

**ANALISIS *CLUSTERING* KEMISKINAN DI INDONESIA
MENGUNAKAN ALGORITMA *FUZZY C-MEANS* DAN *FUZZY
POSSIBILISTIC C-MEANS***

(Skripsi)

Oleh

**VIRDA KURNIAWATI
1817031040**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2022**

ABSTRACT

ANALYSIS OF POVERTY CLUSTERING IN INDONESIA USING FUZZY C-MEANS AND FUZZY POSSIBILISTIC C-MEANS ALGORITHMS

By

VIRDA KURNIAWATI

Cluster analysis is a process of grouping data into groups based on their degree of similarity. The purpose of cluster analysis is to group a certain amount of data into groups so that each group formed will be filled with data that has the same characteristics. In this study, a grouping of 2021 poverty data in Indonesia will be carried out based on the district level. The clustering algorithms used are the Fuzzy C-Means (FCM) and Fuzzy Possibilistic C-Means (FPCM) algorithms. This study aims to determine the comparison of grouping results using the FCM and FPCM algorithms for district poverty data in Indonesia based on the Partition Entropy (PE) index value. The results showed that cluster analysis using the FPCM algorithm is the best algorithm to group poverty in Indonesia based on the PE validity index. Furthermore, the optimum number of clusters produced is as many as 2 clusters.

Keywords: Cluster Analysis, Fuzzy C-means, Fuzzy Possibilistic C-Means, Partition Entropy

ABSTRAK

ANALISIS *CLUSTERING* KEMISKINAN DI INDONESIA MENGUNAKAN ALGORITMA *FUZZY C-MEANS* DAN *FUZZY POSSIBILISTIC C-MEANS*

Oleh

VIRDA KURNIAWATI

Analisis kluster merupakan suatu proses mengelompokkan data ke dalam suatu kelompok berdasarkan tingkat kesamaannya. Tujuan dari analisis kluster adalah untuk mengelompokkan sejumlah data ke dalam kelompok sehingga setiap kelompok yang terbentuk akan terisi data yang memiliki karakteristik yang sama. Dalam penelitian ini akan dilakukan pengelompokan data kemiskinan tahun 2021 di Indonesia berdasarkan tingkat kabupaten. Algoritma pengelompokan yang digunakan adalah algoritma *Fuzzy C-Means* (FCM) dan *Fuzzy Possibilistic C-Means* (FPCM). Penelitian ini bertujuan untuk mengetahui perbandingan hasil pengelompokan menggunakan algoritma FCM dan FPCM untuk data kemiskinan kabupaten yang ada di Indonesia berdasarkan nilai indeks *Partition Entrhopy* (PE). Hasil yang diperoleh menunjukkan bahwa analisis kluster menggunakan algoritma FPCM merupakan algoritma terbaik untuk mengelompokkan kemiskinan di Indonesia berdasarkan indeks validitas PE. Lebih jauh, jumlah kluster optimum yang dihasilkan adalah sebanyak 2 kluster.

Kata Kunci: Analisis Kluster, *Fuzzy C-means*, *Fuzzy Possibistic C-Means*, *Partition Entrhopy*

**ANALISIS *CLUSTERING* KEMISKINAN DI INDONESIA
MENGUNAKAN ALGORITMA *FUZZY C-MEANS* DAN *FUZZY
POSSIBILISTIC C-MEANS***

Oleh

VIRDA KURNIAWATI

Skripsi

Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA MATEMATIKA

Pada

**Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2022**

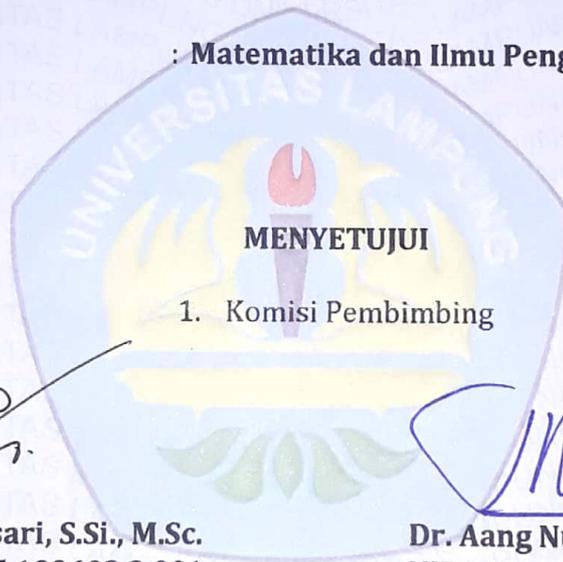
Judul Skripsi : **ANALISIS *CLUSTERING* KEMISKINAN DI INDONESIA
MENGUNAKAN ALGORITMA *FUZZY C-MEANS*
DAN *FUZZY POSSIBILISTIC C-MEANS***

Nama Mahasiswa : ***Virda Kurniawati***

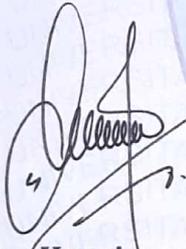
Nomor Pokok Mahasiswa : **1817031040**

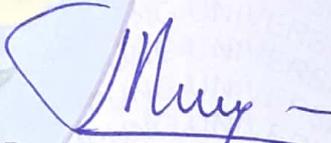
Jurusan : **Matematika**

Fakultas : **Matematika dan Ilmu Pengetahuan Alam**

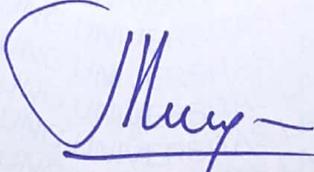


1. Komisi Pembimbing


Dian Kurniasari, S.Si., M.Sc.
NIP 19690305 199603 2 001


Dr. Aang Nuryaman, S.Si., M.Si.
NIP 19740316/200501 1 001

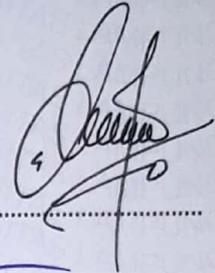
2. Ketua Jurusan Matematika


Dr. Aang Nuryaman, S.Si., M.Si.
NIP 19740316 200501 1 001

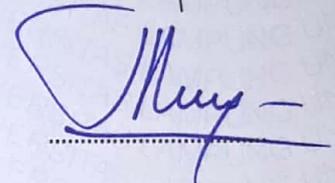
MENGESAHKAN

1. Tim Penguji

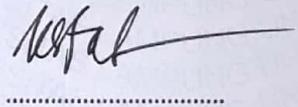
Ketua : **Dian Kurniasari, S.Si., M.Sc.**



Sekretaris : **Dr. Aang Nuryaman, S.Si., M.Si.**



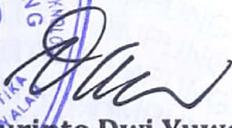
Penguji
Bukan Pembimbing : **Prof. Drs. Mustofa Usman, M.A., Ph.D.**



2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



Dr. Eng. Satripto Dwi Yuwono, S.Si., M.T.
NIP 19740705 200003 1 001



Tanggal Lulus Ujian Skripsi : 04 Juli 2022

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan di bawah ini :

Nama Mahasiswa : Virda Kurniawati

Nomor Pokok Mahasiswa : 1817031040

Jurusan : Matematika

Judul Skripsi : **ANALISIS *CLUSTERING* KEMISKINAN
DI INDONESIA MENGGUNAKAN
ALGORITMA *FUZZY C-MEANS* DAN
*FUZZY POSSIBILISTIC C-MEANS***

Dengan ini menyatakan bahwa penelitian ini adalah hasil pekerjaan saya sendiri dan apabila kemudian hari terbukti bahwa skripsi ini merupakan salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, Juli 2022

Penulis,



Virda Kurniawati

RIWAYAT HIDUP

Penulis bernama Virda Kurniawati lahir di Desa Marga Kencana, Kecamatan Tulang Bawang Udik, Kabupaten Tulang Bawang Barat pada 11 Mei 2000. Penulis merupakan anak ke dua dari dua bersaudara, dengan pasangan Bapak Boniran dan Ibu Bibit Suprihatin.

