

Perbandingan Algoritma K-Means dan K-Medoids dalam Pengelompokan Kabupaten/Kota Berdasarkan Indeks Pembangunan Manusia di Provinsi Jawa Tengah

Nur Wahyu Hidayat^{#1}, Umar Ghoni^{#2}, Fitri Ayuning Tyas ^{#3}

Informatic Engineering, STMIK Muhammadiyah Paguyangan Brebes
Address Including Country Name

¹ wahyu@stmikmpb.ac.id

² ganicomp84@gmail.com

³ tyas_fa @stmikmpb.ac.id

Received on 30-10-2023, revised on dd-mm-yyyy, accepted on 30-10-2023

Abstract

Indek Pembangunan Manusia (IPM) adalah ukuran statistik yang digunakan untuk mengevaluasi kemajuan dan kualitas hidup manusia di suatu Negara. Jawa Tengah adalah salah satu provinsi di Indonesia dengan klasifikasi IPM tinggi. Namun, bila dibandingkan dengan provinsi-provinsi lain di Pulau Jawa menduduki peringkat kedua terendah dari bawah sehingga diperlukan pengelompokan daerah kabupaten/ kota sehingga bisa dijadikan sebagai salah satu alternatif solusi bagi Pemprov Jawa Tengah guna menentukan arah kebijakan terkait pembangunan manusia. Metode klustering adalah salah satu metode dalam data mining yang berguna dalam pengelompokan. Di antara metode klustering tersebut adalah K-Means dan K-Medoids. Proses data mining yang digunakan dalam penelitian ini adalah metode CRISP-DM. Tujuan dari penelitian ini adalah membandingkan kinerja algoritma K-Means dengan K-Medoids dalam pengelompokan kabupaten/kota berdasarkan capaian IPM tiap kabupaten/kota tersebut. Berdasarkan nilai *Davies Bouldin Index* (DBI) yang diperoleh pada masing-masing algoritma tersebut didapatkan bahwa kinerja algoritma K-Medoids mengungguli algoritma K-Means dengan capaian DBI sebesar 0,436 pada jumlah kluster 4, sedangkan DBI K-Means hanya mencapai 0,472 pada jumlah kluster 3.

Keywords: IPM, K-Means, K-Medoids

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Nur Wahyu Hidayat

Teknik Informatika, STMIK Muhammadiyah Paguyangan Brebes

Jalan Pangeran Diponegoro Grengseng No. 184, Taraban, Kec. Paguyangan, Kab. Brebes, Jawa Tengah 52276

Email: wahyu@stmikmpb.ac.id

I. PENDAHULUAN

A. Latar Belakang

Ada perubahan konsep pembangunan manusia yang signifikan dari waktu ke waktu [1]. Pada zaman dulu, pembangunan manusia hanya berfokus pada kebutuhan fisik guna kelangsungan hidup manusia. Pada zaman pertengahan, pembangunan manusia lebih berfokus pada kebutuhan moral dan spiritual agar tercapainya kebahagiaan akhirat. Pada zaman modern, pembangunan manusia lebih berpusat pada kemajuan intelektual agar terciptam manusia terdidik dan rasional. Pada awal abad modern, pembangunan manusia mulai terpusat pada pembangunan ekonomi agar dapat meningkatkan taraf hidup manusia. Akan tetapi, pada saat ini pandangan terhadap pembangunan manusia semakin luas dan berkembang untuk mencapai kesejahteraan hidup manusia secara menyeluruh dan berkelanjutan.

Indek Pembangunan Manusia (IPM) adalah ukuran statistik yang digunakan untuk mengevaluasi kemajuan dan kualitas hidup manusia di suatu Negara [1]. Ada 3 dimensi utama [1] yang mendasari IPM, yaitu umur panjang dan hidup sehat, pengetahuan dan standar hidup layak. IPM memberikan informasi penting bagi pembuat kebijakan dalam membuat kebijakan pembangunan yang lebih baik dan berkelanjutan. Dengan informasi IPM, kemajuan pembangunan manusia di berbagai negara dapat

dibandingkan. Selain itu, membantu dalam identifikasi masalah dan tantangan khusus yang dihadapi suatu negara tertentu dalam mencapai pembangunan berkelanjutan.

Jawa Tengah adalah salah satu provinsi di Indonesia yang memiliki IPM dengan klasifikasi tinggi, yaitu 72,79 di tahun 2022 [2]. Namun demikian, bila dibandingkan dengan provinsi-provinsi lain di Pulau Jawa, Jawa Tengah menduduki peringkat kedua terendah setelah Jawa Timur. Berdasarkan data dari BPS pada tahun 2022, IPM DKI Jakarta adalah 81,65, IPM Yogyakarta adalah 80,64, IPM Banten adalah 73,32, IPM Jawa Barat adalah 73,12, baru kemudian Jawa Tengah. Oleh karena itu, diperlukan pengelompokan kabupaten/kota di Jawa Tengah berdasarkan IPM sebagai salah satu alternatif solusi bagi Pemerintah guna merencanakan kebijakan yang tepat terkait pembangunan manusia di wilayah Provinsi Jawa Tengah.

Klastering adalah proses pengelompokan objek-objek ke dalam beberapa grup yang berbeda, atau tepatnya mempartisi dataset menjadi subsets (klaster) [3]. Klastering merupakan metode data mining yang *unsupervised* karena tidak ada satu atributpun yang digunakan untuk memandu proses pembelajaran. Ada banyak metode dalam klastering, di antaranya K-Means, K-Medoids, Fuzzy C-means, dan lain-lain. K-Means adalah salah satu dari algoritma pembelajaran tanpa guru (*unsupervised learning*) yang paling sederhana yang digunakan untuk menyelesaikan berbagai masalah pengelompokan [4]. K-Medoids adalah algoritma klastering turunan dari K-Means yang didasarkan pada penggunaan medoids bukan dari pengamatan mean yang dimiliki oleh setiap klaster [5].

