ANALISIS CLUSTER DENGAN ALGORITMA K-MEANS, FUZZY C-MEANS DAN HIERARCHICAL CLUSTERING (Studi Kasus: Indeks Pembangunan Manusia tahun 2019)

Rizqina Rahmati[1],**Arie Wahyu Wijayanto**[2]
[1,2]Program Studi Komputasi Statistik Peminatan Sains Data, Politeknik Statistika STIS
Email: 221810583@stis.ac.id[1], ariewahyu@stis.ac.id[2],

Abstrak

Analisis cluster adalah suatu metode data mining untuk mengelompokkan data atau objek yang didasrkan pada informasi yang ada untuk menggambarkan relasi yang terdapat antara objek tersebut. Analisis cluster bertujuan untuk membuat objek yang digabungkan dalam cluster memiliki persamaan yang tinggi dan berbeda antar cluster. Pembangunan IPM di setiap Kabupaten/Kota sangat tidak merata. Pengelompokan IPM ini dilakukan untuk mengetahui variable IPM yang harus di prioritaskan dalam pembangunan. Dalam penelitian ini digunakan tiga metode pengelompokan yaitu pengelompokkan dengan metode K-Means, Fuzzy C-Means dan Hierarchical clustering. Penentuan jumlah cluster yang optimal dan metode pengelompokan terbaik dengan membandingkan Indeks Silhouette, Davis Bouldin dan Calinski Harabasz dari ketiga metode pengelompokkan. Metode yang memiliki indeks optimal akan dipilih sebagai metode terbaik. Hasil yang didapat untuk pengelompokan data IPM Kabupaten/Kota tahun 2019 adalah terdapat 2 jumlah cluster optimal untuk metode K-Means dan Hierarchical dan 3 jumlah cluster untuk metode Fuzzy C-Means. Dengan membandingkan nilai validasi antar ketiga metode, didapat bahwa metode K-Means adalah metoode terbaik untuk pengelompokan data IPM Kabupaten/Kota tahun 2019.
Kata Kunci: *IPM, K-Means, Fuzzy C-Means, Hierarchical, Silhouette, Davies Bouldin, Calinski Harabasz*

Abstract

Cluster analysis is data mining method that for identify a group of object based on existing information to describe thee relationship between object. Cluster analysis aims to make objects that are combined in clusters have high and different similarities between clusters.The HDi development in each regecy in Indonesia is very uneven. This HDI grouping is carried out to determine the HDI variables that must be prioritized in development. In this case, three methods of grouping were used, K-Means, Fuzzy C-Means, and Hierarchical Clustering. Determination of the optimal number of clusters by comparing the Silhouette, Davies Bouldin, and Calinski Harabasz indexs. The method that has optimal index will be chosen aas the best method. The result, there are 2 optimal clusters for K-Means and Hierarchical methods and 3 optimal cluster for the Fuzzy C-means method. It is obtain that K-Means is the best method for grouping HDI Indonesia per district data in 2019.
KeyWords : *IPM, K-Means, Fuzzy C-Means, Hierarchical, Silhouette, Davies Bouldin, Calinski Harabasz*

I. PENDAHULUAN

Pembangunan nasional Indonesia menempatkan rakyat sebagai titik acuan pembangunan. Indeks Pembangunan Manusia (IPM) merupakan indeks yang dipergunakan untuk hasil pencapaian dari pembangunan suatu wilayah dalam tuga dimensi dasar yaitu Umur panjang dan hidup sehat, Pengetahuan, dan Standar hidup layak [3]. Ketiga dimensi diukur dari empat komponen yakni Harapan Lama Sekolah, Rata-rata Lama Sekolah, Pengeluaran per Kapita Disesuaikan, dan Umur Harapan Hidup Saat Lahir.

Badan Pusat Statistik (BPS) kembali meliris IPM periode 2019. Hasilnya, IPM tahun 2019 naik 0.74% dibanding tahun 2018 [3]. Namun, jika dilihat dari peringkat ASEAN dan dunia, Indonesia berada diperingkat 6 ASEAN dan 111 di dunia dari 189 negara [5]. Di kawasan Asia Tenggara dalam hal IPM Indonesia masih tertinggal dari Singapura, Brunei Darussalam, Malaysia, Thailand, dan Filipina. Oleh karena itu, Indonesia harus terus meningkatkan kualitas pengembangan sumber daya manusianya agar angka IPM terus naik.

Secara rata-rata nilai IPM dari kota di provinsi memiliki angka IPM yang lebih tinggi dibandingkan nilai IPM di kabupaten di provinsi. Bahkan beberapa diantara memiliki nilai IPM yang sangat berbeda jauh. Hal ini menunjukkan bahwa upaya peningkatan kualias sumber daya manusia di Indonesia belum merata dan perlu beberapa perbaikan untuk sebagian kabupaten yang tertinggal. Oleh karena itu perlu dilakukan pengelompokkan wilayah untuk mengetahui komponen-komponen apa yang perlu dibenahi terlebih dahulu dari kabupaten yang masih tertinggal.