Penulis menempuh pendidikan pertama kali di Taman Kanak-Kanak (TK) Darmawanita pada tahun 2004-2006. Kemudian melanjutkan pendidikannya di Sekolah dasar (SD) yaitu SDN 1 Marga Kencana pada tahun 2006 – 2012. Selanjutnya, penulis melanjutkan ke Sekolah Menengah Pertama di SMPN 1 Tulang Bawang Udik pada tahun 2012-2015, dan melanjutkan Sekolah Menengah Atas di SMAN 1 Tumijajar pada tahun 2015-2018. Pada tahun 2018, melalui jalur Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN) penulis melanjutkan pendidikannya ke perguruan tinggi negeri untuk mencapai asa sebagai mahasiswi Program Studi S1 Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.

Selama menjadi mahasiswi penulis aktif dalam mengikuti kegiatan yang di selenggarakan oleh pengurus organisasi Himpunan Mahasiswa Matematika (HIMATIKA) FMIPA UNILA. Pada Tahun 2021 penulis melakukan Kuliah Kerja Nyata (KKN) di Desa Marga Kencana selama 40 hari sebagai bentuk pengabdian sebagai seorang mahasiswi kepada masyarakat dan lingkungan sekitar. Penulis juga melaksanakan Kerja Praktik (KP) di Badan Pusat Statistik Kota Metro selama kurang lebih 40 hari sebagai bentuk penerapan ilmu yang telah dipelajari selama proses perkuliahan.

KATA INSPIRASI

“Jika kamu tidak sanggup menahan lelahnya belajar, maka kamu harus siap menanggung pedihnya kebodohan.”

(Imam Syafi’i)

“Karena sesungguhnya sesudah kesulitan itu ada kemudahan.”

(Q.S Al-Insyirah : 5)

“Allah tidak membebani seseorang melainkan sesuai dengan kemampuannya”

(Q.S Al-Baqarah : 286)

“Sesungguhnya jika kamu bersyukur, niscaya aku akan menambah (nikmat) kepadamu”

(Q.S Ibrahim: 7)

“Bertikirlah positif, tidak peduli seberapa keras kehidupanmu”

(Ali bin Abi Thalib)

“Bukan tentang siapa yang paling cepat mencapai di garis *finish* tetapi, tentang siapa yang terus melangkah dan tidak berhenti berjuang”

(Virda Kurniawati)

PERSEMBAHAN

Alhamdulillah rabbilalamin, puji dan syukur kepada Allah SWT atas nikmat serta hidayahnya yang selalu terlimpahkan sehingga skripsi ini dapat terselesaikan dengan baik dan tepat pada waktunya. Oleh karena itu, dengan rasa syukur dan bahagia saya persembahkan rasa terimakasih saya kepada :

Diri Sendiri

Terimakasih sudah berjuang sampai titik ini. Kamu adalah orang yang luar biasa. Tetaplah menjadi manusia yang selalu bersyukur, selalu sabar dengan semua jalan-Nya, karena ketetapan Allah SWT adalah sebaik-baiknya ketetapan.

Bapak Boniran dan Ibu Bibit Suprihatin

Terimakasih kepada Bapak dan Mamak untuk semua pengorbanan, motivasi, doa dan ridho kalian serta dukungannya selama ini untuk terus melanjutkan pendidikan dan menjadi seorang sarjana. Terimakasih telah memberikan pelajaran berharga kepada anakmu ini tentang makna hidup yang sebenarnya sehingga kelak bisa menjadi orang yang bermanfaat bagi semua orang.

Dosen Pembimbing dan Pembahas

Terimakasih kepada dosen pembimbing dan pembahas yang sudah sangat membantu, memberikan motivasi, memberikan arahan serta ilmu yang berharga.

Sahabat-sahabatku

Terimakasih kepada semua orang-orang baik yang telah memberikan pengalaman, semangat, motivasinya, serta doa-doanya dan senantiasa memberikan dukungan dalam hal apapun

Almamater Tercinta Universitas Lampung

SANWACANA

Assalamu 'alaikum warrahmatullahi wabarakatuh

Segala puji dan syukur penulis ucapkan kepada Allah SWT atas segala nikmat dan karunia-Nya yang tak terhingga, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Analisis *Clustering* Kemiskinan di Indonesia Menggunakan Algoritma *Fuzzy C-Means* dan *Fuzzy Possibilistic C-Means*” dengan baik. Dalam penulisan skripsi ini tidak dapat terselesaikan tanpa adanya doa, bimbingan, bantuan dan dukungan dari berbagai pihak. Sehingga, dalam kesempatan ini penulis mengucapkan terimakasih kepada:

1. Ibu Dian Kurniasari, S.Si., M.Sc., selaku dosen pembimbing I yang senantiasa membimbing, memberi masukan saran serta mendukung penulis dalam menyelesaikan skripsi ini.
2. Bapak Dr. Aang Nuryaman, S.Si., M.Si., selaku dosen pembimbing II yang senantiasa memberikan bimbingan, pengarahan, serta saran sehingga penulis dapat menyelesaikan skripsi ini.
3. Bapak Prof. Drs. Mustofa M.A., Ph.D. selaku dosen penguji yang telah memberikan kritik dan saran yang membangun sehingga skripsi ini dapat terselesaikan dengan baik.
4. Bapak Dr.Aang Nuryaman, S.Si., M.Si., selaku Ketua Jurusan Matematika sekaligus dosen pembimbing akademik yang telah memberikan bimbingan dan arahan selama masa perkuliahan.
5. Bapak Dr. Eng. Suropto Dwi Yuwono, S.Si., M.T. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
6. Kedua Orang Tuaku, Bapak Boniran dan Ibu Bibit Suprihatin yang selalu memberikan motivasi, doa, dan dukungannya.
7. Seluruh dosen, staff, karyawan Jurusan Matematika Fakultas Matematika dan

Ilmu Pengetahuan Alam Universitas Lampung.

8. Keluarga sekalian yang selalu memberikan semangat kepada penulis serta doa-doanya.
9. Susanto yang selalu menemani dan memberikan motivasi untuk penulis.
10. Teman-teman kosanku yaitu Intan, Maziatun, Meydia, Muflihah, Risha, dan Silvi yang selalu memberikan dukungan semangat, doa, dan memberikan pengalaman yang luar biasa.
11. Teman-teman seperbimbingan yang selalu memberikan dukungan dan motivasi serta doa-doanya (Alifia, Dalfa, Farel, Febi, Ferzy, Joshua, Luthfia, Maydia, Oktina, Putri, Rekti, Shavira, Sulis, Zaenal).
12. Teman-temanku yaitu Anggelia, Faricha, Sabrina, Sofalina, Yusuf, Zamhara, teman kelas A dan semua teman satu jurusan matematika 2018 yang telah membantu serta memberikan semangat kepada penulis yang tidak bisa disebutkan satu persatu.
13. Orang-orang baik yang namanya tidak bisa saya sebutkan satu persatu yang telah menjadi teman terbaik penulis dimana selalu memberikan semangat dan menemani penulis dalam keadaan apapun, serta telah memberikan pengalaman dan banyak cerita selama masa perkuliahan.

Penulis menyadari bahwa masih banyak kekurangan dalam penulisan skripsi ini. Oleh karena itu, penulis mengharapkan masukan berupa kritik dan saran untuk dijadikan pelajaran kedepannya.

