; 1 \leq k \leq c; 1 \leq j \leq m \quad (3.18)$$

5. Memperbaiki matriks kekhasan relatif ( $\mu_{ik}$ ) dengan persamaan sebagai berikut:

$$\mu_{ik} = \begin{bmatrix} \mu_{11} & \dots & \mu_{1c} \\ \vdots & \ddots & \vdots \\ \mu_{n1} & \dots & \mu_{nc} \end{bmatrix} \quad (3.18)$$

Dimana elemen matriksnya sebagai berikut:

$$\mu_{ik} = \frac{[\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-1/w-1}}{\sum_{i=1}^c [\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-1/w-1}}; 1 \leq i \leq n; 1 \leq k \leq c; 1 \leq j \leq m \quad (3.19)$$

Serta memperbaiki matriks kekhasan absolut ( $t_{ik}$ ) dengan persamaan sebagai berikut:

$$t_{ik} = \left[ 1 + \left( \frac{b \cdot [\sum_{j=1}^m (X_{ij} - V_{kj})^2]}{\gamma_k} \right)^{1/(\eta-1)} \right]^{-1} \quad (3.20)$$

Dengan  $1 \leq i \leq n; 1 \leq k \leq c; 1 \leq j \leq m$

Nilai  $\gamma_k$  dan  $d_{ik}$  dapat dicari dengan persamaan sebagai berikut:

$$\gamma_k = K \frac{\sum_{i=1}^n ((\mu_{ik})^w * [\sum_{j=1}^m (X_{ij} - V_{kj})^2])}{\sum_{i=1}^n (\mu_{ik})^w} \quad (3.21)$$

Dengan  $1 \leq i \leq n; 1 \leq k \leq c; 1 \leq j \leq m$

6. Menentukan kriteria berhenti:

- Jika :  $(|P_t - P_{t-1}| < \zeta)$  atau  $(t > \text{MaksIter})$  maka berhenti ;
- Jika tidak :  $t = t+1$ , ulangi langkah 4

### 3.6 Noisy Data

*Noisy data* merupakan data yang masih mengandung *error*. *Noisy data* juga sering disebut sebagai data korup. Namun begitu, maknanya telah diperluas yakni data tidak dapat dipahami dan ditafsirkan dengan benar oleh mesin, seperti teks tidak terstruktur. Beberapa contoh *noisy data* yang sangat umum adalah:

- a. Data memiliki tipe yang salah (data *string* di atribut numerik)

- b. Data memiliki nilai yang sangat berbeda dengan nilai data lainnya
- c. Data yang salah di tulis seperti 0.52 yang harusnya 0.25
- d. Data *missing*
- e. Dan lain sebagainya (Sunitha, Raju, & Srinivas, 2013)

*Outlier* kadang juga disebut sebagai *noise* (Prasetyo, 2014). *Outlier* dapat didefinisikan sebagai titik yang terletak sangat jauh dari rata-rata data lainnya. Atau dengan kata lain merupakan data yang secara nyata berbeda dengan data-data yang lain (Muhajir, 2017). Adakalanya memang data-data seperti ini dicari dan diolah karena memungkinkan menghasilkan suatu informasi tersendiri nantinya.

### 3.7 Validitas Cluster

Validitas cluster merupakan suatu cara untuk mengevaluasi kualitas dari *cluster* dan menentukan seberapa baik data diwakili oleh *cluster* tersebut. Pada metode fuzzy terdapat dua kategori untuk menghitung indeks validitas *cluster* yakni berdasarkan nilai-nilai keanggotaan serta menggunakan data itu sendiri. Indeks-indeks validasi yang hanya melibatkan nilai keanggotaan salah satunya adalah indeks *Modified Partition Coefficient* (MPC). *Modified Partition Coefficient* (MPC) merupakan indeks validitas hasil perbaikan dari metode *Partition Coefficient* (PC). Metode *Partition Coefficient* (PC) cenderung mengalami perubahan yang monoton terhadap nilai  $c$  (Xie, Hu, Luktarhan, & Zhao, 2011). Metode *Modified Partition Coefficient* (MPC) didefinisikan dengan persamaan:

$$MPC(c) = 1 - \frac{c}{c-1} (1 - PC(c)) \quad (3.22)$$

Adapun persamaan nilai indeks PC ini adalah:

$$PC(c) = \frac{1}{n} \sum_{k=1}^c \sum_{i=1}^n u_{ik}^2 \quad (3.23)$$

Dimana:

$n$  = banyak objek penelitian

$c$  = banyak kelompok

$u_{ik}$  = nilai keanggotaan objek ke- $i$  dengan pusat kelompok ke- $k$ .

Nilai MPC berkisar antara  $0 \leq MPC (c) \leq 1$ . *Cluster* dikatakan optimal jika nilai MPC mendekati 1 atau yang paling besar. Semakin dekat nilai MPC ke 0 maka semakin kabur keakuratannya.

### 3.8 Kruskal Wallis

Uji Kruskal Wallis merupakan uji non parametrik yang digunakan untuk menentukan adakah perbedaan yang signifikan pada dua lebih kelompok. Uji Kruskal Wallis identik dengan Uji *One Way Anova* pada pengujian parametris, sehingga dapat dikatakan bahwa uji ini merupakan alternatif dari uji *One Way Anova* ketika asumsi normalitas tidak terpenuhi. Selain sebagai uji alternatif.

Asumsi yang dibutuhkan dalam melakukan pengujian Kruskal Wallis antara lain sebagai berikut:

1. Variabel independen berskala kategorik lebih dari 2 kategori.
2. Variabel dependen berskala *numeric* (interval/rasio) atau skala ordinal.
3. Independen artinya sampel di tiap kategori harus bebas satu sama lain, yaitu tidak boleh ada sampel yang berada pada 2 kategori atau lebih.

Langkah-langkah perhitungan pada uji Kruskal Wallis adalah:

1. Seluruh data hasil pengamatan dari beberapa kelompok digabung, kemudian dibuat peringkat
2. Menghitung jumlah peringkat pada setiap kelompok
3. Menghitung nilai uji Kruskal Wallis dengan rumus:

$$KW = \left[ \frac{12}{n(n+1)} \sum_{k=1}^c \frac{R_k^2}{n_k} \right] - 3(n+1) \quad (3.24)$$

Dimana:

$n$  = jumlah objek penelitian

$R_k$  = jumlah peringkat pada kelompok ke-k

$n_k$  = jumlah sampel pada kelompok ke-k

4. Menentukan nilai  $\chi^2_{\alpha (k-1)}$  dari tabel *chi-square*
5. Mengambil kesimpulan dengan membandingkan nilai uji Kruskal Wallis dengan nilai tabel *chi-square*. Jika nilai Kruskal Wallis  $\geq \chi^2_{\alpha (k-1)}$  maka tolak  $H_0$ , sebaliknya  $H_0$  diterima (Supranto J. , 2009). Pengambilan

keputusan juga dapat didasarkan pada nilai *p-value* yang dibandingkan dengan  $\alpha$ . Jika *p-value*  $< \alpha$  maka tolak  $H_0$ , sebaliknya  $H_0$  diterima.

Ketika tolak  $H_0$  dilanjutkan dengan uji perbandingan ganda. Ketika membandingkan semua kemungkinan pasangan perlakuan pada taraf nyata  $\alpha$ , hasil dikatakan berbeda secara nyata jika:

$$|\bar{R}_k - \bar{R}_k| > Z_{\frac{\alpha}{c(c-1)}} \sqrt{\frac{n(n+1)}{12} \left( \frac{1}{n_k} + \frac{1}{n_k} \right)} \quad (3.25)$$

Dengan:

$\bar{R}_k$  = rata-rata ranking kelompok ke k

$c$  = banyak kelompok

$n$  = banyak objek penelitian

$n_k$  = banyak objek penelitian kelompok ke k

### 3.9 Sumber Air Minum Layak

Akses terhadap sumber air minum yang layak merupakan hak asasi manusia dan juga merupakan kebutuhan dasar hidup sehat. Sumber air minum yang tidak layak akan berakibat pada timbulnya penyakit diare, kolera, tipus, dan lain sebagainya yang mana dapat berakibat kepada kematian. Indikator ini digunakan untuk memantau akses penduduk terhadap sumber air berkualitas berdasarkan asumsi bahwa sumber air berkualitas menyediakan air yang aman untuk diminum bagi masyarakat karena air yang tidak berkualitas adalah penyebab langsung berbagai sumber penyakit. Adapun sumber air minum yang layak adalah sebagai berikut:

#### 1. Air Kemasan

Air kemasan merupakan air yang didistribusikan oleh suatu perusahaan dalam bentuk botol, gelas dan sebagainya. Misalnya ialah AQUA, VIT, Aguaria dan lain sebagainya.

#### 2. Ledeng Sampai Rumah

Air ledeng merupakan air yang dialirkan melalui saluran air hingga rumah melalui proses penjernihan dan penyehatan. Sumber air ini diusahakan oleh

PAM (Perusahaan Air Minum), PDAM (Perusahaan Daerah Air Minum) dan lain sebagainya.

3. Pompa

Pengambilan air tanah yang dilakukan dengan menggunakan pompa listrik, pompa tangan atau kincir angin.

4. Sumur

Pengambilan air tanah yang dilakukan dengan menggunakan ember atau gayung baik menggunakan perantara katrol maupun tidak. Air sumur dikategorikan menjadi dua yakni sumur terlindung dan sumur tidak terlindung. Sumur terlindung merupakan sumur yang memiliki lingkaran sumur berupa tembok dengan ukuran paling sedikit 0.8 m di atas tanah, 3 meter kebawah tanah dan 1 m lantai semen dari lingkaran sumur. Sementara sumur tidak terlindung merupakan sumur yang tidak memiliki lingkaran sumur berupa tembok serta lantai semen. Dari kedua jenis sumur tersebut yang menjadi sumber air minum yang layak adalah jenis sumur yang terlindung.

5. Mata air

Mata air merupakan sumber air permukaan tanah yang timbul dengan sendirinya. Mata air terbagi menjadi dua yakni mata air terlindung dan mata air tidak terlindung. Mata air terlindung merupakan mata air yang terlindung dari air bekas pakai, bekas mandi, bekas cuci dan lain sebagainya. Sementara mata air tak terlindung merupakan mata air yang tidak terlindung dari air bekas pakai, bekas mandi, bekas cuci dan lain sebagainya. Dari kedua jenis mata air tersebut yang menjadi sumber air minum yang layak adalah jenis mata air yang terlindung (BPS, 2013).

Namun begitu, perlu dicatat bahwa sumber air minum yang layak juga memiliki kemungkinan untuk dapat terkontaminasi dan menyebabkan air menjadi tidak aman untuk diminum.

Rumus yang digunakan untuk mengetahui persentase sumber air minum layak ialah banyaknya rumah tangga dengan sumber air minum layak dibagi dengan jumlah penduduk dikali seratus persen. Semakin besar persentase rumah tangga yang menggunakan sumber air minum layak menunjukkan semakin baiknya

kondisi rumah tangga di suatu daerah. Pada tahun 2016, rumah tangga yang memiliki akses sumber air minum layak sebesar 74.14% (Moeloek, 2017).

### **3.10 Sanitasi Layak**

Sanitasi layak merupakan fasilitas sanitasi yang dimiliki oleh suatu anggota rumah tangga saja atau milik bersama dengan jenis toilet berupa leher angsa atau plengsengan dengan tutup serta tempat pembuangan akhir tinja (TPAT) berupa tangki septik atau Instalasi Pembuangan Air Limbah (IPAL). Maksud dari toilet leher angsa ialah kloset yang di bawah dudukannya terdapat saluran berbentuk huruf "U" (seperti leher angsa) dengan maksud menampung air untuk menahan agar bau tinja tidak keluar (BPS, 2007). Sementara plengsengan adalah jamban/kakus yang di bawah dudukannya terdapat saluran rata yang dimiringkan ke pembuangan kotoran (BPS, 2013). Adanya fasilitas sanitasi yang layak dapat menurunkan risiko penyakit diare dan secara signifikan dapat mengurangi dampak kesehatan yang berakibat kematian dan penyakit yang menyerang jutaan anak di negara berkembang.

Namun begitu berdasarkan fakta sebagian rumah tangga masih memiliki perilaku buang air besar sembarangan (BABS) atau *open defecation*. Hal ini berdasarkan tindakan membuang kotoran atau tinja di ladang, hutan, semak-semak, sungai, pantai atau area terbuka lainnya sehingga menimbulkan risiko mengkontaminasi lingkungan, tanah, udara dan air.

Rumus yang digunakan untuk mengetahui persentase rumah tangga dengan sanitasi layak ialah banyaknya rumah tangga dengan akses terhadap sanitasi layak dibagi dengan jumlah penduduk dikali seratus persen. Pada tahun 2016, rumah tangga yang memiliki sanitasi layak sebesar 67.80% (Moeloek, 2017). Hal ini menunjukkan bahwa masih terdapat 32% persen masyarakat Indonesia yang belum memiliki fasilitas terhadap sanitasi layak.

### **3.11 Penerangan Dengan Listrik**

Penerangan merupakan salah satu komponen dari fasilitas perumahan. Sumber penerangan terbagi menjadi beberapa yakni:

1. Listrik PLN adalah sumber penerangan listrik yang dikelola oleh PLN. Rumah tangga dikatakan menggunakan listrik PLN baik menggunakan meteran maupun tidak menggunakan meteran (volumetrik).
2. Listrik non-PLN adalah sumber penerangan listrik yang dikelola oleh instansi/ pihak lain selain PLN termasuk yang menggunakan sumber penerangan dari accu (aki), generator, dan pembangkit listrik tenaga surya (yang tidak dikelola oleh PLN).
3. Petromak/ aladin
4. Pelita/ sentir/ obor
5. Dan lain-lain seperti lampu karbit, lilin, biji jarak, dan kemiri .

Sumber penerangan dikatakan layak jika menggunakan listrik sebagai penerangannya. Namun begitu, masih banyak daerah-daerah yang tidak mendapatkan akses akan penerangan dengan listrik tersebut yang mana pada hakekatnya sumber penerangan utama dengan listrik merupakan salah satu kunci strategis dalam pemerataan keadilan sosial yang berefek kepada peningkatan taraf hidup masyarakat.



## BAB IV

### METODOLOGI PENELITIAN

#### 4.1 Studi Analisis

Data yang digunakan dalam penelitian ini ialah *noisy* data yang berupa data *outlier*. Jenis data yang digunakan dalam penelitian ini merupakan data sekunder, yakni data perumahan. Data ini diperoleh dari website Badan Pusat Statistik (BPS) yakni:

<https://www.bps.go.id/subject/29/perumahan.html#subjekViewTab3>

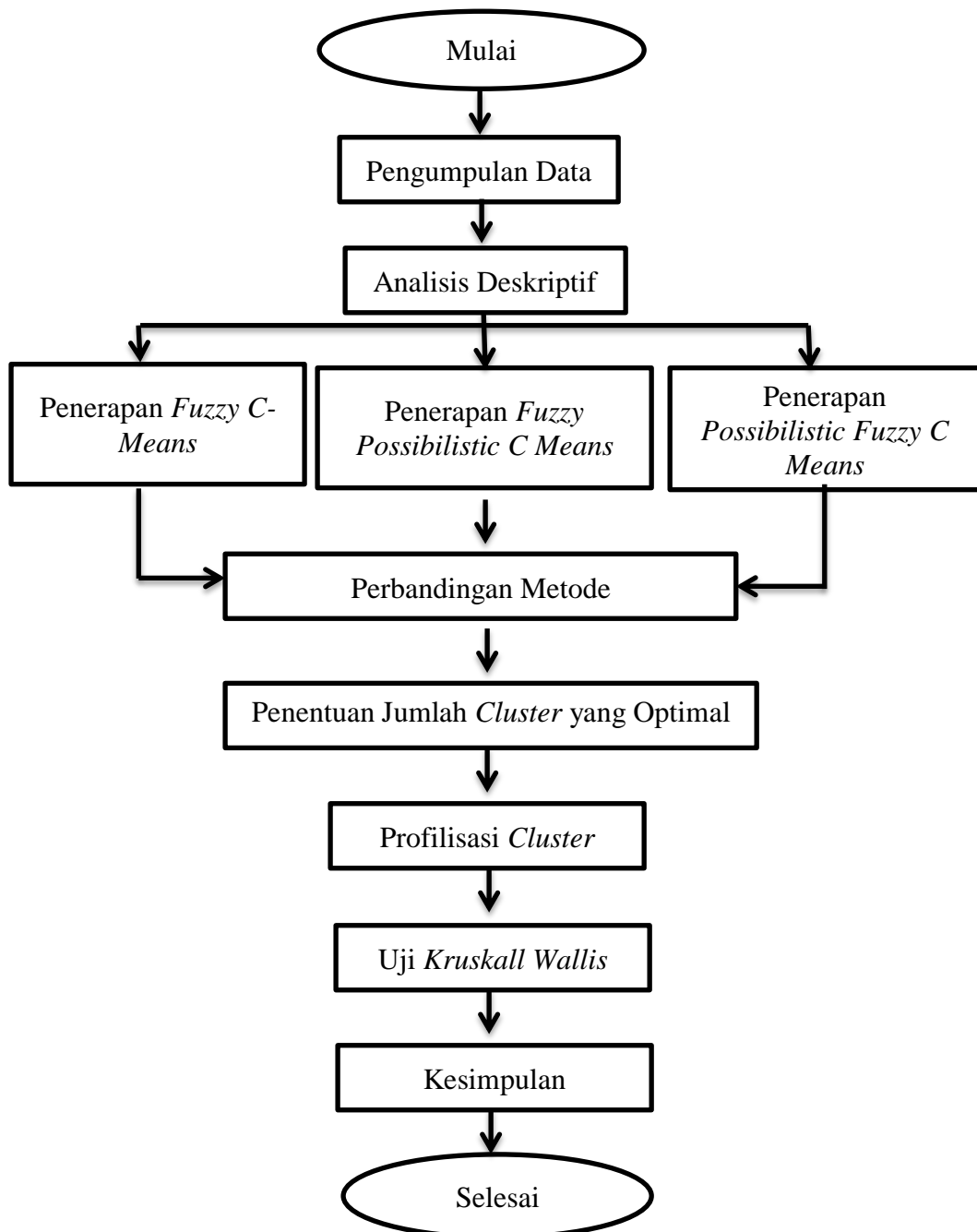
Variabel data yang digunakan dalam penelitian ini adalah data: sumber air minum layak, sumber penerangan dari listrik serta akses sanitasi layak pada tahun 2017. Data tersebut berjumlah 34 sesuai dengan jumlah provinsi di Indonesia. Untuk variabel yang digunakan, lebih detailnya seperti pada **Tabel 2**.

**Tabel 4.1** Penjelasan Variabel

Variabel	Keterangan
Sumber air minum layak	Persentase rumah tangga menurut provinsi berdasarkan sumber air minum layak tahun 2017
Sumber penerangan dari listrik	Persentase rumah tangga menurut provinsi berdasarkan sumber penerangan dari listrik tahun 2017
Akses sanitasi layak	Persentase rumah tangga menurut provinsi berdasarkan akses terhadap sanitasi layak pada 2017

#### 4.2 Metode Penelitian

Analisis awal yang digunakan dalam penelitian ini ialah analisis deskriptif. Selanjutnya ialah analisis *clustering* menggunakan metode *Fuzzy C Means* (FCM), *Fuzzy Possibilistic C Means* (FPCM) dan *Possibilistic Fuzzy C Means* (PFCM). Hasil dari ketiga metode tersebut kemudian dibandingkan keakuratannya berdasarkan nilai indeks *Modified Partition Entrophy* (MPC). Kemudian menentukan jumlah *cluster* yang optimal dari metode yang terbaik. Berikut adalah diagram alir metode penelitian:



**Gambar 4.1** Diagram Alir

## BAB V

### PEMBAHASAN

#### 5.1 Statistika Deskriptif

Analisis deskriptif digunakan untuk memberikan gambaran umum terkait data. Adapun gambaran umum yang akan ditampilkan adalah *mean*, *median*, modus, standar deviasi, nilai minimum dan nilai maksimum:

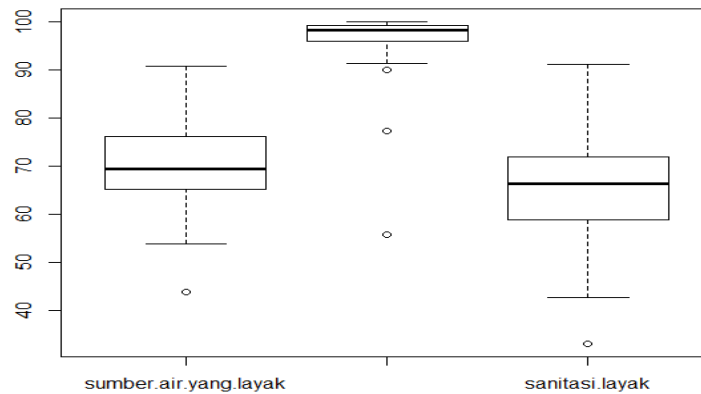
**Tabel 5.1** Statistik Deskriptif

	Sumber Air Minum Layak	Sumber Penerangan dari Listrik	Akses Sanitasi Layak
Mean	70.5159	95.7429	65.7521
Median	69.4500	98.2050	66.2700
Mode	65.73	99.91	33.06 <sup>a</sup>
Std. Deviation	9.68152	8.27459	13.65496
Minimum	43.83	55.81	33.06
Maximum	90.85	100.00	91.13

Berdasarkan **Tabel 5.1**, rata-rata dari variabel sumber air minum layak sebesar 70.5159, sumber penerangan dari listrik sebesar 95.7429, serta akses sanitasi layak sebesar 65.7521. Median dari variabel sumber air minum layak sebesar 69.4500, sumber penerangan dari listrik sebesar 98.2050, serta akses sanitasi layak sebesar 66.2700. Standar deviasi dari variabel sumber air minum layak sebesar 9.68152, sumber penerangan dari listrik sebesar 8.27459, serta akses sanitasi layak sebesar 13.65496. Nilai minimum dari variabel sumber air minum layak sebesar 43.83, sumber penerangan dari listrik sebesar 55.81, serta akses sanitasi layak sebesar 33.06. Sementara nilai maksimum dari variabel sumber air minum layak sebesar 90.85, sumber penerangan dari listrik sebesar 90.85, serta akses sanitasi layak sebesar 91.13.

## 5.2 Noisy Data

Data yang digunakan dalam penelitian ini ialah *noisy* data yang berupa data *outlier*. Terkait dengan data *outlier* yang digunakan, berikut adalah bukti dengan menggunakan *boxplot*:



**Gambar 5.2** Boxplot

Berdasarkan **Gambar 5.2**, data *outlier* pada variabel sumber air minum layak dan akses sanitasi layak berjumlah masing-masing 1. Sementara data *outlier* pada variabel sumber penerangan dari listrik berjumlah 3.

## 5.3 Perbandingan *Fuzzy C Means*, *Fuzzy Possibilistics C Means* dan *Possibilistics Fuzzy C Means*

Pada metode *Fuzzy C Means* (FCM) nilai pembobot ( $w$ ) yang digunakan adalah 2 dan nilai maksimum iterasi yang digunakan adalah 1000. Pada metode *Fuzzy Possibilistics C Means* (FPCM), nilai pembobot untuk *Fuzzy C Means* ( $w$ ) adalah 2, nilai pembobot untuk *Possibilistics C Means* ( $\eta$ ) adalah 2 dan nilai maksimum iterasi adalah 1000. Sementara pada metode *Possibilistics Fuzzy C Means* (PFCM), nilai pembobot untuk *Fuzzy C Means* ( $w$ ) adalah 2, nilai pembobot untuk *Possibilistics C Means* ( $\eta$ ) adalah 2, nilai maksimum iterasi adalah 1000, nilai koefisien pembobot untuk kekhasan relatif ( $a$ ) adalah 5, serta nilai koefisien pembobot untuk kekhasan absolut ( $b$ ) adalah 5.

