

KLASTERING KOTA DAN KABUPATEN DI INDONESIA BERDASARKAN UMUR HARAPAN HIDUP SAAT LAHIR DENGAN K-MEDOIDS

Astri Charolina¹⁾, Dyah Ruswanti²⁾

^{1), 2)} Program Studi Informatika, Fakultas Sains, Teknologi dan Kesehatan, Universitas Sahid Surakarta

Jl. Adi Sucipto 154, Jajar, Surakarta, 57144, Telp. (0271) 743493, 743494

Email: charolina@gmail.com, dyahruswanti@usahidsolo.ac.id

Abstract

One of the uses of cluster analysis is to predict the state of objects, and in this study, 514 provinces in Indonesia are grouped into four clusters based on Life Expectancy at birth. The historical data used is sourced from BPS for a period of 11 years, from 2010 to 2019. The purpose of this provincial grouping is to provide input to local governments and policy makers regarding provincial clusters in their area. With the hope of making improvements or increasing efforts to increase life expectancy at birth and reduce mortality at birth. The algorithm used is K-Medoids which can perform clusters with the advantage of being able to overcome noise and outlier in large data. The results obtained are Cluster 1 as many as 125 provinces, Cluster 2 as many as 119 provinces, Cluster 3 as many as 137 provinces and Cluster 4 as many as 133 provinces. Life expectancy is one of the components in calculating the Human Development Index.

Keywords: cluster, K Medoids, life expectancy, HDI

Pendahuluan

Latar Belakang

Mengutip isi *Human Development Report* (HDR) pertama tahun 1990, pembangunan manusia adalah suatu proses untuk memperbanyak pilihan-pilihan yang dimiliki oleh manusia. Diantara banyak pilihan tersebut, pilihan yang terpenting adalah untuk berumur panjang dan sehat, untuk berilmu pengetahuan, dan untuk mempunyai akses terhadap sumber daya yang dibutuhkan agar dapat hidup secara layak.

Indeks Pembangunan Manusia (IPM) mengukur capaian pembangunan manusia berbasis sejumlah komponen dasar kualitas hidup. Sebagai ukuran kualitas hidup, IPM dibangun melalui pendekatan tiga dimensi dasar. Dimensi tersebut mencakup umur panjang dan sehat; pengetahuan, dan kehidupan yang layak. Ketiga dimensi tersebut memiliki pengertian sangat luas karena terkait banyak faktor. Untuk mengukur dimensi kesehatan, digunakan angka harapan hidup waktu lahir. Selanjutnya untuk mengukur dimensi pengetahuan digunakan gabungan indikator angka melek huruf dan rata-rata lama sekolah. Adapun untuk mengukur dimensi hidup layak digunakan indikator kemampuan daya beli masyarakat terhadap sejumlah kebutuhan pokok yang dilihat dari rata-rata besarnya pengeluaran per kapita sebagai pendekatan pendapatan yang mewakili capaian pembangunan untuk hidup layak.

Selama periode 2010 hingga 2018, Indonesia telah berhasil meningkatkan Umur Harapan Hidup saat lahir sebesar 1,39 tahun atau tumbuh sebesar 0,25 persen per tahun. Pada tahun 2010, Umur Harapan Hidup saat lahir di Indonesia hanya sebesar 69,81 tahun, dan pada tahun 2018 telah mencapai 71,20 tahun. Dalam penyusunan IPM, ada 4 komponen yaitu:

- a. Angka Harapan Hidup. Merupakan rata-rata perkiraan banyak tahun yang dapat ditempuh oleh selama hidup.
- b. Angka Melek Huruf. Angka melek huruf adalah persentase penduduk usia 15 tahun keatas yang dapat membaca dan menulis huruf latin dan atau huruf lainnya.
- c. Rata-Rata Lama Sekolah. Rata-rata lama sekolah menggambarkan jumlah tahun yang digunakan oleh penduduk usia 15 tahun keatas dalam menjalani pendidikan formal.
- d. Pengeluaran, merupakan ril per Kapita yang disesuaikan UNDP untuk mengukur standar hidup layak menggunakan Produk Domestik Bruto (PDB) riil yang disesuaikan, sedangkan BPS dalam standar hidup layak menggunakan rata-rata pengeluaran per kapita riil yang disesuaikan dengan formula Atkinson.