Beberapa penelitian terdahulu terkait penerapan algoritma K-Means dan K-Medoids, yaitu [6], [7], [8]. Penelitian [6] terkait perbandingan algoritma K-Means dengan K-Medoid untuk mengelompokkan data obat dengan *Silhouette Coefficient*. Hasil penelitian tersebut menyebutkan bahwa algoritma K-Means menghasilkan *Silhouette Coefficient* lebih tinggi sebesar 0,627 dibandingkan dengan algoritma K-Medoids sebesar 0,536 yang artinya K-Means lebih berkualitas dibandingkan dengan K-Medoids. Penelitian lain [7] terkait K-Means dan K-Medoids pada Pemetaan Daerah Penanganan Diare pada Balita di Kabupaten Kuningan menunjukkan bahwa algoritma K-Means lebih baik daripada algoritma K-Medoids. Hasil penelitian tersebut merekomendasikan pemetaan daerah prioritas penanganan penyakit diare menggunakan algoritma K-Means dengan jumlah 3 cluster. Penelitian terkait komparasi algoritma K-Means dan K-Medoids untuk mengklaster penyebaran wabah COVID-19 di Indonesia menyatakan bahwa metode K-Means mendapatkan nilai terkecil pada K-5 sebesar 0,064, sedangkan K-Medoids pada nilai k-2 sebesar 0,411 sehingga menghasilkan kesimpulan bahwa metode terbaik untuk mengelompokkan penyebaran wabah virus corona di Indonesia adalah metode K-Means. Penelitian yang lainnya [9] terkait pengelompokan armada truk berdasarkan produktivitas menyatakan bahwa *Davies Bouldin Index* sebagai metode dalam analisis klaster menghasilkan nilai validitas sebesar 0,67 untuk K-Means clustering dan 1,78 untuk K-Medoids. Ada penelitian lain [10] yang berkaitan dengan pengelompokan kabupaten/kota di Jawa Timur berdasarkan Indeks Pembangunan Manusia (IPM) menggunakan algoritma C-Means. Hasil penelitian tersebut menyatakan bahwa dengan metode C-Means didapatkan bahwa jumlah kelompok optimum yang terbentuk adalah sebanyak 4 kelompok. Penelitian yang lain [11] tentang permasalahan optimasi jumlah cluster pada data stunting menyatakan bahwa perubahan jumlah cluster turut mempengaruhi informasi yang dihasilkan dan DBI terbukti menghasilkan jumlah cluster optimal yang berisikan informasi dengan pola yang lebih baik karena memiliki nilai intra cluster kecil

Dalam penelitian ini, klastering/ pengelompokan dilakukan menggunakan metode K-Means dan K-Medoids. Tujuan dari penelitian ini adalah untuk mengetahui algoritma mana yang memiliki kinerja yang lebih tinggi dalam mengelompokkan kabupaten/kota berdasarkan Indeks Pembangunan Manusia (IPM) di wilayah Propinsi Jawa Tengah menggunakan *Davis Bouldin Index* di antara algoritma K-Means dan K-Medoids.

B. Tinjauan Pustaka

Pada bagian ini, dijelaskan beberapa hal terkait indek pembangunan manusia (IPM), klastering, algoritma k-Means dan algoritma k-Medoids serta *Davies Bouldin Index*

1. Indeks Pembangunan Manusia

Pembangunan manusia [1] adalah upaya untuk meningkatkan kualitas hidup manusia melalui berbagai aspek, seperti pendidikan, kesehatan, akses sumber daya dan keamanan. Tujuan dari pembangunan manusia adalah meningkatkan kesejahteraan manusia secara merata, menyeluruh, dan berkesinambungan. Konsep ini diakui secara internasional dan telah digunakan berbagai organisasi untuk mengukur kemajuan pembangunan di berbagai negara. Pembangunan manusia diartikan sebagai perluasan pilihan individu dalam memenuhi kebutuhan dasarnya. Pada pembangunan manusia tidak hanya dilihat dari

kemajuan ekonomi, tetapi juga dilihat dari aspek kemampuan individu dalam mengakses sumberdaya dan pemenuhan kebutuhan dasar secara layak

Indek Pembangunan Manusia (IPM) [1] adalah ukuran statistik yang digunakan untuk mengevaluasi kemajuan dan kualitas hidup manusia di suatu Negara. Ada 3 dimensi utama yang mendasari IPM, yaitu umur panjang dan hidup sehat, pengetahuan dan standar hidup layak. IPM memberikan informasi penting bagi pembuat kebijakan dalam membuat kebijakan pembangunan yang lebih baik dan berkelanjutan. Dengan informasi IPM, kemajuan pembangunan manusia di berbagai negara dapat dibandingkan. Selain itu, membantu dalam identifikasi masalah dan tantangan khusus yang dihadapi suatu negara tertentu dalam mencapai pembangunan berkelanjutan.

2. Klastering

Klastering adalah proses membuat pengelompokan sehingga semua anggota dari setiap partisi mempunyai persamaan berdasarkan matriks tertentu [12]. Tujuan dari klastering ini adalah untuk mengelompokkan objek-objek dalam beberapa kelompok dimana setiap objek memiliki sifat yang berbeda antar kelompok dan setiap objek memiliki sifat yang homogen dalam satu kelompok.