Perbandingan metode *Fuzzy C Means* (FCM), *Fuzzy Possibilistics C Means* (FPCM) dan *Possibilistics Fuzzy C Means* (PFCM) dapat dilakukan dalam

berbagai cara. Berikut adalah hasil perbandingan ketiga metode berdasarkan waktu komputasi:

**Tabel 5.2** Perbandingan Waktu Komputasi

Cluster	Waktu (Detik)		
	FCM	FPCM	PFCM
2	0.24	0.17	0.22
3	0.61	0.45	0.95
4	1.05	0.88	1.06
5	0.83	0.73	0.71

Berdasarkan **Tabel 5.2**, waktu komputasi metode *Fuzzy C Means* (FCM) ialah 0.24, 0.61, 1.05, dan 0.83 detik. Pada metode *Fuzzy Possibilistics C Means* (FPCM), waktu komputasinya adalah 0.17, 0.45, 0.88 dan 0.73 detik. Sementara pada metode *Possibilistics Fuzzy C Means* (PFCM), waktu komputasinya adalah 0.22, 0.95, 1.06 dan 0.71 detik. Berdasarkan waktu komputasi, dapat disimpulkan bahwa pada studi kasus ini metode *Fuzzy Possibilistics C Means* (FPCM) memberikan waktu komputasi yang lebih cepat dibandingkan metode *Fuzzy C Means* (FCM) dan *Possibilistics Fuzzy C Means* (PFCM) pada hampir keseluruhan jumlah *cluster*.

**Tabel 5.3** Perbandingan Jumlah Iterasi dan Fungsi Objektif Minimum

Cluster	Jumlah Iterasi			Fungsi Objektif Minimum		
	FCM	FPCM	PFCM	FCM	FPCM	PFCM
2	60	47	57	4749.232	4755.07	34095.25
3	125	87	174	2839.475	2850.724	27859.29
4	162	119	148	1761.847	1779.072	21098.15
5	101	79	78	1165.686	1178.984	15778.39

Berdasarkan **Tabel 5.3**, jumlah iterasi yang dilakukan oleh metode *Fuzzy C Means* (FCM) untuk meminimumkan fungsi objektif adalah 60, 125, 162, dan 101. Pada metode *Fuzzy Possibilistics C Means* (FPCM), jumlah iterasinya adalah 47, 87, 119 serta 79. Sementara pada metode *Possibilistics Fuzzy C Means* (PFCM), jumlah iterasinya ialah 57, 174, 148 dan 78. Berdasarkan jumlah iterasi, dapat disimpulkan bahwa pada studi kasus ini metode *Fuzzy Possibilistics C Means* (FPCM) memiliki jumlah iterasi yang paling sedikit pada hampir

keseluruhan *cluster* untuk meminimumkan fungsi objektif. Hal inilah yang membuat waktu komputasi metode *Fuzzy Possibilistics C Means* (FPCM) lebih cepat dibandingkan dengan kedua metode lainnya.

Dari segi fungsi objektif, fungsi objektif pada metode *Fuzzy C Means* (FCM) secara berturut-turut adalah 4749.232, 2839.475, 1761.847 dan 1165.686. Pada metode *Fuzzy Possibilistics C Means* (FPCM), fungsi objektifnya secara berturut-turut adalah 4755.07, 2850.724, 1779.072, dan 1178.984. Sementara pada metode *Possibilistics Fuzzy C Means* (PFCM) ialah 34095.25, 27859.29, 21098.15 dan 15778.39. Berdasarkan fungsi objektif, nilai yang terendah adalah metode *Fuzzy C Means* (FCM). Namun begitu, ketiga metode ini tidak dapat dibandingkan berdasarkan fungsi objektif lantaran perhitungan fungsi objektif ketiganya berbeda.

Penentuan metode terbaik dapat dilihat berdasarkan nilai indeks validitas *cluster*. Indeks validitas *cluster* yang digunakan yakni indeks *Modified Partition Coefficient* (MPC). Indeks tersebut berdasarkan nilai-nilai keanggotaan. Pada indeks *Modified Partition Coefficient* (MPC) mengindikasikan bahwa nilai yang terbesar merupakan nilai yang optimal. Berikut adalah hasil indeks validitas *cluster* metode *Fuzzy C Means* (FCM), *Fuzzy Possibilistics C Means* (FPCM) dan *Possibilistics Fuzzy C Means* (PFCM).

**Tabel 5.4** Penentuan Metode Terbaik

C	MPC		
	FCM	FPCM	PFCM
2	0.461864	0.461917	0.458076
3	0.484375	0.488855	0.463409
4	0.525552	0.527508	0.53025
5	0.527822	0.52942	0.525586

Berdasarkan **Tabel 5.4**, nilai indeks *Modified Partition Coefficient* (MPC) tertinggi pada jumlah *cluster* dua, tiga dan lima adalah metode *Fuzzy Possibilistics C Means* (FPCM). Sementara nilai indeks *Modified Partition Coefficient* (MPC) tertinggi pada jumlah *cluster* empat adalah metode *Possibilistics Fuzzy C Means* (PFCM). Berdasarkan indeks validitas *Modified Partition Coefficient* (MPC),

metode *Fuzzy Possibilistic C Means* (FPCM) merupakan metode yang terbaik dibandingkan dengan metode *Fuzzy C Means* (FCM) dan metode *Possibilistics Fuzzy C Means* (PFCM). Hal tersebut didasarkan pada nilai indeks *Modified Partition Coefficient* (MPC) yang tertinggi pada hampir keseluruhan *cluster*.

#### 5.4 Penentuan Jumlah *Cluster* Optimal

Pada metode *Fuzzy Possibilistics C Means* (FPCM), jumlah *cluster* yang optimal dapat dilihat berdasarkan nilai indeks *Modified Partition Coefficient* (MPC) yang tertinggi. Berikut adalah hasil indeks *Modified Partition Coefficient* (MPC) pada metode *Fuzzy Possibilistics C Means* (FPCM).

**Tabel 5.5** Penentuan Jumlah *Cluster*

C	MPC
2	0.46192
3	0.48885
4	0.52751
5	0.52942
6	0.45158
7	0.49298

Berdasarkan **Tabel 5.5**, nilai indeks *Modified Partition Coefficient* (MPC) tertinggi pada metode *Fuzzy Possibilistics C Means* (FPCM) ialah dengan jumlah *cluster* lima. Nilainya adalah 0.52942. Berdasarkan hal tersebut, *cluster* yang optimal untuk digunakan adalah sebanyak lima *cluster*.

#### 5.5 Profilisasi *Fuzzy Possibilistics C Means* Terhadap Hasil *Cluster*

Profilisasi *cluster* meliputi penggambaran karakteristik masing-masing *cluster*. Profilisasi *cluster* ini dilakukan dengan menghitung rata-rata sumber air minum layak, sumber penerangan dari listrik serta akses sanitasi layak pada masing-masing *cluster*. Berikut adalah hasil profilisasi *cluster* :

**Tabel 5.6** Profilisasi *Fuzzy Possibilistics C Means* Terhadap Hasil *Cluster*

Cluster	Rata-Rata			Banyak Anggota
	Sumber Air Yang Layak	Penerangan Dengan Listrik	Sanitasi Layak	
1	62.29 %	94.58 %	51.68 %	9

2	59.09 %	55.81 %	33.06 %	1
3	67.28 %	96.19 %	65.10 %	9
4	85.23 %	99.71 %	89.34 %	4
5	75.58 %	98.52 %	72.19 %	11

Berdasarkan **Tabel 5.6**, *Cluster 1* memiliki rata-rata nilai persentase sumber air minum layak sebesar 62.29 %, sumber penerangan dari listrik sebesar 94.58 % serta nilai persentase akses sanitasi layak sebesar 51.68 %. Jumlah anggota *cluster 1* adalah 9 provinsi. Pada *cluster 2*, rata-rata nilai persentase sumber air minum layak sebesar 59.09 %, sumber penerangan dari listrik sebesar 55.81 % serta akses sanitasi layak sebesar 33.06 %. Jumlah anggota *cluster 2* adalah 1 provinsi.

Pada *cluster 3*, rata-rata nilai persentase sumber air minum layak sebesar 67.28 %, sumber penerangan dari listrik sebesar 96.19 % serta akses sanitasi layak sebesar 65.10 %. Jumlah anggota *cluster 3* adalah 9 provinsi. Pada *cluster 4*, rata-rata nilai persentase sumber air minum layak sebesar 85.23 %, sumber penerangan dari listrik sebesar 99.71 % serta akses sanitasi layak sebesar 89.34 %. Jumlah anggota *cluster 4* adalah 4 provinsi. Pada *cluster 5*, rata-rata nilai persentase sumber air minum layak sebesar 75.58 %, sumber penerangan dari listrik sebesar 98.52 % serta akses sanitasi layak sebesar 72.19 %. Jumlah anggota *cluster 5* adalah 11 provinsi.

Berdasarkan hasil profilisasi tersebut, rata-rata nilai persentase tertinggi adalah *cluster 4*. Semakin tinggi nilai persentase menunjukkan *cluster* tersebut memiliki fasilitas tempat tinggal yang semakin layak. Berikut adalah hasil *cluster*:

**Tabel 5.7** Hasil *Cluster*

Cluster	Provinsi
1	Sumatera Barat, Bengkulu, Lampung, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Gorontalo, dan Sulawesi Barat
2	Papua
3	Aceh, Jambi, Sumatera Selatan, Jawa Barat, Banten, Sulawesi Tengah, Maluku, Maluku Utara, dan Papua Barat



4	Kepulauan Riau, DKI Jakarta, DI Yogyakarta, dan Bali
5	Sumatera Utara, Riau, Kepulauan Bangka Belitung, Jawa Tengah, Jawa Timur, Nusa Tenggara Barat, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Selatan, dan Sulawesi Tenggara

### 5.6 Uji *Kruskall Wallis*

Uji *Kruskall Wallis* dilakukan guna melihat apakah terdapat perbedaan antara kelima *cluster* tersebut. Penggunaan uji *Kruskall Wallis* ini didasarkan pada pengujian normalitas bahwa data tersebut tidak berdistribusi normal. Berikut adalah hasil uji *Kruskall Wallis* dengan menggunakan software SPSS:

**Tabel 5.8** Uji *Kruskall Wallis*

	Sumber Air Minum Layak	Sumber Penerangan dari Listrik	Akses Sanitasi Layak
Chi-Square	23.19	11.399	29.673
Df	4	4	4
Asymp. Sig.	0.000	0.022	0.000

Pengujian hipotesis untuk penelitian ini adalah sebagai berikut:

i. Hipotesis

$H_0$ : Tidak terdapat perbedaan rata-rata antara *cluster 1*, *cluster 2*, *cluster 3*, *cluster 4* dan *cluster 5*.

$H_1$ : Minimal terdapat satu rata-rata yang berbeda

ii. Tingkat Signifikansi ( $\alpha$ ) = 5% = 0.05

iii. Daerah Kritis: Tolak  $H_0$  jika  $p\text{-value} < \alpha$

iv. Statistika Uji

Nilai  $p\text{-value}$  pada variabel sumber air minum layak = 0.000

Nilai  $p\text{-value}$  pada variabel sumber penerangan dari listrik = 0.022

Nilai  $p\text{-value}$  pada variabel akses sanitasi layak = 0.000

v. Keputusan

Pada variabel sumber air minum layak, sumber penerangan dari listrik dan akses sanitasi layak, nilai  $p\text{-value} < \alpha$  (0.05) sehingga tolak  $H_0$ .

## vi. Kesimpulan

Berdasarkan keputusan tersebut, maka minimal terdapat satu rata-rata yang berbeda pada keseluruhan variabel.

Berdasarkan hasil analisis menggunakan uji *Kruskall Wallis*, didapatkan hasil bahwa minimal terdapat satu rata-rata yang berbeda. Untuk mengetahui rata-rata kelompok yang berbeda, maka dilakukan uji perbandingan ganda. Berikut adalah hasil ujiperbandingan ganda:

## i. Hipotesis

$H_0$ : Tidak terdapat perbedaan rata-rata antara kedua *cluster*

$H_1$ : Terdapat perbedaan rata-rata antara kedua *cluster*

ii. Tingkat Signifikansi ( $\alpha$ ) = 10% = 0.1

## iii. Daerah Kritis

Tolak  $H_0$  jika:  $|\bar{R}_k - \bar{R}_k| > Z_{\frac{\alpha}{c(c-1)}} \sqrt{\frac{n(n+1)}{12} \left( \frac{1}{n_k} + \frac{1}{n_k} \right)}$

## iv. Statistika Uji

Selisih rata-rata peringkat yang didapatkan pada kedua kelompok *cluster* ialah:

**Tabel 5.9** Selisih *Mean Ranks*

Kelompok 1 (1)	Kelompok 2 (2)	Sumber Air Minum Layak (3)	Sumber Penerangan dari Listrik (4)	Akses Sanitasi Layak (5)
<i>Cluster 1</i>	<i>Cluster 2</i>	6.22	11.55	5
	<i>Cluster 3</i>	3.78	2.72	9.55
	<i>Cluster 4</i>	22.53	15.69	26.5
	<i>Cluster 5</i>	14.87	8.39	18.54
<i>Cluster 2</i>	<i>Cluster 3</i>	10	14.28	14.55
	<i>Cluster 4</i>	28.75	27.25	31.5
	<i>Cluster 5</i>	21.09	19.95	23.54
<i>Cluster 3</i>	<i>Cluster 4</i>	18.75	12.97	16.94
	<i>Cluster 5</i>	11.09	5.68	8.99

<i>Cluster 4</i>	<i>Cluster 5</i>	7.66	7.29	7.95
------------------	------------------	------	------	------

v. Keputusan

Berikut adalah keputusan hasil uji perbandingan ganda:

**Tabel 5.10** Uji Perbandingan Ganda

(1)	(2)	(3)	(4)	(5)	Z tabel	Keputusan
1	2	9.22	11.55	5	27.03	Gagal tolak $H_0$ pada keseluruhan variabel
	3	3.78	2.72	9.55	12.09	Gagal tolak $H_0$ pada keseluruhan variabel
	4	22.53	15.69	26.5	15.41	Tolak $H_0$ pada keseluruhan variabel
	5	14.87	8.39	18.54	11.52	Tolak $H_0$ pada variabel sumber air minum layak dan akses sanitasi layak
2	3	10	14.28	14.55	27.03	Gagal tolak $H_0$ pada keseluruhan variabel
	4	28.75	27.25	31.5	28.67	Tolak $H_0$ pada variabel sumber air minum layak
	5	21.09	19.95	23.54	26.78	Gagal tolak $H_0$ pada keseluruhan variabel
3	4	18.75	12.97	16.94	15.41	Tolak $H_0$ pada variabel sumber air minum layak
	5	11.09	5.68	8.99	11.52	Gagal tolak $H_0$ pada keseluruhan variabel
4	5	7.66	7.29	7.95	14.97	Gagal tolak $H_0$ pada keseluruhan variabel

vi. Kesimpulan

Pada keseluruhan variabel, rata-rata kelompok yang berbeda adalah *cluster* 1 dan *cluster* 4. Pada variabel sumber air minum layak dan akses sanitasi layak, rata-rata kelompok yang berbeda adalah *cluster* 1 dan *cluster* 5. Sementara pada variabel sumber air minum layak, rata-rata kelompok yang berbeda adalah *cluster* 2 dan *cluster* 4 serta *cluster* 3 dan *cluster* 4.

Berdasarkan uji perbandingan ganda metode *kruskal wallis*, dapat disimpulkan bahwa tidak semua *cluster* adalah berbeda. *Cluster* yang memiliki

perbedaan sangat tinggi dibandingkan dengan *cluster* yang lainnya adalah *cluster* 4 yang terdiri dari propinsi Kepulauan Riau, DKI Jakarta, DI Yogyakarta dan Bali. *Cluster* 4 mengindikasikan propinsi yang memiliki fasilitas tempat tinggal paling layak dibandingkan dengan *cluster* yang lainnya.

## BAB VI

### PENUTUP

#### 6.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan, diperoleh kesimpulan bahwa:

1. Hasil perbandingan indeks validitas *cluster Modified Partition Coefficient* (MPC) pada metode *Fuzzy C Means*, *Fuzzy Possibilistic C Means* dan *Possibilistic Fuzzy C Means* pada *noisy* data ini ialah metode *Fuzzy Possibilistic C Means* (FPCM) merupakan metode yang terbaik dibandingkan dengan metode *Fuzzy C Means* (FCM) dan metode *Possibilistics Fuzzy C Means* (PFCM). Hal tersebut berdasarkan nilai indeks *Modified Partition Coefficient* (MPC) yang tertinggi pada hampir keseluruhan *cluster*.
2. Hasil profilisasi pada pengelompokkan *noisy* data pada kasus fasilitas tempat tinggal ialah bahwa *cluster* 4 merupakan propinsi yang memiliki rata-rata tertinggi pada variabel sumber air minum layak, sumber penerangan dari listrik dan akses sanitasi layak. Semakin tinggi nilai rata-ratanya menunjukkan semakin layak fasilitas tempat tinggal suatu propinsi.

#### 6.2 Saran

Berdasarkan kesimpulan dari hasil analisis, maka diberikan saran sebagai berikut:

1. Untuk peneliti selanjutnya, diharapkan dapat menerapkan metode *Fuzzy Possibilistics C Means* (FPCM), terutama dalam kasus *noisy* data.
2. Untuk peneliti selanjutnya, diharapkan dapat menambahkan indeks validitas *cluster*.
3. Untuk pihak instansi pemerintahan, diharapkan dapat menentukan kebijakan yang tepat berdasarkan hasil *clustering*. Salah satunya ialah memberikan perhatian terhadap propinsi yang memiliki fasilitas tempat tinggal kurang layak seperti Papua.

## DAFTAR PUSTAKA

- Bankers, M. J., & Miller, J. R. 1996. Definition of Climate Region in the Northern Plains Using an Objective Cluster Modification Technique. *Journal of CI*, 130-146.
- BPS. 2007. *SUSENAS*. <https://www.rand.org/content/dam/rand/www/external/labor/bps/manualpdf/susenas/susenas07manualbukuiii.pdf>, Diunduh 1 Mei 2017, Pukul 19.20 WIB
- BPS. 2013. *Panduan Pelaksanaan Studi Pengukuran Tingkat Kebahagiaan (SPTK) 2013*. [https://sirusa.bps.go.id/webadmin/pedoman/Panduan%20Pelaksanaan%20Studi%20Pengukuran%20Tingkat%20Kebahagiaan%20\(SPTK\)%202013.pdf](https://sirusa.bps.go.id/webadmin/pedoman/Panduan%20Pelaksanaan%20Studi%20Pengukuran%20Tingkat%20Kebahagiaan%20(SPTK)%202013.pdf), Diunduh 1 Mei 2018, Pukul 19.32 WIB
- Davies, dan Beynon, P. 2004. *Database Systems*. New York: Palgrave Macmillan.
- Fajeri, M. (2013). Modifikasi Algoritma Fuzzy C Means (Studi Kasus: Pengelompokan Propinsi Berdasarkan Kualitas Pendidikan Madrasah). *Tesis UGM*
- Johnson, R. A. dan Wichern, D. W. 1998. *Applied Multivariate Statistical Analysis*. New York: Prentice Hall.
- Kusuma, D. T. dan Nazori, A. 2015. Prototipe Komparasi Model Clustering Menggunakan Metode K-Means dan FCM Untuk Menentukan Strategi Promosi: Studi Kasus Sekolah Tinggi Teknik PLN Jakarta. *Jurnal TICOM*, 1-10.
- Kusumadewi, S., Hartati, S., Harjoko, A., dan Wardoyo, R. 2006. *Fuzzy Multi-Attribute Decision Making (FUZZY MADM)*. Yogyakarta: Graha Ilmu.
- Larose, D. T. 2005. *Discovering Knowledge in Data: An Introduction to Data Mining*. Canada: John Willey & Son, Inc.
- Luthfi, E. T. 2017. Fuzzy C Means Untuk Clustering Data. *Seminar Nasional Teknologi*.
- Moeloek, N. F. 2017. *Bappenas*. [https://www.bappenas.go.id/files/6315/1004/2136/MENKES\\_AIR-SANITASI-Bappenas-FINAL-edit7nov17.pdf](https://www.bappenas.go.id/files/6315/1004/2136/MENKES_AIR-SANITASI-Bappenas-FINAL-edit7nov17.pdf). Diunduh 28 April 2018, Pukul 19.44 WIB

- Muhajir, M. 2017. *Modul Praktikum Statistika Multivariat Terapan*. Yogyakarta.
- Musdholifah, A. 2006. *Fuzzy Relational Clustering (FRC) Dengan Menggunakan Indeks Validasi Partition Coefficient (PC), Partition Entrophy (PE) dan Xie Beni (XB) Studi Kasus: Rangkuman Laporan Keuangan Perusahaan-Perusahaan di BEJ*. Tesis UGM
- Naba, A. 2009. *Belajar Cepat Fuzzy Logic Menggunakan Matlab*. Yogyakarta: C.V Andi Offset.
- Ned, G. dan Roger, J. 2000. *Fuzzy Logic Toolbox*. USA: Mathwork, inc.
- Ochtario, R. dan Karmila, S. 2013. Sistem Data Mining Untuk Mengetahui Tingkat Kecenderungan Memilih Menu Makanan. *Jurnal Informatika*, 80-89.
- Pal, N. R., Kal, P., dan Bezdek, J. C. 1997. A Mixed C Means Clustering Model. *IEEE*, 11-21.
- Prasetyo, E. 2014. *Data Mining Mengolah Data Menjadi Informasi Menggunakan MATLAB*. Yogyakarta: ANDI.
- Santosa, B. 2007. *Data Mining Teknik Pemanfaatan Data Untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Siyamto, Y. 2017. Pemanfaatan Data Mining Dengan Metode Clustering Untuk Evaluasi Biaya Dokumen Ekspor di PT Winstar Batam. *Media Informatika Budidarman*, 28-31.
- Sunitha, L., Raju, M. B., dan Srinivas, B. S. 2013. A Comparative Study Between Noisy Data and Outlier Data in Data Mining. *International Journal of Current Engineering and Technology*, 575-577.
- Supranto, J. (2009). *Statistik Teori dan Aplikasi*. Jakarta: Erlangga.
- Susanto, S. dan Suryadi, D. 2010. *Pengantar Data Mining Menggali Pengetahuan dari Bongkahan Data*. Yogyakarta: C.V. Andi Offset.
- Wahyono, A., Purnomo, M. H., & Sumpeno, S. 2015. Sistem Pengambilan Keputusan Berbasis Visualisasi Data Multidimensi Menggunakan Metode Fuzzy C Means. *Prosiding Seminar Nasional Manajemen Teknologi XXII*.

Xie, N., Hu, L., Luktarhan, N., dan Zhao, K. 2011. A Classification of Cluster Validity Indexes Based on Membership Degree and Application. *Web Information System and Mining*, 43-50.