Salah satu faktor penting penghitung IPM adalah angka umur harapan hidup. Semakin tinggi angka harapan hidup akan menyumbang kenaikan angka IPM. Untuk itu sebagai salah satu cara memprediksi angka harapan hidup adalah dengan melakukan klastering semua Kota dan Kabupaten di Indonesia berdasarkan angka harapan hidup ini.

K Medoids merupakan salah satu algoritma untuk melakukan klaster yang sangat baik dalam menangani noise dan juga outlier. Biasanya pada data yang besar akan muncul ketimpangan dalam distribusi daya yang menyebabkan muncul noise maupun outlier.

Berdasarkan latar belakang masalah, maka dirumuskan masalah yaitu Bagaimana mengelompokkan Kota dan Kabupaten di Indosensia berdasarkan angka umur harapan hidup menggunakan algortima K medoids. Batasan masalah dari penelitian ini adalah: menggunakan Algoritma K Medoids dan Data yang digunakan merupakan data tahunan pada setiap Kota dan Kabupaten dalam kurun waktu 11 tahun (tahun 2010-2019).

Permasalahan

Berdasarkan latar belakang diatas maka dapat disimpulkan adanya permasalahan yaitu bagaimanakah proses klastering menggunakan K Medoids diterapkan sehingga dihasilkan klaster atau kelas-kelas tertentu.

Tujuan Penulisan

Tujuan dari penelitian ini adalah untuk membangun sebuah sistem mengklaster kumpulan Kota/Kabupaten di Indonesia dan dapat mengklasifikasikan pada setiap topik atau Klaster.

Landasan Teori

Tinjauan Pustaka

Mengutip pernetitian terdahulu yang dilakukan oleh Arora, Varshney dan Deepali, tahun 2016 melakukan penelitian penggunaan klastering pada big data. Algoritma yang digunakan adalah K-Means dan K-Medoids yang dievaluasi pada set data transaksi 10k dari data umum di situs KEEL. Masukan ke algoritma ini adalah titik data terdistribusi secara acak dan berdasarkan pada kelompok kesamaan mereka telah

dihadarkan. Hasil perbandingan menunjukkan bahwa waktu yang diambil dalam pemilihan kepala klaster dan kompleksitas ruang tumpang tindih klaster jauh lebih baik di K-Medoids daripada K-Means. Juga K-Medoids lebih baik dalam hal waktu eksekusi, tidak sensitif terhadap outlier dan mengurangi noise dibandingkan dengan K-Means karena meminimalkan jumlah perbedaan objek data (BPS,2014).

Penelitian yang dilakukan oleh Wira, Budianto dan Wiguna (2018) dengan judul Implementasi Metode K-Medoids Clustering Untuk Mengetahui Pola Pemilihan Program Studi Mahasiswa Baru Tahun 2018 Di Universitas Kanjuruhan Malang. Dimana Setiap tahun Universitas Kanjuruhan Malang menerima hampir 2.000 mahasiswa yang tersebar diberbagai program studi. Oleh karenanya data yang telah ditampung pastinya banyak sekali, dari data tersebut dapat dilihat pola - pola pemilihan program studi berdasarkan nilai tes, asal sekolah, dan program studi. Penelitian ini menggunakan metode K-Medoids agar dapat diketahui pola pemilihan program studi bagi mahasiswa baru. K-Medoids merupakan metode partisional clustering dimana bertujuan untuk menemukan satu set k-cluster di antara data yang paling mencirikan objek dalam kumpulan suatu data. Hasil penelitian pengelompokan mahasiswa baru menunjukkan bahwa mahasiswa baru yang berasal dari SMA/SMK dengan nilai ujian diatas 70 mengambil jurusan TI, sedangkan mahasiswa baru yang berasal dari SMK dengan nilai ujian dibawah 70 dan SMA dengan nilai ujian dibawah 50 mengambil jurusan SI, dan sisanya mengambil jurusan NON TI/SI. Kualitas cluster yang dihasilkan berdasarkan proses pengujian yang dilakukan didapatkan hasil nilai Silhouette Coefficient terbaik yaitu 0.690754 dengan jumlah cluster sebanyak tiga dan jumlah data sebanyak 15

Klastering

Klastering adalah suatu metode pengelompokan berdasarkan ukuran kedekatan (kemiripan). Klastering adalah proses mengelompokkan objek berdasarkan informasi yang diperoleh dari data yang menjelaskan hubungan dari setiap obyek, yang memaksimalkan kesamaan antar anggota, dan meminimalkan kemiripan dalam klaster yang berbeda. Tujuannya menemukan klaster yang berkualitas dalam waktu yang layak. Klastering dalam data mining berguna untuk menemukan pola distribusi di dalam sebuah data set yang berguna untuk proses analisa data (J.Han, M.Kamber and J.Pei, 2012).