Dalam klastering, biasanya terbagi dalam 2 (dua) metode [12], yaitu metode hierarki dan metode non hierarki. Dalam metode hierarki, pengelompokkan dimulai dari dua atau lebih objek yang memiliki kesamaan yang paling dekat kemudian dikelompokkan lagi objek-objek yang memiliki kedekatan kedua dan seterusnya sehingga membentuk seperti pohon dimana ada tingkatan yang jelas antar objek. Adapun dalam metode non hierarki, pengelompokkan dimulai dengan menentukan jumlah kluster terlebih dahulu kemudian proses kluster dilakukan tanpa mengikuti proses hierarki. Pengelompokkan metode non hierarki selanjutnya dikenal dengan nama K-Means.

3. Algoritma K-Means

Algoritma K-Means dapat dikelompokkan sebagai *partitional method* atau *hard clustering* [3]. Tujuan algoritma ini adalah untuk mempartisi sejumlah n objek ke k buah kluster, yang jarak antara objek dengan titik tengah reratanya paling dekat (nearest). Algoritma ini menerima masukan berupa data tanpa label kelas sehingga dikenal sebagai salah satu algoritma *unsupervised learning*. Diagram alir algoritma K-Means dapat dilihat pada Gambar 1.

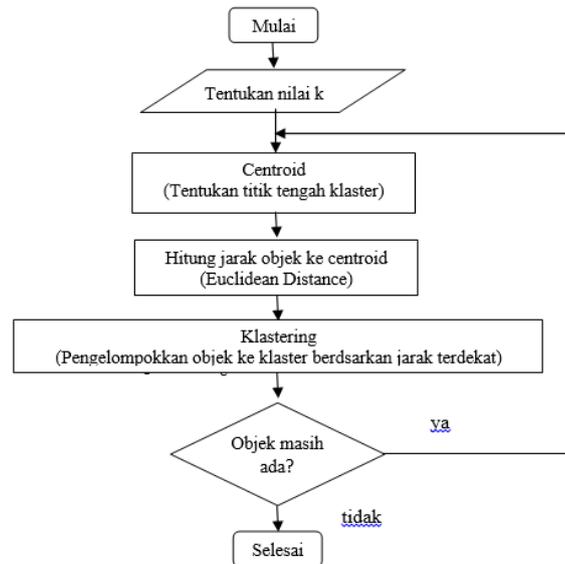
K-Means ditemukan [3] oleh beberapa orang, yaitu Liloyd pada tahun 1957, Forgey pada tahun 1965, Friedman dan Rubin pada tahun 1967 dan McQueen pada tahun 1967. Ide awal dari Liloyd pada tahun 1957, namun baru dipublikasi pada tahun 1982.

Adapun langkah-langkah algoritma k-Means adalah sebagai berikut:

- a) Tentukan k buah kluster
- b) Pilih sejumlah k buah objek secara acak yang akan dijadikan sebagai titik centroid cluster
- c) Tentukan k buah centroid (titik tengah)
- d) Kelompokkan objek ke centroid cluster terdekat berdasarkan *Euclidean distance*
- e) Hitung kembali semua nilai titik centroidnya
- f) Ulangi langkah c) sampai e) sehingga titik centroid tidak berubah lagi.

Untuk perhitungan *Euclidean distance* [13] digunakan persamaan (1)

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$



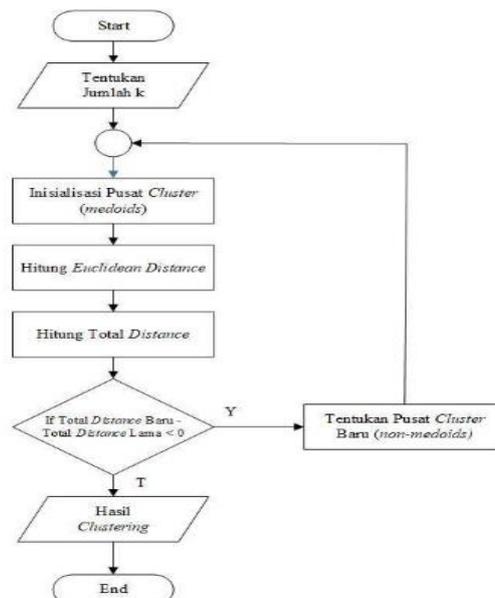
Gambar 1. Diagram Alir K-Means

4. Algoritma K-Medoids

Algoritma K-Medoids adalah algoritma klustering yang menggunakan objek sebagai perwakilan (medoid) sebagai pusat kluster untuk setiap kluster. Algoritma ini mampu mengatasi kekurangan algoritma K-Means yang sensitive terhadap *noise* dan *outlier* [14]. Diagram alir untuk algoritma K-Medoids dapat dilihat pada Gambar 2.

Adapun langkah-langkah dari algoritma K-Medoid adalah sebagai berikut:

- Tentukan jumlah kluster k
- Inisialisasi pusat kluster (medoids) dari tiap kluster secara acak
- Tetapkan setiap objek pengamatan ke kluster terdekat menggunakan persamaan *Euclidean distance*
- Pilih secara acak objek tiap kluster sebagai calon medoid baru
- Hitung jarak tiap objek di setiap kluster terhadap hasil medoid yang baru
- Hitung pada simpangan total (S) dengan menghitung nilai jarak total *distance* baru-*distance* lama. Bila simpangan total $S < 0$, maka menukar objek dengan data kluster untuk membentuk satu set k objek baru menjadi medoid
- Ulangi langkah d) sampai f) sehingga tidak terdapat perubahan medoids.