# LAMPIRAN

**Lampiran 1 Data**

<b>Provinsi</b>	<b>Sumber Air Minum Layak</b>	<b>Penerangan Dari Listrik</b>	<b>Sanitasi Layak</b>
Aceh	64.85	99.02	63.38
Sumatera Utara	70.07	97.41	73.00
Sumatera Barat	68.83	97.81	52.77
Riau	75.12	96.69	70.04
Jambi	65.73	97.96	64.20
Sumatera Selatan	64.02	98.30	66.36
Bengkulu	43.83	98.11	42.71
Lampung	53.79	99.06	52.89
Kepulauan Bangka Belitung	68.14	99.17	83.56
Kepulauan Riau	83.95	99.10	86.33
DKI Jakarta	88.93	100.00	91.13
Jawa Barat	70.50	99.91	64.40
Jawa Tengah	76.09	99.91	71.84
DI Yogyakarta	77.19	99.90	89.40
Jawa Timur	75.54	99.92	68.83
Banten	66.11	99.95	71.68
Bali	90.85	99.85	90.51
Nusa Tenggara Barat	70.48	99.43	69.25
Nusa Tenggara Timur	65.20	77.36	45.31
Kalimantan Barat	68.77	91.51	49.65
Kalimantan Tengah	63.90	96.00	45.46
Kalimantan Selatan	60.62	99.16	58.09
Kalimantan Timur	82.75	99.14	72.83
Kalimantan Utara	83.78	97.62	66.59
Sulawesi Utara	73.29	99.30	71.93
Sulawesi Tengah	67.10	94.85	61.12
Sulawesi Selatan	76.34	98.53	76.73

Sulawesi Tenggara	79.83	96.55	69.52
Gorontalo	75.00	96.86	58.75
Sulawesi Barat	60.66	95.39	59.48
Maluku	68.34	91.35	63.29
Maluku Utara	65.73	94.23	66.18
Papua Barat	73.12	90.10	65.30
Papua	59.09	55.81	33.06
<b>Total</b>	<b>72.04</b>	<b>98.14</b>	<b>67.89</b>

**Lampiran 2** Statistika Deskriptif

		Sumber_Air_Layak	Penerangan_Listrik	Sanitasi_Layak
N	Valid	34	34	34
	Missing	0	0	0
Mean		70.5159	95.7429	65.7521
Median		69.4500	98.2050	66.2700
Mode		65.73	99.91	33.06 <sup>a</sup>
Std. Deviation		9.68152	8.27459	13.65496
Minimum		43.83	55.81	33.06
Maximum		90.85	100.00	91.13

### Lampiran 3 Hasil Analisis FCM, FPCM dan PFCM

#### 1. Fuzzy C Means (2 cluster)

```
> data=read.delim("clipboard")
> data
  sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
1          64.85          99.02          63.38
2          70.07          97.41          73.00
3          68.83          97.81          52.77
4          75.12          96.69          70.04
5          65.73          97.96          64.20
6          64.02          98.30          66.36
7          43.83          98.11          42.71
8          53.79          99.06          52.89
9          68.14          99.17          83.56
10         83.95          99.10          86.33
11         88.93         100.00          91.13
12         70.50          99.91          64.40
13         76.09          99.91          71.84
14         77.19          99.90          89.40
15         75.54          99.92          68.83
16         66.11          99.95          71.68
17         90.85          99.85          90.51
18         70.48          99.43          69.25
19         65.20          77.36          45.31
20         68.77          91.51          49.65
21         63.90          96.00          45.46
22         60.62          99.16          58.09
23         82.75          99.14          72.83
24         83.78          97.62          66.59
25         73.29          99.30          71.93
26         67.10          94.85          61.12
27         76.34          98.53          76.73
28         79.83          96.55          69.52
29         75.00          96.86          58.75
30         60.66          95.39          59.48
31         68.34          91.35          63.29
32         65.73          94.23          66.18
33         73.12          90.10          65.30
34         59.09          55.81          33.06
```

```
> u0 <- inaparc::imembrand(nrow(data), k=2)$u
> u0
```

```
      Cl.1      Cl.2
1 0.16216216 0.83783784
2 0.46875000 0.53125000
3 0.22429907 0.77570093
4 0.57017544 0.42982456
5 0.83018868 0.16981132
6 0.56382979 0.43617021
7 0.55932203 0.44067797
8 0.36363636 0.63636364
9 0.52432432 0.47567568
10 0.30000000 0.70000000
11 0.21600000 0.78400000
12 0.81034483 0.18965517
13 0.50000000 0.50000000
14 0.84615385 0.15384615
15 0.80281690 0.19718310
16 0.22222222 0.77777778
17 0.79661017 0.20338983
18 0.45121951 0.54878049
19 0.72289157 0.27710843
20 0.81927711 0.18072289
21 0.88000000 0.12000000
22 0.75714286 0.24285714
23 0.98630137 0.01369863
24 0.04761905 0.95238095
25 0.05263158 0.94736842
26 0.83185841 0.16814159
27 0.64423077 0.35576923
28 0.98888889 0.01111111
29 0.35294118 0.64705882
30 0.72857143 0.27142857
31 0.09375000 0.90625000
32 0.31707317 0.68292683
33 0.17391304 0.82608696
34 0.37735849 0.62264151
```

```

> #fcm cluster = 2
> res.fcm.2<- fcm(data, centers=2,membership=u0)
> summary(res.fcm.2)
Summary for 'res.fcm.2'

Number of data objects: 34

Number of clusters: 2

Crisp clustering vector:
 [1] 2 1 2 1 2 2 2 2 1 1 1 1 1 1 1 1 1 1 2 2 2 2 1 1 1 2 1 1 2 2 2 1 2

Initial cluster prototypes:
      sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1          77.19                99.90                89.40
Cluster 2          68.83                97.81                52.77

Final cluster prototypes:
      sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1          76.29368            98.12327            74.30532
Cluster 2          63.92950            93.69514            55.72681

Distance between the final cluster prototypes
      Cluster 1
Cluster 2  517.6423

Difference between the initial and final cluster prototypes
      sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1          -0.8963181            -1.776729            -15.094680
Cluster 2          -4.9005048            -4.114859             2.956815

Root Mean Squared Deviations (RMSD): 11.8638
Mean Absolute Deviation (MAD): 44.60986

Membership degrees matrix (top and bottom 5 rows):
      Cluster 1 Cluster 2
1  0.25899498  0.74100502
2  0.89522727  0.10477273
3  0.08728767  0.91271233
4  0.94004593  0.05995407
5  0.30370515  0.69629485
...
      Cluster 1 Cluster 2
30  0.05537224  0.9446278
31  0.26278389  0.7372161
32  0.36914022  0.6308598
33  0.54860357  0.4513964
34  0.34244334  0.6575567

Descriptive statistics for the membership degrees by clusters
      Size      Min      Q1      Mean      Median      Q3      Max
Cluster 1   18  0.5379343  0.8095809  0.8420989  0.8616543  0.9339557  0.9903188
Cluster 2   16  0.6142338  0.6866103  0.7878802  0.7815467  0.9027474  0.9446278

Dunn's Fuzziness Coefficients:
dunn_coef normalized
0.7309322  0.4618644

Within cluster sum of squares by cluster:
      1      2
2285.093 4048.904
(between_SS / total_SS = 40.91%)

Available components:
 [1] "u"      "v"      "v0"     "d"      "x"
 [6] "cluster" "csize"  "sumsqrs" "k"      "m"
[11] "iter"    "best.start" "func.val" "comp.time" "inpargs"
[16] "algorithm" "call"

```

```

> res.fcm.2$comp.time
 [1] 0.24
> res.fcm.2$iter
 [1] 60
> res.fcm.2$func.val
 [1] 4749.232

```

```

> res.fcm.2$u
  Cluster 1 Cluster 2
1 0.25899498 0.741005023
2 0.89522727 0.104772732
3 0.08728767 0.912712331
4 0.94004593 0.059954072
5 0.30370515 0.696294851
6 0.38576616 0.614233842
7 0.22415945 0.775840550
8 0.12630717 0.873692833
9 0.84294304 0.157056963
10 0.87002032 0.129979683
11 0.81125751 0.188742489
12 0.53793428 0.462065718
13 0.97955491 0.020445091
14 0.85328829 0.146711708
15 0.91088791 0.089112092
16 0.72366849 0.276331512
17 0.80511259 0.194887413
18 0.80902208 0.190977925
19 0.21274723 0.787252768
20 0.08422092 0.915779077
21 0.10057431 0.899425690
22 0.08344532 0.916554682
23 0.93775663 0.062243373
24 0.81994721 0.180052791
25 0.95963806 0.040361941
26 0.13073993 0.869260067
27 0.99031881 0.009681188
28 0.92255290 0.077447100
29 0.36622881 0.633771190
30 0.05537224 0.944627762
31 0.26278389 0.737216108
32 0.36914022 0.630859778
33 0.54860357 0.451396426
34 0.34244334 0.657556656

> fpc.2 <- PC(res.fcm.a$U)
> fpc.2
[1] 0.7309322
> fmpc.2 <- MPC(res.fcm.a$U)
> fmpc.2
[1] 0.4618644
>

```

## 2. Fuzzy C Means (3 cluster)

```

> u0 <- inaparc::imembrand(nrow(data), k=3)$u
> u0
  Cl.1 Cl.2 Cl.3
1 0.10526316 0.543859649 0.35087719
2 0.38857143 0.137142857 0.47428571
3 0.41139241 0.310126582 0.27848101
4 0.08737864 0.514563107 0.39805825
5 0.46478873 0.366197183 0.16901408
6 0.25403226 0.391129032 0.35483871
7 0.20689655 0.482758621 0.31034483
8 0.45794393 0.439252336 0.10280374
9 0.25373134 0.253731343 0.49253731
10 0.20224719 0.640449438 0.15730337
11 0.15483871 0.541935484 0.30322581
12 0.12765957 0.393617021 0.47872340
13 0.39735099 0.152317881 0.45033113
14 0.23076923 0.676923077 0.09230769
15 0.37323944 0.119718310 0.50704225
16 0.01176471 0.047058824 0.94117647
17 0.01986755 0.357615894 0.62251656
18 0.15447154 0.544715447 0.30081301
19 0.87254902 0.009803922 0.11764706
20 0.23913043 0.554347826 0.20652174
21 0.05172414 0.500000000 0.44827586
22 0.37837838 0.108108108 0.51351351
23 0.21978022 0.362637363 0.41758242
24 0.38888889 0.569444444 0.04166667
25 0.44000000 0.493333333 0.06666667
26 0.11656442 0.343558282 0.53987730
27 0.54545455 0.363636364 0.09090909
28 0.31835206 0.318352060 0.36329588
29 0.10691824 0.402515723 0.49056604
30 0.28971963 0.074766355 0.63551402
31 0.25000000 0.270270270 0.47972973
32 0.46596859 0.062827225 0.47120419
33 0.30578512 0.342975207 0.35123967
34 0.10476190 0.571428571 0.32380952

```

```

> #fcm cluster = 3
> res.fcm.3 <- fcm(data, centers=3, membership=u0)
> summary(res.fcm.3)
Summary for 'res.fcm.3'

Number of data objects: 34

Number of clusters: 3

Crisp clustering vector:
[1] 1 1 3 1 1 1 3 3 2 2 2 1 1 2 1 1 2 1 1 2 1 3 3 3 3 2 1 1 1 2 1 1 3 1 1 1 3

Initial cluster prototypes:
      sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1          60.62                99.16                58.09
Cluster 2          90.85                99.85                90.51
Cluster 3          43.83                98.11                42.71

Final cluster prototypes:
      sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1          70.24860                97.06318                66.60723
Cluster 2          82.63893                99.14665                85.28590
Cluster 3          61.44288                90.93720                48.66816

Distance between the final cluster prototypes
      Cluster 1 Cluster 2
Cluster 2  506.7538
Cluster 3  436.8787 1857.5268

Difference between the initial and final cluster prototypes
      sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1          9.628598                -2.0968191                8.517232
Cluster 2         -8.211074                -0.7033542               -5.224097
Cluster 3         17.612881                -7.1727995                5.958156

Root Mean Squared Deviations (RMSD): 14.85507
Mean Absolute Deviation (MAD): 65.12501

Membership degrees matrix (top and bottom 5 rows):
      Cluster 1 Cluster 2 Cluster 3
1  0.8316834 0.04531527 0.12300130
2  0.8407548 0.11055903 0.04868617
3  0.3583038 0.05563078 0.58606543
4  0.8521312 0.10298622 0.04488258
5  0.8893932 0.03282877 0.07777804
...
      Cluster 1 Cluster 2 Cluster 3
30 0.4576979 0.05727004 0.4850321
31 0.8039026 0.05074889 0.1453485
32 0.8868144 0.03760362 0.0755820
33 0.8042095 0.08217999 0.1136105
34 0.2807153 0.16057696 0.5587077

Descriptive statistics for the membership degrees by clusters
      Size   Min      Q1     Mean   Median    Q3     Max
Cluster 1  18 0.5775652 0.7891109 0.8132556 0.8351377 0.8566626 0.9495831
Cluster 2   7 0.5211426 0.5284034 0.7521761 0.8847983 0.9043487 0.9937876
Cluster 3   9 0.4850321 0.5587077 0.6689653 0.6981642 0.7518398 0.9029961

Dunn's Fuzziness Coefficients:
dunn_coef normalized
0.6562503 0.4843754

Within cluster sum of squares by cluster:
      1      2      3
939.7065 679.6085 2726.9233
(between_SS / total_SS = 63%)

Available components:
[1] "u"      "v"      "v0"     "d"      "x"
[6] "cluster" "csize"  "sumsqrs" "k"      "m"
[11] "iter"   "best.start" "func.val" "comp.time" "inpargs"
[16] "algorithm" "call"

```

```

> res.fcm.3$comp.time
[1] 0.61
> res.fcm.3$iter
[1] 125
> res.fcm.3$func.val
[1] 2839.475

```



```

> res.fcm.3$u
      Cluster 1 Cluster 2 Cluster 3
1  0.831683433 0.04531527 0.123001298
2  0.840754796 0.11055903 0.048686172
3  0.358303786 0.05563078 0.586065434
4  0.852131201 0.10298622 0.044882580
5  0.889393188 0.03282877 0.077778037
6  0.858173093 0.04912109 0.092705821
7  0.218314296 0.08352155 0.698164153
8  0.222301681 0.05470058 0.722997743
9  0.382781745 0.53195136 0.085266897
10 0.004809664 0.99378757 0.001402762
11 0.070212445 0.90428934 0.025498217
12 0.948656099 0.02117573 0.030168166
13 0.717333713 0.22265029 0.060015996
14 0.074119642 0.90440809 0.021472273
15 0.842000615 0.10754578 0.050453609
16 0.838591931 0.09353842 0.067869649
17 0.083943520 0.88479826 0.031258219
18 0.949583120 0.02963032 0.020786556
19 0.181817630 0.06634252 0.751839847
20 0.142012214 0.02993924 0.828048541
21 0.077544115 0.01945981 0.902996076
22 0.450715973 0.06244750 0.486836528
23 0.405684520 0.52114257 0.073172908
24 0.577565155 0.29994311 0.122491737
25 0.821769357 0.13165738 0.046573258
26 0.784180279 0.04174211 0.174077613
27 0.419434259 0.52485543 0.055710310
28 0.663552341 0.25350180 0.082945862
29 0.728305675 0.08002087 0.191673456
30 0.457697878 0.05727004 0.485032083
31 0.803902574 0.05074889 0.145348532
32 0.886814375 0.03760362 0.075582005
33 0.804209477 0.08217999 0.113610537
34 0.280715312 0.16057696 0.558707725

> res.fcm.b <- ppclust2(res.fcm.3, "fclust") # optimal
> fpc.3 <- FC(res.fcm.b$U)
> fpc.3
[1] 0.6562503
> fmpc.3 <- MPC(res.fcm.b$U)
> fmpc.3
[1] 0.4843754
>

```

### 3. Fuzzy C Means (4 cluster)

```

> u0 <- inaparc::imembrand(nrow(data), k=4)$u
> u0
      Cl.1      Cl.2      Cl.3      Cl.4
1  0.07531381 0.389121339 0.25104603 0.284518828
2  0.10859729 0.375565611 0.29411765 0.221719457
3  0.29931973 0.061224490 0.36054422 0.278911565
4  0.35869565 0.282608696 0.13043478 0.228260870
5  0.39591837 0.359183673 0.07346939 0.171428571
6  0.11203320 0.406639004 0.39004149 0.091286307
7  0.23287671 0.232876712 0.45205479 0.082191781
8  0.31843575 0.078212291 0.13407821 0.469273743
9  0.33333333 0.085106383 0.26241135 0.319148936
10 0.36144578 0.138554217 0.40963855 0.090361446
11 0.36666667 0.050000000 0.44166667 0.141666667
12 0.45859873 0.006369427 0.02547771 0.509554140
13 0.01764706 0.317647059 0.55294118 0.111764706
14 0.34536082 0.190721649 0.45876289 0.005154639
15 0.11538462 0.211538462 0.49038462 0.182692308
16 0.02631579 0.254385965 0.22807018 0.491228070
17 0.11034483 0.524137931 0.13793103 0.227586207
18 0.20879121 0.307692308 0.45054945 0.032967033
19 0.39053254 0.437869822 0.05917160 0.112426036
20 0.19718310 0.309859155 0.29577465 0.197183099
21 0.04982206 0.302491103 0.30249110 0.345195730
22 0.08947368 0.336842105 0.41052632 0.163157895
23 0.05228758 0.444444444 0.24183007 0.261437908
24 0.27099237 0.339694656 0.04580153 0.343511450
25 0.29249012 0.328063241 0.33596838 0.043478261
26 0.33333333 0.188888889 0.33333333 0.144444444
27 0.24867725 0.328042328 0.35978836 0.063492063
28 0.18103448 0.215517241 0.27586207 0.327586207
29 0.30522088 0.257028112 0.11244980 0.325301205
30 0.44545455 0.050000000 0.10000000 0.404545455
31 0.04972376 0.480662983 0.14917127 0.320441989
32 0.44500000 0.180000000 0.08000000 0.295000000
33 0.13513514 0.554054054 0.28378378 0.027027027
34 0.29573171 0.262195122 0.27743902 0.164634146

```

```

> #fcm cluster = 4
> res.fcm.4<- fcm(data, centers=4,membership=u0)
> summary(res.fcm.4)
Summary for 'res.fcm.4'

Number of data objects: 34

Number of clusters: 4

Crisp clustering vector:
[1] 2 1 2 1 1 1 2 2 1 4 4 1 1 4 1 1 4 1 3 2 2 2 1 1 1 2 1 1 1 2 1 1 1 3

Initial cluster prototypes:
      sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1          66.11                99.95                71.68
Cluster 2          63.90                96.00                45.46
Cluster 3          59.09                55.81                33.06
Cluster 4          90.85                99.85                90.51

Final cluster prototypes:
      sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1          72.89669                97.88102                69.23534
Cluster 2          63.18358                96.19473                55.82167
Cluster 3          59.70715                60.71960                35.48187
Cluster 4          84.31286                99.54545                87.81336

Distance between the final cluster prototypes
      Cluster 1 Cluster 2 Cluster 3
Cluster 2  277.1145
Cluster 3 2694.2324 1684.2780
Cluster 4  478.2421 1481.1418 4851.4732

Difference between the initial and final cluster prototypes
      sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1          6.7866876                -2.0689770                -2.444660
Cluster 2         -0.7164165                 0.1947270                10.361671
Cluster 3          0.6171456                 4.9095978                 2.421867
Cluster 4         -6.5371436                -0.3045469                -2.696639

Root Mean Squared Deviations (RMSD): 7.821139
Mean Absolute Deviation (MAD): 30.04506

Membership degrees matrix (top and bottom 5 rows):
      Cluster 1 Cluster 2 Cluster 3 Cluster 4
1 0.3809718 0.56304150 0.016826105 0.03916064
2 0.8886224 0.05782514 0.006952203 0.04660025
3 0.1261669 0.82851137 0.020646196 0.02467552
4 0.9614917 0.01954053 0.002472482 0.01649528
5 0.4806911 0.46217037 0.016406642 0.04073186
...
      Cluster 1 Cluster 2 Cluster 3 Cluster 4
30 0.073352672 0.90294042 0.01035466 0.013352234
31 0.477233815 0.44534839 0.02639015 0.051027645
32 0.570629177 0.35907767 0.02009979 0.050193358
33 0.674804923 0.22745270 0.02657023 0.071172152
34 0.009024003 0.01362177 0.97203564 0.005318585

Descriptive statistics for the membership degrees by clusters
      Cluster 1 Cluster 2 Cluster 3 Cluster 4
Size      19      9      2      4
Min      0.4617943 0.5502448 0.7114195 0.6748049
Q1      0.5470083 0.6737834 0.7482713 0.7728727
Mean    0.4175212 0.5561498 0.6947784 0.6947784
Median  0.4175212 0.5561498 0.6947784 0.8334070
Q3      0.4175212 0.5561498 0.6947784 0.8334070
Max     0.8506705 0.8953147 0.9224785 0.9238238

Dunn's Fuzziness Coefficients:
dunn_coeff normalized
0.6441637 0.5255516

Within cluster sum of squares by cluster:
      1      2      3      4
1277.9208 988.1555 325.8985 125.7030
(between_SS / total_SS = 76.96%)

Available components:
[1] "u"      "v"      "v0"     "d"      "x"
[6] "cluster" "csize" "sumsqrs" "x"      "m"
[11] "iter"   "best.start" "func.val" "comp.time" "inpargs"

> res.fcm.4$comp.time
[1] 1.05
> res.fcm.4$iter
[1] 162
> res.fcm.4$func.val
[1] 1761.847

```

```

> res.fcm.4$u
  Cluster 1 Cluster 2 Cluster 3 Cluster 4
1  0.380971755 0.563041497 0.0168261047 0.039160643
2  0.888622407 0.057825144 0.0069522034 0.046600246
3  0.126166912 0.828511369 0.0206461958 0.024675523
4  0.961491704 0.019540534 0.0024724825 0.016495279
5  0.480691125 0.462170373 0.0164066421 0.040731861
6  0.529860423 0.397840388 0.0193865860 0.052912603
7  0.194336768 0.547008252 0.1767734041 0.081881576
8  0.130259931 0.785728969 0.0456493963 0.038361703
9  0.461794268 0.132000182 0.0274469482 0.378758602
10 0.006033307 0.001830931 0.0005400392 0.991595723
11 0.041151276 0.015847078 0.0055501183 0.937451527
12 0.772666977 0.182261853 0.0103207357 0.034750434
13 0.892559923 0.043095942 0.0060237517 0.058320383
14 0.105814284 0.033951571 0.0095636150 0.850670529
15 0.940352913 0.031670850 0.0036676936 0.024308543
16 0.757538276 0.155612774 0.0147601508 0.072088799
17 0.058550631 0.023006122 0.0082471435 0.910196103
18 0.950451097 0.032093654 0.0028428710 0.014612378
19 0.160090866 0.359123111 0.4175212309 0.063264792
20 0.155178778 0.750355941 0.0556215243 0.038843757
21 0.128365456 0.772872732 0.0612438018 0.037518010
22 0.067498969 0.910198791 0.0093828148 0.012919425
23 0.592385481 0.097101406 0.0194318653 0.291081247
24 0.644812839 0.149265832 0.0278226158 0.178098714
25 0.948980294 0.024102471 0.0029813673 0.023935868
26 0.280427655 0.673783413 0.0162334249 0.029555507
27 0.668257545 0.074289970 0.0134227590 0.244029725
28 0.792277239 0.085083840 0.0138912853 0.108747636
29 0.511559577 0.397176503 0.0283647003 0.062899220
30 0.073352672 0.902940417 0.0103546772 0.013352234
31 0.477233815 0.445348387 0.0263901538 0.051027645
32 0.570629177 0.359077671 0.0200997942 0.050193358
33 0.674804923 0.227452696 0.0265702289 0.071172152
34 0.009024003 0.013621775 0.9720356368 0.005318585

> res.fcm.c <- ppclust2(res.fcm.4, "fclust") # optimal
> fpc.4 <- PC(res.fcm.c$U)
> fpc.4
[1] 0.6441637
> fmpc.4 <- MPC(res.fcm.c$U)
> fmpc.4
[1] 0.5255516