Pengelompokan wilayah bertujuan untuk membagi wilayah-wilayah kedalam kelompok yang memiliki persamaan karateristik yang tinggi dalam kelompok dan memiliki perbedaan antar kelompok. Salah satu analisis yang dapat diunakan dalam pengelompokan adalah analisis cluster. Terdapat dua pendekatan dalam clustering, yaitu patisioning dan hirarki. Pada penelitian ini digunakan dua metode dari patrisioning yaitu K-Means dan Fuzzy C-Means. Sedangkan metode yang digunakan untuk Hierarchical Clustering adalah *Agglomerative Hierarchical Clustering* (AGNES).

Metode Fuzzy C-Means dan K-Means sederhana dan mudah digunakan. Fuzzy C-Means dan K-Means cocok untuk data besar[6] .Metode Hierachical yang mudah dipahami dan mudah untuk diimplementasikan. Alasan tersebut adalah mengapa di penelitian ini penelti memilih ketiga metode tersebut untuk mengelompokkan IPM 293 kabupaten/kota tahun 2019.

II. METODE

A. Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang didapatkan dari situs Badan Pusat Statistik (bps.go.id). Variabel yang digunakan adalah indikator IPM kabupaten/ kota yang memiliki nilai IPM dibawah 70 tahun 2019 yang meliputi Harapan Lama Sekolah (HLS), Pengeluaran per Kapita Yang disesuaikan, Rata-rata Lama Sekolah (RLS), dan Umur Harapan Hidup saat Lahir.

B. K-Means

K-Means merupakan salah satu metode analisis cluster non hirarki untuk memartisi objek yang ada ke dalam satu atau lebih cluster atau kelompok objek berdasarkan karakteristiknya. Sehingga objek yang mempunyai karakteristik yang berbeda di kelompokkan ke dalam cluster lainnya. Tujuan pengelompokan adalah menimimalkan variasi dalam satu cluster dan memaksimalkan variasi antar cluster.

Dalam algoritma K-Means setiap data harus termasuk ke dalam cluster tertentu pada suatu tahapan proses. Pada awl algoritmanya, K-Means mengambil sebagian dari banyaknya komponen dari populasi untuk dijadikan pusat cluster awal. Pada tahap ini, pusat cluster dipilih secara acak dari sekumpulan populasi data. Berikutnya K-Means menguji masing-masing omponen di dalam populasi data dan menandai komponen tersebut ke salah satu pusat cluster yang telah didefinisikan tergantung dari jarak minimum antar komponen pada setiap pusat cluster. Posisi pusat cluster dihitung kembali hingga semua komponen data digolongkan dalam tiap-tiap cluster dan posisi pusat cluster akan dihitung kembali hingga seluruh komponen data digolongkan ke dalam tiap-tiap cluster dan terakhir akan membentuk posisi cluster baru.

K-Means pada dasarnya melakukan dua proses yakni proses pendeteksian lokasi pusat cluster dan proses pencarian anggota tiap-tiap cluster. Proses algoritmanya adalah sebagai berikut:

1) Tentukan nilai k sebagai jumlah cluster yang ingin dibentuk.
2) Bangkitkan titik pusat cluster (centroid) sebanyak k secara random.
3) Hitung jarak setiap data ke masing-masing centroid menggunakan rumus korelasi antara dua objek yaitu *Euclidean Distance*.

$$d_v = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{1}$$

4) Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan centroidnya.
5) Lakukan iterasi, kemudian tentukan posisi centroid baru.
6) Ulangi langkah 3 jika posisi centroid tidak sama.

C. Fuzzy C-Means

Logika fuzzy merupakan seuatu logika yang memiliki nilai yang samar atau kekaburan (fuzziness) antara benar atau salah. Logika Fuzzy pertama kali diperkenalkan oleh Prof. Lotfi A. Zadeh pada tahun 1965. Dalam teori logika fuzzy suatu nilai dapat bernilai benar atau salah secara bersama. Fuzzy dinyatakan dalam derajat dari suatu keanggotaan dan derajat dari kebenaran. Oleh sebab itu sesuatu dapat dikatakan sebagian benar dan sebagian salah pada waktu yang sama. Namun besarnya keberadaan dan kesalahan suatu tergantung pada bobot keanggotaan yang dimilikinya.

Logika Fuzzy memungkinkan nilai keanggotaan antara 0 dan 1, tingkat keabuan dan juga hitam dan putih, dan dalam bentuk linguistik, konsep tidak pasti seperti ”sedikit”, ”lumayan” dan ”sangat”. Nilai keanggotaan atau derajat keanggotaan atau membership function menjadi ciri utama dalam penalaran dengan logika fuzzy tersebut.