Gambar 2. Diagram Alir K-Means

5. Davies Bouldin Index.

Davies Bouldin Index (DBI) adalah sebuah metode yang digunakan untuk mengukur validitas kluster dalam pengelompokan data. DBI pertama kali diusulkan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979 [11]. Semakin kecil nilai DBI maka semakin baik pula kluster yang dihasilkan [13]. Senada dengan [13], nilai DBI yang lebih kecil [15] menunjukkan bahwa kualitas kluster yang dihasilkan lebih optimal. DBI [11] sangat membantu dalam memastikan jumlah cluster yang paling optimal dari sejumlah k yang diujikan pada data.

Ada beberapa hal [13] yang diperlukan untuk menghitung DBI, yaitu: *Sum of square within Cluster* (SSW), *Sum of Square between Cluster* (SSB), Rasio, dan DBI itu sendiri. SSW digunakan untuk mengetahui matrik kohesi dalam sebuah kluster ke- i . SSW dihitung dengan persamaan (2)

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i) \quad (2)$$

di mana:

m = nilai rata-rata sebuah kluster

$d(x_j, c_i)$ = *Euclidean distance* jarak data terhadap titik pusat kluster

SSB digunakan untuk mengetahui sparasi antar kluster. SSB dihitung dengan persamaan (3)

$$SSB_{i,j} = d(C_j, C_i) \quad (3)$$

di mana:

$d(C_j, C_i)$ = *Euclidean distance* jarak titik utama kluster terhadap titik pusat kluster yang lain.

Rasio adalah perbandingan antara kluster ke- i dan kluster ke- j . Perhitungan rasio tiap-tiap kluster dihitung dengan persamaan (4).

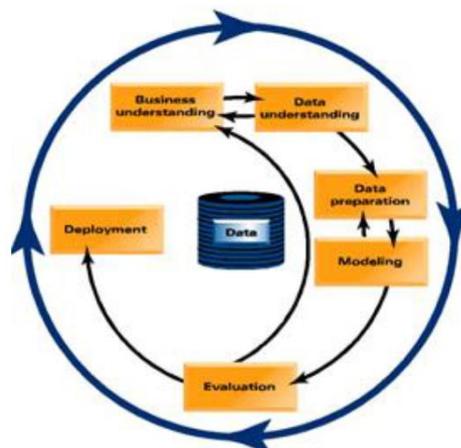
$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (4)$$

DBI dihitung dengan persamaan (5)

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad (5)$$

II. METODE PENELITIAN

Penelitian ini merupakan penelitian jenis kuantitatif dan proses data mining yang digunakan didasarkan pada metode CRISP-DM. Metode Crisp-DM adalah metode yang digunakan dalam proses data mining di mana di dalamnya terdapat 6 tahapan pengembangan, yaitu *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation* dan *deployment* [16], [17].



Gambar 3. Model CRISP-DM

Adapun 6 (enam) tahapan pengembangan model CRISP-DM, yaitu sebagai berikut:

Pertama yaitu *bussiness understanding* (pemahaman konsep). Pada tahap ini adalah pemahaman tujuan dan sudut pandang bisnis selanjutnya pengetahuan tersebut diterjemahkan ke dalam masalah data mining. Jawa Tengah adalah salah satu provinsi terbesar di Indonesia dengan IPM tergolong dalam klasifikasi tinggi. Namun demikian, bila dibandingkan dengan provinsi-provinsi lain di Pulau Jawa, IPM Jawa Tengah menduduki peringkat kedua dari bawah. Oleh karena itu, agar mempermudah Pemprov dalam memprioritaskan kebijakan terkait pembangunan manusia diperlukan metode klastering data mining.

Kedua yaitu *data understanding* (pemahaman data). Data penelitian ini adalah data sekunder yang diperoleh Badan Pusat Statistik (BPS) Provinsi Jawa Tengah data Indeks Pembangunan Manusia kabupaten/ kota se-Jawa Tengah dari tahun 2010 sampai dengan 2022. Data tersebut dapat diunduh melalui link <https://jateng.bps.go.id/indicator/26/83/1/indeks-pembangunan-manusia-metode-baru-.html>. Jumlah objek dalam data tersebut adalah sebanyak 35 (karena Provinsi Jawa Tengah memiliki 35 kabupaten/kota).

Ketiga yaitu *data preparation*. Pada tahap ini, mencakup penggabungan data (data integration) dan pembersihan data (data cleaning) sehingga menghasilkan data yang siap untuk diproses pada tahap selanjutnya.

Keempat yaitu *modeling* (pemodelan). Proses pemodelan dilakukan dengan membandingkan dua buah algoritma klastering, yaitu K-Means dan K-Medoids. Pemodelan dilakukan secara berurutan, yaitu yang pertama menggunakan algoritma K-Means dengan mengujicobakan nilai k yang bervariasi mulai dari k=2 sampai dengan k=5 dan yang selanjutnya menggunakan algoritma K-Medoids.

Kelima yaitu *evaluation* (evaluasi). Pada tahap ini, kedua algoritma klastering, baik K-Means maupun K-Medoids dievaluasi. Pada penelitian ini untuk mengevaluasi kedua algoritma klastering tersebut adalah dengan menghitung nilai *Davies Bouldin Index* (DBI). Suatu klaster akan dianggap memiliki hasil klastering yang optimal manakala memiliki DBI yang minimal [18].[13]

Keenam yaitu *deployment* (pengembangan). Pada tahap ini, pengetahuan dan informasi yang diperoleh dari hasil pengolahan data mining dipresentasikan dalam bentuk laporan sehingga mempermudah pembaca dalam mengambil kesimpulan. Pengetahuan yang diperoleh dari hasil pengolahan data menggunakan algoritma klastering ini berupa klaster/ pemetaan daerah yang dapat dijadikan sebagai masukan bagi Pemerintah Provinsi Jawa Tengah dalam memprioritaskan kebijakan terkait pembangunan manusia, khususnya di Jawa Tengah.