```

#### 4. Fuzzy C Means (5 cluster)

```

> u0 <- inaparc::imembrand(nrow(data), k=5)$u
> u0
  Cl.1 Cl.2 Cl.3 Cl.4 Cl.5
1  0.068441065 0.35361217 0.22813688 0.25855513 0.09125475
2  0.332000000 0.26000000 0.19600000 0.17600000 0.03600000
3  0.172638436 0.13355049 0.32247557 0.25407166 0.11726384
4  0.204545455 0.31493506 0.28571429 0.05844156 0.13636364
5  0.092465753 0.33561644 0.32191781 0.07534247 0.17465753
6  0.213389121 0.41422594 0.07531381 0.23849372 0.05857741
7  0.117647059 0.41176471 0.23039216 0.05882353 0.18137255
8  0.213270142 0.28436019 0.10900474 0.32227488 0.07109005
9  0.229166667 0.03125000 0.27604167 0.08854167 0.37500000
10 0.007042254 0.02816901 0.56338028 0.02112676 0.38028169
11 0.307189542 0.06209150 0.21895425 0.12091503 0.29084967
12 0.009523810 0.11428571 0.20952381 0.48571429 0.18095238
13 0.023076923 0.22307692 0.20000000 0.43076923 0.12307692
14 0.340807175 0.08968610 0.14798206 0.17040359 0.25112108
15 0.344537815 0.02521008 0.27731092 0.31092437 0.04201681
16 0.062706271 0.18481848 0.29042904 0.27722772 0.18481848
17 0.046979866 0.28523490 0.28523490 0.32550336 0.05704698
18 0.257028112 0.31325301 0.12449799 0.03212851 0.27309237
19 0.148594378 0.16064257 0.28514056 0.35742972 0.04819277
20 0.262390671 0.21574344 0.24198251 0.24781341 0.03206997
21 0.264317181 0.14977974 0.26431718 0.11453744 0.20704846
22 0.264957265 0.29059829 0.05128205 0.17948718 0.21367521
23 0.207792208 0.24675325 0.24675325 0.20779221 0.09090909
24 0.269102990 0.32558140 0.03654485 0.07308970 0.29568106
25 0.033333333 0.32222222 0.10000000 0.21481481 0.32962963
26 0.169014085 0.07511737 0.27699531 0.09389671 0.38497653
27 0.131250000 0.01250000 0.30312500 0.26875000 0.28437500
28 0.335403727 0.27950311 0.36645963 0.01242236 0.00621118
29 0.257372654 0.16890080 0.15817694 0.24932976 0.16621984
30 0.198113208 0.30188679 0.15408805 0.22012579 0.12578616
31 0.028735632 0.02873563 0.40229885 0.10344828 0.43678161
32 0.068376068 0.08974359 0.30769231 0.29059829 0.24358974
33 0.069565217 0.15217391 0.01304348 0.42608696 0.33913043
34 0.170403587 0.20627803 0.28699552 0.26457399 0.07174888

```

```

> #fcm cluster = 5
> res.fcm.5<- fcm(data, centers=5,membership=u0)
> summary(res.fcm.5)
Summary for 'res.fcm.5'

Number of data objects: 34

Number of clusters: 5

Crisp clustering vector:
[1] 3 5 1 5 3 3 1 1 5 4 4 3 5 4 5 3 4 5 1 1 1 3 5 5 5 3 5 5 3 3 3 3 2

Initial cluster prototypes:
      sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1          68.83                97.81                52.77
Cluster 2          65.20                77.36                45.31
Cluster 3          68.14                99.17                83.56
Cluster 4          88.93                100.00               91.13
Cluster 5          76.34                98.53                76.73

Final cluster prototypes:
      sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1          61.56990                95.29674                49.53832
Cluster 2          59.34234                57.64635                33.98936
Cluster 3          66.27916                96.50010                63.48665
Cluster 4          85.30456                99.62232                88.70126
Cluster 5          75.32641                98.30732                70.88147

Distance between the final cluster prototypes
      Cluster 1 Cluster 2 Cluster 3 Cluster 4
Cluster 2 1664.2844
Cluster 3 218.1809 2427.8236
Cluster 4 2115.7805 5429.4117 1007.4910
Cluster 5 653.8348 3269.8328 139.8021 418.8379

Difference between the initial and final cluster prototypes
      sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1          -7.260099                -2.5132577                -3.231675
Cluster 2          -5.857661                -19.7136510                -11.320640
Cluster 3          -1.860840                -2.6698981                -20.073352
Cluster 4          -3.625436                -0.3776765                -2.428738
Cluster 5          -1.013591                -0.2226782                -5.848533

Root Mean Squared Deviations (RMSD): 14.7551
Mean Absolute Deviation (MAD): 52.81064

Membership degrees matrix (top and bottom 5 rows):
      Cluster 1 Cluster 2 Cluster 3 Cluster 4 Cluster 5
1 0.035318866 0.002930410 0.9086913 0.007204727 0.04585474
2 0.036397059 0.007093422 0.2159562 0.047210566 0.69334272
3 0.546404621 0.018463585 0.3084307 0.024242869 0.10245823
4 0.005333676 0.001051578 0.0266750 0.007016332 0.95992342
5 0.011840978 0.001099132 0.9634756 0.002873853 0.02071047
...
      Cluster 1 Cluster 2 Cluster 3 Cluster 4 Cluster 5
30 0.29093172 0.013967680 0.593489307 0.019604641 0.082006652
31 0.08992873 0.010855100 0.731237685 0.022484838 0.145493649
32 0.03711407 0.004538834 0.862620107 0.011923582 0.083803412
33 0.09853486 0.018118543 0.442466870 0.051204004 0.389675719
34 0.00232682 0.994079361 0.001623192 0.000749492 0.001221136

Descriptive statistics for the membership degrees by clusters
      Size Min Q1 Mean Median Q3 Max
Cluster 1 6 0.3894718 0.5343690 0.6152948 0.6064723 0.6665636 0.8924239
Cluster 2 1 0.9940794 0.9940794 0.9940794 0.9940794 0.9940794 0.9940794
Cluster 3 12 0.4424669 0.4866571 0.6899044 0.6804379 0.8705394 0.9634756
Cluster 4 4 0.7565614 0.8718447 0.8935318 0.9277841 0.9494712 0.9619975
Cluster 5 11 0.4159381 0.6585091 0.7633121 0.7528141 0.9273502 0.9609130

Dunn's Fuzziness Coefficients:
dunn_coeff normalized
0.6222577 0.5278221

Within cluster sum of squares by cluster:
      1 2 3 4 5
926.2408 0.0000 493.0213 125.7030 484.3260
(between_SS / total_SS = 80.87%)

Available components:
[1] "u" "v" "v0" "d" "x"
[6] "cluster" "csize" "sumsqrs" "k" "m"
[11] "iter" "best.start" "func.val" "comp.time" "inpargs"
[16] "algorithm" "call"
> res.fcm.5$comp.time
[1] 0.83
> res.fcm.5$iter
[1] 101
> res.fcm.5$func.val
[1] 1165.686

```

```

> res.fcm.5$u
      Cluster 1 Cluster 2 Cluster 3 Cluster 4 Cluster 5
1  0.035318866 0.002930410 0.908691260 0.0072047268 0.045854738
2  0.036397059 0.007093422 0.215956236 0.0472105664 0.693342717
3  0.546404621 0.018463585 0.308430696 0.0242428693 0.102458229
4  0.005333676 0.001051578 0.026674999 0.0070163321 0.959923415
5  0.011840978 0.001099132 0.963475563 0.0028738529 0.020710473
6  0.046765592 0.005118808 0.839523094 0.0146090228 0.093983483
7  0.530357098 0.100221358 0.208725332 0.0510293666 0.109666845
8  0.666540056 0.027230979 0.208393433 0.0251637391 0.072671793
9  0.072924804 0.020815545 0.214374741 0.2759467776 0.415938133
10 0.003979092 0.001468705 0.008844484 0.9619974558 0.023710263
11 0.007252640 0.003056071 0.014064504 0.9452957992 0.030330986
12 0.059222526 0.006723157 0.629638207 0.0235408845 0.280875225
13 0.005361808 0.001117724 0.022016020 0.0105914051 0.960913042
14 0.027098615 0.009710466 0.062640440 0.7565613748 0.143989104
15 0.010760261 0.001941178 0.050270819 0.0129198640 0.924107877
16 0.067841138 0.011095477 0.456944666 0.0548819469 0.409236772
17 0.012131780 0.005196939 0.023057924 0.9102724161 0.049340941
18 0.035944335 0.005599920 0.293264343 0.0291459935 0.636045409
19 0.389471839 0.249319163 0.196875407 0.0493778368 0.114955755
20 0.666571390 0.029794764 0.198251802 0.0236683325 0.081713712
21 0.892423897 0.012400346 0.060839831 0.0085986291 0.025737297
22 0.379692716 0.014648262 0.495079682 0.0218395123 0.088739828
23 0.040349049 0.010742013 0.111025087 0.1569109922 0.680972858
24 0.071109243 0.017231035 0.176996179 0.1133525441 0.621310999
25 0.008856887 0.001721330 0.045207286 0.0136219535 0.930592543
26 0.048791656 0.003690467 0.894297433 0.0072172485 0.046003195
27 0.027439139 0.007013444 0.094614939 0.1181183632 0.752814115
28 0.028185391 0.006476169 0.094058764 0.0508080017 0.820471674
29 0.169993380 0.018990909 0.461389485 0.0450112107 0.304615015
30 0.290931721 0.013967680 0.593489307 0.0196046409 0.082006652
31 0.089928728 0.010855100 0.731237685 0.0224848375 0.145493649
32 0.037114065 0.004538834 0.862620107 0.0119235821 0.083803412
33 0.098534864 0.018118543 0.442466870 0.0512040039 0.389675719
34 0.002326820 0.994079361 0.001623192 0.0007494918 0.001221136

> res.fcm.d <- ppclust2(res.fcm.5, "fclust") # optimal
> fpc.5 <- FC(res.fcm.d$U)
> fpc.5
[1] 0.6222577
> fmpc.5 <- MPC(res.fcm.d$U)
> fmpc.5
[1] 0.5278221
>

```

## 5. Fuzzy Possibilistics C Means (2 cluster)

```

> #fpcm cluster = 2
> res.fpcm.2 <- fpcm(data, centers=res.fcm.2$v, memberships=res.fcm.2$u)
> summary(res.fpcm.2)
Summary for 'res.fpcm.2'

Number of data objects: 34

Number of clusters: 2

Crisp clustering vector:
[1] 2 1 2 1 2 2 2 2 1 1 1 2 1 1 1 1 1 1 2 2 2 2 1 1 1 2 1 1 2 2 2 2 2

Initial cluster prototypes:
      sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1      76.29368                98.12327      74.30532
Cluster 2      63.92950                93.69514      55.72681

Final cluster prototypes:
      sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1      76.30034                98.13027      74.31781
Cluster 2      63.94282                93.72045      55.77541

Distance between the final cluster prototypes
      Cluster 1
Cluster 2 515.9759

Difference between the initial and final cluster prototypes
      sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1      0.006661246                0.006997074      0.01249315
Cluster 2      0.013321046                0.025308820      0.04859038

Root Mean Squared Deviations (RMSD): 0.04140303
Mean Absolute Deviation (MAD): 0.1700576

```

```

Membership degrees matrix (top and bottom 5 rows):
      [,1]      [,2]
[1,] 0.006695882 0.04770969
[2,] 0.041007080 0.01189692
[3,] 0.003237625 0.08336274
[4,] 0.077376137 0.01227142
[5,] 0.007865581 0.04491182
...
      [,1]      [,2]
[30,] 0.003566171 0.151649114
[31,] 0.007292723 0.050823813
[32,] 0.008719172 0.037044799
[33,] 0.010802072 0.022006538
[34,] 0.000444467 0.002093469

Descriptive statistics for the membership degrees by clusters
      Size      Min      Q1      Mean      Median      Q3
Cluster 1  16 0.003532166 0.01031793 0.05667524 0.03253400 0.05660948
Cluster 2  18 0.002093469 0.02727400 0.04822846 0.03706197 0.05994019
      Max
Cluster 1 0.2816499
Cluster 2 0.1516491

Dunn's Fuzziness Coefficients:
dunn_coef  normalized
0.006065586 -0.987868828

Within cluster sum of squares by cluster:
      1      2
1911.340 4400.912
(between_SS / total_SS = 40.91%)

Available components:
 [1] "u"      "t"      "v"      "v0"     "d"
 [6] "x"      "cluster" "csize"  "sumsqrs" "k"
[11] "m"      "eta"    "iter"   "best.start" "func.val"
[16] "comp.time" "inpargs" "algorithm" "call"

> res.fpcm.2$comp.time
[1] 0.17
> res.fpcm.2$iter
[1] 47
> res.fpcm.2$func.val
[1] 4755.07

> res.fpcm.2$u
      Cluster 1 Cluster 2
1  0.25641229 0.743587708
2  0.89439169 0.105608307
3  0.08711192 0.912888084
4  0.93936586 0.060634140
5  0.30084807 0.699151934
6  0.38312709 0.616872909
7  0.22451897 0.775481030
8  0.12643238 0.873567615
9  0.84256257 0.157437434
10 0.86993783 0.130062172
11 0.81111510 0.188884899
12 0.53526784 0.464732162
13 0.97937014 0.020629864
14 0.85314580 0.146854200
15 0.91014590 0.089854096
16 0.72206486 0.277935137
17 0.80496646 0.195033542
18 0.80748379 0.192516212
19 0.21340999 0.786590010
20 0.08481309 0.915186914
21 0.10121267 0.898787326
22 0.08267168 0.917328315
23 0.93762805 0.062371954
24 0.81931945 0.180680547
25 0.95922164 0.040778359
26 0.12866437 0.871335634
27 0.99038425 0.009615745
28 0.92207075 0.077929252
29 0.36460487 0.635395125
30 0.05462266 0.945377336
31 0.26065898 0.739341015
32 0.36640702 0.633592980
33 0.54669855 0.453301448
34 0.34281830 0.657181703

```

```

> res.fpcm.2$t
      [,1]      [,2]
[1,] 0.0066958822 0.047709693
[2,] 0.0410070795 0.011896922
[3,] 0.0032376250 0.083362737
[4,] 0.0773761371 0.012271423
[5,] 0.0078655814 0.044911816
[6,] 0.0078644378 0.031111859
[7,] 0.0008202423 0.006960906
[8,] 0.0017422291 0.029576645
[9,] 0.0110017915 0.005050969
[10,] 0.0082663311 0.003036554
[11,] 0.0037793177 0.002162388
[12,] 0.0124599516 0.026579903
[13,] 0.1801114331 0.009321724
[14,] 0.0072787282 0.003078393
[15,] 0.0496872666 0.012052508
[16,] 0.0147596971 0.013958866
[17,] 0.0035321660 0.002102704
[18,] 0.0274989430 0.016108494
[19,] 0.0012064304 0.010925502
[20,] 0.0023754357 0.062978986
[21,] 0.0016994261 0.037079150
[22,] 0.0033007019 0.089986998
[23,] 0.0375690657 0.006140371
[24,] 0.0145289705 0.007872242
[25,] 0.1044047531 0.010905272
[26,] 0.0062475423 0.103954154
[27,] 0.2816498700 0.006718839
[28,] 0.0443523056 0.009209964
[29,] 0.0068560543 0.029356286
[30,] 0.0035661706 0.151649114
[31,] 0.0072927225 0.050823813
[32,] 0.0087191716 0.037044799
[33,] 0.0108020724 0.022006538
[34,] 0.0004444671 0.002093469

> fppc.2<- PC(res.fpcm.2$u)
> fppc.2
[1] 0.7309582
> fmpc.2 <- MPC(res.fpcm.2$u)
> fmpc.2
[1] 0.4619165
>

```

## 6. Fuzzy Possibilistics C Means (3 cluster)

```

> #fpcm cluster = 3
> res.fpcm.3 <- fpcm(data, centers=res.fcm.3$V, memberships=res.fcm.3$U)
> summary(res.fpcm.3)
Summary for 'res.fpcm.3'

Number of data objects: 34

Number of clusters: 3

Crisp clustering vector:
[1] 1 1 3 1 1 1 3 3 1 2 2 1 1 2 1 1 2 1 3 3 3 3 1 1 1 3 1 1 3 3 1 1 1 3

Initial cluster prototypes:
      sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1      70.24860                97.06318      66.60723
Cluster 2      82.63893                99.14665      85.28590
Cluster 3      61.44288                90.93720      49.66816

Final cluster prototypes:
      sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1      70.37917                97.09452      66.75259
Cluster 2      83.02097                99.15518      85.70245
Cluster 3      61.57163                91.19892      48.88904

Distance between the final cluster prototypes
      Cluster 1 Cluster 2
Cluster 2  523.1584
Cluster 3  431.4376 1878.6036

Difference between the initial and final cluster prototypes
      sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1      0.1305737                0.031343472      0.1453593
Cluster 2      0.3820406                0.008538047      0.4165459
Cluster 3      0.1287488                0.261718355      0.2208805

Root Mean Squared Deviations (RMSD): 0.4051998
Mean Absolute Deviation (MAD): 1.725749

```

```

Membership degrees matrix (top and bottom 5 rows):
      [,1]      [,2]      [,3]
[1,] 0.03995092 0.001334113 0.02736147
[2,] 0.04649818 0.003327974 0.01144421
[3,] 0.00919191 0.000858330 0.06920757
[4,] 0.05453249 0.003522157 0.01166789
[5,] 0.06315545 0.001449083 0.02593330
...
      [,1]      [,2]      [,3]
[30,] 0.012138445 0.000919703 0.059076723
[31,] 0.037110915 0.001419296 0.030460840
[32,] 0.060497957 0.001569230 0.023700025
[33,] 0.031153563 0.001853650 0.019097514
[34,] 0.000614719 0.000211634 0.005111242

Descriptive statistics for the membership degrees by clusters
      Size      Min      Q1      Mean      Median      Q3      Max
Cluster 1  19 0.006250296 0.02310476 0.0463719 0.03995092 0.05068979 0.1558669
Cluster 2   4 0.013020577 0.01599128 0.2325423 0.01995020 0.23650122 0.8772482
Cluster 3  11 0.005111242 0.03018924 0.0635543 0.05178747 0.06414215 0.1917311

Dunn's Fuzziness Coefficients:
dunn_coeff normalized
0.02715236 -0.45927146

Within cluster sum of squares by cluster:
      1      2      3
1252.910 125.703 3202.010
(between_SS / total_SS = 57.72%)

Available components:
[1] "u"      "t"      "v"      "v0"     "d"
[6] "x"      "cluster" "csize"  "sumsqrs" "k"
[11] "m"      "eta"     "iter"   "best.start" "func.val"
[16] "comp.time" "inpargs" "algorithm" "call"

> res.fpcm.3$comp.time
[1] 0.45
> res.fpcm.3$iter
[1] 87
> res.fpcm.3$func.val
[1] 2850.724
. .
> res.fpcm.3$u
      Cluster 1 Cluster 2 Cluster 3
1  0.821658339 0.04527689 0.133064773
2  0.851225966 0.10053291 0.048241125
3  0.340780028 0.05250999 0.606709986
4  0.864175031 0.09210317 0.043721803
5  0.881089159 0.03335968 0.085551163
6  0.852226305 0.04869288 0.099080811
7  0.218177679 0.08274597 0.699076347
8  0.215056894 0.05247867 0.732464434
9  0.397776716 0.51353196 0.088691319
10 0.002198549 0.99716208 0.000639371
11 0.062855324 0.91434465 0.022800024
12 0.946910953 0.02088028 0.032208764
13 0.736633950 0.20321351 0.060152540
14 0.076630789 0.90120346 0.022165747
15 0.853017695 0.09732637 0.049655932
16 0.842467702 0.08831712 0.069215174
17 0.076647403 0.89484216 0.028510439
18 0.954042332 0.02608174 0.019875928
19 0.185548426 0.06705460 0.747396969
20 0.134686695 0.02814603 0.837167279
21 0.073574132 0.01831228 0.908113588
22 0.432903570 0.05976860 0.507327827
23 0.425495725 0.49834630 0.076157976
24 0.587988399 0.28724824 0.124763365
25 0.836222160 0.11803432 0.045743521
26 0.769347620 0.04171709 0.188935293
27 0.449746776 0.49134162 0.058911606
28 0.679187161 0.23693338 0.083879461
29 0.722892274 0.07759686 0.199510862
30 0.439394242 0.05493616 0.505669598
31 0.795420136 0.05019812 0.154381746
32 0.880719367 0.03769667 0.081583961
33 0.804416510 0.07898062 0.116602872
34 0.282948044 0.16074380 0.556308156

```



```

> res.fpcm.3$t
      [,1]      [,2]      [,3]
[1,] 0.0399509191 0.0013341131 0.027361472
[2,] 0.0464981809 0.0033279739 0.011144212
[3,] 0.0091919100 0.0008583296 0.069207574
[4,] 0.0545324884 0.0035221570 0.011667895
[5,] 0.0631554477 0.0014490834 0.025933297
[6,] 0.0433783513 0.0015019800 0.021327921
[7,] 0.0014205553 0.0003264940 0.019249226
[8,] 0.0038704709 0.0005723660 0.055749097
[9,] 0.0062502959 0.0048900045 0.005893634
[10,] 0.0031916290 0.8772482348 0.003925274
[11,] 0.0019263183 0.0169815161 0.002955032
[12,] 0.1353425758 0.0018085988 0.019468846
[13,] 0.0274592235 0.0045906021 0.009482686
[14,] 0.0032158325 0.0229188874 0.003933810
[15,] 0.0468470869 0.0032391842 0.011532847
[16,] 0.0360035636 0.0022872714 0.012509340
[17,] 0.0018403510 0.0130205768 0.002894996
[18,] 0.1558668744 0.0025822798 0.013732653
[19,] 0.0020819372 0.0004559524 0.035465211
[20,] 0.0055900761 0.0007079306 0.146942061
[21,] 0.0036731331 0.0005540316 0.191731093
[22,] 0.0104492793 0.0008742766 0.051787472
[23,] 0.0093939893 0.0066675557 0.007110680
[24,] 0.0101392795 0.0030017621 0.009098426
[25,] 0.0454355594 0.0038865393 0.010511015
[26,] 0.0383842476 0.0012613191 0.039864320
[27,] 0.0132994884 0.0088050329 0.007367304
[28,] 0.0187502988 0.0039639259 0.009792988
[29,] 0.0213450378 0.0013885070 0.024913271
[30,] 0.0121384452 0.0009197029 0.059076723
[31,] 0.0371109146 0.0014192963 0.030460840
[32,] 0.0604979572 0.0015692305 0.023700025
[33,] 0.0311535626 0.0018536504 0.019097514
[34,] 0.0006147194 0.0002116338 0.005111242

> fppc.3<- FC(res.fpcm.3$u)
> fppc.3
[1] 0.6592364
> fpmpc.3 <- MPC(res.fpcm.3$u)
> fpmpc.3
[1] 0.4888546
>