Klastering beda dengan group, dimana group berarti kelompok yang sama, kondisinya kalau tidak "Ya" dapat dipastikan bukan sebagai anggota kelompoknya. Namun, pada klaster, anggota tidak harus sama tetapi pengelompokannya berdasarkan pada kedekatan dari suatu karakteristik sampel yang ada, salah satunya dengan menggunakan rumus jarak ecludian. Aplikasi klaster ini sangat banyak, karena dalam mengidentifikasi permasalahan atau pengambilan keputusan selalu tidak sama persis akan tetapi cenderung memiliki kemiripan saja. Prinsip dasar dari Klastering adalah Similarity Measures (ukuran kedekatan) dan Distances dan Similarity Coefficients untuk beberapa sepasang dari item Ecluidean Distance (BPS,2014).

Beberapa manfaat klastering untuk Identifikasi Obyek (Recognition) misalnya dalam bidang image processing, Computer Vision atau robot vision. Untuk Decission Support System, misalnya untuk Segementasi Pasar, pemetaan wilayah, Manajemen marketing. Terdapat berbagai macam metode untuk melakukan klastering, sebagai berikut :

- a. Berbasis Metode Statistik
 - 1) *Hierarchical clustering method* : pada kasus untuk jumlah kelompok belum ditentukan terlebih dulu, contoh data-data hasil survei kuisioner. Macam-metode jenis ini: Single Linkage, Complete Linkage, Average Linkage dll.
 - 2) *Non Hierarchical clustering method*: Jumlah kelompok telah ditentukan terlebih dulu. Metode yang digunakan : K-Means.
- b. Berbasis Fuzzy : Fuzzy C-Means
- c. Berbasis Neural Network : Kohonen SOM, LVQ
- d. Metode lain untuk optimasi centroid atau lebar klaster : Genetik Algoritma (GA)

K-Medoids

Metode K-medoid setiap klaster dipresentasikan dari sebuah objek di dalam *klaster* yang disebut dengan medoid. Tujuannya adalah menemukan kelompok K-klaster (jumlah *klaster*) diantara semua objek data di dalam sebuah kelompok data. Klasternya dibangun dari hasil mencocokkan setiap objek data yang paling dekat dengan *klaster* yang dianggap sebagai medoid sementara. Langkah-langkah menghitung medoids, yaitu (BPS,2014).

1. Pilih point k sebagai inisial *centroid*/nilai tengah (medoids) sebanyak k *klaster*.
2. Cari semua point yang paling dekat dengan medoid, dengan cara menghitung jarak vektor antar dokumen. (menggunakan Euclidian distance)
3. Secara random, pilih point yang bukan medoid.
4. Hitung total distance
5. If TD baru < TD awal, tukar posisi medoid dengan medoids baru, jadilah medoid yang baru.
6. ulangi langkah 2 - 5 sampai medoid tidak berubah.

Diberikan himpunan data sejumlah n titik $P = \{p_1, p_2, \dots, p_n\}$ dalam ruang R^2 , tujuan klustering K-medoid adalah untuk membagi titik-titik data ke dalam k klaster sehingga fungsi objektif berikut minimal.

$$J(M) = J(m_1, \dots, m_k) = \sum_{i=1}^k \sum_{p \in C_i} d(p, m_i)$$

Pada $J(C)$, m_i merupakan titik perwakilan (representative point) bagi klaster C_i .

1. Jika m_i dibatasi harus merupakan anggota dari P , maka m_i disebut medoid.
2. Sebaliknya, jika m_i adalah rataan dari titik-titik klaster dan tidak harus merupakan anggota dari P , maka m_i disebut mean.

Sehingga algoritma K-mean dan K-medoid mempunyai hubungan yang sangat erat. K-medoid memberikan karakteristik *K klaster*, dan setiap titik dalam P menjadi milik medoid terdekat. Karena kita telah membatasi dalam ruang R^2 , maka fungsi jarak d biasanya adalah jarak Euclidean.

$$d(p, m_i) = \sqrt{(p(x) - m_i(x))^2 + (p(y) - m_i(y))^2}$$

Salah satu kekurangan pendekatan K-medoid adalah algoritma tersebut menghasilkan “*hard*” *klaster*, yaitu masing-masing titik secara unik diberikan satu dan hanya satu *klaster* (BPS,2014).