Bandar Lampung, Juli 2022

Penulis,

Virda Kurniawati

DAFTAR ISI

Halaman

DAFTAR TABEL	xiv
DAFTAR GAMBAR	xvi
I. PENDAHULUAN	1
1.1 Latar Belakang dan Masalah	1
1.2 Tujuan Penelitian.....	4
1.3 Manfaat Penelitian.....	5
II. TINJAUAN PUSTAKA	6
2.1 Data Mining	6
2.2 <i>Statistical Learning</i>	7
2.3 Analisis Kluster	10
2.4 Himpunan <i>Fuzzy</i>	11
2.5 Metode <i>Fuzzy C-Means</i> (FCM).....	12
2.6 Metode <i>Fuzzy Possibilistic C-Means</i> (FPCM).....	15
2.7 Validitas <i>Cluster</i>	17
III. METODOLOGI PENELITIAN	18
3.1 Tempat dan Waktu Penelitian	18
3.2 Data Penelitian.....	18
3.3 Metode Penelitian	19
IV. HASIL DAN PEMBAHASAN	21
4.1 Penginputan Data	21
4.2 Standarisasi Data	22
4.3 Analisis Kluster Menggunakan Metode <i>Fuzzy C-Means</i> (FCM)	22
4.4 Analisis Kluster Menggunakan Metode <i>Fuzzy Possibilistic C-Means</i> (FPCM).....	33
4.5 Penentuan Metode Terbaik.....	44
4.6 Penentuan Jumlah Kluster Optimal	45
V. KESIMPULAN	47
DAFTAR PUSTAKA	48
LAMPIRAN	

DAFTAR TABEL

Tabel	Halaman
1. Data Kemiskinan Tingkat Kabupaten di Indonesia	21
2. Standarisasi Data Kemiskinan Tingkat Kabupaten di Indonesia	22
3. Hasil Pusat Klaster Metode FCM k=2	23
4. Hasil Akhir Fungsi Objektif Metode FCM k=2.....	24
5. Hasil Derajat Keanggotaan Metode FCM k=2	24
6. Hasil Pusat Klaster Metode FCM k=3	26
7. Hasil Akhir Fungsi Objektif Metode FCM k=3.....	26
8. Hasil Derajat Keanggotaan Metode FCM k=3	26
9. Hasil Pusat Klaster Metode FCM k=4	28
10. Hasil Akhir Fungsi Objektif Metode FCM k=4.....	29
11. Hasil Derajat Keanggotaan Metode FCM k=4	29
12. Hasil Pusat Klaster Metode FCM k=5	31
13. Hasil Akhir Fungsi Objektif Metode FCM k=5.....	31
14. Hasil Derajat Keanggotaan Metode FCM k=5	32
15. Hasil Derajat Keanggotaan Metode FPCM k=2	33
16. Hasil Pusat Klaster Metode FPCM k=2	34
17. Hasil Akhir Fungsi Objektif Metode FPCM k=2.....	34
18. Nilai Kekhasan Absolut Metode FPCM k=2	35
19. Hasil Derajat Keanggotaan Metode FPCM k=3	36
20. Hasil Pusat Klaster Metode FPCM k=3	36
21. Hasil Akhir Fungsi Objektif Metode FPCM k=3.....	36
22. Nilai Kekhasan Absolut Metode FPCM k=3	37
23. Hasil Derajat Keanggotaan Metode FPCM k=4	39
24. Hasil Pusat Klaster Metode FPCM k=4.....	39

25. Hasil Akhir Fungsi Objektif Metode FPCM k=4.....	40
26. Nilai Kekhasan Absolut Metode FPCM k=4	40
27. Hasil Derajat Keanggotaan Metode FPCM k=5	42
28. Hasil Pusat Klaster Metode FPCM k=5	42
29. Hasil Akhir Fungsi Objektif Metode FPCM k=5.....	42
30. Nilai Kekhasan Absolut Metode FPCM k=5	43
31. Hasil Indeks Validitas <i>Cluster Partition Entrophy</i> (PE).....	45
32. Hasil Indeks Validitas PE Metode FCM dan FPCM	45

DAFTAR GAMBAR

Gambar	Halaman
1. <i>Flowchart</i> Metode FCM dan FPCM.....	20
2. <i>Cluster Plot</i> Metode FCM pada $k=2$	25
3. <i>Cluster Plot</i> Metode FCM pada $k=3$	27
4. <i>Cluster Plot</i> Metode FCM pada $k=4$	30
5. <i>Cluster Plot</i> Metode FCM pada $k=5$	32
6. <i>Cluster Plot</i> Metode FPCM pada $k=2$	35
7. <i>Cluster Plot</i> Metode FPCM pada $k=3$	38
8. <i>Cluster Plot</i> Metode FPCM pada $k=4$	41
9. <i>Cluster Plot</i> Metode FPCM pada $k=5$	44

I PENDAHULUAN

Bab ini akan dijelaskan mengenai pendahuluan dari penelitian, dibagi menjadi beberapa pokok bahasan penting seperti latar belakang dan masalah, tujuan dari penelitian, dan manfaat penelitian. Berikut merupakan penjabarannya.

1.1 Latar Belakang dan Masalah

Kemiskinan merupakan suatu masalah utama di Indonesia yang proses penanggulangannya masih terus diupayakan. Beberapa kondisi kemiskinan yaitu ketidakmampuan akses ekonomi, budaya, keadaan geografis, dan juga keadaan dalam masyarakat (Probosiwi, 2016). Menurut Alkire dan Foster (2011), mengemukakan kemiskinan pada umumnya dapat diukur dari kondisi ekonomi, tetapi untuk menggambarkan kemiskinan yang sesungguhnya, tidak cukup apabila hanya diukur melalui kondisi ekonomi saja. Kemiskinan sendiri tidak hanya diukur melalui indikator pendapatan, tetapi juga dapat diukur dari berbagai dimensi lain yang mencerminkan kemiskinan sesungguhnya, seperti terbatasnya akses terhadap air bersih dan sanitasi, pendidikan dan pelayanan kesehatan, serta pencatatan kelahiran. Penduduk miskin sendiri pada umumnya tidak menghasilkan pendapatan yang tetap dan harus memiliki rasa puas dengan pendapatan yang tidak menentu. Pendapatan yang dihasilkan juga biasanya masih berada dibawah batas pendapatan minimal. Sebagian besar penduduk miskin biasanya tidak memiliki jam kerja secara penuh, tidak memperoleh jaminan kesehatan, jaminan hari tua, perlindungan kerja, tidak memperoleh tingkat upah kerja minimum yang ditetapkan oleh pemerintah daerah dari berbagai jenis pekerjaan dan berbagai kota/daerah, tidak memperoleh upah lembur, tidak memperoleh hak cuti, memperoleh sedikit pendidikan/latihan-latihan dan memiliki etos kerja yang sangat rendah.

Menurut hasil pencatatan Badan Pusat Statistik (2021), diperoleh persentase penduduk miskin di Indonesia pada bulan Maret 2021 mencapai angka sebesar 10,14 persen, apabila diperbandingkan dengan Maret 2020, persentase penduduk miskin diperoleh angka peningkatannya sebesar 0,36 persen. Jumlah penduduk miskin pada Maret 2021 sebesar 27,54 juta orang, mengalami peningkatan 1,12 juta orang terhadap maret 2020, Garis kemiskinan pada Maret 2021 sebesar Rp. 472.525,00 per kapita/bulan dengan komposisi garis kemiskinan makanan sebesar Rp. 349.474,00 (73,96 persen) dan garis kemiskinan non makanan sebesar Rp. 123.051,00 (26,04 persen). Tercatat juga secara rata-rata rumah tangga miskin di Indonesia mempunyai 4,49 orang anggota rumah tangga. Dengan demikian, dapat diketahui besarnya garis kemiskinan per rumah tangga miskin secara rata-rata adalah sebesar Rp. 2.121.637,00 per rumah tangga/bulan. Berdasarkan pencatatan Badan Pusat Statistik, dapat disimpulkan bahwa hasil yang diperoleh tergolong tinggi dan merupakan salah satu masalah dalam perekonomian. Kemiskinan merupakan salah satu masalah yang harus segera ditangani oleh pemerintah. Salah satu cara untuk menangani masalah ini yaitu dengan melakukan pemerataan pembangunan. Pemerataan pembangunan sendiri akan lebih mudah dilaksanakan dengan mengelompokkan setiap daerah-daerahnya. Analisis yang dapat digunakan untuk membantu penanggulangan kemiskinan yaitu analisis kluster.

Analisis kluster merupakan suatu analisis yang digunakan untuk mengelompokkan objek tertentu dan membentuk suatu kelompok berdasarkan karakteristik yang sama (Mattjik dan Sumertajaya, 2011). Analisis kluster lebih tepat digunakan dalam pengelompokan suatu data, karena pada analisis ini memiliki beberapa kelebihan, diantaranya dapat melakukan pengelompokan data dalam jumlah yang besar dengan variabel yang cukup banyak, selain itu, dapat melakukan proses pengelompokan pada data berskala interval dan juga rasio (Yana, dkk. 2018). Istilah lain dari pengelompokan yaitu *unsupervised learning*. *Unsupervised learning* merupakan suatu analisis dimana pada proses pengelompokannya tidak dibutuhkan suatu variabel target (Harrington, 2012). Analisis kluster terdapat beberapa jenis metode, salah satunya adalah metode *partition clustering*. Penelitian ini akan menggunakan metode *partition clustering* yaitu metode *fuzzy*. Metode

fuzzy memiliki salah satu kelebihan, yaitu dapat mengolah data dalam ukuran yang cukup besar. Data dikatakan besar apabila jumlahnya atau n nya melebihi 30. Metode ini memiliki keunggulan daripada metode klasik yaitu dapat memproses pemikiran logis secara bahasa, sehingga dalam proses perancangannya tidak memerlukan persamaan matematika dari objek yang dikendalikan.