```

## 7. Fuzzy Possibilistics C Means (4 cluster)

```

> #fpcm cluster = 4
> res.fpcm.4 <- fpcm(data, centers=res.fcm.4$v, memberships=res.fcm.4$u)
> summary(res.fpcm.4)
Summary for 'res.fpcm.4'

Number of data objects: 34

Number of clusters: 4

Crisp clustering vector:
[1] 2 1 2 1 2 2 2 2 1 4 4 1 1 4 1 1 4 1 3 2 2 2 1 1 1 2 1 1 2 2 2 2 1 3

Initial cluster prototypes:
      sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1      72.89669                97.88102      69.23534
Cluster 2      63.18358                96.19473      55.82167
Cluster 3      59.70715                60.71960      35.48187
Cluster 4      84.31286                99.54545      87.81336

Final cluster prototypes:
      sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1      72.71155                97.83808      69.08016
Cluster 2      63.02263                96.08364      55.41586
Cluster 3      59.31247                57.80122      34.03300
Cluster 4      84.13921                99.45776      87.43281

Distance between the final cluster prototypes
      Cluster 1 Cluster 2 Cluster 3
Cluster 2 283.6662
Cluster 3 3010.7892 1936.5361
Cluster 4 470.0346 1482.3797 5203.1742

Difference between the initial and final cluster prototypes
      sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1      -0.1851385                -0.04294065      -0.1551776
Cluster 2      -0.1609509                -0.11108318      -0.4058071
Cluster 3      -0.3946805                -2.91837860      -1.4488633
Cluster 4      -0.1736414                -0.08769522      -0.3805489

```

```

Root Mean Squared Deviations (RMSD): 1.674641
Mean Absolute Deviation (MAD): 4.848679

Membership degrees matrix (top and bottom 5 rows):
      [,1]      [,2]      [,3]      [,4]
[1,] 0.01407853 0.05001858 0.001794153 0.001251343
[2,] 0.05980658 0.01045638 0.001451283 0.002898778
[3,] 0.00479284 0.08627674 0.002276027 0.000827042
[4,] 0.16756341 0.01045775 0.001519743 0.003038902
[5,] 0.01856359 0.04284489 0.001813003 0.001350633
...
      [,1]      [,2]      [,3]      [,4]
[30,] 0.005535064 0.166996637 0.002254041 0.000881855
[31,] 0.014221361 0.033463968 0.002253269 0.001324589
[32,] 0.019198934 0.029777887 0.001935515 0.001454659
[33,] 0.018123683 0.016015016 0.002101968 0.001702479
[34,] 0.000414609 0.001764373 0.936968027 0.000216758

Descriptive statistics for the membership degrees by clusters
      Size      Min      Q1      Mean      Median      Q3
Cluster 1  14 0.005798485 0.01825993 0.06126934 0.03449243 0.09645537
Cluster 2  14 0.007062804 0.03086846 0.05943576 0.04029631 0.06830405
Cluster 3   2 0.008539020 0.24064627 0.47275352 0.47275352 0.70486078
Cluster 4   4 0.021767903 0.02248527 0.23472269 0.02747740 0.23971482
      Max
Cluster 1 0.1786077
Cluster 2 0.1684345
Cluster 3 0.9369680
Cluster 4 0.8621680

Dunn's Fuzziness Coefficients:
dunn_coeff normalized
0.05318513 -0.26241983

Within cluster sum of squares by cluster:
      1      2      3      4
749.9130 1579.3509 325.8985 125.7030
(between_SS / total_SS = 78.49%)

Available components:
 [1] "u"      "c"      "v"      "r0"     "d"
 [6] "x"      "cluster" "csize"  "sumsqrs" "k"
[11] "m"      "eta"     "iter"   "best.start" "func.val"
[16] "comp.time" "inpargs" "algorithm" "call"
> res.fpcm.4$comp.time
[1] 0.88
> res.fpcm.4$iter
[1] 119
> res.fpcm.4$func.val
[1] 1779.072

> res.fpcm.4$u
      Cluster 1      Cluster 2      Cluster 3      Cluster 4
1  0.415472329 0.5273676973 0.0153451343 0.0418148391
2  0.889379456 0.0555541803 0.0062548436 0.0488115203
3  0.128794460 0.8283144158 0.0177258844 0.0251652394
4  0.956515261 0.0213278981 0.0025142547 0.0196425861
5  0.516723112 0.4260811415 0.0146258449 0.0425699012
6  0.558608080 0.3695976222 0.0171575481 0.0546367492
7  0.197034966 0.5644953432 0.1553696966 0.0830999946
8  0.128561948 0.7953725071 0.0382686713 0.0377968734
9  0.456080201 0.1280072646 0.0249786651 0.3909338694
10 0.003229045 0.0009793739 0.0002721712 0.9955194099
11 0.045590243 0.0175179978 0.0058078416 0.9310839175
12 0.793706460 0.1631192668 0.0087483870 0.0344258864
13 0.882120786 0.0452030187 0.0059053101 0.0667708855
14 0.102518997 0.0327117835 0.0086907247 0.8560784945
15 0.935801373 0.0330198154 0.0035716333 0.0276071783
16 0.767994524 0.1456247721 0.0129894000 0.0733913043
17 0.062418841 0.0244802929 0.0083082528 0.9047926132
18 0.955809953 0.0279229795 0.0023272297 0.0139398379
19 0.179077251 0.4074438764 0.3423766347 0.0711022375
20 0.153547664 0.7626904791 0.0452521043 0.0385097525
21 0.123681210 0.7910640671 0.0490598921 0.0361948306
22 0.075125082 0.9017872681 0.0088122604 0.0142753896
23 0.575271592 0.0955231182 0.0178121349 0.3113931547
24 0.639213960 0.1474754896 0.0253701301 0.1879404203
25 0.943385230 0.0257249520 0.0029798831 0.0279099350
26 0.312607012 0.6401052471 0.0149716782 0.0323160624
27 0.648601563 0.0734062170 0.0124094352 0.2655827843
28 0.783364783 0.0854308334 0.0129318829 0.1182725008
29 0.526689534 0.3832175014 0.0249584215 0.0651345433
30 0.082798442 0.8924923781 0.0097720950 0.0149370849
31 0.501990317 0.4220159672 0.0230511968 0.0529425188
32 0.599109498 0.3319863529 0.0175045888 0.0513995606
33 0.686963279 0.2168760936 0.0230908209 0.0730698069
34 0.001519598 0.0023103496 0.9952704855 0.0008995672

```

```

> res.fpcm.4$t
      [,1]      [,2]      [,3]      [,4]
[1,] 0.0140785268 0.050018582 0.0017941532 0.0012513429
[2,] 0.0598065771 0.010456380 0.0014512827 0.0028987781
[3,] 0.0047928400 0.086276736 0.0022760268 0.0008270421
[4,] 0.1675634149 0.010457746 0.0015197435 0.0030389024
[5,] 0.0185635846 0.042844888 0.0018130035 0.0013506331
[6,] 0.0162010915 0.030003296 0.0017169830 0.0013994349
[7,] 0.0008807596 0.007062804 0.0023963713 0.0003280544
[8,] 0.0021671931 0.037528264 0.0022258826 0.0005626934
[9,] 0.0057984845 0.004555233 0.0010957620 0.0043894223
[10,] 0.0031665437 0.002688207 0.0009209314 0.8621680414
[11,] 0.0017869738 0.001921912 0.0007854793 0.0322304141
[12,] 0.0433358601 0.024928463 0.0016481214 0.0016599822
[13,] 0.0577620870 0.008284857 0.0013342317 0.0038612929
[14,] 0.0030814294 0.002752041 0.0009013170 0.0227243904
[15,] 0.1086716294 0.010732722 0.0014311102 0.0028312962
[16,] 0.0245842624 0.013047793 0.0014347025 0.0020747915
[17,] 0.0017004054 0.001866621 0.0007809445 0.0217679031
[18,] 0.1786077257 0.014604718 0.0015005195 0.0023004733
[19,] 0.0012944063 0.008243293 0.0085390203 0.0004538834
[20,] 0.0031105438 0.043245823 0.0031630467 0.0006889610
[21,] 0.0021085319 0.037747730 0.0028858699 0.0005449464
[22,] 0.0050131328 0.168434514 0.0020290151 0.0008412847
[23,] 0.0115614012 0.005373395 0.0012351714 0.0055268463
[24,] 0.0104630508 0.006756700 0.0014328756 0.0027168337
[25,] 0.1271749230 0.009706658 0.0013860706 0.0033227833
[26,] 0.0129811411 0.074399203 0.0021451534 0.0011851229
[27,] 0.0186686553 0.005913854 0.0012324258 0.0067509634
[28,] 0.0256490022 0.007829316 0.0014609721 0.0034199598
[29,] 0.0119321723 0.024300371 0.0019509928 0.0013031877
[30,] 0.0055350638 0.166996637 0.0022540413 0.0008818548
[31,] 0.0142213607 0.033463968 0.0022532688 0.0013245892
[32,] 0.0191989335 0.029777887 0.0019355150 0.0014546586
[33,] 0.0181236835 0.016015016 0.0021019677 0.0017024787
[34,] 0.0004146088 0.001764373 0.9369680273 0.0002167576

> fppc.4<- PC(res.fpcm.4$t)
> fppc.4
[1] 0.6456312
> fpmpc.4 <- MPC(res.fpcm.4$t)
> fpmpc.4
[1] 0.5275083
`

```

## 8. Fuzzy Possibilistics C Means (5 cluster)

```

> #fpcm cluster = 5
> res.fpcm.5 <- fpcm(data, centers=res.fcm.5$t, memberships=res.fcm.5$t)
> summary(res.fpcm.5)
Summary for 'res.fpcm.5'

Number of data objects: 34

Number of clusters: 5

Crisp clustering vector:
[1] 3 5 1 5 3 3 1 1 4 4 4 3 5 4 5 3 4 5 1 1 1 1 5 5 5 3 5 5 1 1 3 3 1 2

Initial cluster prototypes:
      sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1      61.56990      95.29674      49.53832
Cluster 2      59.34234      57.64635      33.98936
Cluster 3      66.27916      96.50010      63.48665
Cluster 4      85.30456      99.62232      88.70126
Cluster 5      75.32641      98.30732      70.88147

Final cluster prototypes:
      sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1      62.01426      95.15676      49.39738
Cluster 2      59.19414      56.65303      33.48176
Cluster 3      66.18389      96.56318      63.49451
Cluster 4      85.15059      99.57881      88.48721
Cluster 5      75.30841      98.30191      70.85653

Distance between the final cluster prototypes
      Cluster 1 Cluster 2 Cluster 3 Cluster 4
Cluster 2 1743.7971
Cluster 3 218.0928 2542.4410
Cluster 4 2082.8591 5541.9590 993.4650
Cluster 5 647.1213 3391.1718 140.4794 409.3400

```

```

Difference between the initial and final cluster prototypes
  sumber.air.yang.layak penerangan.dengan.liistrik sanitasi.layak
Cluster 1      0.44436082      -0.139979084      -0.140944061
Cluster 2      -0.14819636      -0.993315589      -0.507596482
Cluster 3      -0.09526578      0.063076910      0.007858159
Cluster 4      -0.15397104      -0.043510269      -0.214051745
Cluster 5      -0.01799587      -0.005407384      -0.024941932

Root Mean Squared Deviations (RMSD): 0.5636874
Mean Absolute Deviation (MAD): 1.800283

Membership degrees matrix (top and bottom 5 rows):
  [,1] [,2] [,3] [,4] [,5]
[1,] 0.026281104 0.000326801 0.130038026 0.003276401 0.005361848
[2,] 0.009156901 0.000266133 0.009587942 0.007238895 0.027076699
[3,] 0.088515774 0.000411889 0.008238018 0.002211109 0.002406910
[4,] 0.009566161 0.000278469 0.008295610 0.007605049 0.269340762
[5,] 0.023846788 0.000330231 0.383442886 0.003523572 0.006528160
...
  [,1] [,2] [,3] [,4] [,5]
[30,] 0.055451642 0.000408140 0.02120534 0.002342154 0.002521995
[31,] 0.023198484 0.000408619 0.03194392 0.003467921 0.005767179
[32,] 0.019377192 0.000352301 0.07914895 0.003782044 0.006827795
[33,] 0.014290353 0.000382438 0.01092937 0.004424365 0.008636029
[34,] 0.003148604 0.988596278 0.00038598 0.000602894 0.000254192

Descriptive statistics for the membership degrees by clusters
  Size  Min  Q1  Mean  Median  Q3  Max
Cluster 1  10  0.014290353 0.01807626 0.07216221 0.05792335 0.0817343 0.2904340
Cluster 2  1  0.988596278 0.98859628 0.98859628 0.98859628 0.9885963 0.9885963
Cluster 3  8  0.012971528 0.03289594 0.10536182 0.07156634 0.1136237 0.3834429
Cluster 4  5  0.010887743 0.05312960 0.17136056 0.09322700 0.1593156 0.5402429
Cluster 5  10 0.009829646 0.02548983 0.09013732 0.03392896 0.1399308 0.2693408

Dunn's Fuzziness Coefficients:
dunn_coeff normalized
0.05223491 -0.18470637

Within cluster sum of squares by cluster:
  1 2 3 4 5
1662.6727 0.0000 167.4165 386.3427 280.7549
(between_SS / total_SS = 80.79%)

Available components:
[1] "u" "c" "v" "v0" "d"
[6] "x" "cluster" "csize" "sumsqrs" "k"
[11] "m" "eta" "iter" "best.start" "func.val"
[16] "comp.time" "inpargs" "algorithm" "call"
> res.fpcm.5$comp.time
[1] 0.73
> res.fpcm.5$iter
[1] 79
> res.fpcm.5$func.val
[1] 1178.984

> res.fpcm.5$u
  Cluster 1 Cluster 2 Cluster 3 Cluster 4 Cluster 5
1 0.0327693134 0.0026313174 0.9145507083 0.0068655922 0.0431830688
2 0.0363349412 0.0069193084 0.2145927424 0.0482731189 0.6939798892
3 0.5641390106 0.0169516249 0.2961432718 0.0236827800 0.0990833127
4 0.0052832803 0.0009931356 0.0258421111 0.0070587068 0.9608227662
5 0.010608444 0.0009533352 0.9668878156 0.0026472927 0.0188507121
6 0.0445840891 0.0047373352 0.8452553910 0.0143407502 0.0910824345
7 0.5219559241 0.0982836328 0.2146362855 0.0526893967 0.1124347609
8 0.6429176828 0.0277300510 0.2244727988 0.0271549533 0.0777245140
9 0.0722601556 0.0200678108 0.2134541968 0.2810953640 0.4131224728
10 0.0032927308 0.0011765511 0.0072634416 0.9688007870 0.0194664896
11 0.0080923892 0.0033099826 0.0155819739 0.9394075469 0.0336081074
12 0.0599382340 0.0064844390 0.6251579579 0.0241036457 0.2843157234
13 0.0055179595 0.0011017560 0.0223330660 0.0111238105 0.9599234081
14 0.0264797479 0.0092214357 0.0610772936 0.7630112008 0.1402103221
15 0.0107507495 0.0018504032 0.0492682352 0.0130939143 0.9250366978
16 0.0670391597 0.0106314122 0.4580981032 0.0557234447 0.4085078803
17 0.0130296871 0.0054155228 0.0245687009 0.9043920610 0.0525940282
18 0.0358567590 0.0053626144 0.2902409480 0.0296533684 0.6388863102
19 0.4046534349 0.2298960113 0.1985004555 0.0504471451 0.1165029532
20 0.6931245470 0.0260798238 0.1826597266 0.0222049835 0.0759309192
21 0.9048352648 0.0104379396 0.0540806286 0.0077248059 0.0229213611
22 0.3643780353 0.0141154125 0.5093602283 0.0223346696 0.0898116543
23 0.0408031466 0.0103871274 0.1102472677 0.1616705313 0.6768919269
24 0.0723468677 0.0166023371 0.1751295715 0.1156171854 0.6203040383
25 0.0088816138 0.0016566118 0.0448416904 0.0139579672 0.9306621168
26 0.0512388010 0.0036629859 0.8893929704 0.0076300663 0.0480851763
27 0.0276535640 0.0068015923 0.0944357319 0.1228304665 0.7482786453
28 0.0286343794 0.0062534738 0.0931983680 0.0522075037 0.8197062752
29 0.1758472234 0.0181822026 0.4539850755 0.0457861423 0.3061993562
30 0.2801706280 0.0133162941 0.6043202851 0.0198875311 0.0823052617
31 0.0931313159 0.0105930357 0.7233325222 0.0233971226 0.1495460037
32 0.0373683502 0.0043872463 0.8609381274 0.0122573561 0.0850489201
33 0.01009039335 0.0174377845 0.4352852133 0.0525016780 0.3938713907
34 0.0004925961 0.9987514159 0.0003406048 0.0001585153 0.0002568679

```

```

> res.fpcm.5$t
      [,1]      [,2]      [,3]      [,4]      [,5]
[1,] 0.026281104 0.0003268014 0.1300380260 0.0032764015 0.0053618485
[2,] 0.009156901 0.0002661331 0.0095879423 0.0072388950 0.0270766993
[3,] 0.088515774 0.0004118891 0.0082380177 0.0022111086 0.0024060908
[4,] 0.009566161 0.0002784691 0.0082956101 0.0076050491 0.2693407625
[5,] 0.023846788 0.0003302314 0.3834428858 0.0035235715 0.0065281604
[6,] 0.019035979 0.0003132304 0.0639837322 0.0036434266 0.0060208070
[7,] 0.014948545 0.0004358938 0.0010898194 0.0008979066 0.0004985290
[8,] 0.060395053 0.0004033955 0.0037384883 0.0015178816 0.0011303912
[9,] 0.004703707 0.0002022906 0.0024633852 0.0108877428 0.0041633691
[10,] 0.003085801 0.0001707485 0.0012068144 0.5402428930 0.0028243866
[11,] 0.002306418 0.0001460903 0.0007873540 0.1593155501 0.0014829623
[12,] 0.017961379 0.0003009141 0.0332132838 0.0042979497 0.0131905039
[13,] 0.007926504 0.0002450885 0.0056877148 0.0095082279 0.2134835939
[14,] 0.003098679 0.0001671074 0.0012671525 0.0531296003 0.0025401978
[15,] 0.009844460 0.0002623940 0.0079984615 0.0071345318 0.1311405104
[16,] 0.010707175 0.0002629491 0.0129715280 0.0052957459 0.0101011668
[17,] 0.002257232 0.0001452837 0.0007545893 0.0932270038 0.0014105983
[18,] 0.011862035 0.0002747256 0.0170228850 0.0058372064 0.0327217500
[19,] 0.016711916 0.0014703099 0.0014534190 0.0012397167 0.0007449115
[20,] 0.097315665 0.0005670359 0.0045467449 0.0018550891 0.0016504976
[21,] 0.290433978 0.0005188324 0.0030775559 0.0014753951 0.0011390508
[22,] 0.061389868 0.0003682752 0.0152144357 0.0022390659 0.0023426207
[23,] 0.005771177 0.0002275101 0.0027645529 0.0136064386 0.0148222746
[24,] 0.007405054 0.0002631553 0.0031780028 0.0070416385 0.0098296456
[25,] 0.008806201 0.0002543618 0.0078825188 0.0082349649 0.1428609500
[26,] 0.035144659 0.0003890721 0.1081522334 0.0031140927 0.0051061827
[27,] 0.005958307 0.0002269430 0.0036074008 0.0157478288 0.0249608779
[28,] 0.007927920 0.0002681188 0.0045747365 0.0086009662 0.0351361758
[29,] 0.022169284 0.0003549744 0.0101471488 0.0034347335 0.0059764761
[30,] 0.055431642 0.0004081404 0.0212053425 0.0023421544 0.0025219947
[31,] 0.023198484 0.0004086192 0.0319439172 0.0034679208 0.0057671792
[32,] 0.019377192 0.0003523006 0.0791489516 0.0037820439 0.0068277952
[33,] 0.014290353 0.0003824376 0.0109293693 0.0044243646 0.0086360287
[34,] 0.003148604 0.9885962775 0.0003859795 0.0006028944 0.0002541919

> fppc.5<- PC(res.fpcm.5$u)
> fppc.5
[1] 0.6235363
> fpmpc.5 <- MPC(res.fpcm.5$u)
> fpmpc.5
[1] 0.5294204

```

## 9. Possibilistics Fuzzy C Means (2 cluster)

```

> #pfc cluster = 2
> res.pfc.2 <- pfc(data, centers=res.fcm.2$w, memberships=res.fcm.2$u,a=5,b=5)
> summary(res.pfc.2)
Summary for 'res.pfc.2'

Number of data objects: 34

Number of clusters: 2

Crisp clustering vector:
[1] 2 1 2 1 2 2 2 2 1 1 1 2 1 1 1 1 1 1 2 2 2 2 1 1 1 2 1 1 2 2 2 2 2

Initial cluster prototypes:
      sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1      76.29368                98.12327      74.30532
Cluster 2      63.92950                93.69514      55.72681

Final cluster prototypes:
      sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1      76.07830                98.16445      73.66066
Cluster 2      64.55387                94.46128      57.41417

Distance between the final cluster prototypes
      Cluster 1
Cluster 2 410.474

Difference between the initial and final cluster prototypes
      sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1      -0.2153855                0.04118301     -0.644665
Cluster 2      0.6243756                0.76614130      1.687359

Root Mean Squared Deviations (RMSD): 1.464186
Mean Absolute Deviation (MAD): 5.968664

```

```

Membership degrees matrix (top and bottom 5 rows):
  Cluster 1 Cluster 2
1 0.12295325 0.4459951
2 0.46763729 0.1387905
3 0.06247924 0.4709153
4 0.66797142 0.1413890
5 0.14219082 0.4323720
...
  Cluster 1 Cluster 2
30 0.068030767 0.69134891
31 0.132247220 0.43707673
32 0.154385041 0.36735789
33 0.184897361 0.22722114
34 0.008660528 0.02102014

Descriptive statistics for the membership degrees by clusters
      Size      Min      Q1      Mean      Median      Q3      Max
Cluster 1   16 0.06063559 0.1574439 0.3943756 0.398704 0.5792206 0.8366906
Cluster 2   18 0.02102014 0.2298625 0.3434156 0.333635 0.4437655 0.6913489

Dunn's Fuzziness Coefficients:
dunn_coeff normalized
0.1963561 -0.6072878

Within cluster sum of squares by cluster:
  1      2
1911.340 4400.912
(between_SS / total_SS = 35.52%)

Available components:
 [1] "u"      "t"      "v"      "v0"     "d"
 [6] "x"      "cluster" "csize"  "sumsqrs" "k"
[11] "m"      "eta"    "a"      "b"      "omega"
[16] "iter"   "best.start" "func.val" "comp.time" "inpargs"
[21] "algorithm" "call"

> res.pfcm.2$comp.time
[1] 0.22
> res.pfcm.2$iter
[1] 57
> res.pfcm.2$func.val
[1] 34095.25

> res.pfcm.2$u
  Cluster 1 Cluster 2
1 0.19539329 0.80460671
2 0.88373564 0.11626436
3 0.09454319 0.90545681
4 0.94455783 0.05544217
5 0.23281556 0.76718444
6 0.32354866 0.67645134
7 0.24803488 0.75196512
8 0.14494430 0.85505570
9 0.81601927 0.18398073
10 0.84672613 0.15327387
11 0.78808021 0.21191979
12 0.48700155 0.51299845
13 0.98313566 0.01686434
14 0.82792349 0.17207651
15 0.91315401 0.08684599
16 0.68915083 0.31084917
17 0.78242673 0.21757327
18 0.79231703 0.20768297
19 0.24487612 0.75512388
20 0.11402276 0.88597724
21 0.13318231 0.86681769
22 0.07303928 0.92696072
23 0.92752199 0.07247801
24 0.80885557 0.19114443
25 0.96260722 0.03739278
26 0.07564902 0.92435098
27 0.98211976 0.01788024
28 0.91909420 0.08090580
29 0.34125862 0.65874138
30 0.04347081 0.95652919
31 0.21489836 0.78510164
32 0.30481037 0.69518963
33 0.51826868 0.48173132
34 0.36199859 0.63800141