Fuzzy C-Means Clustering (FCM) yang juga dikenal sebagai Fuzzy Isodata merupakan salah satu metode clustering yang merupakan bagian dari metode Hard K-Means. Konsep dasar FCM, pertama kali adalah menentukan centroid yang akan menandai lokasi rata- rata untuk tiap-tiap cluster. Pada kondisi awal centroid ini masih belum akurat. Tiap-tiap data memiliki persentase keanggotaan untuk tiap-tiap cluster. Dengan memperbaiki pusat cluster dan nilai keanggotaan tiap-tiap data secara berulang, maka dapat dilihat bahwa pusat cluster akan mengarah ke lokasi yang tepat. Perulangan ini didasarkan pada minimasi fungsi obyektif yang menggambarkan jarak dari titik data yang diberikan ke pusat cluster yang berbobot oleh derajat keanggotaan titik data tersebut.

Algoritma yang digunakan pada Fuzzy C-Means adalah sebagai berikut [12]:

1) Tentukan jumlah cluster (k),pangkat (w), maksimum iterasi, error terkecil yang diharapkan, fungsi objektif awal $P_o = 0$, dan iterasi awal berupa $t = 1$.
2) Bangkitkan bilangan random μ_{ik}.

3) Hitung pusat cluster ke-k, V_{kj}

$$V_{kj} = \frac{\sum_{i-1}^{n} ((\mu_{ik})^{w} X_{ij})}{\sum_{i-1}^{n} (\mu_{ik})^{w}} \tag{2}$$

4) Hitung fungsi objektif pada iterasi ke-t, P_t

$$P_t = \sum_{i-1}^{m} \sum_{k=1}^{c} \left(\left[\sum_{j=1}^{m} (x_{ij} - V_{kj})^2 \right] (\mu_{ik})^{w} \right) \tag{3}$$

5) Hitung perubahan matriks partisi

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

6) Cek kondisi berhenti
- jika: $(|P_t - P_t - 1| < \xi)$ atau $(t > MaxIter)$ maka berhenti
- jika tidak: $t = t + 1$, ulangi langkah ke-4

D. Hierarchical Clustering

Terdapat dua tipe Hierarchical clustering[1].

1) Agglomerative Hierarchical clustering atau AGNES

Dengan pendekatan bottom-up, algoritmanya sebagai berikut [1].

- Metode hierarchical agglomeratif clustering, mengasumsikan setiap data yang ada sebagai cluster di awal proses. Jika jumlah data adalah n, dan jumlah cluster adalah k, maka besarnya n = k.
- Kemudian dihitung jarak antar clusternya dengan menggunakan Euclidean distance berdasarkan jarak rata-rata antar objek.
- Selanjutnya, dari hasil perhitungan jarak dipilih jarak yang paling minimal dan digabungkansehingga besarnya n = n -1.
- Ketika dua cluster digabungkan, jarak antara dua cluster yang digabungkan dengan cluster yang lain di-update.
- Penggabungan cluster akan terus dilakukan dan akan berhenti jika memenuhi kondisi jumlah k = 1.

2) Divisive Hierarchical clustering atau DIANA

Pendekatan dengan Top-Down, algoritmanya sebagai berikut [14].

- Mulai dari atas dengan semua objek dalam satu cluster
- Cluster dibagi dengan menggunakan algoritma flat clustering
- Terapkan secara berulang.

Penelitian ini menggunakan *Agglomerative Hierarchical clustering* dengan metode ward sebagai metode update jarak. Metode Ward dapat membentuk cluster berdasarkan jumlah total kuadrat deviasi tiap pengamatan dari rata-rata cluster yang menjadi anggotanya [13]. Metode Ward berusaha untuk meminimalkan variasi antar objek dalam satu cluster dan memaksimalkan variasi dengan objek yang ada di cluster lainnya[15].

E. Evaluasi pengelompokan

Tingkat keberhasilan usaha ditentukan berdasarkan penilaian kinerja ketiga metode tersebut. Evaluasi internal yang digunakan di penelitian ini ada tiga, yaitu *Silhouette, Davis Bouldin* dan *Calinski Harabasz*. Clustering menggunakan K-means menunjukkan bahwa indeks validitas *Silhouette, Davies Bouldin*, dan *Calinski Harabasz* memiliki hasil yang lebih baik dibandingkan indeks validitas *Dunn* [8]. Pengujian model dilakukan untuk mengetahui seberapa dekat objek di dalam sebuah cluster dan seberapa jauh cluster terpisah dari cluster lainnya.