Analisis kluster memiliki banyak algoritma, salah satunya yaitu algoritma *K-means*, merupakan suatu algoritma yang cukup sederhana, mudah untuk dipahami, dan proses *running* program lebih cepat dari yang lainnya. Selain terdapat kelebihan, algoritma ini juga memiliki beberapa kekurangan diantaranya yaitu ketika telah ditentukan titik-titik pusat kluster, diperoleh hasil kluster belum cukup baik. Kekurangan tersebut dapat diatasi dengan menggunakan algoritma lain pada pengelompokan, salah satunya yaitu dengan algoritma *Fuzzy C-Means* (FCM). Algoritma FCM merupakan algoritma yang menggabungkan prinsip *fuzzy* dengan metode *K-Means*. Data yang akan menjadi anggota dari setiap kluster ditentukan oleh nilai derajat keanggotaan. Kelebihan dari algoritma ini yaitu menghasilkan karakteristik kluster yang lebih bagus daripada algoritma *K-means* (Sivarathi dan Govardhan, 2014). Namun, algoritma ini memiliki kekurangan yaitu sangat sensitif terhadap pencilan data. Salah satu algoritma yang dapat mengatasi hal tersebut yaitu algoritma *Fuzzy Possibilistic C-Means* (FPCM). Algoritma ini merupakan metode perluasan dari algoritma FCM dan *Possibilistic C-Means* (PCM) dengan menggabungkan prinsip *fuzzy* dan *possibilistic* yang dapat mengurangi kelemahan dari metode tersebut. Metode ini memiliki dua jenis keanggotaan yaitu kekhasan relatif (*fuzzy*) atau derajat keanggotaan dan kekhasan absolut (*possibilistic*). Kelebihan dari algoritma ini yaitu dapat mengatasi permasalahan pada data yang memuat *error* (Grover, 2014).

Penelitian terkait penerapan algoritma FCM dan FPCM telah banyak dilakukan. Misalnya Mauliyadi., dkk. (2013), memperoleh bahwa metode *Fuzzy Gustafson Kessel* (FGK) pada klusterisasi data Citra Satelit *Quickbird* lebih akurat dibandingkan dengan metode FCM. Kemudian, Putri., dkk (2014), mengimplementasikan algoritma FCM dan FPCM untuk klusterisasi data *tweets*

akun Twitter Tokopedia dengan indeks validitas *Modified Partition Coefficient* (MPC) dimana hasil klasterisasi metode FCM lebih baik daripada metode FPCM. Berdasarkan uraian di atas, penulis tertarik untuk mengusulkan analisis *clustering* dengan menggunakan algoritma FCM dan algoritma FPCM dengan menggunakan dataset yang berbeda yaitu tentang kemiskinan di Indonesia. Variabel data kemiskinan yang digunakan meliputi garis kemiskinan, persentase penduduk miskin, indeks kedalaman kemiskinan, indeks keparahan kemiskinan, dan jumlah penduduk miskin. Garis kemiskinan merupakan suatu nilai yang menyatakan pengeluaran minimum (rupiah) yang dibutuhkan oleh seseorang dalam memenuhi kebutuhan pokok hidupnya selama satu bulan. Persentase penduduk miskin merupakan suatu persentase banyaknya jumlah penduduk yang berada dibawah garis kemiskinan. Indeks kedalaman kemiskinan merupakan suatu ukuran rata-rata dari ketidakseimbangan pengeluaran penduduk yang miskin terhadap suatu garis kemiskinan. Selanjutnya, indeks keparahan kemiskinan merupakan sesuatu yang menggambarkan penyebaran pengeluaran dari penduduk miskin. Kedua algoritma yang digunakan pada penelitian ini akan dibandingkan keakuratannya berdasarkan indeks validitasnya yaitu menggunakan *Partition Entropy Index* (PEI) sehingga akan diperoleh algoritma yang tepat untuk mengelompokan kabupaten/kota di Indonesia.

1.2 Tujuan Penelitian

Adapun tujuan penelitian ini adalah sebagai berikut:

1. Mengetahui perbandingan hasil klasterisasi menggunakan dari algoritma FCM dan FPCM pada data kemiskinan kabupaten yang ada di Indonesia.
2. Mengetahui jumlah klaster optimum pada data kemiskinan tingkat kabupaten di Indonesia berdasarkan nilai indeks PE dari kedua algoritma yang digunakan.

1.3 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut ini:

1. Dapat dijadikan sebagai referensi bahan bacaan apabila akan melaksanakan suatu penelitian tentang analisis kluster dengan menggunakan algoritma FCM dan algoritma FPCM.
2. Dapat mengetahui algoritma yang paling baik digunakan untuk mengelompokkan kabupaten/kota dan mengelompokkan kabupaten di Indonesia pada data kemiskinan.
3. Dapat dijadikan acuan pengambilan keputusan dan kebijakan pemerintah di Indonesia untuk lebih melakukan pemerataan pembangunan dengan tujuan menurunkan persentase kemiskinan yang ada di Indonesia.

II. TINJAUAN PUSTAKA

Bab ini akan dijelaskan mengenai tinjauan pustaka sebagai penunjang materi dari penelitian. Adapun materi penunjang yang digunakan pada penelitian ini antara lain: data mining, *statistical learning*, analisis kluster, himpunan *fuzzy*, metode FCM, metode FPCM, dan validitas kluster. Berikut ini merupakan penjabaran lengkap pada bab ini.

2.1 Data Mining

Data mining merupakan proses yang menggunakan teknik matematika, statistik, *machine learning*, dan kecerdasan buatan untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai data yang besar (Turban. dkk, 2005). Definisi lain menyatakan bahwa data mining adalah analisis pengamatan database untuk menemukan hubungan yang tidak terduga dan untuk meringkas data dengan metode baru yang dapat dimengerti dan bermanfaat. Pengaplikasian data mining yaitu pada proses *Knowledge Discovery in Database* (KDD). Ciri-ciri data mining adalah berpautan dengan penemuan pola suatu data yang tersembunyi dan tidak terduga, data yang digunakan pada data mining yaitu suatu data yang berukuran sangat besar dengan tujuan untuk menghasilkan kesimpulan yang baik dimana berkaitan dengan pola data. Tujuan dari data mining untuk pengambilan suatu keputusan yang baik, terutama yang bersifat penting. Fungsi-fungsi dari data mining adalah fungsi deskripsi, fungsi pengelompokan, fungsi klasifikasi, fungsi prediksi, fungsi estimasi, dan fungsi asosiasi (Davies, 2004).

Menurut Fayyad (1996), tahapan yang dilakukan pada proses data mining sebagai berikut.

1. *Data selection*

Data selection adalah proses pengambilan data yang berhubungan dengan analisis dari basis data dan dilakukan teknik perolehan sebuah pengurangan representasi dari data untuk meminimalkan hilangnya informasi dari data. Pemilihan data atau seleksi data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi ini yang akan digunakan dalam proses data mining.

2. *Pre-processing/cleaning*

Sebelum proses data mining dilakukan, perlu dilakukan *proses cleaning* pada data yang menjadi fokus KDD. Proses ini mencakup antara lain membuang duplikasi data, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan pada data.