```

```

> res.pfcm.2$t
      Cluster 1 Cluster 2
1  0.122953247 0.44599506
2  0.467637292 0.13879050
3  0.062479243 0.47091526
4  0.667971421 0.14138901
5  0.142190821 0.43237204
6  0.140907670 0.32350526
7  0.016052251 0.06452197
8  0.033896122 0.22398333
9  0.167476100 0.05948648
10 0.127347464 0.03552943
11 0.064375853 0.02515269
12 0.213705725 0.28533002
13 0.836690574 0.10917587
14 0.114538063 0.03613698
15 0.549637022 0.13931240
16 0.234367508 0.16145890
17 0.060635593 0.02441988
18 0.383501928 0.18526010
19 0.023490880 0.09374935
20 0.046114950 0.34376480
21 0.033229773 0.23778647
22 0.063290588 0.54458840
23 0.413906045 0.07145684
24 0.229210911 0.08925122
25 0.729935862 0.12771450
26 0.115804292 0.69056495
27 0.772058700 0.07918345
28 0.490719853 0.10577183
29 0.126438150 0.28037867
30 0.068030767 0.69134891
31 0.132247220 0.43707673
32 0.154385041 0.36735789
33 0.184897361 0.22722114
34 0.008660528 0.02102014

> pfpc.2<- PC(res.pfcm.2$u)
> pfpc.2
[1] 0.7290381
> pfmpc.2 <- MPC(res.pfcm.2$u)
> pfmpc.2
[1] 0.4580762
>

```

## 10. Possibilistic Fuzzy C Means (3 cluster)

```

> #pfcm cluster = 3
> res.pfcm.3 <- pfcm(data, centers=res.fcm.3$v, memberships=res.fcm.3$u,a=5,b=5)
> summary(res.pfcm.3)
Summary for 'res.pfcm.3'

Number of data objects: 34

Number of clusters: 3

Crisp clustering vector:
[1] 3 1 3 1 1 1 3 3 2 2 1 1 2 1 1 2 1 3 3 3 3 2 1 1 3 2 1 3 3 3 1 1 3

Initial cluster prototypes:
      sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1      70.24860      97.06318      66.60723
Cluster 2      82.63893      99.14665      85.28590
Cluster 3      61.44288      90.93720      48.66816

Final cluster prototypes:
      sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1      71.26094      97.36056      67.76322
Cluster 2      83.22298      99.13229      86.14011
Cluster 3      63.16952      93.80353      53.02287

Distance between the final cluster prototypes
      Cluster 1 Cluster 2
Cluster 2 483.9396
Cluster 3 295.4015 1527.2887

Difference between the initial and final cluster prototypes
      sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1      1.0123442      0.29737548      1.1559909
Cluster 2      0.5840584      -0.01435523      0.8542087
Cluster 3      1.7266397      2.86632921      4.3547162

Root Mean Squared Deviations (RMSD): 3.350685
Mean Absolute Deviation (MAD): 12.86602

```

```

Membership degrees matrix (top and bottom 5 rows):
  Cluster 1 Cluster 2 Cluster 3
1 0.20997543 0.02945144 0.2718432
2 0.36753500 0.06931082 0.1003199
3 0.06767889 0.01925445 0.5156091
4 0.44953246 0.07276725 0.1041902
5 0.27747484 0.03186820 0.2562752
...
  Cluster 1 Cluster 2 Cluster 3
30 0.08312987 0.020608820 0.50369124
31 0.20584495 0.031275195 0.27060625
32 0.28095796 0.034415617 0.22180244
33 0.21217437 0.040320399 0.16287804
34 0.00541470 0.004896268 0.02683892

Descriptive statistics for the membership degrees by clusters
      Size      Min      Q1      Mean      Median      Q3      Max
Cluster 1    14 0.09582003 0.2386293 0.3276162 0.2792164 0.3973019 0.7023688
Cluster 2     7 0.09981371 0.1438498 0.3253905 0.2502595 0.3306415 0.9786780
Cluster 3    13 0.02683892 0.2196598 0.3192264 0.3071620 0.4571596 0.5164135

Dunn's Fuzziness Coefficients:
dunn_coeff normalized
0.1570631 -0.2644053

Within cluster sum of squares by cluster:
      1      2      3
641.6012 679.6085 3559.2846
(between_SS / total_SS = 59.14%)

Available components:
 [1] "u"      "t"      "v"      "v0"     "d"
 [6] "x"      "cluster" "csize"  "sumsqrs" "k"
 [11] "m"      "eta"     "a"      "b"      "omega"
 [16] "iter"   "best.start" "func.val" "comp.time" "inpargs"
 [21] "algorithm" "call"

> res.pfcm.3$comp.time
[1] 0.95
> res.pfcm.3$iter
[1] 174
> res.pfcm.3$func.val
[1] 27859.29

> res.pfcm.3$u
  Cluster 1 Cluster 2 Cluster 3
1 0.652304961 0.04808152 0.2996135203
2 0.872994551 0.07222905 0.0547764036
3 0.167517368 0.02924914 0.8032334916
4 0.902027256 0.05596335 0.0420093978
5 0.741430938 0.04102885 0.2175402168
6 0.740341770 0.05380331 0.2058549177
7 0.239875614 0.09630398 0.6638204029
8 0.171041638 0.04591479 0.7830435670
9 0.418838983 0.46963474 0.1115262793
10 0.001110183 0.99852989 0.0003599309
11 0.061495060 0.91380678 0.0246981632
12 0.898981207 0.02603043 0.0749883610
13 0.790764459 0.14346092 0.0657746243
14 0.083101219 0.88969256 0.0272062167
15 0.883852417 0.06395795 0.0521896320
16 0.819815698 0.07923611 0.1009481963
17 0.076475463 0.89209886 0.0314256801
18 0.965018081 0.01530883 0.0196730937
19 0.238172990 0.09087664 0.6709503714
20 0.112261297 0.02588077 0.8618579363
21 0.098821479 0.02683657 0.8743419484
22 0.216797310 0.03508955 0.7481131411
23 0.474671416 0.43039547 0.0949331161
24 0.600689238 0.24690835 0.1524024110
25 0.880699534 0.07394509 0.0453553791
26 0.526303665 0.03943008 0.4342662557
27 0.517811299 0.40868969 0.0734990100
28 0.713975747 0.18715124 0.0988730088
29 0.609658057 0.07072138 0.3196205651
30 0.207897417 0.03115035 0.7609522333
31 0.645804122 0.05193160 0.3022642800
32 0.770762960 0.04539053 0.1838465060
33 0.747966508 0.07533279 0.1767006973
34 0.308642645 0.18008878 0.5112685797

```



```

> res.pfcm.3$t
  Cluster 1 Cluster 2 Cluster 3
1  0.20997543 0.029451435 0.27184316
2  0.36753500 0.069310821 0.10031992
3  0.06767889 0.019254452 0.51560908
4  0.44953246 0.072767246 0.10419019
5  0.27747484 0.031868199 0.25627522
6  0.23266121 0.033004405 0.20497522
7  0.01199479 0.007493055 0.09317049
8  0.03069429 0.012995722 0.30716196
9  0.06001136 0.099813712 0.04941803
10 0.03189581 0.978678000 0.03163234
11 0.01900609 0.308407346 0.02324300
12 0.47684466 0.039274457 0.18864852
13 0.26521546 0.092087969 0.08409237
14 0.03183586 0.352875679 0.03187248
15 0.39199245 0.067393177 0.10427986
16 0.25653338 0.049118975 0.11499181
17 0.01813866 0.250259509 0.02268848
18 0.70236885 0.054808380 0.12825348
19 0.01750301 0.010419010 0.13305358
20 0.04350562 0.015982570 0.51641350
21 0.02939522 0.012579003 0.45038148
22 0.07390644 0.019614711 0.45715959
23 0.09438135 0.127680573 0.05992174
24 0.09582003 0.063206947 0.07597682
25 0.39907169 0.079500051 0.09468553
26 0.19834212 0.027909850 0.38435405
27 0.13481948 0.160018943 0.06335567
28 0.17844482 0.081041281 0.08423639
29 0.14935265 0.030582464 0.21965982
30 0.08312987 0.020608820 0.50369124
31 0.20584495 0.031275195 0.27060625
32 0.28095796 0.034415617 0.22180244
33 0.21217437 0.040320399 0.16287804
34 0.00541470 0.004896268 0.02683892

> pfpc.3<- PC(res.pfcm.3$u)
> pfpc.3
[1] 0.6422729
> pfmpc.3 <- MPC(res.pfcm.3$u)
> pfmpc.3
[1] 0.4634093
> |

```

## 11. Possibilistics Fuzzy C Means (4 cluster)

```

> res.pfcm.4 <- pfcm(data, centers=res.fcm.4$v, memberships=res.fcm.4$u,a=5,b=
> summary(res.pfcm.4)
Summary for 'res.pfcm.4'

Number of data objects: 34

Number of clusters: 4

Crisp clustering vector:
[1] 2 1 2 1 2 2 2 2 1 4 4 1 1 4 1 1 4 1 1 4 1 3 2 2 2 1 1 1 2 1 1 2 2 2 1 3

Initial cluster prototypes:
  sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1             72.89669                97.88102            69.23534
Cluster 2             63.18358                96.19473            55.82167
Cluster 3             59.70715                60.71960            35.48187
Cluster 4             84.31286                99.54545            87.81336

Final cluster prototypes:
  sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1             73.63436                98.11701            69.79885
Cluster 2             63.92655                96.30757            58.02401
Cluster 3             59.51395                59.59840            34.89978
Cluster 4             84.37930                99.48508            87.70034

Distance between the final cluster prototypes
  Cluster 1 Cluster 2 Cluster 3
Cluster 2  236.1624
Cluster 3 2901.0146 1901.7643
Cluster 4  437.7887 1309.0959 4997.1323

Difference between the initial and final cluster prototypes
  sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1             0.73766765                0.23598795            0.5635119
Cluster 2             0.74296926                0.11284160            2.2023401
Cluster 3            -0.19320019                -1.12119859           -0.5820868
Cluster 4             0.06644194                -0.06037162           -0.1130201

```

```

Root Mean Squared Deviations (RMSD): 1.41302
Mean Absolute Deviation (MAD): 5.048728

Membership degrees matrix (top and bottom 5 rows):
  Cluster 1 Cluster 2 Cluster 3 Cluster 4
1 0.09097499 0.36581209 0.01298176 0.014470151
2 0.33713548 0.07480212 0.01041003 0.032512233
3 0.03669095 0.28305836 0.01658951 0.009660214
4 0.73494610 0.07310980 0.01091005 0.034040334
5 0.11276127 0.32538150 0.01311646 0.015586192
...
  Cluster 1 Cluster 2 Cluster 3 Cluster 4
30 0.040546009 0.609559398 0.01641411 0.010278396
31 0.093101146 0.228675887 0.01636693 0.015303796
32 0.116241373 0.223155756 0.01400521 0.016753130
33 0.123341849 0.107882896 0.01520071 0.019547413
34 0.003546646 0.009221562 0.63731145 0.002575753

Descriptive statistics for the membership degrees by clusters
  Size      Min      Q1      Mean      Median      Q3      Max
Cluster 1  14 0.05128128 0.1320323 0.3194994 0.2230596 0.4919371 0.7349461
Cluster 2  14 0.03210442 0.1488722 0.2763447 0.2266479 0.3557044 0.6095594
Cluster 3   2 0.06455841 0.2077467 0.3509349 0.3509349 0.4941232 0.6373115
Cluster 4   4 0.20696152 0.2186897 0.3998531 0.2632163 0.4443796 0.8660182

Dunn's Fuzziness Coefficients:
dunn_coeff normalized
0.1505875 -0.1325500

Within cluster sum of squares by cluster:
  1      2      3      4
749.9130 1579.3509 325.8985 125.7030
(between_SS / total_SS = 77.11%)

Available components:
 [1] "u"      "t"      "v"      "v0"     "d"
 [6] "x"      "cluster" "csize"  "sumsqrs" "k"
 [11] "m"      "eta"     "a"      "b"      "omega"
 [16] "iter"   "best.start" "func.val" "comp.time" "inpargs"
 [21] "algorithm" "call"

> res.pfcm.4$comp.time
[1] 1.06
> res.pfcm.4$iter
[1] 148
> res.pfcm.4$func.val
[1] 21098.15

> res.pfcm.4$u
  Cluster 1 Cluster 2 Cluster 3 Cluster 4
1 0.227147567 0.733722023 0.0113098353 0.027820575
2 0.867990778 0.077329331 0.0068017104 0.047878180
3 0.139063437 0.807869830 0.0233348033 0.029731929
4 0.972706259 0.015507340 0.0014660106 0.010320391
5 0.305762613 0.650322924 0.0121143881 0.031800075
6 0.387461992 0.550046002 0.0161134561 0.046378550
7 0.205079303 0.518453842 0.1858303281 0.090636527
8 0.150332533 0.748559613 0.0534669131 0.047640941
9 0.461698854 0.150259866 0.0254185911 0.362622689
10 0.005760219 0.001812177 0.0004565626 0.991971042
11 0.044099730 0.017604179 0.0053962945 0.932899796
12 0.662649004 0.288668203 0.0106862011 0.037996592
13 0.922439749 0.035164312 0.0037897365 0.038606203
14 0.114556813 0.039057929 0.0093389810 0.837046277
15 0.952264278 0.028136891 0.0024547694 0.017144061
16 0.694884334 0.215798236 0.0145923555 0.074725074
17 0.061993202 0.025185584 0.0079288147 0.904892400
18 0.914415222 0.061270377 0.0037932806 0.020521120
19 0.168363801 0.355147338 0.4066078085 0.069881053
20 0.175842790 0.713764830 0.0629730343 0.047419345
21 0.162181533 0.710149263 0.0770472343 0.050621970
22 0.057195068 0.922240650 0.0083601437 0.012204138
23 0.623572713 0.100071220 0.0164300122 0.259926054
24 0.652899207 0.157886710 0.0243815769 0.164832506
25 0.962343719 0.020108753 0.0018588494 0.015688678
26 0.140869059 0.832350796 0.0091204589 0.017659686
27 0.702496650 0.076650366 0.0109934857 0.209859498
28 0.810563674 0.086119397 0.0111423537 0.092174575
29 0.452811021 0.460398795 0.0258697402 0.060920443
30 0.045332228 0.938585235 0.0067822516 0.009300286
31 0.356339816 0.576741490 0.0218820548 0.045036639
32 0.421399355 0.515787126 0.0172412054 0.045572314
33 0.609173706 0.293441313 0.0253199525 0.072065029
34 0.005259979 0.007708691 0.9838453118 0.003186018

```

```

> res.pfcm.4$t
  Cluster 1 Cluster 2 Cluster 3 Cluster 4
1 0.090974989 0.365812089 0.012981765 0.014470151
2 0.337135478 0.074802119 0.010410035 0.032512233
3 0.036690949 0.283058362 0.016589509 0.009660214
4 0.734946097 0.073109803 0.010910048 0.034040334
5 0.112761272 0.325381497 0.013116461 0.015586192
6 0.102627948 0.224619851 0.012397922 0.016133056
7 0.007299545 0.032104415 0.017282837 0.003877669
8 0.017223501 0.134729426 0.016185503 0.006608661
9 0.051281279 0.030433617 0.007793460 0.048392124
10 0.030382434 0.017285487 0.006512683 0.866018201
11 0.016931979 0.012119339 0.005532085 0.303833398
12 0.220416788 0.180173185 0.011891613 0.019049754
13 0.470725181 0.057044274 0.009552191 0.042683491
14 0.028953931 0.017816366 0.006375041 0.206961523
15 0.603976183 0.074421883 0.010270285 0.031841981
16 0.158103825 0.094253271 0.010301814 0.023618713
17 0.016112828 0.011732202 0.005498073 0.222599131
18 0.499007706 0.106412106 0.010788254 0.026076996
19 0.010710711 0.039154254 0.064558413 0.005353939
20 0.024582119 0.154354104 0.023267436 0.008074908
21 0.016953667 0.118743702 0.021167420 0.006406710
22 0.037334005 0.527364798 0.014741728 0.009815059
23 0.113319718 0.035303807 0.008809687 0.059984009
24 0.095112999 0.043386241 0.010254136 0.030807152
25 0.663109127 0.068369857 0.009935448 0.037014406
26 0.084821405 0.494221252 0.015591582 0.013726613
27 0.176812503 0.040138683 0.008793940 0.071373769
28 0.225702497 0.052366460 0.010465582 0.038189260
29 0.086778525 0.147044897 0.014126886 0.015082768
30 0.040546009 0.609559398 0.016414107 0.010278396
31 0.093101146 0.228675887 0.016366930 0.015303796
32 0.116241373 0.223155756 0.014005210 0.016753130
33 0.123341849 0.107882896 0.015200709 0.019547413
34 0.003546646 0.009221562 0.637311454 0.002575753

> pfpc.4<- PC(res.pfcm.4$t)
> pfpc.4
[1] 0.6476873
> pfmpc.4 <- MPC(res.pfcm.4$t)
> pfmpc.4
[1] 0.5302497

```

## 12. Possibilistics Fuzzy C Means (5 cluster)

```

> #pfcm cluster = 5
> res.pfcm.5 <- pfcm(data, centers=res.fcm.5$t, memberships=res.fcm.5$u,a=5,b=5)
> summary(res.pfcm.5)
Summary for 'res.pfcm.5'

Number of data objects: 34

Number of clusters: 5

Crisp clustering vector:
 [1] 3 5 1 5 3 3 1 1 5 4 4 3 5 4 5 3 4 5 1 1 1 1 5 5 3 5 5 1 1 3 3 3 2

Initial cluster prototypes:
  sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1             61.56990                95.29674           49.53832
Cluster 2             59.34234                57.64635           33.98936
Cluster 3             66.27916                96.50010           63.48665
Cluster 4             85.30456                99.62232           88.70126
Cluster 5             75.32641                98.30732           70.88147

Final cluster prototypes:
  sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1             63.54611                95.20014           50.55695
Cluster 2             59.19503                56.80459           33.54988
Cluster 3             66.03236                96.80413           63.77506
Cluster 4             85.26020                99.58191           88.50521
Cluster 5             75.21171                98.37569           70.82522

Distance between the final cluster prototypes
  Cluster 1 Cluster 2 Cluster 3 Cluster 4
Cluster 2 1782.3903
Cluster 3 183.4728 2560.2739
Cluster 4 1930.7727 5529.3811 989.0063
Cluster 5 556.9732 3374.1411 136.4349 415.0094

```

```

Difference between the initial and final cluster prototypes
      sumber.air.yang.layak penerangan.dengan.listrik sanitasi.layak
Cluster 1      1.97620643          -0.09660555      1.01862404
Cluster 2     -0.14731316          -0.84175797     -0.43948161
Cluster 3     -0.24679930          0.30402568      0.28841588
Cluster 4     -0.04436098          -0.04041712     -0.19604765
Cluster 5     -0.11469996           0.06836320     -0.05624638

Root Mean Squared Deviations (RMSD): 1.11133
Mean Absolute Deviation (MAD): 3.527619

Membership degrees matrix (top and bottom 5 rows):
      Cluster 1 Cluster 2 Cluster 3 Cluster 4 Cluster 5
1 0.11530896 0.004835650 0.59022544 0.009137753 0.04799552
2 0.04098727 0.003937916 0.08382021 0.019908319 0.20403686
3 0.37280283 0.006091777 0.06685828 0.006197244 0.02192672
4 0.04367542 0.004119921 0.07098847 0.020930862 0.70358629
5 0.10606813 0.004886095 0.85272750 0.009818726 0.05786559
...
      Cluster 1 Cluster 2 Cluster 3 Cluster 4 Cluster 5
30 0.21117685 0.006036746 0.158831149 0.006553949 0.023009331
31 0.10539588 0.006041911 0.208673202 0.009670770 0.050921573
32 0.08617300 0.005211311 0.426862722 0.010527908 0.060106552
33 0.06568900 0.005654313 0.087165984 0.012306785 0.073718367
34 0.01238994 0.913748483 0.003472287 0.001700395 0.002346806

Descriptive statistics for the membership degrees by clusters
      Size      Min      Q1      Mean      Median      Q3      Max
Cluster 1      9 0.04883665 0.1048622 0.2249742 0.2111769 0.3608095 0.4683002
Cluster 2      1 0.91374848 0.9137485 0.9137485 0.9137485 0.9137485 0.9137485
Cluster 3      9 0.08716598 0.2086732 0.3745696 0.4178991 0.4367430 0.8527275
Cluster 4      4 0.12769765 0.1930172 0.3134779 0.2674430 0.3879037 0.5913279
Cluster 5     11 0.03722508 0.1517419 0.3291726 0.2378568 0.5737879 0.7035863

Dunn's Fuzziness Coefficients:
dunn coeff normalized
0.17052811 -0.03683986

Within cluster sum of squares by cluster:
      1      2      3      4      5
1372.0106 0.0000 247.5223 125.7030 484.3260
(between_SS / total_SS = 80.1%)

Available components:
[1] "u"      "t"      "v"      "v0"     "d"
[6] "x"      "cluster" "csize" "sumsqrs" "k"
[11] "m"      "eta"     "a"      "b"      "omega"
[16] "iter"   "best.start" "func.val" "comp.time" "inpargs"
[21] "algorithm" "call"

> res.pfcm.5$comp.time
[1] 0.71
> res.pfcm.5$iter
[1] 78
> res.pfcm.5$func.val
[1] 15778.39

> res.pfcm.5$u
      Cluster 1 Cluster 2 Cluster 3 Cluster 4 Cluster 5
1 0.0329987524 0.0022055114 0.922567059 0.0056895478 0.0365391290
2 0.0401458155 0.0066576187 0.217411571 0.0464953051 0.6892896900
3 0.6863550558 0.0126880426 0.209305304 0.0175467899 0.0741048074
4 0.0064383020 0.0010455496 0.027252428 0.0073441048 0.9579196160
5 0.0079230396 0.0005877919 0.978135065 0.0016135309 0.0117405724
6 0.0431604576 0.0039704262 0.863833578 0.0118778269 0.0771577117
7 0.4775661384 0.1081330611 0.233476688 0.0574164730 0.1234076397
8 0.5972659455 0.0315236436 0.252325470 0.0305424535 0.0883424876
9 0.0775342158 0.0199523560 0.217465061 0.2747918762 0.4102564906
10 0.0037745829 0.0012451612 0.007735233 0.9668869725 0.0203580507
11 0.0083540170 0.0031864675 0.015066527 0.9413468681 0.0320461201
12 0.0707247360 0.0063118918 0.618128395 0.0232053972 0.2816295796
13 0.0062997984 0.0011038812 0.022659771 0.0110093901 0.9589271595
14 0.0290804373 0.0094206343 0.063147807 0.7562001253 0.1421509960
15 0.0119892489 0.0017785311 0.047709872 0.0124453196 0.9260770287
16 0.0716209849 0.0101219810 0.470264975 0.0523771939 0.3956148654
17 0.0135577022 0.0052532818 0.023912745 0.9067546014 0.0505216700
18 0.0397541306 0.0051304914 0.291499041 0.0280257203 0.6355906167
19 0.4021642096 0.2348338910 0.194753182 0.0507480890 0.1175006281
20 0.7664793207 0.0205638630 0.136079883 0.0173186178 0.0595583154
21 0.8770555792 0.0137954111 0.068859292 0.0101016832 0.0301880350
22 0.3987189954 0.0134649561 0.480795894 0.0210728171 0.0859473374
23 0.0466234363 0.0105518681 0.111867591 0.1627513475 0.6682057576
24 0.0835125490 0.0167008016 0.173526200 0.1154547355 0.6108057140
25 0.0094669293 0.0015477838 0.042920808 0.0128736566 0.9331908225
26 0.0816806062 0.0044736459 0.845732684 0.0092143207 0.0588987428
27 0.0312310406 0.0068861588 0.096920530 0.1224282516 0.7425340191
28 0.0340953448 0.0064751003 0.095610728 0.0535270832 0.8102917434
29 0.2201565497 0.0177832659 0.418910573 0.0443538561 0.2987957557
30 0.3187833207 0.0129653622 0.568827151 0.0191431428 0.0802810235
31 0.1210537224 0.0111973385 0.685480541 0.0244507613 0.1578176372
32 0.0429238103 0.0042749432 0.857668633 0.0118019867 0.0833306270
33 0.1211638681 0.0175686171 0.416318863 0.0523263003 0.3926223518
34 0.0006594226 0.9983051064 0.000463346 0.0002181698 0.0003539552