1) Silhouette

Metode *Silhouette* mengukur seberaa dekat relasi antara objek dalam sebuah *cluster* seberapa jauh sebuah *cluster* terpisah dari *cluster* lainnya. Tahapan perhitungannya adalah sebagai berikut [6]:

(a) Hitung rata-rata jarak objek dengan semua objek lain yang berada di dalam satu *cluster*

$$a(i) = \frac{1}{n_k - 1} \sum_{i \epsilon I_K, i' \neg i} d(M_i, M_i') \tag{5}$$

(b) Hitung rata-rata jarak objek dengan semua objek lain yang berada pada cluster lain, lalu ambil nilai paling minimum

$$b(i) = (\frac{1}{n_k'} \sum_{i \epsilon I_k} d(M_i, M_i')) \tag{6}$$

(c) Hitung nilai silhouette dengan persamaan

$$s(i) = \frac{b(a) - a(i)}{max(a(i), b(i))} \tag{7}$$

Nilai indeks *silhouette* berkisar antara -1 sampai 1. Jika nilai silhouette ¡0 maka objek berada dalam kelompok yang salah. Sedangkan jika nilai silhouette ¿0 maka objek berada dalam kelompok yang benar dan jika nilai silhouette = 0 maka objek berada diantara dua cluster sehinnga belum dapat ditentukan masuk ke dalam kelompok yang benar atau salah [11].

2) Davies Bouldin

Indeks *Davies Bouldin* menghitung rata-rata nilai setiap titik pada himpunan data. Perhitungan nilai setiap titik adalah jumlah nilai *compactness* yang dibagi dengan jarak antara kedua titik pusat *cluster* sebagai *separation*. Jumlah klaster terbaik ditunjukkan sengan nilai indeks *Davies Bouldin* yang semakin kecil [9].

3) Calinski Harabasz

Perhitungan indeks Calinski Harabasz adalah sebagai berikut [6]:

$$C = \frac{\frac{BGSS}{K-1}}{\frac{WGSS}{K-1}} = \frac{N-K}{K-1}\frac{BGSS}{WGSS} \tag{8}$$

Dengan,

$$BGSS = \sum_{k=1}^{k} n_k \sum_{j=0}^{p} (\mu_j^k - \mu_j)^2 = \sum_{k=1}^{k} n_k \left\| G^k - G \right\|^2$$

$$WGSS = \sum_{k=1}^{k} n_k \sum_{i \epsilon I_k} \left\| M_i^k - G^k \right\|^2$$

III. HASIL

A. Penentuan Jumlah Cluster Optimum

Jumlah *cluster optimum* diantara ketiga metode yaitu *K-Means, Fuzzy C-Means* dan *AGNES* dengan membandingkan indeks validitas dari tiap metode.

B. K-Means

Pemilihan jumlah *cluster* yang optimum menggunakan nilai indeks *Silhouette*, *Davies Bouldin*, dan *Calinski Harabasz*. Berkut adalah grafik dari ketiga indeks tersebut untuk tiap jumlah *cluster* untuk data IPM yang digunakan.

Tabel I: Indeks Validasi Metode K-Means Tiap Jumlah Cluster

	Kmeans		
Cluster	**Silhouette**	**Davies Bouldin**	**Calinski Hrabasz**
2	0.53	0.67	163
3	0.341	1.169	153.95
4	0.3005	1.155	143.15
5	0.2858	1.147	138.67
6	0.2647	1.176	129.49
7	0.2498	1.152	121.71

Dari ketiga metode validasi yang ditunjukkan pada Tabel I, dapat disimpulkan bahwa jumlah cluster optimum untuk metode K-Means adalah 2.

1) Fuzzy C-Means

Tabel II: Indeks Validasi Metode Fuzzy C-Means Tiap Jumlah Cluster

	Fuzzy C-Means		
Cluster	**Silhouette**	**Davies Bouldin**	**Calinski Hrabasz**
2	0.21341	1.6415	78.516
3	0.3071	1.2207	152.9
4	0.29565	1.166	142.29
5	0.24216	1.3269	126.06
6	0.23974	1.224	123.87
7	0.20281	1.376	111.19

Dari Tabel II, terlihat bahwa jumlah cluster dengan indeks optimum adalah 3 dan 4. Karena dari penelitian sebelumnya didapat bahwa metode validasi calinski Hrabasz adalah metode yang memiliki peforma yang paling baik [4], maka cluster optimum dapat dilihat dari nilai Calinski Hrabasz yaitu 3.