3. Transformasi

Transformasi adalah proses dalam pemilihan data yang telah dipilih, sehingga data tersebut sesuai dengan proses data mining. Definisi transformasi dalam KDD adalah proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data. Salah satu teknik pada transformasi yaitu standarisasi data. Standarisasi data merupakan proses yang dilakukan ketika diantara variabel-variabel yang diteliti terdapat perbedaan ukuran satuan yang besar. Perbedaan satuan yang mencolok dapat mengakibatkan perhitungan pada analisis kluster tidak valid. Standarisasi yang paling umum digunakan adalah konversi setiap variabel terhadap nilai standar (dikenal dengan *z-score* dengan melakukan substraksi nilai tengah dan membaginya dengan standar deviasi tiap variabel). Menurut Walpole dan Myers (1995), persamaan standarisasi data sebagai berikut:

$$z = \frac{x_i - \mu}{\sigma} \quad (2.1)$$

dimana:

z = standarisasi variabel

x_i = data ke- i

μ = rata-rata keseluruhan data setiap variabel

σ = standar deviasi

4. Data mining

Data mining adalah proses mencari pola yang sesuai dalam data yang ingin ditampilkan dengan menggunakan teknik atau metode tertentu. Metode atau algoritma dalam data mining sangat beragam. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

5. Interpretasi atau Evaluasi

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut dengan interpretasi. Tahap ini mencakup pemeriksaan apakah informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

2.2 *Statistical Learning*

Statistical Learning merupakan suatu pembelajaran mesin yang menggunakan ilmu bidang statistik. Pembelajaran dalam *statistical learning* merupakan perkembangan dari bidang statistik dan ilmu komputer yang dapat disebut dengan *machine learning* (James, *et al.*, 2017). Menurut Russell dan Norvig (2016), *machine learning* adalah pembelajaran mesin yang termasuk kedalam pendekatan *artificial intelligence*. Pembelajaran *machine learning* menggunakan prinsip statistik. *Statistical learning* dibagi menjadi tiga yaitu *supervised learning*, *unsupervised learning* dan *semi-supervised learning*.

Menurut Brownlee (2016), penjelasan tentang *supervised learning*, *unsupervised learning* dan *semi-supervised learning* sebagai berikut ini:

a. *Supervised Learning*

Supervised learning atau proses pelatihan terawasi adalah pendekatan pada *machine learning* dimana menggunakan data berlabel. Pelatihan *supervised*

learning membutuhkan *training dataset*. *Supervised learning* dikelompokkan menjadi dua yaitu klasifikasi dan regresi. Klasifikasi merupakan pembelajaran ketika variabel *output* berbentuk kategori, seperti merah atau biru, ada penyakit dan tidak ada penyakit. Regresi merupakan ketika variabel *output* adalah bernilai real, seperti dollar atau berat. Pelatihan *supervised learning* dilakukan dengan merancang data yang digunakan untuk mengawasi algoritma dalam melakukan klasifikasi atau prediksi data secara akurat. Keuntungan dari pelatihan ini adalah proses yang sederhana dan mudah dipahami. Adapun keterbatasan dari pelatihan ini yaitu membutuhkan dalam jumlah besar sehingga bisa memakan waktu yang cukup lama karena harus diberi label manual oleh manusia, serta membutuhkan validasi berulang dengan dataset untuk melatih ketepatan model.

b. *Unsupervised Learning*

Unsupervised learning atau proses pelatihan tak terawasi adalah pendekatan pada *machine learning* dimana data yang akan diolah tidak mempunyai label. Tujuan dari metode ini adalah untuk mengelompokkan unit yang memiliki kemiripan pada area tertentu. *Unsupervised learning* tidak membutuhkan data *training*. *Unsupervised learning* dikelompokkan menjadi dua yaitu *clustering* dan *association rule*. *Clustering* adalah suatu proses pengelompokan data. Salah satu contoh dari *clustering* yaitu mengelompokkan pelanggan berdasarkan perilaku pembelian. Sedangkan, *association rule* atau asosiasi adalah aturan yang menggambarkan sebagian besar data, seperti orang yang membeli A juga cenderung membeli B. Keuntungan dari pelatihan ini adalah karena data tidak membutuhkan label, algoritma lebih leluasa untuk mencari pola yang sebelumnya belum diketahui. Selain memiliki keuntungan, pelatihan ini juga memiliki beberapa kekurangan dimana sulitnya untuk menemukan pola awal untuk membentuk algoritma.

c. *Semi-supervised Learning*

Semi-supervised learning adalah pelatihan yang tidak bisa dikategorikan ke dalam *supervised* atau *unsupervised learning*. Pelatihan ini cocok digunakan untuk sejumlah data berukuran besar, kemudian dibagi menjadi dua bagian yaitu data yang diberi label dan tidak diberi label. Keuntungan dari pelatihan

ini adalah membutuhkan biaya yang lebih murah karena hanya setengah data yang diberi label dan tidak memerlukan tenaga ahli untuk mengolahnya. Pelatihan ini sangat berguna ketika sulit untuk mendapatkan fitur yang relevan dari data dan ketika data yang dimiliki berjumlah besar.

2.3 Analisis Klaster

Analisis klaster adalah suatu prosedur untuk mengelompokkan data sehingga terbentuk kelompok dengan ciri yang sama. Data-data yang memiliki kemiripan ciri-cirinya akan berkumpul menjadi satu kelompok, sementara data yang memiliki perbedaan ciri-cirinya, akan berkumpul menjadi kelompok yang lain. Tujuan dari analisis klaster adalah untuk mengelompokkan sejumlah data ke dalam kelompok sehingga, hasil anggota dari setiap kelompok data akan memiliki karakteristik yang sama. Analisis klaster merupakan bagian dari ilmu data mining yang bersifat tanpa arahan atau pengawasan (*unsupervised learning*). *Unsupervised learning* adalah suatu pendekatan yang tidak memiliki data latih atau dengan kata lain tidak ada fase *learning* pada data tersebut (Siyamto, 2017).

Menurut Johson dan Wichern (2007), analisis klaster bertujuan untuk mengidentifikasi sekelompok objek yang mempunyai kemiripan karakteristik tertentu yang dapat dipisahkan dengan kelompok objek yang lainnya. Oleh karena itu, objek yang berada dalam kelompok yang sama relatif lebih homogen dibandingkan objek yang berada pada kelompok yang berbeda. Prinsip pada analisis klaster adalah memaksimalkan homogenitas (kesamaan) dalam satu kelompok serta memaksimalkan heterogenitas (ketidaksamaan) antar kelompok. Analisis klaster memiliki dua metode yaitu *hierarchical* dan *partitioning*. Metode *hierarchical* terdiri dari metode *complete linkage clustering*, *single linkage clustering*, *average linkage clustering* dan *centroid linkage clustering*. Sedangkan metode *partitioning*, terdiri dari *K-means* dan FCM. Pengelompokan dengan metode hierarki, kelompok disusun dengan dendogram. Teknik pada klaster metode hierarki dibagi menjadi dua yaitu *agglomerative* dan *devisive*. Metode *agglomerative* merupakan suatu metode dimana terdapat penggabungan dua grup

yang terdekat setiap iterasinya, sedangkan metode *devisive* adalah suatu metode dimana terdapat pembagian dari seluruh set data ke dalam klaster. Analisis klaster dengan metode partisi, klaster disusun dengan penentuan titik pusat klaster (*centroid*) yang bertujuan untuk meminimumkan jarak dari seluruh pusat *centroid* masing-masing (Alfina, dkk., 2012). *Clustering* dengan menggunakan metode partisi terdapat tiga jenis yaitu partisi klasik, partisi *fuzzy*, dan partisi *possibilistic*. Partisi klasik adalah suatu data secara eksklusif menjadi anggota hanya pada satu klaster saja. Partisi *fuzzy* merupakan suatu metode yang memiliki nilai derajat keanggotaan setiap data pada semua klaster, dimana jumlah dari nilai derajat keanggotaannya adalah satu. Sedangkan, partisi *possibilistic* merupakan suatu metode yang memiliki jumlah nilai keanggotaan suatu data pada semua klaster tidak harus satu (Kusumadewi dan Purnomo, 2010).