```

```

> res.pfcm.5$t
  Cluster 1 Cluster 2 Cluster 3 Cluster 4 Cluster 5
1 0.115308956 0.004835650 0.590225437 0.009137753 0.047995516
2 0.040987267 0.003937916 0.083820213 0.019908319 0.204036862
3 0.372802829 0.006091777 0.066858281 0.006197244 0.021926718
4 0.043675419 0.004119921 0.070988469 0.020930862 0.703586287
5 0.106068133 0.004886095 0.852727497 0.009818726 0.057865593
6 0.083196360 0.004634878 0.417899082 0.010144375 0.053630474
7 0.048836651 0.006442959 0.009824580 0.002526792 0.004613395
8 0.169364902 0.005967046 0.032928075 0.004260587 0.010425410
9 0.020489362 0.002993607 0.022665214 0.029524913 0.037225085
10 0.013577970 0.002526428 0.011027117 0.591327865 0.025278396
11 0.010078687 0.002161496 0.007205787 0.320095711 0.013459468
12 0.082439491 0.004452742 0.236868964 0.011952821 0.111094534
13 0.035903956 0.003626741 0.050285748 0.026012756 0.664452823
14 0.013532818 0.002472785 0.011638045 0.127697655 0.022889064
15 0.044987458 0.003882748 0.068985260 0.019671726 0.559677577
16 0.047216385 0.003891412 0.113960648 0.014654248 0.087276917
17 0.009868337 0.002149503 0.006900621 0.214790365 0.012807782
18 0.053661401 0.004065368 0.141152193 0.016139789 0.240522276
19 0.063304305 0.021538365 0.012771289 0.003487452 0.006850397
20 0.360809535 0.008376749 0.038103987 0.005206763 0.015090910
21 0.468300217 0.007668300 0.026606208 0.004145671 0.010479009
22 0.225310646 0.005448741 0.121749565 0.006268352 0.021430915
23 0.026057669 0.003366199 0.024746897 0.036911632 0.118125732
24 0.033849206 0.003892889 0.027970325 0.019489043 0.082158449
25 0.039781117 0.003763977 0.069113562 0.022596841 0.587898214
26 0.159278201 0.005754626 0.436742963 0.008694218 0.045548548
27 0.026665354 0.003358085 0.032513267 0.042210911 0.185358043
28 0.036230652 0.003966480 0.040002380 0.023646300 0.237856786
29 0.104862238 0.005250489 0.080974386 0.009592185 0.052617132
30 0.211176855 0.006036746 0.158831149 0.006553949 0.023009331
31 0.105395878 0.006041911 0.208673202 0.009670770 0.050921573
32 0.086172998 0.005211311 0.426862722 0.010527908 0.060106552
33 0.065688997 0.005654313 0.087165984 0.012306785 0.073718367
34 0.012389938 0.913748483 0.003472287 0.001700395 0.002346806
- - - - -
> pfpc.5<- PC(res.pfcm.5$u)
> pfpc.5
[1] 0.6204691
> pfmpc.5 <- MPC(res.pfcm.5$u)
> pfmpc.5
[1] 0.5255864

```

#### Lampiran 4 Uji *Kruskall Wallis*

##### Ranks

	Cluster	N	Mean Rank
Sumber_Air_Layak	Cluster 1	9	9.22
	Cluster 2	1	3.00
	Cluster 3	9	13.00
	Cluster 4	4	31.75
	Cluster 5	11	24.09
	Total	34	
Penerangan_Listrik	Cluster 1	9	12.56
	Cluster 2	1	1.00
	Cluster 3	9	15.28
	Cluster 4	4	28.25
	Cluster 5	11	20.95
	Total	34	
Sanitasi_Layak	Cluster 1	9	6.00
	Cluster 2	1	1.00
	Cluster 3	9	15.56
	Cluster 4	4	32.50
	Cluster 5	11	24.55
	Total	34	

##### Test Statistics<sup>a,b</sup>

	Sumber_Air_Layak	Penerangan_Listrik	Sanitasi_Layak
Chi-Square	23.190	11.399	29.673
df	4	4	4
Asymp. Sig.	.000	.022	.000

a. Kruskal Wallis Test

b. Grouping Variable: Cluster

**PERBANDINGAN METODE *FUZZY C MEANS*, *FUZZY POSSIBILISTICS C MEANS* DAN *POSSIBILISTICS FUZZY C MEANS* PADA *NOISY DATA***  
(Studi Kasus: Fasilitas Tempat Tinggal Setiap Propinsi Tahun 2017)

Khusnul Hajar Nuansari<sup>1</sup>, Kariyam<sup>2</sup>

<sup>1</sup>Mahasiswa Jurusan Statistika, FMIPA, Universitas Islam Indonesia

<sup>2</sup>Dosen Jurusan Statistika, FMIPA, Universitas Islam Indonesia

e-mail: [chusnul.sari@gmail.com](mailto:chusnul.sari@gmail.com), [kariyam@uii.ac.id](mailto:kariyam@uii.ac.id)

### ABSTRACT

In this digital age, data has become a very important part of human civilization. The abundance of current data will result in an information that is easy to understand, meaningful and useful if done processing of the data, including in it is data processing error (noisy data). One case of noisy data in the form of outliers that need to be processed is data of residence facilities every province in 2017. The data comes from the BPS website. In an effort to overcome the case of noisy data in the form of outlier data on residential facilities of each province in 2017, it can be done by using clustering method. The clustering method used in this research is Fuzzy C Means method, Fuzzy Possibilistics C Means and Possibilistics Fuzzy C Means. The three methods compare their accuracy in classifying the noisy data based on the Modified Partition Coefficient (MPC) index. Based on the Modified Partition Coefficient (MPC) index, it is found that Fuzzy Possibilistics C Means (FPCM) method is the best method compared to Fuzzy Possibilistics C Means (FPCM) and Possibilistics Fuzzy C Means (PFCM) method. The optimal number of clusters in the Fuzzy Possibilistics C Means (FPCM) method is 5 clusters. Of the five clusters, cluster 4 is a cluster with the most decent living facilities compared to others.

**Keyword:** Fuzzy C Means, Fuzzy Possibilistics C Means, Possibilistics Fuzzy C Means, Noisy Data, Modified Partition Coefficient

### ABSTRAK

Di era digital ini, data telah menjadi bagian yang sangat penting dalam peradaban manusia. Melimpahnya data saat ini akan menghasilkan suatu informasi yang mudah dimengerti, bermakna dan bermanfaat apabila dilakukan pengolahan terhadap data tersebut, termasuk didalamnya adalah pengolahan data error (noisy data). Salah satu kasus noisy data dalam bentuk outlier yang perlu diolah adalah data fasilitas tempat tinggal setiap propinsi tahun 2017. Data tersebut berasal dari website BPS. Dalam upaya mengatasi kasus noisy data yang berupa outlier pada data fasilitas tempat tinggal setiap propinsi tahun 2017, maka dapat dilakukan dengan menggunakan metode *clustering*. Metode clustering yang digunakan dalam penelitian ini adalah metode Fuzzy C Means, Fuzzy Possibilistics C Means dan Possibilistics Fuzzy C Means. Ketiga metode tersebut dibandingkan keakuratannya dalam mengelompokkan *noisy* data berdasarkan indeks *Modified Partition Coefficient* (MPC). Berdasarkan indeks *Modified Partition Coefficient* (MPC) didapatkan hasil bahwa metode *Fuzzy Possibilistics C Means* (FPCM) merupakan metode terbaik dibandingkan dengan metode *Fuzzy Possibilistics C Means* (FPCM) dan *Possibilistics Fuzzy C Means* (PFCM). Jumlah cluster yang optimal pada metode *Fuzzy Possibilistics C Means* (FPCM) ialah 5 *cluster*. Dari kelima *cluster* tersebut, *cluster* 4

merupakan *cluster* dengan fasilitas tempat tinggal paling layak dibandingkan dengan yang lainnya.

**Kata Kunci :** *Fuzzy C Means, Fuzzy Possibilistics C Means, Possibilistics Fuzzy C Means, Noisy Data, Modified Partition Coefficient*

## Pendahuluan

Di era digital ini, data telah menjadi bagian yang sangat penting dalam peradaban manusia. Semua orang menghasilkan data bahkan untuk hal kecil sekalipun. Data dibuat dan berada dimana-mana, di berbagai sektor dan bidang kehidupan. Kumpulan dari berbagai macam data tersebut biasanya disimpan dalam suatu tempat yang namanya *database*. *Database* berfungsi sebagai tempat untuk menyimpan dan mengelompokkan data secara sistematis dan terorganisir sehingga lebih memudahkan dalam mengakses data-data, *maintenance* data, serta pengelolaan data-data yang berada dalam *database* tersebut.

Melimpahnya data saat ini akan menghasilkan suatu informasi yang mudah dimengerti, bermakna dan bermanfaat apabila dilakukan pengolahan terhadap data tersebut. Namun begitu, sering kita jumpai bahwa data yang digunakan mengandung *error (noisy)* seperti data yang memiliki tipe salah (data string di atribut numerik), data dengan nilai yang sangat berbeda dengan nilai data lainnya, data yang salah diinputkan, data *missing*, data *outlier* dan lain sebagainya yang mana akan berakibat pada hasil keputusan. Di sisi lain, terkadang data tersebut juga mengandung makna yang berarti. Salah satu kasus *noisy* data dalam bentuk *outlier* yang perlu diolah adalah data fasilitas

tempat tinggal pada indikator perumahan.

Perumahan merupakan salah satu kebutuhan pokok manusia. Fungsi rumah sebagai tempat tinggal selayaknya memiliki bentuk fisik dan fasilitas yang memadai. Dengan demikian, rumah sebagai sarana pembinaan keluarga dapat memperoleh hasil yang maksimal yaitu tercapainya peningkatan kualitas sumber daya manusia. Salah satu indikator perumahan dan permukiman adalah fasilitas tempat tinggal seperti sumber air minum, sumber penerangan dan akses terhadap sanitasi. Untuk mengatasi kasus *noisy* data yang berupa data *outlier* pada data fasilitas tempat tinggal maka dapat dilakukan dengan menggunakan *clustering*.

*Clustering* merupakan proses pembagian data ke dalam suatu kelompok berdasarkan tingkat kesamaannya sehingga data-data yang memiliki kemiripan karakteristik akan berkumpul dalam satu kelompok. Sementara data-data yang memiliki perbedaan karakteristik, akan berkumpul dalam kelompok yang berbeda (Siyamto, 2017). Dalam kaitannya dengan *noisy* data khususnya *outlier*, metode *clustering* dapat digunakan sebagai *outlier detection*.

Salah satu jenis metode *clustering* ialah metode *partition clustering*. *Partition clustering* merupakan suatu metode penyusunan partisi dari objek ke dalam suatu



*cluster*. Dengan kata lain, jumlah *cluster* dalam hal ini ditentukan terlebih dahulu.

Pada penelitian ini, analisis difokuskan pada metode *partition clustering* yakni *Fuzzy. Fuzzy clustering* merupakan suatu metode pengelompokan dimana suatu data dapat berada pada lebih dari satu *cluster*. Metode fuzzy yang digunakan untuk membandingkan pengelompokan *noisy* data ialah *Fuzzy C Means* (FCM), *Fuzzy Possibilistics C Means* (FPCM) dan *Possibilistics Fuzzy C Means* (PFCM). Ketiga metode tersebut akan dibandingkan keakurasiannya berdasarkan indeks validitas *Modified Partition Coefficient* (MPC) sehingga akan didapatkan metode yang tepat dalam pengelompokan *noisy* data.

### Landasan Teori *Fuzzy C Means*

*Fuzzy C Means* (FCM) merupakan salah satu jenis *clustering* non hirarki yang termasuk dalam teknik fuzzy *clustering*. Teknik ini pertama kali diperkenalkan oleh Jim Bezdek pada tahun 1981. *Fuzzy C Means* (FCM) merupakan suatu teknik pengelompokan data dimana keberadaan tiap-tiap titik data dalam suatu kelompok ditentukan oleh derajat keanggotaan data tersebut (Kusumadewi dkk. 2006). Pada *Fuzzy C Means* (FCM), data dapat menjadi anggota dari semua kelompok yang terbentuk dengan derajat keanggotaan yang berada antara 0 hingga 1 (Luthfi, 2017).

Pengelompokan data dengan metode *Fuzzy C Means* (FCM) dimulai dengan tebakan awal tentang pusat *cluster*. Tiap titik data akan

diberi suatu derajat keanggotaan terhadap *cluster* tebakan. Selanjutnya dilakukan iterasi guna memperbaharui pusat-pusat *cluster* dan derajat keanggotaan tiap titik data hingga sedekat mungkin dengan pusat-pusat *cluster* yang benar (Naba, 2009). Perulangan iterasi ini didasarkan pada minimasi fungsi obyektif (Ned & Roger, 2000). Hal tersebut menggambarkan jarak dari titik data yang diberikan ke pusat *cluster* yang terbobot oleh derajat keanggotaan titik data tersebut. Fungsi obyektif yang dilambangkan dengan  $P_t$  pada metode *Fuzzy C Means* (FCM) adalah:

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

Adapun langkah- langkah *Fuzzy C Means* (FCM) adalah sebagai berikut:

5. Memasukkan data ( $X_{ij}$ ) yang akan di*cluster* ke dalam sebuah matriks, dimana matriks berukuran  $n \times m$ , dengan  $n$  adalah jumlah data yang akan di*cluster* dan  $m$  adalah jumlah atribut atau variabel.
6. Menentukan
  - g. Jumlah *cluster* =  $c$  ( $\geq 2$ );
  - g. Pangkat/pembobot =  $w$  ( $> 1$ );
  - h. Maksimum iterasi
  - i. Kriteria penghentian =  $\xi$ ;
  - j. Fungsi Obyektif awal =  $P_0 = 0$ ;
  - k. Iterasi awal =  $t = 1$ ;
7. Membangkitkan bilangan acak sebagai elemen matriks partisi awal  $U$  atau matriks kekhasan relatif. Matriks kekhasan relatif dilambangkan dengan  $\mu_{ik}$ . Berikut adalah bentuk matriks partisi awal:

$$\mu_{ik} = \begin{bmatrix} \mu_{11} & \dots & \mu_{1c} \\ \vdots & \ddots & \vdots \\ \mu_{n1} & \dots & \mu_{nc} \end{bmatrix} \quad (2)$$

$$1 \leq i \leq n; 1 \leq k \leq c$$

Jumlah setiap nilai elemen kolom dalam satu baris adalah 1.

8. Menghitung pusat *cluster*. Pusat cluster dilambangkan dengan  $V_{kj}$ .

$$V_{kj} = \frac{\sum_{i=1}^n (\mu_{ik})^w X_{ij}}{\sum_{i=1}^n (\mu_{ik})^w} \quad (3)$$

9. Menghitung perubahan derajat keanggotaan setiap data pada setiap *cluster*/ memperbaiki matriks kekhasan relatif ( $\mu_{ik}$ ) dengan:

$$\mu_{ik} = \frac{[\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-\frac{1}{w-1}}}{\sum_{k=1}^c [\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-\frac{1}{w-1}}} \quad (4)$$

10. Mengecek kondisi berhenti:
  - Jika : (  $|P_t - P_{t-1}| < \xi$  ) atau (  $t > \text{MaksIter}$  ) maka berhenti;
  - Jika tidak :  $t = t+1$ , ulangi langkah 4

### Fuzzy Possibilistics C Means

*Fuzzy Possibilistics C Means* (FPCM) merupakan suatu algoritma pengembangan dari *Fuzzy C Means* (FCM) dan *Possibilistics C Means* (PCM). Pada algoritma FCM, nilai  $\mu_{ik}$  dipengaruhi oleh  $X_{ij}$  dan semua pusat *cluster*. Sementara pada algoritma PCM, nilai  $t_{ik}$  dipengaruhi oleh  $X_{ij}$ , pusat cluster ke  $k$  serta  $\gamma_k$  (Kusumadewi dkk. 2006). Adapun fungsi objektif pada metode *Fuzzy Possibilistic C-Means* (FPCM) ialah:

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left( [\sum_{j=1}^m (X_{ij} - V_{kj})^2] (\mu_{ik}^w + t_{ik}^\eta) \right) \quad (5)$$

Langkah- langkah *Fuzzy Possibilistic C Means* (FPCM) adalah:

8. Memasukkan data ( $X_{ij}$ ) yang akan *dicluster* ke dalam sebuah

matriks, dimana matriks berukuran  $n \times m$ , dengan  $n$  adalah jumlah data yang akan *dicluster* dan  $m$  adalah jumlah atribut atau variabel.

9. Menentukan
  - a. Jumlah *cluster* =  $c$  ( $\geq 2$ );
  - b. Pangkat/ pembobot =  $w$  ( $> 1$ ) dan  $\eta$  ( $> 1$ );
  - c. Maksimum iterasi Error yang diharapkan =  $\xi$ ;
  - d. Fungsi Objektif awal =  $P_0 = 0$ ;
  - e. Iterasi awal =  $t = 1$ ;
10. Memanggil hasil akhir yang berupa matriks kekhasan relatif ( $\mu_{ik}$ ) dan pusat *cluster* ( $V_{kj}$ ) pada algoritma FCM untuk menghitung matriks kekhasan absolut ( $t_{ik}$ ). Matriks kekhasan absolut dilambangkan dengan dengan persamaan sebagai berikut:

$$t_{ik} = \begin{bmatrix} t_{11} & \dots & t_{1c} \\ \vdots & \ddots & \vdots \\ t_{n1} & \dots & t_{nc} \end{bmatrix} \quad (6)$$

Dimana elemen matriksnya ialah:

$$t_{ik} = \frac{[\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-1/\eta-1}}{\sum_{i=1}^n [\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-1/\eta-1}} \quad (7)$$

Dengan  $i = 1, 2, \dots, n$ ;  $k = 1, 2, \dots, c$ ;  $j = 1, 2, \dots, m$

11. Memperbaiki pusat *cluster* ( $V_{kj}$ ) dengan perhitungan sebagai berikut:

$$V_{kj} = \frac{\sum_{i=1}^n (\mu_{ik}^w + t_{ik}^\eta) X_{ij}}{\sum_{i=1}^n (\mu_{ik}^w + t_{ik}^\eta)} \quad (8)$$

12. Memperbaiki matriks kekhasan relatif ( $\mu_{ik}$ ) dengan perhitungan sebagai berikut:

$$\mu_{ik} = \frac{[\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-1/w-1}}{\sum_{i=1}^c [\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-1/w-1}} \quad (9)$$

13. Memperbaiki matriks kekhasan absolut ( $t_{ik}$ ) dengan perhitungan sebagai berikut:

$$t_{ik} = \frac{[\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-1/\eta-1}}{\sum_{i=1}^n [\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-1/\eta-1}} \quad (10)$$

14. Mengecek kondisi berhenti:
- Jika :  $(|P_t - P_{t-1}| < \xi)$  atau  $(t > \text{MaksIter})$  maka berhenti ;
  - Jika tidak :  $t = t+1$ , ulangi langkah 4

### Possibilistic Fuzzy C-Means

*Possibilistic Fuzzy C Means* (PFCM) merupakan pengembangan dari *Fuzzy Possibilistic C Means* yang menggabungkan algoritma FCM dan algoritma PCM (Pal, Kal, & Bezdek, 1997). Dengan kata lain yakni algoritma clustering yang menggabungkan nilai keanggotaan *Fuzzy C-Means* (FCM) dan nilai kesesuaian *Possibilistic C-Means* (PCM). Pengembangan metode *Fuzzy Possibilistics C Means* tersebut dilakukan dengan memberikan derajat kepentingan yang berbeda antara matriks kekhasan relatif dan matriks kekhasan absolut (Kusumadewi dkk. 2006). Adapun fungsi obyektif ialah:

$$P_t = \sum_{i=1}^n \sum_{k=1}^c (\alpha \mu_{ik}^w + b t_{ik}^\eta) [\sum_{j=1}^m (X_{ij} - V_{kj})^2] + \sum_{k=1}^c \gamma_k \sum_{i=1}^n (1 - t_{ik})^\eta \quad (11)$$

Langkah- langkah Possibilistic Fuzzy C-Means (PFCM)

1. Memasukkan data ( $X_{ij}$ ) yang akan dicluster ke dalam sebuah matriks.
2. Menentukan
  - j. Jumlah cluster =  $c \geq 2$ ;

- k. Pangkat/pembobot untuk FCM ( $w = w > 1$ );
- l. Pangkat/pembobot untuk PCM ( $\eta = \eta > 1$ );
- m. Maksimum iterasi;
- n. Kriteria penghentian =  $\xi$ ;
- o. Iterasi awal  $t = 1$ ;
- p. Koefisien pembobot untuk kekhasan relatif =  $a (> 0)$
- q. Koefisien pembobot untuk kekhasan absolut =  $b (> 0)$
- r. Koefisien  $\gamma = K$  (biasanya  $K=1$ )

3. Memanggil hasil akhir yang berupa matriks kekhasan relatif ( $\mu_{ik}$ ) dan pusat cluster ( $V_{kj}$ ) pada algoritma FCM yang bertujuan untuk mencari nilai matriks kekhasan absolut. Elemen matriksnya sebagai berikut:

$$t_{ik} = \left[ 1 + \left( \frac{b [\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{1/(\eta-1)}}{\gamma_k} \right)^{-1} \right]^{-1} \quad (12)$$

Nilai  $\gamma_k$  dan  $d_{ik}$  dapat dicari dengan persamaan sebagai berikut:

$$\gamma_k = K \frac{\sum_{i=1}^n ((\mu_{ik})^w \cdot [\sum_{j=1}^m (X_{ij} - V_{kj})^2])}{\sum_{i=1}^n (\mu_{ik})^w} \quad (13)$$

4. Memperbaiki pusat cluster ( $V_{kj}$ ) dengan rumus sebagai berikut:

$$V_{kj} = \frac{\sum_{i=1}^n (\alpha \mu_{ik}^w + b t_{ik}^\eta) X_{ij}}{\sum_{i=1}^n (\alpha \mu_{ik}^w + b t_{ik}^\eta)} \quad (14)$$

5. Memperbaiki matriks kekhasan relatif ( $\mu_{ik}$ ). Elemen matriksnya sebagai berikut:

$$\mu_{ik} = \frac{[\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-1/w-1}}{\sum_{i=1}^c [\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-1/w-1}} \quad (15)$$

Serta memperbaiki matriks kekhasan absolut ( $t_{ik}$ )

6. Menentukan kriteria berhenti:

- Jika :  $(|P_t - P_{t-1}| < \xi)$  atau  $(t > \text{MaksIter})$  maka berhenti ;
- Jika tidak :  $t = t+1$ , ulangi langkah 4

### Validitas Cluster

Validitas cluster merupakan suatu cara untuk mengevaluasi kualitas dari *cluster* dan menentukan seberapa baik data diwakili oleh *cluster* tersebut. Pada metode fuzzy terdapat dua kategori untuk menghitung indeks validitas *cluster* yakni berdasarkan nilai-nilai keanggotaan serta menggunakan data itu sendiri. Indeks-indeks validasi yang hanya melibatkan nilai keanggotaan salah satunya adalah indeks *Modified Partition Coefficient* (MPC). *Modified Partition Coefficient* (MPC) merupakan indeks validitas hasil perbaikan dari metode *Partition Coefficient* (PC). Metode *Partition Coefficient* (PC) cenderung mengalami perubahan yang monoton terhadap nilai  $c$  (Xie, Hu, Luktarhan, & Zhao, 2011). Metode *Modified Partition Coefficient* (MPC) didefinisikan dengan persamaan

$$\text{MPC}(c) = 1 - \frac{c}{c-1} (1 - \text{PC}(c)) \quad (16)$$

Adapun persamaan nilai indeks PC ini adalah:

$$\text{PC}(c) = \frac{1}{n} \sum_{k=1}^c \sum_{i=1}^n u_{ik}^2 \quad (17)$$

Dimana:

$n$  = banyak objek penelitian

$c$  = banyak kelompok

$u_{ik}$  = nilai keanggotaan objek ke- $i$  dengan pusat kelompok ke- $k$ .