III.HASIL DAN PEMBAHASAN

Bussiness understanding adalah tahap pertama dalam metode CRISP-DM. Berdasarkan data BPS Jawa Tengah, Provinsi Jawa Tengah pada tahun 2022 memiliki IPM yang masuk dalam kategori tinggi (menurut standar UNDP dengan IPM 70-79). Namun demikian, bila dibandingkan dengan provinsi-provinsi lain di Pulau Jawa. IPM Jawa tengah menduduki peringkat kedua terendah. Terkait dengan peran data mining, ada salah satu peran data mining yang berguna untuk mengelompokkan data yang dikenal dengan nama klastering. Hasil dari pengelompokkan data IPM bisa menjadi alternatif solusi terkait arah kebijakan dalam

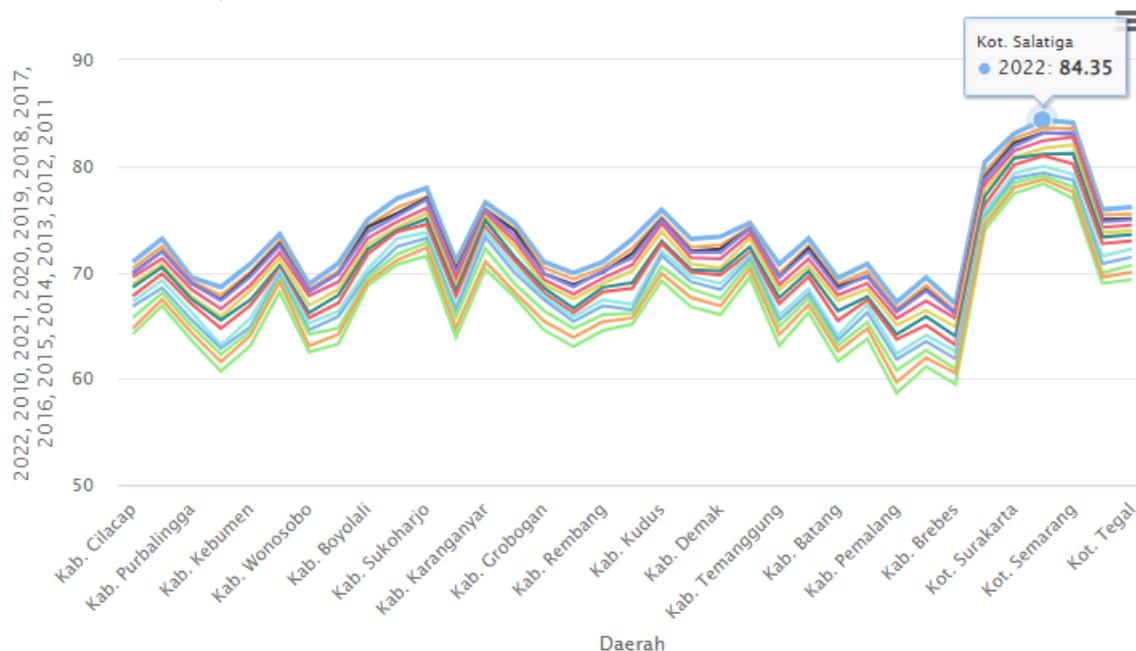
Dalam *data understanding* ini, dataset yang digunakan untuk klastering adalah data Indeks Pembangunan Manusia (IPM) yang diperoleh dari Badan Pusat Statistik (BPS) Jawa Tengah. Data tersebut diambil dari periode 2010 sampai dengan 2022. Karena Provinsi Jawa Tengah memiliki 35 kabupaten/kota sehingga jumlah objek yang diambil sebanyak 35 objek. Karena yang ingin diamati tren/ kecenderungan IPM dari tahun ke tahun maka setiap objek memiliki 13 atribut yang berisi IPM dari tahun 2010 sampai dengan tahun 2022.

Selanjutnya adalah *data preparation*. Pada tahap ini meliputi *data integration*, *data transformation*, dan *data cleaning*. Dataset yang terpisah-pisah tersebut digabungkan menjadi satu kesatuan atau dikenal dengan nama *data integration*. Karena dataset yang ada dari BPS Provinsi Jawa Tengah masih terpisah-pisah, maka perlu dilakukan *data integration*. *Data transformation* juga dilakukan agar dataset yang ada sesuai dengan format yang diperlukan. Karena semua dataset sudah lengkap, pada tahap ini *data cleaning* (pembersihan data) tidak diperlukan. Hasil dari data preparation menghasilkan dataset baru yang siap untuk diproses pada tahap selanjutnya. Dataset baru dapat dilihat pada Tabel I.

Table I. INDEKS PEMBANGUNAN MANUSIA (METODE BARU)