2.4 Himpunan *Fuzzy*

Himpunan *fuzzy* diperkenalkan oleh Lotfi A. Zadeh pada tahun 1965. Munculnya himpunan *fuzzy* ini mematahkan teori probabilitas. Menyatakan bahwa tidak hanya teori probabilitas saja yang dapat digunakan dalam hal memecahkan masalah yang mengandung ketidakpastian. Namun, teori himpunan *fuzzy* tidak bisa dikatakan sebagai pengganti dari teori probabilitas. Himpunan *fuzzy* merupakan suatu nilai derajat keanggotaan dan dinyatakan dalam suatu bilangan riil pada interval tertutup yaitu $[0,1]$. Dasar dari himpunan *fuzzy* berasal dari himpunan klasik, dimana dalam himpunan klasik, misalkan keberadaan suatu elemen himpunan A, hanya akan memiliki dua kemungkinan keanggotaannya yaitu, suatu elemen tersebut akan masuk sebagai anggota dari A atau tidak menjadi anggota dari A. Keanggotaan dari himpunan *fuzzy* yaitu ditentukan berdasarkan nilai derajat keanggotaannya. Konsep dari himpunan *fuzzy* untuk menjelaskan ketidakpastian dan ketidaktepatan. Atribut pada himpunan *fuzzy* yaitu linguistik dan numeris. Linguistik yaitu penamaan suatu grup yang mewakili suatu keadaan tertentu menggunakan bahasa, seperti: muda, paruh baya, dan tua. Atribut numeris

merupakan suatu nilai yang mewujudkan ukuran dari suatu variabel, seperti: 5, 10, dan 15 (Kusumadewi dan Purnomo, 2010).

2.5 Metode *Fuzzy C-means* (FCM)

Menurut Kusumadewi dan Purnomo (2010), metode FCM adalah adalah suatu metode dalam pengklasteran data dimana setiap data yang akan menjadi anggota masing-masing kluster ditentukan oleh nilai keanggotaan. Metode ini merupakan suatu metode yang pertama kali diperkenalkan oleh Jim Besdek pada tahun 1981. Kelebihan dari metode ini adalah penempatan pusat kluster yang lebih tepat dibandingkan dengan metode kluster yang lainnya. Langkah pertama pada metode ini yaitu menentukan pusat kluster. Langkah ini dilakukan untuk menandai lokasi rata-rata dari masing-masing kluster. Hasil pusat kluster yang diperoleh dalam langkah pertama ini masih belum akurat. Sehingga, untuk memperoleh hasil pusat kluster yang tepat, dibutuhkan perbaikan pusat kluster secara berulang-ulang hingga pusat kluster akan berada pada titik yang tepat. Setiap data akan memiliki derajat keanggotaan untuk tiap-tiap kluster. Hasil derajat keanggotaan yang tepat akan dibutuhkan perbaikan nilai keanggotaan sehingga data akan menempati kluster yang tepat. Perulangan ini didasarkan pada minimasi fungsi objektif. Fungsi objektif pada FCM sebagai berikut:

$$J_w(\mathbf{U}, \mathbf{V}) = \sum_{i=1}^n \sum_{k=1}^c ((\mu_{ik})^w d_{ik}^2) \quad (2.2)$$

dimana:

$J_w(\mathbf{U}, \mathbf{V})$	= fungsi objektif terhadap \mathbf{U} dan \mathbf{V}
k	= jumlah kluster
i	= jumlah data yang diproses
w	= pangkat (pembobot), $w > 1$
μ_{ik}	= elemen-elemen matriks partisi \mathbf{U} atau fungsi keanggotaan data ke- i pada kluster ke- k
d_{ik}	= jarak objek ke- i dengan pusat kluster ke- k

Adapun persamaan jarak yang digunakan sebagai berikut.

$$d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - v_{kj})^2} \quad (2.3)$$

dimana:

- m : jumlah variabel
- x_{ij} : data dari objek ke- i pada variabel ke- j
- v_{kj} : pusat kluster ke- k pada variabel ke- j

\mathbf{U} merupakan suatu matriks partisi awal. Dalam metode FCM penentuan matriks partisi awal digunakan untuk mencari nilai awal pusat kluster, dimana nilai awal pusat kluster akan didapatkan nilai yang belum akurat. Sehingga, dibutuhkan perbaikan nilai pusat kluster untuk mendapatkan nilai yang tepat. Dibawah ini merupakan persamaan matriks partisi awal.

$$\mathbf{U} = \begin{bmatrix} \mu_{11} & \mu_{12} & \dots & \mu_{1k} \\ \mu_{21} & \mu_{22} & \dots & \mu_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{i1} & \mu_{i2} & \dots & \mu_{ik} \end{bmatrix} \quad (2.4)$$

Matriks partisi awal adalah matriks dengan elemen-elemennya yaitu nilai derajat keanggotaan. Nilai derajat keanggotaan merupakan nilai digunakan untuk menentukan anggota setiap kluster berdasarkan perbandingan nilai derajat keanggotaan pada setiap data. Semakin besar nilai derajat keanggotaannya maka, data tersebut akan masuk ke dalam anggota kluster tersebut.

\mathbf{V} merupakan matriks pusat kluster. berikut ini merupakan persamaan matriks dari pusat kluster.

$$\mathbf{V} = \begin{bmatrix} v_{11} & v_{12} & \dots & v_{1m} \\ v_{21} & v_{22} & \dots & v_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ v_{k1} & v_{k2} & \dots & v_{km} \end{bmatrix} \quad (2.5)$$

Berikut ini merupakan persamaan untuk mencari nilai derajat keanggotaan pada metode *Fuzzy C-Means*.

$$\mu_{ik} = \left[\frac{[\sum_{j=1}^m (x_{ij} - v_{kj})^2]^{\frac{1}{w-1}}}{\sum_{k=1}^c [\sum_{j=1}^m (x_{ij} - v_{kj})^2]^{\frac{1}{w-1}}} \right]^{-1} \quad (2.6)$$

dimana:

- μ_{ik} = elemen-elemen matriks partisi U atau nilai derajat keanggotaan data ke- i ($i = 1, 2, \dots, n$) pada kluster ke- k ($k = 1, 2, \dots, c$)
- m = jumlah variabel setiap data
- k = jumlah kluster
- x_{ij} = data ke- i pada variabel ke- j
- v_{kj} = nilai pusat kluster ke- k pada variabel ke- j
- w = pangkat pembobot; $w > 1$

sedangkan, untuk mencari nilai pusat kluster pada algoritma FCM dapat menggunakan persamaan sebagai berikut:

$$v_{kj} = \frac{\sum_{k=1}^n ((\mu_{ik})^w x_{ij})}{\sum_{k=1}^n (\mu_{ik})^w} \quad (2.7)$$

dimana:

- v_{kj} = nilai pusat kluster ke- k pada variabel ke- j
- i = jumlah data yang diproses
- μ_{ik} = elemen-elemen matriks partisi U atau derajat keanggotaan data ke- i pada kluster ke- k
- x_{ij} = data ke- i pada variabel ke- j
- w = pangkat pembobot; $w > 1$

2.6 Metode *Fuzzy Possibilistic C-Means* (FPCM)

Menurut Zarandi., dkk (2021), algoritma FCM dan beberapa algoritma pengembangannya adalah suatu metode pengelompokan yang paling terkenal dalam analisis kluster. Algoritma FCM adalah algoritma yang pertama kali diperkenalkan oleh Dunn, kemudian dikembangkan lagi oleh Bezdek. Algoritma FCM diketahui memiliki beberapa kekurangan dan masalah, masalah yang paling utama adalah sangat rentan terhadap *outlier*. Kekurangan tersebut dapat diatasi menggunakan pengembangan dari algoritma FCM yaitu algoritma *Possibilistic C-Means* (PCM). Kelebihan dari algoritma PCM ini bisa mengatasi keberadaan *outlier* pada data. Selain memiliki kelebihan, algoritma PCM memiliki beberapa kekurangan, salah satunya yaitu bergantung pada nilai partisi awal yang baik dan tidak menghasilkan kluster yang tepat. Kemudian, Pal dan Bezdek mendefinisikan algoritma FPCM untuk mengatasi kekurangan tersebut. Algoritma FPCM merupakan pengembangan dari algoritma dari FCM dan PCM. Algoritma ini terdapat beberapa nilai penting yaitu nilai pusat kluster, nilai fungsi objektif dan nilai kekhasan. Adapun persamaan untuk mencari nilai fungsi objektif dari algoritma FPCM adalah sebagai berikut.

$$J_{w,\eta}(U, T, V) = \sum_{k=1}^c \sum_{i=1}^n (\mu_{ik}^w + t_{ik}^\eta) d_{ik}^2 \quad (2.8)$$

dimana:

$J_{w,\eta}(U, T, V)$	= fungsi objektif terhadap U , T dan V
k	= jumlah kluster
i	= jumlah data yang diproses
w	= pangkat pembobot pada matriks kekhasan relatif atau derajat keanggotaan; $w > 1$
η	= pangkat pembobot pada matriks kekhasan absolut; $\eta > 1$
μ_{ik}	= elemen-elemen matriks partisi U atau fungsi keanggotaan data ke- i ($i = 1, 2, \dots, n$) pada kluster ke- k ($k = 1, 2, \dots, c$)
t_{ik}	= elemen-elemen matriks kekhasan absolut data ke- i pada kluster ke- k

Matriks kekhasan absolut atau dilambangkan dengan T adalah matriks yang dihasilkan dari nilai akhir derajat keanggotaan (μ_{ik}) dan pusat kluster (v_{kj}) pada algoritma FCM. Berikut ini merupakan persamaan dari matriks kekhasan absolut pada metode FPCM.

$$\mathbf{T} = \begin{bmatrix} t_{11} & t_{12} & \dots & t_{1k} \\ t_{21} & t_{22} & \dots & t_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ t_{i1} & t_{i2} & \dots & t_{ik} \end{bmatrix} \quad (2.9)$$

Pada persamaan matriks kekhasan absolut diatas, terdapat elemen-elemen matriks yaitu $t_{11}, t_{12}, \dots, t_{ik}$. Nilai-nilai pada elemen matriks kekhasan absolut dapat diperoleh dengan persamaan berikut ini.

$$t_{ik} = \left[\frac{[\sum_{j=1}^m (x_{ij} - v_{kj})^2]^{\frac{1}{\eta-1}}}{\sum_{i=1}^n [\sum_{j=1}^m (x_{ij} - v_{kj})^2]^{\frac{1}{\eta-1}}} \right]^{-1} \quad (2.10)$$

dimana :

t_{ik} = elemen-elemen matriks kekhasan absolut data ke- i pada kluster ke- k

Selain metode FCM, pada metode FPCM juga terdapat pusat kluster yang dapat diperoleh dengan menggunakan persamaan sebagai berikut.

$$v_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik}^w + t_{ik}^\eta) x_{ij})}{\sum_{i=1}^n \mu_{ik}^w + t_{ik}^\eta} \quad (2.11)$$

dimana:

v_{kj} = nilai pusat kluster ke- k pada variabel ke- j

i = jumlah data yang diproses

w = pangkat pembobot pada matriks derajat keanggotaan; $w > 1$

η = pangkat pembobot pada matriks kekhasan absolut; $\eta > 1$

μ_{ik} = elemen-elemen matriks partisi U atau fungsi keanggotaan data ke- i ($i = 1, 2, \dots, n$) pada kluster ke- k ($k = 1, 2, \dots, c$)

t_{ik} = elemen-elemen matriks kekhasan absolut data ke- i pada kluster ke- k

x_{ij} = data ke- i pada variabel ke- j

2.7 Validitas Cluster

Menurut Xie, dkk (2011), validitas *cluster* merupakan suatu cara untuk mengevaluasi kualitas dari *cluster* dan menentukan seberapa baik data diwakili oleh *cluster* tersebut. Validitas *cluster* dapat dilakukan dengan menggunakan *Cluster Validity Index* (CVI). CVI merupakan ukuran validitas yang digunakan untuk menemukan jumlah *cluster* optimal yang sepenuhnya dapat menjelaskan struktur data dan mengukur derajat kekompakan dan separasi struktur data pada seluruh *cluster*. Salah satu CVI ini meliputi metode *Partition Entropy Index* (PEI). Adapun persamaan PEI didefinisikan seperti berikut ini.

$$PEI = -\frac{1}{n} \sum_{k=1}^c \sum_{i=1}^n \mu_{ik} \ln(\mu_{ik}) \quad (2.12)$$

dimana:

i = jumlah data yang diproses

k = jumlah klaster

μ_{ik} = nilai keanggotaan objek ke- i dengan pusat kelompok ke- k

Partition Entropy Index merupakan evaluasi yang mengukur tingkat keaburan dari partisi *cluster*, klaster optimal didapatkan jika nilai yang diperoleh semakin kecil (angka mendekati 0). Validitas klaster dengan menggunakan PEI lebih mudah digunakan dan merupakan metode yang baik untuk menentukan jumlah klaster yang optimal.

III. METODOLOGI PENELITIAN

Pada bab ini akan dijelaskan mengenai metodologi dari penelitian. Metodologi penelitian meliputi waktu dan tempat penelitian, data yang digunakan dalam penelitian dan langkah-langkah pada metode yang digunakan. Berikut ini merupakan penjelasannya.

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan pada semester genap tahun ajaran 2021/2022 bertempat di Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.

3.2 Data Penelitian

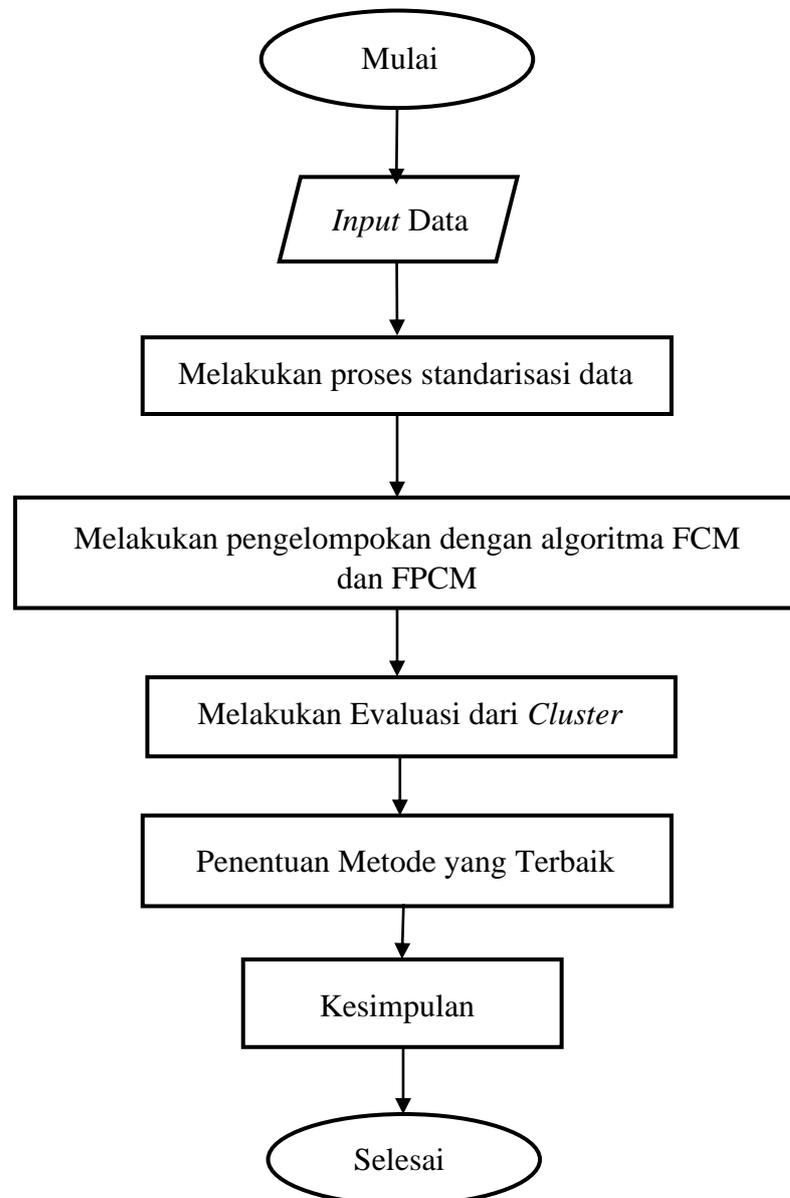
Data yang di gunakan pada penelitian ini adalah data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) Indonesia yaitu data kemiskinan setiap kabupaten/kota di Indonesia. Jumlah data kemiskinan ditingkat kabupaten/kota yaitu sebanyak 514 data. Data kemiskinan ini merupakan data kemiskinan di tahun 2021. Adapun variabel-variabel yang digunakan pada penelitian ini yaitu X_1 = garis kemiskinan (rupiah), X_2 = persentase penduduk miskin (persen), X_3 = indeks kedalaman kemiskinan (persen), X_4 = indeks keparahan kemiskinan (persen), X_5 = jumlah penduduk miskin (juta).