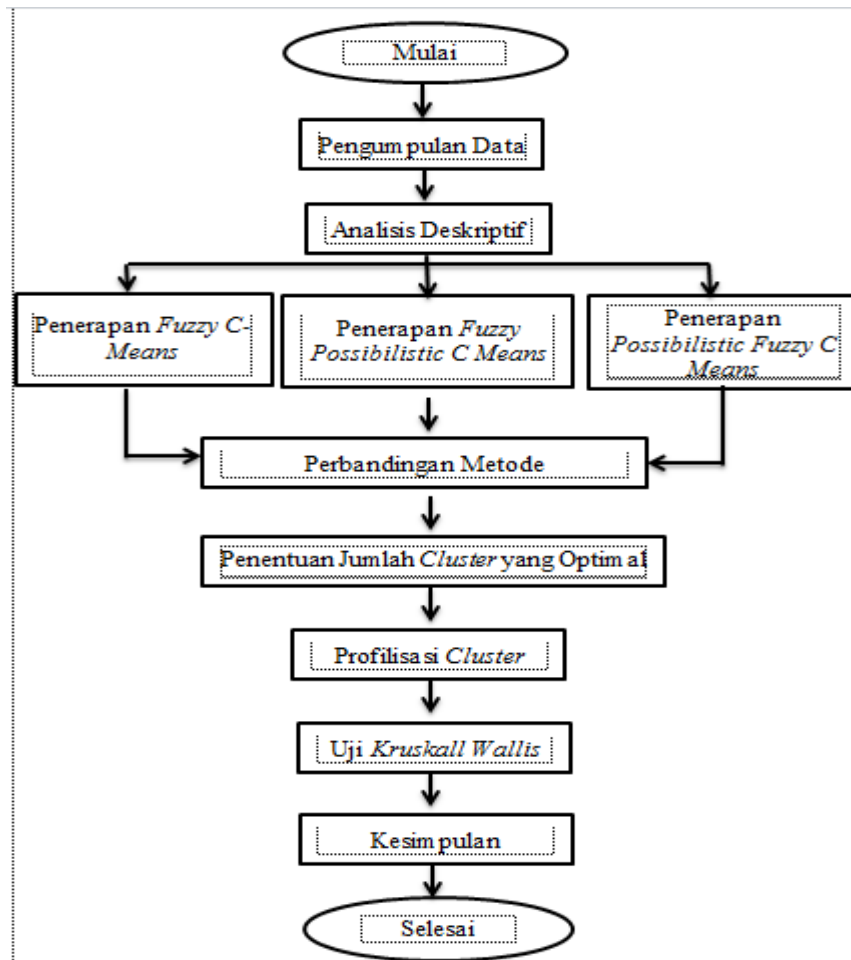
Nilai MPC berkisar antara  $0 \leq \text{MPC}(c) \leq 1$ . *Cluster* dikatakan optimal jika nilai MPC mendekati 1 atau yang paling besar. Semakin dekat nilai MPC ke 0 maka semakin kabur keakuratannya.

### Metodologi Penelitian

Data yang digunakan dalam penelitian ini ialah *noisy* data yang berupa data *outlier*. Jenis data yang digunakan dalam penelitian ini merupakan data sekunder, yakni data perumahan. Data ini diperoleh dari website Badan Pusat Statistik (BPS) yakni:

<https://www.bps.go.id/subject/29/perumahan.html#subjekViewTab3>

Variabel data yang digunakan dalam penelitian ini adalah data: sumber air minum layak, sumber penerangan dari listrik serta akses sanitasi layak pada tahun 2017. Data tersebut berjumlah 34 sesuai dengan jumlah provinsi di Indonesia. Berikut adalah diagram alir metode penelitian:



Gambar 1. Diagram Alir

## Hasil dan Pembahasan

### Perbandingan Metode *Fuzzy C Means*, *Fuzzy Possibilistics C Means* dan *Possibilistics Fuzzy C Means*

Pada metode *Fuzzy C Means* (FCM) nilai pembobot ( $w$ ) yang digunakan adalah 2 dan nilai maksimum iterasi yang digunakan adalah 1000. Pada metode *Fuzzy Possibilistics C Means* (FPCM), nilai pembobot untuk *Fuzzy C Means* ( $w$ ) adalah 2, nilai pembobot untuk *Possibilistics C Means* ( $\eta$ ) adalah 2 dan nilai maksimum iterasi adalah 1000. Sementara pada metode *Possibilistics Fuzzy C Means* (PFCM), nilai pembobot untuk

*Fuzzy C Means* ( $w$ ) adalah 2, nilai pembobot untuk *Possibilistics C Means* ( $\eta$ ) adalah 2, nilai maksimum iterasi adalah 1000, nilai koefisien pembobot untuk kekhasan relatif ( $\alpha$ ) adalah 5, serta nilai koefisien pembobot untuk kekhasan absolut ( $\beta$ ) adalah 5.

Perbandingan metode *Fuzzy C Means* (FCM), *Fuzzy Possibilistics C Means* (FPCM) dan *Possibilistics Fuzzy C Means* (PFCM) dapat dilakukan dalam berbagai cara. Berikut adalah hasil perbandingan ketiga metode berdasarkan waktu komputasi:

**Tabel 1** Perbandingan Waktu Komputasi

Cluster	Waktu (Detik)		
	FCM	FPCM	PFCM
2	0.24	0.17	0.22
3	0.61	0.45	0.95
4	1.05	0.88	1.06
5	0.83	0.73	0.71

Berdasarkan **Tabel 1**, waktu komputasi metode *Fuzzy C Means* (FCM) ialah 0.24, 0.61, 1.05, dan 0.83 detik. Pada metode *Fuzzy Possibilistics C Means* (FPCM), waktu komputasinya adalah 0.17, 0.45, 0.88 dan 0.73 detik. Sementara pada metode *Possibilistics Fuzzy C*

*Means* (PFCM), waktu komputasinya adalah 0.22, 0.95, 1.06 dan 0.71 detik. Berdasarkan waktu komputasi, dapat disimpulkan bahwa pada studi kasus ini metode *Fuzzy Possibilistics C Means* (FPCM) memberikan waktu komputasi yang lebih cepat dibandingkan metode *Fuzzy C Means* (FCM) dan *Possibilistics Fuzzy C Means* (PFCM) pada hampir keseluruhan jumlah *cluster*.

Selanjutnya adalah perbandingan metode *Fuzzy C Means* (FCM), *Fuzzy Possibilistics C Means* (FPCM) dan *Possibilistics Fuzzy C Means* (PFCM).

**Tabel 2** Perbandingan Jumlah Iterasi dan Fungsi Objektif Minimum

Cluster	Jumlah Iterasi			Fungsi Objektif Minimum		
	FCM	FPCM	PFCM	FCM	FPCM	PFCM
2	60	47	57	4749.232	4755.07	34095.25
3	125	87	174	2839.475	2850.724	27859.29
4	162	119	148	1761.847	1779.072	21098.15
5	101	79	78	1165.686	1178.984	15778.39

Berdasarkan **Tabel 2**, jumlah iterasi yang dilakukan oleh metode *Fuzzy C Means* (FCM) untuk meminimumkan fungsi objektif adalah 60, 125, 162, dan 101. Pada metode *Fuzzy Possibilistics C Means* (FPCM), jumlah iterasinya adalah 47, 87, 119 serta 79. Sementara pada metode *Possibilistics Fuzzy C Means* (PFCM), jumlah iterasinya ialah 57, 174, 148 dan 78. Berdasarkan jumlah iterasi, dapat disimpulkan bahwa pada studi kasus ini metode *Fuzzy Possibilistics C Means* (FPCM) memiliki jumlah iterasi yang paling sedikit pada hampir keseluruhan *cluster* untuk meminimumkan fungsi objektif. Hal inilah yang membuat waktu komputasi metode *Fuzzy*

*Possibilistics C Means* (FPCM) lebih cepat dibandingkan dengan kedua metode lainnya.

Dari segi fungsi objektif, fungsi objektif pada metode *Fuzzy C Means* (FCM) secara berturut-turut adalah 4749.232, 2839.475, 1761.847 dan 1165.686. Pada metode *Fuzzy Possibilistics C Means* (FPCM), fungsi objektifnya secara berturut-turut adalah 4755.07, 2850.724, 1779.072, dan 1178.984. Sementara pada metode *Possibilistics Fuzzy C Means* (PFCM) ialah 34095.25, 27859.29, 21098.15 dan 15778.39. Berdasarkan fungsi objektif, nilai yang terendah adalah metode *Fuzzy C Means* (FCM). Namun begitu,

ketiga metode ini tidak dapat dibandingkan berdasarkan fungsi objektif lantaran perhitungan fungsi objektif ketiganya berbeda.

Penentuan metode terbaik dapat dilihat berdasarkan nilai indeks validitas *cluster*. Indeks validitas *cluster* yang digunakan yakni indeks *Modified Partition Coefficient* (MPC). Indeks tersebut berdasarkan nilai-nilai keanggotaan. Pada indeks *Modified Partition Coefficient* (MPC) mengindikasikan bahwa nilai yang terbesar merupakan nilai yang optimal. Berikut adalah hasil indeks validitas *cluster* metode *Fuzzy C Means* (FCM), *Fuzzy Possibilistics C Means* (FPCM) dan *Possibilistics Fuzzy C Means* (PFCM).

**Tabel 3** Penentuan Metode Terbaik

C	MPC		
	FCM	FPCM	PFCM
2	0.461864	0.461917	0.458076
3	0.484375	0.488855	0.463409
4	0.525552	0.527508	0.53025
5	0.527822	0.52942	0.525586

Berdasarkan **Tabel 3**, nilai indeks *Modified Partition Coefficient* (MPC) tertinggi pada jumlah *cluster* dua, tiga dan lima adalah metode *Fuzzy Possibilistics C Means* (FPCM). Sementara nilai indeks *Modified Partition Coefficient* (MPC) tertinggi pada jumlah *cluster* empat adalah metode *Possibilistics Fuzzy C Means* (PFCM). Berdasarkan indeks validitas *Modified Partition Coefficient* (MPC), metode *Fuzzy Possibilistic C Means* (FPCM) merupakan metode yang terbaik dibandingkan dengan metode *Fuzzy C Means* (FCM) dan metode *Possibilistics Fuzzy C Means* (PFCM). Hal tersebut didasarkan

pada nilai indeks *Modified Partition Coefficient* (MPC) yang tertinggi pada hampir keseluruhan *cluster*.

#### Penentuan Jumlah *Cluster* Optimal

Pada metode *Fuzzy Possibilistics C Means* (FPCM), jumlah *cluster* yang optimal dapat dilihat berdasarkan nilai indeks *Modified Partition Coefficient* (MPC) yang tertinggi. Berikut adalah hasil indeks *Modified Partition Coefficient* (MPC) pada metode *Fuzzy Possibilistics C Means* (FPCM).

**Tabel 4** Penentuan Jumlah *Cluster*

C	MPC
2	0.46192
3	0.48885
4	0.52751
5	0.52942
6	0.45158
7	0.49298

Berdasarkan **Tabel 4**, nilai indeks *Modified Partition Coefficient* (MPC) tertinggi pada metode *Fuzzy Possibilistics C Means* (FPCM) ialah dengan jumlah *cluster* lima. Nilainya adalah 0.52942. Berdasarkan hal tersebut, *cluster* yang optimal untuk digunakan adalah sebanyak lima *cluster*.

#### Profilisasi *Fuzzy Possibilistics C Means* Terhadap Hasil *Cluster*

Profilisasi *cluster* meliputi penggambaran karakteristik masing-masing *cluster*. Profilisasi *cluster* ini berdasarkan rata-rata sumber air minum layak, sumber penerangan dari listrik serta akses sanitasi layak pada masing-masing *cluster*.

**Tabel 5** Profilisasi *Fuzzy Possibilistics C Means* Terhadap Hasil *Cluster*

Cluster	Rata-Rata			Banyak Anggota
	Sumber Air Yang Layak	Penerangan Dengan Listrik	Sanitasi Layak	
1	62.29 %	94.58 %	51.68 %	9
2	59.09 %	55.81 %	33.06 %	1
3	67.28 %	96.19 %	65.10 %	9
4	85.23 %	99.71 %	89.34 %	4
5	75.58 %	98.52 %	72.19 %	11

Berdasarkan **Tabel 5**, *Cluster 1* memiliki rata-rata nilai persentase sumber air minum layak sebesar 62.29 %, sumber penerangan dari listrik sebesar 94.58 % serta nilai persentase akses sanitasi layak sebesar 51.68 %. Jumlah anggota *cluster 1* adalah 9 provinsi. Pada *cluster 2*, rata-rata nilai persentase sumber air minum layak sebesar 59.09 %, sumber penerangan dari listrik sebesar 55.81 % serta akses sanitasi layak sebesar 33.06 %. Jumlah anggota *cluster 2* adalah 1 provinsi.

Pada *cluster 3*, rata-rata nilai persentase sumber air minum layak sebesar 67.28 %, sumber penerangan dari listrik sebesar 96.19 % serta akses sanitasi layak sebesar 65.10

%. Jumlah anggota *cluster 3* adalah 9 provinsi. Pada *cluster 4*, rata-rata nilai persentase sumber air minum layak sebesar 85.23 %, sumber penerangan dari listrik sebesar 99.71 % serta akses sanitasi layak sebesar 89.34 %. Jumlah anggota *cluster 4* adalah 4 provinsi. Pada *cluster 5*, rata-rata nilai persentase sumber air minum layak sebesar 75.58 %, sumber penerangan dari listrik sebesar 98.52 % serta akses sanitasi layak sebesar 72.19 %. Jumlah anggota *cluster 5* adalah 11 provinsi.

Rata-rata nilai persentase tertinggi adalah *cluster 4*. Semakin tinggi nilai persentase menunjukkan *cluster* tersebut memiliki fasilitas tempat tinggal yang semakin layak. Berikut adalah hasil *cluster*:

**Tabel 6** Hasil *Cluster*

Cluster	Provinsi
1	Sumatera Barat, Bengkulu, Lampung, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Gorontalo, dan Sulawesi Barat
2	Papua
3	Aceh, Jambi, Sumatera Selatan, Jawa Barat, Banten, Sulawesi Tengah, Maluku, Maluku Utara, dan Papua Barat
4	Kepulauan Riau, DKI Jakarta, DI Yogyakarta, dan Bali
5	Sumatera Utara, Riau, Kepulauan Bangka Belitung, Jawa Tengah, Jawa Timur, Nusa Tenggara Barat, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Selatan, dan Sulawesi Tenggara



### Uji *Kruskall Wallis*

Uji *Kruskall Wallis* dilakukan guna melihat apakah terdapat perbedaan antara kelima *cluster* tersebut. Penggunaan uji *Kruskall Wallis* ini didasarkan pada pengujian

normalitas bahwa data tersebut tidak berdistribusi normal. Berikut adalah hasil uji *Kruskall Wallis* dengan menggunakan software SPSS:

**Tabel 7** Uji *Kruskall Wallis*

	Sumber Air Minum Layak	Sumber Penerangan dari Listrik	Akses Sanitasi Layak
Chi-Square	23.19	11.399	29.673
Df	4	4	4
Asymp. Sig.	0.000	0.022	0.000

Pengujian hipotesis untuk penelitian ini adalah sebagai berikut:

- i. Hipotesis
  - $H_0$ : Tidak terdapat perbedaan rata-rata antara *cluster* 1, *cluster* 2, *cluster* 3, *cluster* 4 dan *cluster* 5.
  - $H_1$ : Minimal terdapat satu rata-rata yang berbeda
- ii. Tingkat Signifikansi ( $\alpha$ ): 5% = 0.05
- iii. Daerah Kritis: Tolak  $H_0$  jika  $p\text{-value} < \alpha$
- iv. Statistika Uji
  - Nilai  $p\text{-value}$  pada variabel sumber air minum layak = 0.000
  - Nilai  $p\text{-value}$  pada variabel sumber penerangan dari listrik = 0.022
  - Nilai  $p\text{-value}$  pada variabel akses sanitasi layak = 0.000
- v. Keputusan
  - Pada variabel sumber air minum layak, sumber penerangan dari listrik dan akses sanitasi layak, nilai  $p\text{-value} < \alpha$  (0.05) sehingga tolak  $H_0$ .
- vi. Kesimpulan

Berdasarkan keputusan tersebut, maka minimal terdapat satu rata-rata yang berbeda pada keseluruhan variabel.

Berdasarkan hasil analisis menggunakan uji *Kruskall Wallis*, didapatkan hasil bahwa minimal terdapat satu rata-rata yang berbeda. Untuk mengetahui rata-rata kelompok yang berbeda, maka dilakukan uji perbandingan ganda. Berikut adalah hasil uji perbandingan ganda:

- i. Hipotesis
  - $H_0$ : Tidak terdapat perbedaan rata-rata antara kedua *cluster*
  - $H_1$ : Terdapat perbedaan rata-rata antara kedua *cluster*
- ii. Tingkat Signifikansi ( $\alpha$ ) = 10%
- iii. Daerah Kritis
  - Tolak  $H_0$  jika:  $|\overline{R}_k - \overline{R}_k| > Z_{\frac{\alpha}{c(c-1)}} \sqrt{\frac{n(n+1)}{12} \left( \frac{1}{n_k} + \frac{1}{n_k} \right)}$
- iv. Statistika Uji
  - Selisih rata-rata peringkat yang didapatkan pada kedua kelompok *cluster* ialah:

**Tabel 8** Selisih *Mean Ranks*

Kelompok 1 (1)	Kelompok 2 (2)	Sumber Air Minum Layak (3)	Sumber Penerangan dari Listrik (4)	Akses Sanitasi Layak (5)
Cluster 1	Cluster 2	6.22	11.55	5
	Cluster 3	3.78	2.72	9.55
	Cluster 4	22.53	15.69	26.5
	Cluster 5	14.87	8.39	18.54
Cluster 2	Cluster 3	10	14.28	14.55
	Cluster 4	28.75	27.25	31.5
	Cluster 5	21.09	19.95	23.54
Cluster 3	Cluster 4	18.75	12.97	16.94
	Cluster 5	11.09	5.68	8.99
Cluster 4	Cluster 5	7.66	7.29	7.95

## vii. Keputusan

**Tabel 9** Uji Perbandingan Ganda

(1)	(2)	(3)	(4)	(5)	Z tabel	Keputusan
1	2	9.22	11.55	5	27.03	Gagal tolak $H_0$ pada keseluruhan variabel
	3	3.78	2.72	9.55	12.09	Gagal tolak $H_0$ pada keseluruhan variabel
	4	22.53	15.69	26.5	15.41	Tolak $H_0$ pada keseluruhan variabel
	5	14.87	8.39	18.54	11.52	Tolak $H_0$ pada variabel sumber air minum layak dan akses sanitasi layak
2	3	10	14.28	14.55	27.03	Gagal tolak $H_0$ pada keseluruhan variabel
	4	28.75	27.25	31.5	28.67	Tolak $H_0$ pada variabel sumber air minum layak
	5	21.09	19.95	23.54	26.78	Gagal tolak $H_0$ pada keseluruhan variabel
3	4	18.75	12.97	16.94	15.41	Tolak $H_0$ pada variabel sumber air minum layak
	5	11.09	5.68	8.99	11.52	Gagal tolak $H_0$ pada keseluruhan variabel
4	5	7.66	7.29	7.95	14.97	Gagal tolak $H_0$ pada keseluruhan variabel

#### viii. Kesimpulan

Padan keseluruhan variabel, rata-rata kelompok yang berbeda adalah *cluster* 1 dan *cluster* 4. Pada variabel sumber air minum layak dan akses sanitasi layak, rata-rata kelompok yang berbeda adalah *cluster* 1 dan *cluster* 5. Sementara pada variabel sumber air minum layak, rata-rata kelompok yang berbeda adalah *cluster* 2 dan *cluster* 4 serta *cluster* 3 dan *cluster* 4.

Berdasarkan uji perbandingan ganda metode *kruskall wallis*, dapat disimpulkan bahwa tidak semua *cluster* adalah berbeda. *Cluster* yang memiliki perbedaan sangat tinggi dibandingkan dengan *cluster* yang lainnya adalah *cluster* 4. *Cluster* 4 mengindikasikan propinsi yang memiliki fasilitas tempat tinggal paling layak dibandingkan dengan *cluster* yang lainnya.

#### Kesimpulan

Berdasarkan hasil analisis dan pembahasan, diperoleh kesimpulan bahwa:

1. Hasil perbandingan indeks validitas *cluster Modified Partition Coefficient* (MPC) ialah metode *Fuzzy Possibilistic C Means* (FPCM) merupakan metode yang terbaik dibandingkan dengan metode *Fuzzy C Means* (FCM) dan metode *Possibilistics Fuzzy C Means* (PFCM) pada kasus *noisy* data ini .
2. Hasil profilisasi pada pengelompokan *noisy* data pada kasus fasilitas tempat

tinggal ialah bahwa *cluster* 4 merupakan propinsi yang memiliki rata-rata tertinggi pada variabel sumber air minum layak, sumber penerangan dari listrik dan akses sanitasi layak. Semakin tinggi nilai rata-ratanya menunjukkan semakin layak fasilitas tempat tinggal suatu propinsi.

#### Daftar Pustaka

- Kusumadewi, S., Hartati, S., Harjoko, A., dan Wardoyo, R. 2006. *Fuzzy Multi-Attribute Decision Making (FUZZY MADM)*. Yogyakarta: Graha Ilmu.
- Luthfi, E. T. 2017. *Fuzzy C Means Untuk Clustering Data. Seminar Nasional Teknologi*.
- Naba, A. 2009. *Belajar Cepat Fuzzy Logic Menggunakan Matlab*. Yogyakarta: C.V Andi Offset.
- Ned, G. dan Roger, J. 2000. *Fuzzy Logic Toolbox*. USA: Mathwork, inc.
- Pal, N. R., Kal, P., dan Bezdek, J. C. 1997. A Mixed C Means Clustering Model. *IEEE*, 11-21.
- Siyamto, Y. 2017. Pemanfaatan Data Mining Dengan Metode Clustering Untuk Evaluasi Biaya Dokumen Ekspor di PT Winstar Batam. *Media Informatika Budidarman*, 28-31.
- Xie, N., Hu, L., Luktarhan, N., dan Zhao, K. 2011. A Classification of Cluster Validity Indexes Based on

Membership Degree and  
Application. *Web Information  
System and Mining*, 43-50