Metode Penelitian

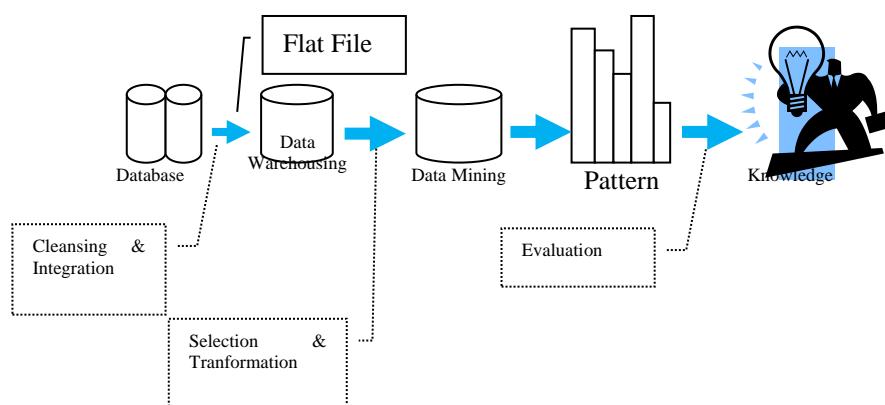
Penelitian ini menggunakan data yang didapatkan dari Badan Pusat Statistik dan data pada tiap wilayah Kota dan Kabupaten. Metode yang digunakan pada penelitian ini adalah metode deskriptif analisis yaitu dengan cara mengumpulkan data, menganalisa data, membuat suatu pemecahan masalah, dan kemudian disusun untuk menarik kesimpulan mengenai masalah tersebut. Ada dua teknik dalam metedologi penelitian yaitu teknik pengumpulan data dan teknik pengembangan data. Proses dimulai dari pengumpulan data, kemudian melakukan cleansing dan melakukan cek pada data kosong, baru dapat dilakukan proses klaster.

Metode Pengumpulan Data

Teknik pengumpulan data dapat diperoleh dari objek penelitian. Cara yang mendukung untuk memperoleh datanya adalah Studi Literatur, merupakan pengumpulan data melalui buku-buku, dokumen, jurnal yang erat kaitannya dengan topik yang dibahas dalam penelitian ini.

Metode Pengembangan Sistem

Proses mencari pola data yang tersembunyi dalam data mining, digunakan algoritma data mining tertentu sehingga dapat menemukan informasi yang tersembunyi dalam data salah satunya dengan klaster. Model pencarian data dengan langkah-langkah dalam proses menemukan pengetahuan seperti yang ada di gambar.1



Gambar 1. Tahap Penemuan Knowledge

Tahap-tahap tersebut, bersifat interaktif dimana pemakai terlibat langsung atau dengan perantaraan knowledge base.

1. Cleansing

Tahap yang pertama dalam proses data mining adalah cleansing atau pembersihan data. Yang termasuk dalam tahap ini adalah penghilangan noise, inkonsistensi data, unrelevant data. Tahapan ini penting karena data mentah yang didapatkan dari sumber data pada umumnya merupakan data noise, ada data yang hilang, data yang kurang atau tidak sempurna, data yang tidak relevan, maupun data yang tidak konsisten. Ketika data yang akan diolah dalam data mining sudah bersih, maka dapat meningkatkan performance dari proses data mining yang dilakukan.

2. Data Integration

Penggabungan data diperlukan saat sumber data berasal dari database maupun file yang berbeda. Penggabungan ini dilakukan dengan cara mengidentifikasi unique entity pada atribut-atribut yang ada dalam database.

3. Data Selection

Tidak semua data yang didapatkan dari sumber data akan dipakai dalam proses data mining, meskipun data tersebut mempunyai korelasi dengan data pada atribut-atribut lainnya dalam database. Sebagai contoh, saat akan dilakukan prediksi terhadap tingkat kelulusan mahasiswa, maka atribut seperti nama, nim, nama orang tua tidak akan digunakan sebagai dasar untuk melakukan proses prediction.

4. Data Transformation

Dalam proses data mining, tidak semua data dalam format tertentu dapat digunakan, untuk itu diperlukan tahapan transformasi data. Misalnya mengubah data dalam bentuk nominal ke dalam bentuk kategorikal.