2) AGNES

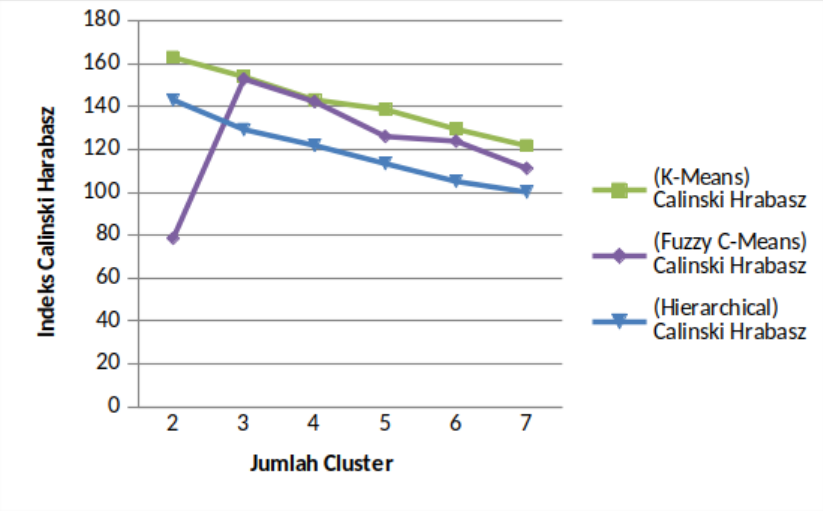
Tabel III: Indeks Validasi Metode AGNES Tiap Jumlah Cluster

	Hierarchical		
Cluster	**Silhouette**	**Davies Bouldin**	**Calinski Hrabasz**
2	0.6571	0.4913	143.12
3	0.3516	1.32213	129.305
4	0.30536	1.21898	121.897
5	0.26653	1.3055	113.414
6	0.24738	1.30986	105.091
7	0.22751	1.32906	100.121

Dari Tabel III, terlihat bahwa jumlah cluster optimum untuk metode AGNES adalah 2.

C. Metode Terbaik

Pemilihan metode terbaik antara K-Means, Fuzzy C-Means dan AGNES untuk pengelompokan IPM kabupaten/kota tahun 2019 dapat dilakukan dengan melihat perbandingan antara ketiga indeks validasi. Tetapi, karena pada data IPM kabupaten/kota tahun 2019 memiliki keambiguan maka pemilihan metode terbaik dilakukan hanya dengan membandingkan nilai indeks Calinski Harabasz. Dipilihnya indeks Calinski Harabasz diantara metode lainnya, karena metode Calinski Harabasz memiliki peforma yang paling baik diantara yang lain [4].



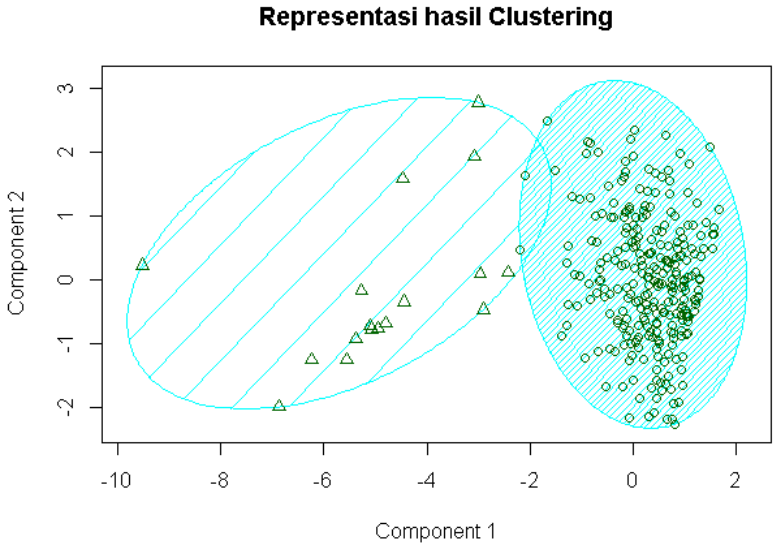
Gambar 1: Grafik indeks Calinski Harabasz untuk Ketiga Metode Clustering

Indeks Calinski Harabast terhadap jumlah cluster optimum metode K-Means memiliki nilai yang lebih tinggi daripada metode lainnya di jumlah cluster yang sama. Sehingga metode K-Means dengan jumlah cluster optimum 2 akan digunakan untuk mengelompokkan data.