1	Daerah	2022	2021	2020	2019	2018	2017	2016	2015	2014	2013	2012	2011	2010
2	Kab. Cilacap	70,99	70,42	69,95	69,98	69,56	68,9	68,6	67,77	67,25	66,8	65,72	64,73	64,18
3	Kab. Banyumas	73,17	72,44	71,98	71,96	71,3	70,75	70,49	69,89	69,25	68,55	68,06	67,45	66,87
4	Kab. Purbalingga	69,54	69,15	68,97	68,99	68,41	67,72	67,48	67,03	66,23	65,53	64,94	64,33	63,61
5	Kab. Banjarnegara	68,61	67,86	67,45	67,34	66,54	65,86	65,52	64,73	63,15	62,84	62,29	61,58	60,7
7	Kab. Purworejo	73,6	72,98	72,68	72,5	71,87	71,31	70,66	70,37	70,12	69,77	69,4	69,11	68,16
33	Kot. Salatiga	84,35	83,6	83,14	83,12	82,41	81,68	81,14	80,96	79,98	79,37	79,1	78,76	78,35
34	Kot. Semarang	84,08	83,55	83,05	83,19	82,72	82,01	81,19	80,23	79,24	78,68	78,04	77,58	76,96
35	Kot. Pekalongan	75,9	75,4	74,98	74,77	74,24	73,77	73,32	72,69	71,53	70,82	69,95	69,54	68,95
36	Kot. Tegal	76,15	75,52	75,07	74,93	74,44	73,95	73,55	72,96	72,2	71,44	70,68	70,03	69,33

Gambar 4. menunjukkan grafik garis (Line Chart) dari data IPM kabupaten/kota se-Jawa Tengah selama periode 2010 sampai dengan 2022. Berdasarkan grafik tersebut menunjukkan bahwa terjadi peningkatan IPM di setiap kabupaten/kota di seluruh wilayah Jawa Tengah dari tahun ke tahun mulai dari tahun 2010 sampai dengan tahun 2022. Pada tahun 2022 daerah dengan IPM tertinggi adalah Kota Salatiga, yaitu 84,35. Adapun daerah dengan IPM terendah pada tahun yang sama adalah Kabupaten Brebes, yaitu 67,03. Selama periode waktu 2010-2022, Kabupaten Pemalang mampu mengejar keteringgalan IPM dari Kabupaten Brebes dengan kenaikan sebesar 8,75, sedangkan kenaikan Kabupaten Brebes adalah sebesar 7,54.



Gambar 4. Grafik Garis IPM Kabupaten/Kota se-Jawa Tengah periode 2010-2022

Setelah proses *data preparation*, langkah selanjutnya adalah *modeling*. Pada tahap ini, algoritma klastering, yaitu k-Means kemudian K-Medoids digunakan untuk memproses data baru yang dihasilkan dari tahap *preparation*. Pemodelan pertama dilakukan dengan menggunakan beberapa operator K-Means sekaligus dimana setiap operator K-Means pada nilai k diberi nilai masing-masing 2, 3, 4, dan 5. Setiap operator k-Means tersebut dihubungkan dengan operator *cluster distance performance* sehingga bisa dilakukan evaluasi. Pemodelan kedua dilakukan dengan menggunakan beberapa operator k-Medoids sekaligus di mana setiap operator K-Medoids pada nilai k diberi nilai masing-masing adalah 2, 3, 4, dan 5. Setiap operator k-Medoids tersebut dihubungkan dengan operator *cluster distance performance* pada Rapidminer untuk proses evaluasi.

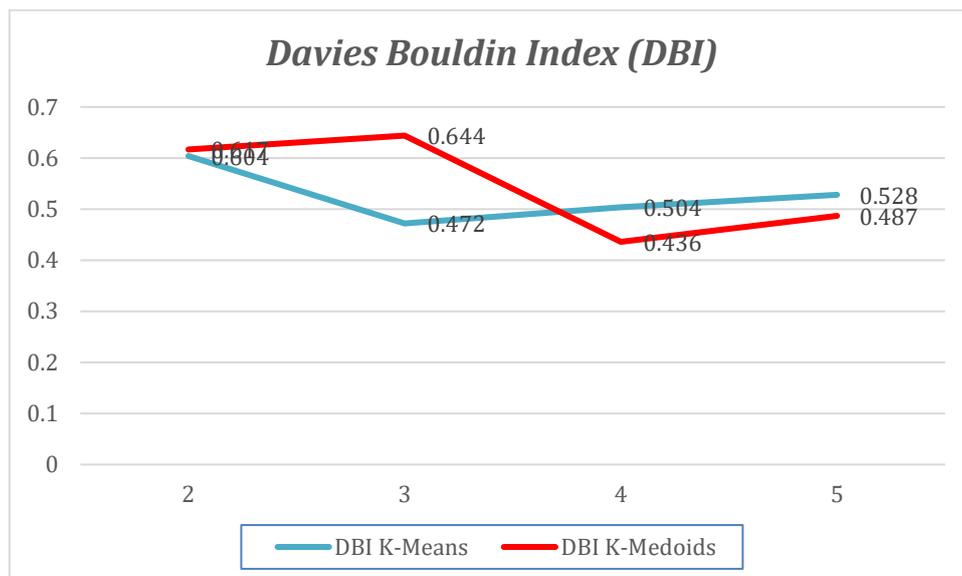
Setelah melalui *modeling*, selanjutnya adalah evaluasi. Evaluasi hasil klastering tersebut dilakukan dengan menggunakan indikator *Davies Bouldin Index* (DBI). Berdasarkan hasil perhitungan algoritma

klustering pada data IPM kabupaten/kota se-Jawa Tengah menggunakan Rapidminer untuk beberapa nilai k dapat dilihat pada Tabel II. Dari Tabel II tersebut, terlihat bahwa nilai DBI untuk setiap jumlah kluster berbeda-beda. Untuk algoritma K-Means dengan jumlah variasi jumlah kluster dari 2 sampai 5 diperoleh nilai DBI terkecil ada pada jumlah kluster 3, yaitu sebesar 0,472. Adapun untuk algoritma K-Medoids diperoleh nilai DBI terkecil ada pada jumlah kluster 4, yaitu sebesar 0,436. Semakin kecil nilai DBI dari suatu algoritma klustering menunjukkan bahwa semakin unggul algoritma tersebut. Oleh karena itu, algoritma K-Medoids memiliki kinerja yang lebih unggul dibandingkan dengan algoritma K-Means dalam pengelompokan/ klustering data IPM Kabupaten/Kota se-Jawa Tengah .