5. Proses mining

Tahapan ini adalah kegiatan utama dalam proses data mining, dimana dalam tahapan ini, algoritma yang digunakan akan menggunakan informasi atau pola yang tersembunyi dalam dataset yang telah dilakukan proses mining.

6. Presentasi pengetahuan

Merupakan tahapan terakhir dalam proses data mining, dimana dalam tahapan ini hasil dari proses mining akan dievaluasi. Hasil dari representasi pengetahuan digunakan sebagai dasar untuk mengambil keputusan terhadap hasil proses mining.

Hasil Dan Pembahasan

Penelitian ini menggunakan data set yang didapatkan dari Badan Pusat statistik (BPS). Terdapat 514 (lima ratus empat belas) Kabupaten dan kota di Indonesia, 4 (empat) atribut pembentuk indek pembangunan manusia (IPM) dan IPM itu sendiri. Data Indek pembangunan Manusia yang diujikan merupakan data pada tahun 2017. Penjelasan dari atribut terdapat di tabel data dasar dan bentuk dataset yang digunakan dalam penelitian disajikan dalam tabel berikutnya.

- Data Cleansing**, dilakukan melalui proses cek data kosong didapatkan ada 28 data dengan nilai Nol, dan diselesaikan dengan metode rata rata sehingga tidak ada lagi data kosong.
- Data Integration**, pada langkah ini data pada setiap tahun dan setiap Kota dan Kabupaten dijadikan dalam satu data sebelum diproses lebih lanjut seperti pada tabel 1.

Tabel.1 Data Dasar

Kota/Kab	Tahun										
	2019	2020	2017	2018	2015	2016	2013	2014	2012	2011	2010
Simeulue	65.22	65.26	64.90	65.00	64.66	64.78	64.23	64.24	64.15	64.22	64.05
Aceh Singkil	67.36	67.39	67.07	67.16	66.97	67.02	66.91	66.94	66.76	66.85	66.71
.
.
Deiyai	65.11	65.24	64.63	64.83	64.47	64.55	64.25	64.27	64.19	64.21	64.07
Kota Jayapura	70.38	70.45	70.00	70.15	69.97	69.99	69.95	69.95	69.90	69.94	69.89

- c) **Data Selection**, Atribut yang digunakan hanya berupa angka umur harapan hidup dan data kota dan kabupaten, tidak menyertakan Kota dan Kabupaten.
- d) **Data Transformation**, untuk dapat dilakukan klastering, data dibuat dalam bentuk numerik.
- e) **Proses Mining**, menghitung nilai jarak Euclidian dari masing masing cluster dan menerapkan K Medoids. Berikut ini adalah Tabel.2 menunjukkan jarak untuk masing masing kalster sebanyak 4 klaster .

Tabel.2 Pusat klaster sebanyak 4 klaster

Klaster	C1	C2	C3	C4
C1	0.06454	0.60822	0.28641	-0.73440
C2	-1.77462	-0.55363	-0.24907	-0.75012
C3	0.69796	0.41337	0.58123	0.76614
C4	0.52571	-0.13508	-0.02645	0.49112

sedangkan tabel.3 dibawah ini menunjukkan hasil dari masing masing klaster untuk setiap tahun data.

Tabel.3 Hasil masing masing klaster

Attribute	cluster_0	cluster_1	cluster_2	cluster_3
Propinsi	513	512	493	490
2019	70.380	65.110	72.270	68.200
2020	70.450	65.240	72.320	68.250
2017	70	64.630	71.930	67.870
2018	70.150	64.830	72.060	68
2015	69.970	64.470	71.890	67.860
2016	69.990	64.550	71.900	67.860
2013	69.950	64.250	71.850	67.840
2014	69.950	64.270	71.870	67.850
2011	69.900	64.190	71.780	67.810
2012	69.940	64.210	71.810	67.820
2010	69.890	64.070	71.740	67.790

- f) **Presentasi pengetahuan**, selanjutnya disajikan tabel hasil klaster, berikut adalah rincian daftar klaster untuk masing masing kota dan kabupaten di Indonesia. Dari hasil perhitungan didapatkan empat klaster yaitu

Cluster Model

Cluster 1 (Tinggi): 125 items

Cluster 2 (Sedang): 119 items

Cluster 3 (Rendah): 137 items

Cluster 4 (Sangat rendah): 133 items

Total number of items: 514

Berikut ini adalah tabel.4 hasil Klaster Kota/Kabupaten di Indonesia:

Tabel.4 Hasil Klaster Kota/Kabupaten di Indonesia

Klaster 1	Klaster 2	Klaster 3	Klaster 4
Aceh Besar	Simeulue	Deli Serdang	Aceh Singkil
Bireuen	Aceh Selatan	Kota Pematang Siantar	Aceh Tenggara
Pidie Jaya	Aceh Barat Daya	Kota Medan	Aceh Timur
Kota Banda Aceh	Gayo Lues	Kota Binjai	Aceh Tengah
Kota Sabang	Kota Subulussalam	Agam	Aceh Barat
Kota Lhokseumawe	Mandailing Natal	Kota Padang	Pidie
Nias	Tapanuli Selatan	Kota Solok	Aceh Utara

Toba Samosir	Pakpak Bharat	Kota Padang Panjang	Aceh Tamiang
Labuhan Batu	Batu Bara	Kota Bukittinggi	Nagan Raya
Simalungun	Kota Tanjung Balai	Kota Payakumbuh	Aceh Jaya
Karo	Kepulauan Mentawai	Kota Pekanbaru	Bener Meriah
Samosir	Sijunjung	Kota Jambi	Kota Langsa
Kota Tebing Tinggi	Tanjung Jabung Timur	Kota Sungai Penuh	Tapanuli Tengah
Kota Gunungsitoli	Lahat	Kota Metro	Tapanuli Utara
Pesisir Selatan	Ogan Komering Ulu Selatan	Belitung Timur	Asahan
Lima Puluh Kota	Ogan Ilir	Kota Pangkal Pinang	Dairi
Dharmasraya	Empat Lawang	Kota Batam	Langkat
Kota Sawah Lunto	Musi Rawas Utara	Kota Tanjung Pinang	Nias Selatan
Kota Pariaman	Kota Pagar Alam	Kota Jakarta Selatan	Humbang Hasundutan
Indragiri Hulu	Kaur	Kota Jakarta Timur	Serdang Bedagai
Pelalawan	Mukomuko	Kota Jakarta Pusat	Padang Lawas Utara
Siak	Lebong	Kota Jakarta Barat	Padang Lawas
Kambar	Pesisir Barat	Kota Jakarta Utara	Labuhan Batu Selatan
Rokan Hulu	Natuna	Bandung	Labuhan Batu Utara
Bengkalis	Lingga	Kuningan	Nias Utara
Rokan Hilir	Bondowoso	Cirebon	Nias Barat
Kota Dumai	Probolinggo	Sumedang	Kota Sibolga
Kerinci	Pandeglang	Subang	Kota Padangsidimpuan
Merangin	Serang	Karawang	Solok
Batang Hari	Kota Cilegon	Bekasi	Tanah Datar
Muaro Jambi	Lombok Barat	Bandung Barat	Padang Pariaman
Tebo	Lombok Tengah	Kota Bogor	Pasaman
Kota Palembang	Lombok Timur	Kota Sukabumi	Solok Selatan
Kota Prabumulih	Sumbawa	Kota Bandung	Pasaman Barat
Kota Bengkulu	Dompu	Kota Cirebon	Kuantan Singingi
Lampung Timur	Bima	Kota Bekasi	Indragiri Hilir
Lampung Tengah	Lombok Utara	Kota Depok	Kepulauan Meranti
Tulangbawang	Sumba Timur	Kota Cimahi	Sarolangun
Tulang Bawang Barat	Kupang	Kota Tasikmalaya	Tanjung Jabung Barat
Kota Bandar Lampung	Timor Tengah Selatan	Cilacap	Bungo
Bangka	Timor Tengah Utara	Banyumas	Ogan Komering Ulu
Belitung	Belu	Purbalingga	Ogan Komering Ilir
Bangka Barat	Alor	Banjarnegara	Muara Enim
Bangka Tengah	Lembata	Kebumen	Musi Rawas
Karimun	Flores Timur	Purworejo	Musi Banyuasin
Bintan	Sikka	Wonosobo	Banyu Asin
Bogor	Ende	Magelang	Ogan Komering Ulu Timur
Sukabumi	Manggarai	Boyolali	Penukal Abab Lematang Ilir
Cianjur	Rote Ndao	Klaten	Kota Lubuklinggau
Garut	Manggarai Barat	Sukoharjo	Bengkulu Selatan
Ciamis	Sabu Raijua	Wonogiri	Rejang Lebong
Majalengka	Malaka	Karanganyar	Bengkulu Utara
Indramayu	Katingan	Sragen	Seluma
Purwakarta	Banjar	Grobogan	Kepahiang
Pangandaran	Barito Kuala	Blora	Bengkulu Tengah
Kota Banjar	Hulu Sungai Selatan	Rembang	Lampung Barat
Tegal	Hulu Sungai Tengah	Pati	Tanggamus
Lumajang	Hulu Sungai Utara	Kudus	Lampung Selatan
Banyuwangi	Bolaang Mongondow Selatan	Jepara	Lampung Utara