D. Hasil Clustering

Tabel IV: Hasil Pengelompokan Kabupaten/Kota Menggunakan Metode K-Means

Kelompok	Anggota Kelompok
1	Simeulue, Aceh Singkil, Aceh Selatan, Aceh Tenggara, Aceh Timur, Aceh Utara, Aceh Barat Daya, Gayo Lues ,Aceh Tamiang, Nagan Raya, Aceh Jaya,Kota Subulussalam, Nias, Mandailing Natal, Tapanuli Selatan, Tapanuli Tengah, Asahan, Nias Selatan, Humbang Hasundutan, Pakpak Bharat, Batu Bara, Padang Lawas Utara, Padang Lawas, Nias Utara, Nias Barat, Kota Tanjung Balai, Kota Gunungsitoli, Kepulauan Mentawai, Solok, Sijunjung, Lima Puluh Kota, Pasaman, Solok Selatan, Pasaman Barat, Indragiri Hilir, Rokan Hulu, Rokan Hilir, Kepulauan Meranti, Merangin, Sarolangun, Batang Hari, Muaro Jambi, Tanjung Jabung Timur, Tanjung Jabung Barat, Tebo, Bungo, Ogan Komering Ulu, Ogan Komering Ilir, Muara Enim,Lahat, Musi Rawas, Musi Banyuasin, Banyu Asin, Ogan Komering Ulu Selatan, Ogan Komering Ulu Timur, Ogan Ilir, Empat Lawang, Penukal Abab Lematang Ilir, Musi Rawas Utara, Kota Pagar Alam, Bengkulu Utara, Kaur,Seluma, Mukomuko,Lebong, Kepahiang, Bengkulu Tengah, Lampung Barat, Tanggamus, Lampung Selatan, Lampung Timur, Lampung Utara, Way Kanan, Tulangbawang, Pesawaran, Pringsewu, Mesuji, Tulang Bawang Barat, Pesisir Barat,Bangka Barat, Bangka Selatan, Lingga, Kepulauan Anambas, Sukabumi, Cianjur, Garut, Tasikmalaya, Kuningan, Cirebon, Majalengka, Indramayu, Subang, Bandung Barat, Pangandaran, Cilacap, Purbalingga, Banjarnegara, Kebumen, Wonosobo, Magelang, Wonogiri, Grobogan, Blora, Temanggung, Batang, Pekalongan, Pemalang, Tegal, Brebes, Gunung Kidul, Pacitan, Trenggalek, Lumajang, Jember, Bondowoso, Situbondo, Probolinggo, Pasuruan, Bojonegoro, Tuban, Bangkalan, Sampang, Pamekasan, Sumenep, Pandeglang, Lebak, Serang, Bangli, Karangasem, Lombok Barat, Lombok Tengah, Lombok Timur, Sumbawa, Dompu, Bima, Lombok Utara, Sumba Barat, Sumba Timur, Kupang, Timor Tengah Selatan, Timor Tengah Utara, Belu, Alor ,Lembata, Flores Timur, Sikka, Ende, Ngada, Manggarai, Rote Ndao, Manggarai Barat, Sumba Tengah, Sumba Barat Daya, Nagekeo, Manggarai Timur, Sabu Raijua, Malaka, Sambas, Bengkayang, Landak, Mempawah, Sanggau, Ketapang, Sintang, Kapuas Hulu, Sekadau, Melawi, Kayong Utara, Kubu Raya, Kapuas, Sukamara, Seruyan, Katingan, Pulang Pisau, Murung Raya, Tanah Laut, Kota Baru, Banjar, Barito Kuala, Hulu Sungai Selatan, Hulu Sungai Tengah, Hulu Sungai Utara, Balangan, Mahakam Ulu, Tana Tidung, Nunukan, Bolaang Mongondow, Kepulauan Talaud, Bolaang Mongondow Utara, Siau Tagulandang Biaro, Bolaang Mongondow Selatan, Bolaang Mongondow Timur, Banggai Kepulauan, Donggala, Toli-Toli, Buol, Parigi Moutong, Tojo Una-Una, Sigi, Banggai Laut, Morowali Utara, Kepulauan Selayar,Bulukumba, Bantaeng, Jeneponto, Takalar, Gowa, Sinjai, Maros, Pangkajene dan Kepulauan, Bone, Soppeng, Wajo,Tana Toraja, Luwu Utara, Toraja Utara, Buton, Muna, Konawe Selatan, Bombana, Wakatobi, Kolaka Utara, Buton Utara, Konawe Utara, Kolaka Timur, Konawe Kepulauan, Muna Barat, Buton Tengah, Buton Selatan, Boalemo, Gorontalo, Pohuwato, Bone Bolango, Gorontalo Utara, Majene, Polewali Mandar, Mamasa, Mamuju, Mamuju Utara, Mamuju Tengah, Maluku Tenggara Barat, Maluku Tenggara, Buru, Kepulauan Aru, Seram Bagian Barat, Seram Bagian Timur, Maluku Barat Daya, Buru Selatan, Kota Tual, Halmahera Barat, Halmahera Tengah, Kepulauan Sula, Halmahera Selatan, Halmahera Utara, Halmahera Timur, Pulau Morotai, Pulau Taliabu, Fakfak, Kaimana, Teluk Wondama, Teluk Bintuni, Sorong Selatan, Sorong, Maybrat, Manokwari Selatan, Merauke, Jayawijaya, Nabire, Kepulauan Yapen , Boven Digoel, Mappi, Sarmi, Keerom, Waropen, Supiori, Mamberamo Raya, Nduga, Lanny Jaya.
2	Tambrauw, , Pegunungan Arfak, Paniai, Puncak Jaya, Asmat, Yahukimo, Pegunungan Bintang, Tolikara, Mamberamo Raya, Nduga, Lanny Jaya, Mamberamo Tengah, Yalimo, Puncak, Dogiyai, Intan Jaya, Deiyai.