Table II. PERBANDINGAN ALGORITMA K-MEANS DAN K-MEDOIDS

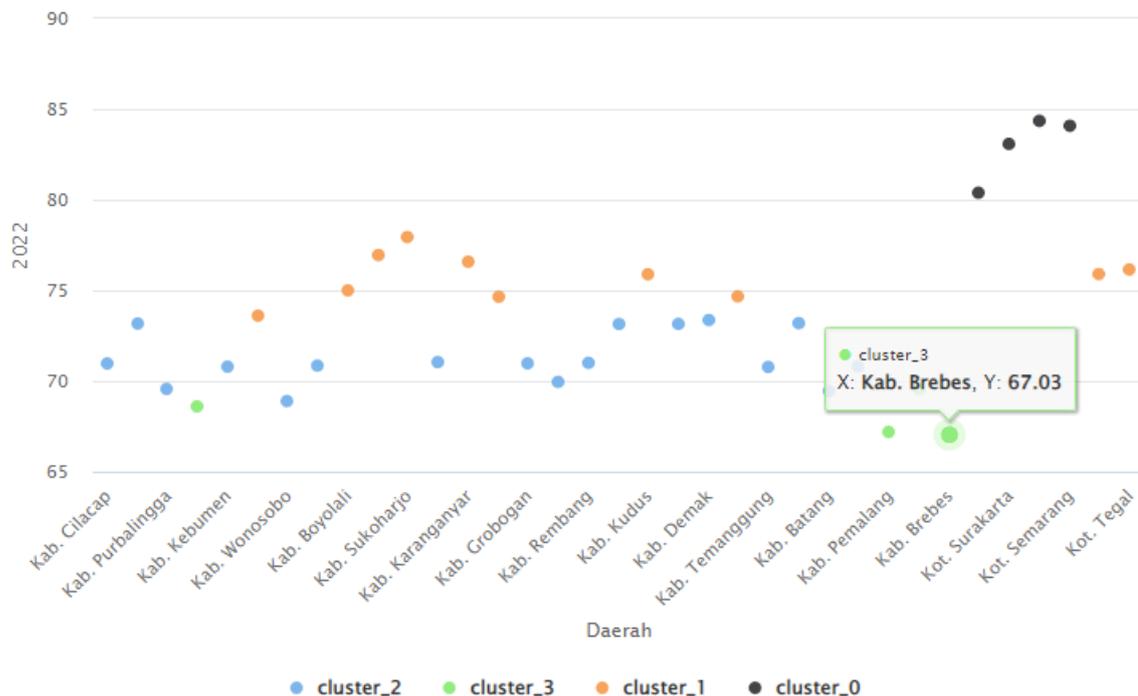
No	Jumlah kluster	DBI K-Means	DBI K-Medoids
1	2	0,604	0,617
2	3	0,472	0,644
3	4	0,504	0,436
4	5	0,528	0,487

Untuk mempermudah visualisasi dari perbandingan capaian *Davies Bouldin Index* dapat dilihat pada Gambar 5. Berdasarkan grafik tersebut nampak bahwa algoritma K-Means lebih unggul dari pada K-Medoids pada jumlah 3 kluster, sedangkan K-Medoids lebih unggul dari pada K-Means pada jumlah 4 kluster. Lihat pada gambar 5.



Gambar 5. Grafik Garis Perbandingan *Davies Bouldin Index* (DBI)

Terkait jumlah kluster 4 pada algoritma K-Medoids dapat dilihat pada grafik scatter pada Gambar 6. Berdasarkan grafik tersebut dapat diamati bahwa ada data IPM kabupaten/ Kota se-Jawa Tengah dapat dikelompokkan menjadi 4 kelompok, yaitu cluster_0 yang disimbolkan dengan titik berwarna hitam, cluster_1 yang disimbolkan dengan titik berwarna orange, cluster_2 yang disimbolkan dengan titik berwarna biru dan cluster_3 yang disimbolkan dengan titik berwarna hijau.



Gambar 6. Grafik Scatter K-Medoids pada 4 Klaster

Berdasarkan capaian IPM, pada setiap klaster memiliki beberapa anggota/ daerah sebagaimana yang terlihat pada Gambar 3. Untuk cluster_0 terdiri dari 4 (empat) daerah. Untuk cluster_1 terdiri dari 10 (sepuluh) daerah. Untuk cluster_2 terdiri dari 17 (tujuh belas) daerah. Adapun untuk cluster_3 terdiri dari 4 (empat) daerah. Adapun detail dari keempat klaster tersebut dapat dilihat pada Tabel III.

Table III. PEMBAGIAN 4 (EMPAT) KLASTER/ KELOMPOK BERDASARKAN IPM METODE K-MEDOIDS

Jumlah Daerah	Nama Klaster	Daerah (Kabupaten/ Kota)
4	Cluster_0	kota Salatiga, kota Semarang, kota Surakarta dan kota Magelang
10	Cluster_1	kab. Sukoharjo, kab. Klaten, kab. Karanganyar, kota Tegal, kota Pekalongan, kab. Kudus, kab. Boyolali, kab. Semarang, kab. Sragen dan kab. Purworejo.
17	Cluster_2	kab. Demak, kab. Kendal, kab. Banyumas, kab. Jepara, kab. Pati, kab. Wonogiri, kab. Rembang, kab. Cilacap, kab. Grobogan, kab. Magelang, kab. Pekalongan, kab. Kebumen, kab. Temanggung, kab. Blora, kab. Purbalingga, kab. Batang, dan Wonosobo
4	Cluster_3	kab. Tegal, kab. Purbalingga, kab. Pemalang dan kab. Brebes