Pasuruan	Banggai Kepulauan	Demak	Way Kanan
Madiun	Donggala	Semarang	Pesawaran
Bojonegoro	Toli-Toli	Temanggung	Pringsewu
Tuban	Parigi Moutong	Kendal	Mesuji
Bangkalan	Tojo Una-Una	Batang	Bangka Selatan
Sumenep	Banggai Laut	Pekalongan	Kepulauan Anambas
Kota Probolinggo	Jeneponto	Pemalang	Kep. Seribu
Kota Pasuruan	Pangkajene dan Kepulauan	Kota Magelang	Tasikmalaya
Tangerang	Bone	Kota Surakarta	Brebes
Klungkung	Pohuwato	Kota Salatiga	Jember
Bangli	Gorontalo Utara	Kota Semarang	Situbondo
Karangasem	Majene	Kota Pekalongan	Sampang
Buleleng	Polewali Mandar	Kota Tegal	Pamekasan
Kota Mataram	Mamuju Utara	Kulon Progo	Lebak
Kota Bima	Maluku Tenggara Barat	Bantul	Kota Serang
Mempawah	Maluku Tenggara	Gunung Kidul	Sumbawa Barat
Sanggau	Maluku Tengah	Sleman	Sumba Barat
Ketapang	Buru	Kota Yogyakarta	Ngada
Kubu Raya	Kepulauan Aru	Pacitan	Sumba Tengah
Kotawaringin Barat	Seram Bagian Barat	Ponorogo	Sumba Barat Daya
Kotawaringin Timur	Seram Bagian Timur	Trenggalek	Nagekeo
Lamandau	Maluku Barat Daya	Tulungagung	Manggarai Timur
Gunung Mas	Buru Selatan	Blitar	Kota Kupang
Murung Raya	Kota Tual	Kediri	Sambas
Tapin	Halmahera Barat	Malang	Kayong Utara
Tabalong	Halmahera Tengah	Sidoarjo	Kapuas
Tanah Bumbu	Kepulauan Sula	Mojokerto	Barito Selatan
Kota Banjarmasin	Halmahera Selatan	Jombang	Seruyan
Penajam Paser Utara	Pulau Morotai	Nganjuk	Pulang Pisau
Malinau	Pulau Taliahu	Magetan	Barito Timur
Tana Tidung	Kaimana	Ngawi	Tanah Laut
Nunukan	Teluk Wondama	Lamongan	Kota Baru
Minahasa	Teluk Bintuni	Gresik	Balangan
Kepulauan Sangihe	Sorong Selatan	Kota Kediri	Bolaang Mongondow
Kepulauan Talaud	Sorong	Kota Blitar	Bolaang Mongondow Utara
Minahasa Selatan	Raja Ampat	Kota Malang	Bolaang Mongondow Timur
Minahasa Utara	Tambrauw	Kota Mojokerto	Morowali
Siau Tagulandang Biaro	Maybrat	Kota Madiun	Buol
Minahasa Tenggara	Jayawijaya	Kota Surabaya	Sigi
Kota Bitung	Paniai	Kota Batu	Morowali Utara
Kota Tomohon	Puncak Jaya	Kota Tangerang	Kepulauan Selayar
Kota Kotamobagu	Boven Digoel	Kota Tangerang Selatan	Bulukumba
Banggai	Mappi	Jembrana	Takalar
Poso	Asmat	Tabanan	Sinjai
Kota Palu	Yahukimo	Badung	Maros
Bantaeng	Pegunungan Bintang	Gianyar	Barru
Gowa	Tolikara	Kota Denpasar	Soppeng
Enrekang	Sarmi	Bengkayang	Wajo
Luwu	Keerom	Landak	Sidenreng Rappang
Luwu Timur	Waropen	Sintang	Pinrang
Kota Parepare	Supiori	Kapuas Hulu	Luwu Utara