Gambar 2: Representasi Hasil Clustering IPM 293 Kabupaten/Kota Tahun 2019

IV. PEMBAHASAN

Indeks Calinski Harbasz metode Fuzzy C-Means dengan cluster optimum sama dengan 3 yaitu 152,9 , sedangkan metode K-means dengan jumlah cluster sama dengan 3 memiliki nilai Indeks Calinski Harabasz yang lebih tinggi yaitu 153,95. Jika dibandingkan dengan metode AGNES yang memiliki jumlah cluster optimum yang sama dengan K-Means yaitu 2, nilai indeks Calinski Harabasz untuk metode AGNES masih lebih rendah dibanding metode K-means. Dengan Indeks untuk metode AGNES sebesar 143.12 dan K-Means sebesar 163.

Dari cluster yang sudah terbentuk, akan dilihat karakteristik tiap kelompoknya dengan membandingkan antara rata-rata masing masing kelompok dengan rata-rata data keseluruhan. Kelompok yang memiliki rata-rata yang kurang dari rata-rata data keseluruhan untuk tiap variabelnya akan diberi kode minus (-) dan kelompok yang memiliki rata-rata yang lebih dari rata-rata keseluruhan untuk tiap variabelnya akan diberi kode plus (+).

Karakteristik tiap kelompok akan ditampilkan di Tabel V berikut:

Tabel V: Karakteristik Kelompok Berdasarkan Rata-rata

Cluster	HLS	Pengeluaran	RLS	HH
1	+	+	+	+
2	-	-	-	-

Terlihat bahwa karakteristik tiap variable untuk kelompok 1 memiliki rrata-rata yang lebih kecil dari rata-rata keseluruhan yang artinya pembangunan IPM di kabupaten/kota anggota kelompok 1 masih sangat kecil. Sedangkan kelompok 2 memiliki karakteristik yang lebih baik, yakni tiap variablenya memiliki rata-rata yang lebih besar dari rata-rata keseluruhan.

V. KESIMPULAN DAN SARAN

A. *Kesimpulan*

Berdasarkan dari hasil dan pembahasan, dapat diambil kesimpulan bahwa:

1) Jumlah cluster optimal antara metode K-Means dan AGNES dengan Fuzzy C-Means untuk data IPM 293 kabupaten/kota tahun 2019 berbeda. Pengelompokan dengan menggunakan metode K-Means dan AGNES menghasilkan cluster optimal 2, sedangkan dengan metode Fuzzy C-Means menghasilkan cluster optimal 3.
2) Dapat disimpulkan bahwa K-Means memiliki nilai indeks Calinski Harabasz yang lebih tinggi dibanding metode lainnya untuk data IPM 293 kabupaten/kota tahun 2019.
3) Dari hasil pengelompokan, terlihat perbedaan yang jauh antara variable pembangun IPM di kelompok 1 dengan variable pembangun IPM di kelompok 2. Kelompok 1 memiliki variable Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah(RLS), Pengeluaran per Kapita Disesuaikan, dan Umur Harapan Hidup (UHH) yang tinggi, sedangkan kelompok 2 memiliki. variable Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah(RLS), Pengeluaran per Kapita Disesuaikan, dan Umur Harapan Hidup (UHH) yang rendah. Hal ini bersesuaian dengan penelitian sebelumnya yang dilakukan Santoso(2019) [18] untuk pengelompokan IPM tahun 2017.
4) Kabupaten/kota anggota kelompok 1 memerlukan perhatian lebih dari pemerintah karena semua variable pembangun dari IPM di kelompok 1 masih kurang dari rata-rata. Bahkan sangat terbanting apabila IPM ibukota 80-an, sedangkan kabupatennya masih ada yang memiliki IPM yang masih di angka 30.

B. *Saran*

1) Hasil clustering dari metode K-Means kurang memuaskan bagi peneliti karna jumlah cluster optimalnya hanya 2. Sangat kurang karena saya ingin medaftarkan kekurangan dari masing-masing hasil pengelompokan agar dapat menjadi saran yang terperinci bagi pemerintah. Saran saya untuk penelitian selanjutnya agar memakai metode lain atau metode perkembangan dari K-Means agar hasil yang didapat lebih optimal dan memuaskan.
2) Harapan kedepannya pemerintah bisa menyetarakan pembangunan IPM tiap kabupaten/kota dengan memfokuskan terlebih dahulu pada daerah daerah yang memiliki angka IPM yang sangat kecil. Analisis cluster kali ini bertujuan untuk memberi tahu pada pemerintah variable apa yang harus difokuskan terlebih dahulu pada saat membangun IPM agar tepat sasaran.