IV. KESIMPULAN

Berdasarkan pada uraian yang telah dipaparkan pada bab hasil dan pembahasan, dapat disimpulkan bahwa kinerja algoritma K-Medoids lebih unggul dari pada kinerja algoritma K-Means dimana bisa dilihat dengan nilai DBI yang dihasilkan sebesar 0,436 pada 4 klaster. Berdasarkan nilai DBI yang diperoleh hasil perhitungan algoritma K-Medoids tersebut yaitu cluster_0 (sangat tinggi), cluster_1 (tinggi), cluster_2 (sedang) dan cluster_3 (rendah). Adapun cluster_3 memiliki 4 anggota, yang meliputi kab. Brebes, kab. Pemalang, kab. Purbalingga dan kab. Tegal. Jadi, keempat anggota cluster_3 tersebut sangat layak untuk diprioritaskan mendapatkan perhatian khusus dari Pemrov. Jawa Tengah terkait arah kebijakan pembangunan manusia di wilayah Provinsi Jawa Tengah.

ACKNOWLEDGMENT

Penulis mengucapkan terima kasih kepada segenap pihak yang telah membantu terselesaikannya penelitian ini, khususnya kepada LPPM STMIK Muhammadiyah Paguyangan Brebes dan Badan Pusat Statistik Provinsi Jawa Tengah.

REFERENSI

- [1] "Indeks Pembangunan Manusia 2022.Pdf."
- [2] "Analisis Kualitas Pembangunan Manusia Provinsi Jawa Tengah 2022.Pdf."
- [3] R. Primartha, *Belajar Machine Learning Teori Dan Praktek*. Penerbit Informatika, 2018.
- [4] "Condrobimo Et Al. - 2016 - The Application Of K-Means Algorithm For LQ45 Inde.Pdf."
- [5] Y. H. Susanti And E. Widodo, "PERBANDINGAN K-MEANS DAN K-MEDOIDS CLUSTERING TERHADAP KELAYAKAN PUSKESMAS DI DIY TAHUN 2015".
- [6] R. A. Farissa, R. Mayasari, and Y. Umaidah, "Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokkan Data Obat dengan Silhouette Coefficient," vol. 5, no. 2.
- [7] T. S. Syamfithriani, N. Mirantika, and R. Trisudarmo, "Perbandingan Algoritma K-Means dan K-Medoids Untuk Pemetaan Daerah Penanganan Diare Pada Balita di Kabupaten Kuningan," *J. Sist. Inf. BISNIS*, vol. 12, no. 2, pp. 132–139, Mar. 2023, doi: 10.21456/vol12iss2pp132-139.
- [8] W. Utomo, "The comparison of k-means and k-medoids algorithms for clustering the spread of the covid-19 outbreak in Indonesia," *Ilk. J. Ilm.*, vol. 13, no. 1, pp. 31–35, Apr. 2021, doi: 10.33096/ilkom.v13i1.763.31-35.
- [9] A. Supriyadi, A. Triayudi, and I. D. Sholihati, "PERBANDINGAN ALGORITMA K-MEANS DENGAN K-MEDOIDS PADA PENGELOMPOKAN ARMADA KENDARAAN TRUK BERDASARKAN PRODUKTIVITAS," vol. 06.
- [10] M. Qori'atunnadyah, "Metode C-Means untuk Pengelompokkan Kabupaten/Kota Provinsi Jawa Timur berdasarkan Indikator Indeks Pembangunan Manusia (IPM)".
- [11] D. Jollyta, "Optimasi Cluster Pada Data Stunting: Teknik Evaluasi Cluster Sum of Square Error dan Davies Bouldin Index," *Pros. Semin. Nas. Ris. Inf. Sci. SENARIS*, Sep. 2019.
- [12] M. W. Talakua, Z. A. Leleury, and A. W. Taluta, "ANALISIS CLUSTER DENGAN MENGGUNAKAN METODE K-MEANS UNTUK PENGELOMPOKKAN KABUPATEN/KOTA DI PROVINSI MALUKU BERDASARKAN INDIKATOR INDEKS PEMBANGUNAN MANUSIA TAHUN 2014," *BAREKENG J. Ilmu Mat. Dan Terap.*, vol. 11, no. 2, pp. 119–128, Dec. 2017, doi: 10.30598/barekengvol11iss2pp119-128.
- [13] F. Tempola, "Penggunaan Internet di Kalangan Siswa SD di Kota Ternate: Suatu Survey, Penerapan Algoritma Clustering dan Validasi DBI," *J. Teknol. Inf. Dan Komput. JTIK*, vol. 7, Desember 2020.
- [14] "Riyanto - 2019 - PENERAPAN ALGORITMA K-MEDOIDS CLUSTERING UNTUK PEN.pdf."
- [15] W. Widiarina, "Algoritma Cluster Dinamik untuk Optimasi Cluster pada Algoritma K-Means dalam Pemetaan Nasabah Potensial," *J. Intell. Syst.*, vol. 1, Feb. 2015.
- [16] "Eckerson et al. - 2000 - DIRECTOR OF EDUCATION AND RESEARCH.pdf."
- [17] I. Budiman, "Data Clustering Menggunakan Metodologi CRISP-DM untuk Pengenalan Pola Proporsi Pelaksanaan Tridharma," *Semin. Nas. Apl. Teknol. Inf.*, Jun. 2012.
- [18] "Kamila et al. - 2019 - Perbandingan Algoritma K-Means dan K-Medoids untuk.pdf."