Kota Palopo	Mamberamo Raya	Sekadau	Buton
Muna	Nduga	Melawi	Bombana
Konawe	Lanny Jaya	Kota Pontianak	Konawe Utara
Kolaka	Mamberamo Tengah	Kota Singkawang	Konawe Kepulauan
Konawe Selatan	Yalimo	Barito Utara	Buton Tengah
Wakatobi	Puncak	Sukamara	Buton Selatan
Kolaka Utara	Dogiyai	Kota Palangka Raya	Boalemo
Buton Utara	Intan Jaya	Kota Banjar Baru	Gorontalo
Muna Barat	Deiyai	Paser	Bone Bolango
Kota Baubau		Kutai Barat	Mamuju
Mamasa		Kutai Kartanegara	Mamuju Tengah
Kota Ambon		Kutai Timur	Halmahera Utara
Kota Ternate		Berau	Halmahera Timur
Kota Sorong		Mahakam Ulu	Kota Tidore Kepulauan
Kota Jayapura		Kota Balikpapan	Fakfak
		Kota Samarinda	Manokwari
		Kota Bontang	Manokwari Selatan
		Bulungan	Pegunungan Arfak
		Kota Tarakan	Merauke
		Kota Manado	Jayapura
		Tana Toraja	Nabire
		Toraja Utara	Kepulauan Yapen
		Kota Makasar	Biak Numfor
		Kolaka Timur	
		Kota Kendari	
		Kota Gorontalo	
		Mimika	

Simpulan

K Medoids dapat digunakan untuk membentuk klaster kota dan kabupaten di Indonesia menurut data angka umur harapan hidup saat lahir yang merupakan salah satu faktor penilai dalam membentuk Indeks Pembangunan Manusia. Algoritma yang digunakan adalah K-Medoids yang dapat melakukan klaster dengan keunggulan mampu mengatasi noise dan outlier pada data yang besar,. Hasil yang didapatkan Cluster 1 sebanyak 125 propinsi, Cluster 2 sebanyak 119 propinsi, Cluster 3 sebanyak 137 propinsi dan Cluster 4 sebanyak 133 propinsi. Umur harapan hidup ini merupakan salah satu komponen dalam menghitung Indeks Pembangunan Manusia. Dengan diketahuinya klaster ke 3 dan ke 4 yaitu rendah dan sangat rendah, diharapkan ada evaluasi terhadap daerah daerah tersebut untuk meningkatkan tingkat harapan hidup saat lahir. Untuk mengetahui efektifitas K Medoids, klustering data dapat dibandingkan dengan algoritma lainnya, dan dilakukan pengukuran misalnya menggunakan Davies Bouldin Index.

Daftar Pustaka

C. C. Aggarwal. 2015. *Data Mining: The Textbook*.

M. Hofmann and R. Klinkenberg, 2013. *Rapid Miner Data Mining Use Cases and Business Analytics Applications*.

P. Arora, Deepali, and S. Varshney, 2016. "Analysis of K-Means and K-Medoids Algorithm for Big Data," *Phys. Procedia*, vol. 78, no. December 2015, pp. 507–512.

S. Mylevaganam, 2017. "The Analysis of Human Development Index (HDI) for

Categorizing the Member States of the United Nations (UN),” pp. 661–690,

J. Han, M. Kamber, and J. Pei, 2012. *Data Mining: Concepts and Techniques*.
BPS, 2014. “Indeks Pembangunan manusia 2014,” *Bps*.

B. P. Statistik, 2017. “Indeks Pembangunan Manusia 2017,” *Buku Lap. Tah*.

T. C. Havens, J. C. Bezdek, C. Leckie, L. O. Hall, and M. Palaniswami, 2012. “Fuzzy c-Means algorithms for very large data,” *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 6, pp. 1130–1146.

S. Harikumar and P. Surya, 2015. “K-Medoid Clustering for Heterogeneous DataSets,” *Procedia Comput. Sci.*, vol. 70, pp. 226–237.

J. Ledolter, 2013. *Data Mining and Business Analytics with R*.

D. Zhu *et al.*, 2016. “A Cluster Separation Measure,” *Procedia Comput. Sci.*, vol. 2, no. 1, pp. 1–6.

D. L. Davies and D. W. Bouldin, 1979. “A Cluster Separation Measure,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 2, pp. 224–227.