PUSTAKA

[1] Aastha Joshi, Rajneet Kaur, " Comparative Study of Various Clustering Techniques in Data Mining", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 3, March 2013 ISSN: 2277 128X
[2] Alfina, T., Santosa.B., Barakbah.A.R., 2012. Analisa Perbandingan Metode Hierarchical Clustering, K-Means dan Gabungan Keduanya dalam Cluster Data (Studi kasus: Problem Kerja Praktek Jurusan Teknik Industri ITS)
[3] Badan Pusat Statistik. 17 Februari 2020. Indeks Pembangunan Manusia 2019 No. 21/02/Th. XXIII. Berita Resmi Statistik.
[4] Calinski T, Harabasz J. 1974. A Dendrite Method for Cluster Analysis. Communications in Statistics. 3(1):1-27
[5] Citradi, Tirta. 17 Februari 2020. "IPM RI Naik, Tapi Masih Kalah Sama Tetangga". CNBC INDONESIA. https://www.cnbcindonesia.com/news/20200217142358-4-138395/ipm-ri-naik-tapi-masih-kalah-sama-tetangga
[6] Desgraupes, Bernard. November 2017. Clustering Indices (Package Clustering for R)

[7] Gelley N,Roger J. 2000. Fuzzy Logic Toolbox. USA: Mathwork Inc.
[8] Khairati, A.F., Adlina, A.A , Hertono,G.F, & Handari,B.D. (2019). Kajian Indeks Validitas pada Algoritma K-means Enhanced dan K-means MMCA. PRISMA, Prosiding Seminar Nasional Matematika 2, 161-170
[9] Kingrani, Suneel Kumar and Levene, Mark and Zhang, Dell (2018). Estimating the number of clusters using diversity. Artificial Intelligence Research 7 (1). Pp. 15-22. ISSN 1927-6974.
[10] Kusumadewi S, Purnomo H. 2004. Aplikasi Logika Fuzzy untuk pendukung keputusan, Yogyakarta: Graha Ilmu.
[11] Lewis, Paul D. 2010. R for Madacine and Biology. Massachusets: Jones and Bartlett Publisher.
[12] Malik, R.A., Defit, S., Yuhandari. Januari 2018. PERBANDINGAN ALGORITMA K-MEANS CLUSTERING DENGAN FUZZY C-MEANS DALAM MENGUKUR TINGKAT KEPUASAN TERHADAP TELEVISI DAKWAH SURAU TV. RABIT (Jurnal Teknologi dan Sistem Informasi Univrab). Vol 3, No.1 : 10-21.
[13] Oktavia, S., Mara, M. N., Satyahadewi, N. 2013. Pengelompokan Kinerja Dosen Jurusan Matematika FMIPA UNTAN Berdasarkan Penilaian Mahasiswa Menggunakan Metode Ward. Buletin Ilmiah Mat. Stat. dan Terapannya (Bimaster) Volume 02, No. 2 (2013), hal 93 – 100. Tanjungpura
[14] Patel, K. M. A., & Thakral, P. (2016). The best clustering algorithms in data mining. 2016 International Conference on Communication and Signal Processing (ICCSP).
[15] Rahmawati, linda, dkk. 2014. ANALISA CLUSTERING MENGGUNAKAN METODE K-MEANS DAN HIERARCHICAL CLUSTERING (STUDI KASUS : DOKUMEN SKRIPSI JURUSAN KIMIA, FMIPA, UNIVERSITAS SEBELAS MARET). ITSMART: Jurnal Teknologi dan Informasi, Vol 3, N0.2.
[16] Risqiyani TA, Kesumawati A. 2016. Pengelompokan Kabupaten Kota di Provinsi Jawa Tengah dengan Fuzzy C-Means Clustering (Studi Kasus : Jumlah Kasus Gizi Buruk, Faktor Sarana dan Tenaga Kesehatan serta Faktor Kependudukan di Jawa Tengah Tahun 2014). SEMINAR NASIONAL MATEMATIKA DAN PENDIDIKAN MATEMATIKA UNY 2016.
[17] Rizal AS, Hakim RBF. 2012. METODE K-MEANS CLUSTER DAN FUZZY C-MEANS CLUSTER (Studi Kasus: Indeks Pembangunan Manusia di Kawasan Indonesia Timur tahun 2012). Prosiding Seminar Nasional Matematika dan Pendidikan Matematika UMS 2015.
[18] Santoso, M. 2019. ANALISIS K-MEANS CLUSTER INDEKS PEMBANGUNAN MANUSIA DI INDONESIA. Tugas Akhir, Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang. Pembimbing Utama Dr. Walid, S.Pd.,M.Si.
[19] Santosa, Budi. 2007. Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis. Yogyakarya: Graha Ilmu.
[20] Singh, Amanpreet, dkk.18 December 2017. How to Perform Hierarchical Clustering using R. R-Bloggers. https://www.r-bloggers.com/2017/12/how-to-perform-hierarchical-clustering-using-r/