3.3 Metode Penelitian

Penelitian ini akan menentukan metode manakah yang terbaik antara algoritma *Fuzzy C Means* dan *Fuzzy Possibilistic C-Means* dengan menggunakan bantuan aplikasi RStudio 4.1.2.

1. langkah-langkah analisis kluster dengan algoritma *Fuzzy C Means* sebagai berikut:
 - a. Menginput data kemiskinan
 - b. Melakukan standarisasi data kemiskinan dengan menggunakan *z-score*. Langkah ini dilakukan karena variabel-variabel pada data memiliki perbedaan ukuran satuan
 - c. Melakukan proses pengelompokan
Proses pengelompokan pada pengelompokan pada algoritma FCM dilakukan dengan menggunakan jumlah kluster yaitu 2 hingga 5, pangkat pembobot yaitu 2.
 - d. Melakukan evaluasi terhadap hasil kluster-kluster yang telah terbentuk dengan menggunakan *Partition Entropy Index (PEI)*.
2. langkah-langkah analisis kluster dengan algoritma *Fuzzy Possibilistic C Means* sebagai berikut:
 - a. Menginput data kemiskinan
 - b. Melakukan standarisasi data kemiskinan dengan menggunakan *z-score*. Proses ini dilakukan karena variabel-variabel pada data memiliki perbedaan ukuran satuan
 - c. Melakukan proses pengelompokan
Proses pengelompokan pada algoritma FPCM ini menggunakan jumlah kluster 2 sampai dengan 5, pangkat pembobot yaitu 2.
 - d. Melakukan evaluasi terhadap hasil kluster-kluster yang telah terbentuk dengan menggunakan *Partition Entropy Index (PEI)*.
3. Membandingkan hasil dari kedua algoritma yang telah dilakukan, setelah itu menentukan algoritma mana yang paling baik digunakan untuk pengelompokan.

Adapun diagram alir langkah-langkah pada algoritma FCM dan *Fuzzy* FPCM:



Gambar 1. *Flowchart* Metode FCM dan FPCM

V. KESIMPULAN

Bab ini akan dijelaskan mengenai kesimpulan dari penelitian yang telah dilakukan. Berikut merupakan kesimpulan yang diperoleh dari penelitian ini.

5.1 Kesimpulan

Berdasarkan hasil dan pembahasan yang telah dijelaskan pada bab sebelumnya, maka dapat diambil kesimpulan sebagai berikut:

1. Hasil perbandingan metode FCM dan metode FPCM diperoleh bahwa metode FPCM merupakan metode yang terbaik dalam mengelompokkan data kemiskinan tingkat kabupaten di Indonesia. Penentuan metode ini merupakan metode terbaik yaitu menggunakan nilai indeks validitas *cluster* PE. Hal ini dilihat dari nilai indeks validitas PE pada metode FPCM lebih kecil dari pada nilai indeks validitas *cluster* PE pada metode FCM.
2. Berdasarkan hasil pengujian indeks validitas pada hasil analisis kluster dengan algoritma FCM dan FPCM, jumlah kluster 2 merupakan jumlah kluster yang optimal karena memiliki nilai indeks validitas yang paling rendah dari pada jumlah kluster yang lainnya yaitu dengan nilai 0,46827 dan 0,02074.

DAFTAR PUSTAKA

- Alfina, T., Santosa, B., dan Barakbah, A.R. Analisa Perbandingan Metode Hierarchical Clustering, K-means dan Gabungan Keduanya dalam Cluster Data (Studi kasus: Problem Kerja Praktek Jurusan Teknik Industri ITS). *Jurnal Teknik ITS*. **1**.
- Alkire, S. dan Foster, J. 2011. Counting and Multidimensional Poverty Measurement. *Journal of Public Economics* **95**. 476-487.
- Badan Pusat Statistik. 2021. *Profil Kemiskinan di Indonesia Maret 2021*. Badan Pusat Statistik.
- Brownlee, J. 2016. *Master Machine Learning Algorithms: Discover How They Work and Implement Them form Scratch*. Melbourne, Australia.
- Davies, P.B. 2004. *Database System Third Edition*. Palgrave Macmillan, New York.
- Ferezagia, D.V. 2018. Analisis Tingkat Kemiskinan di Indonesia. *Jurnal Sosial Humaniora aterapan*. **1**(1).
- Fayyad, U. 1996. *Advances in Knowledge Discovery and Data Mining*. MIT Press, USA.
- Grover, N. 2014. A Study of Various Fuzzy Clustering Algorithms. *International Journal of Engineering Research*. **3**(3): 177-181.
- Harrington, P. 2012. *Machine Learning in Action*. Manning Publications Co, Shelter Island, New York.
- James, G., Witten, D., Hastie, T., and Tibshirani, R. 2013. *An Introduction to Statistical Learning with Aplications in R*. Springer, New York.
- Johnson, R.A. dan Wichern, D.W. 2007. *Applied Multivariate Statistical Analysis*. 6th Edition. Pearson Pretice Hall, New Jersey.
- Kusumadewi, S. dan Purnomo 2010. *Aplikasi Logika Fuzzy untuk Pendukung Keputusan*. Graha Ilmu, Yogyakarta.

- Mattjik, A.A. dan Sumertajaya, I.M. 2011. *Sidik Peubah Ganda dengan Menggunakan SAS*. Institut Pertanian Bogor Press, Bogor.
- Mauliyadi, A., Sofyan, H., dan Subianto, M. 2013. Perbandingan Metode Fuzzy C-Means (FCM) dan Fuzzy Gustafson-Kessel (FGK) Menggunakan Data Citra Satelit Quickbird. *Jurnal Matematika*. 1-5.
- Putri, G.N.S., Ispriyanti, D., dan Widiharih, T. 2022. Implementasi Algoritma Fuzzy C-Means dan Fuzzy Possibilistic C-Means untuk Klasterisasi Data Tweets pada Akun Twitter Tokopedia. *Jurnal Gaussian*. **11**(1): 86-98.
- Probosiwi, R. 2016. Pengangguran dan Pengaruhnya terhadap Tingkat Kemiskinan. *Journal PKS*. **15**(2): 89-100.
- Russel, S.J. dan Norvig, P. 2016. *Artificial Intelligence: A Modern Approach*. Pearson Education Limited, Malaysia.
- Sivarathri, S. dan Govardhan, A. 2014. Experiments on Hypothesis “Fuzzy K-Means is Better than K-Means for Clustering”. *International Journal of Data Mining & Knowledge Process (IJDKP)*. **4**(5): 21-34.
- Siyanto, Y. 2017. Pemanfaatan Data Mining dengan Metode Clustering untuk Evaluasi Biaya Dokumen Ekspor di PT Winstar Batam. *Media Informatika Budidarma*. **1**(2): 28-31.
- Turban, E., Aronson, J., dan Liang, T.P. 2005. *Decision Support Systems and Intelligent Systems Seventh Edition*. Prentice-Hall, New Delhi.
- Walpole, R.E. dan Myers, R.H. 1995. *Ilmu Peluang dan Statistika untuk Insinyur dan Imuan*. Ed. Ke-4. ITB, Bandung.
- Xie, N., Hu, L., Luktarhan, N., dan Zhao, K. 2011. *Web Information Systems and Mining*. Springer.
- Yacoub, Y. 2012. Pengaruh Tingkat Pengangguran terhadap Tingkat Kemiskinan Kabupaten/Kota di Provinsi Kalimantan Barat. *Jurnal EKSOS*. **8**(3): 176-185.
- Yana, M.S., Setiawan, L., Ulfa, E.M., dan Rusyana, A. 2018. Penerapan Metode K-Means dalam Pengelompokan Wilayah Menurut Intensitas Kejadian Bencana Alam di Indonesia Tahun 2013-2018. *Journal of Data Analysis*. **1**(2): 93-102.
- Zarandi . M.H.F., Sotudian, S., dan Castillo, O. 2021. A New Validity Index for Fuzzy -Possibilistic C-Means Clustering. *Scientia Iranica E*. **28**(4): 2277-2293.