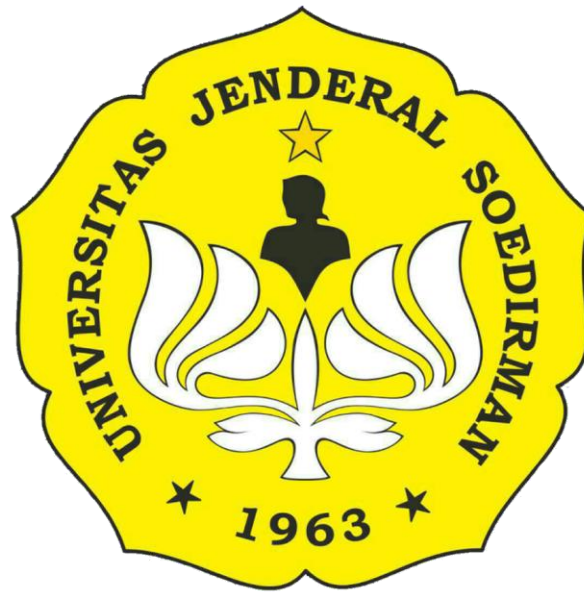


MAKALAH
KLASTERING DAERAH BERPOTENSI MAJU BERDASAR
INDEKS PEMBANGUNAN MANUSIA DI INDONESIA
MENGGUNAKAN METODE K-MEANS



Disusun oleh :
Tim Warehouse

Ketua Tim : Taufik Satria Nugraha
Anggota Tim : 1. Ali Murtadho
2. Muhammad Ilham Kurniawan

UNIVERSITAS JENDERAL SOEDIRMAN
Jl. Profesor DR. HR Boenyamin No. 708, Dukuhbandong,
Grendeng, Kec. Purwokerto Utara
Kabupaten Banyumas
Jawa Tengah 53122
2022

ABSTRAK

Sumber Daya Manusia merupakan faktor penting dalam kemajuan daerah terhadap negaranya. Hal tersebut pula yang menjadikan Badan Pusat Statistik memasukan nilai-nilai dari setiap aspek Sumber Daya Manusia pada setiap daerah di Indonesia untuk dimasukkan ke dalam indikator Indeks Pembangunan Manusia. Pemerataan nilai sumber daya manusia pada setiap daerah terbilang penting mengingat salah satu aspek dari negara maju ialah minimnya daerah terbelakang. Namun banyak dari masyarakat yang masih bingung untuk menentukan daerah maju hingga daerah terbelakang di Indonesia. Pada penelitian ini guna mendapatkan hasil klasterisasi golongan daerah-daerah, diambil nilai Indeks Pembangunan Manusia yang mencangkup seluruh kabupaten/kota di seluruh Indonesia pada tahun 2021. Untuk kami lanjutkan memproses data tersebut dengan data mining dengan metode K-Means Clustering. Hasil dari klasterisasi yang kami uji yaitu mendapati daerah berpotensi maju di Indonesia dibagi menjadi 3 macam cluster yaitu daerah berpotensi maju rendah, daerah berpotensi maju sedang, dan daerah berpotensi maju kuat. Dimana terdapat 547 daerah kabupaten/kota dan provinsi di Indonesia yang dilakukan proses clustering, hasil dari clustering yaitu pada cluster rendah terdiri atas 142 daerah atau sebesar 25,9%, cluster sedang terdiri dari 304 daerah atau sebesar 55,5%, dan cluster tinggi terdiri 102 daerah atau sebesar 18,6%. Dari hasil yang kami dapatkan, sebanyak 25,9% atau 142 daerah berpotensi rendah agar kedepannya bisa lebih diutamakan untuk penunjang fasilitas ataupun program yang bisa menunjang pemerataan agar menjadikan negara Indonesia maju.

Kata kunci : Clustering, Indonesia, IPM, K-Means, Maju

BAB I PENDAHULUAN

Tidak bisa ditampik lagi bahwa kemajuan bangsa merupakan salah tujuan dari terbentuknya bangsa dan negara untuk melangsungkan kehidupan bernegaranya agar lebih teratur dan lancar. Kemajuan suatu negara tentunya tidak lepas dari beberapa faktor internal maupun eksternal dari negara itu sendiri, salah satunya dari faktor internal, yaitu keunggulan dari SDM (Sumber Daya Manusia) yang mencakup Inteligensi, Ekonomi, dan Indeks Kehidupan dari masyarakat dapat menjadikan sebuah negara menjadi negara maju.

Dalam hal ini salah satu indikator terpenting dalam menggambarkan kemajuan dalam suatu negeri adalah peningkatan kualitas Sumber Daya Alam, itulah sebabnya perencanaan pembangunan dewasa ini, khususnya dari ketiga aspek: Intelegensi, Ekonomi, dan Indeks Kehidupan senantiasa menjadi fokus utama. Ketiga faktor dari Sumber Daya Manusia tersebut dikemas dalam satu Indeks, yaitu Indeks Pembangunan Manusia oleh Badan Pusat Statistik untuk dijadikan indikator penilaian aspek kualitas dari pembangunan, untuk mengklasifikasikan apakah sebuah negara tersebut negara maju, negara berkembang, atau negara terbelakang, dan juga sekaligus untuk mengukur pengaruh dari kebijakan ekonomi terhadap kualitas hidup[1].

Pada penelitian ini guna mendapatkan nilai Indeks Pembangunan Manusia digunakan nilai Harapan Lama Sekolah, Pengeluaran per Kapita, Rata-rata Lama

Sekolah, dan Umur Harapan Hidup. Data tersebut diperoleh dari Badan Pusat Statistik yang mencakup seluruh kabupaten/kota di seluruh Indonesia pada tahun 2021. Setelah data didapatkan dilanjutkan mengelompokkan data dengan data mining.

Data mining adalah metode pengolahan data yang difokuskan untuk mencari pola tersembunyi dalam data tersebut [2]. Salah satu metode yang terdapat pada Data Mining untuk menentukan pola-pola penggalian informasi dengan menggunakan metode Clustering. Clustering dapat diartikan sebagai metode untuk melakukan pengelompokan dengan kriteria tertentu berdasarkan kemiripan karakteristik dari setiap objek yang akan dikelompokkan. Sehingga objek yang memiliki karakteristik yang hampir sama akan terkumpul dalam satu kelompok. Sementara itu, Sifat objek antar kelompok memiliki kesamaan yang rendah. Dengan kata lain, keragaman dalam suatu kelompok kecil sedangkan keragaman antar kelompok tinggi. [3] Untuk tujuan ini, maka digunakan sebuah metode klastering non-hirarki yaitu K-Means. Algoritma K-Means clustering merupakan salah satu algoritma pengelompokan data dengan sistem partisi[4]. Diharapkan bahwa hasil clustering dapat menunjukkan persebaran kelompok wilayah yang berbeda secara signifikan dilihat dari parameter nilai IPM yang didapat. Nantinya terdapat 3 cluster yaitu daerah berpotensi rendah untuk maju, berpotensi sedang untuk maju, berpotensi kuat untuk maju.

Pentingnya informasi tentang data cluster ini dapat membantu pemerintah untuk membuat kebijakan yang sesuai pada masing-masing wilayah berdasarkan informasi hasil clustering wilayah kota/kabupaten di seluruh Indonesia. Tujuan dari penelitian ini adalah untuk mengetahui bagaimana hasil pengelompokan wilayah kabupaten/kota di seluruh wilayah Indonesia berdasarkan nilai IPM masing-masing dengan menggunakan metode K-Means .

BAB II

STUDI LITERATUR

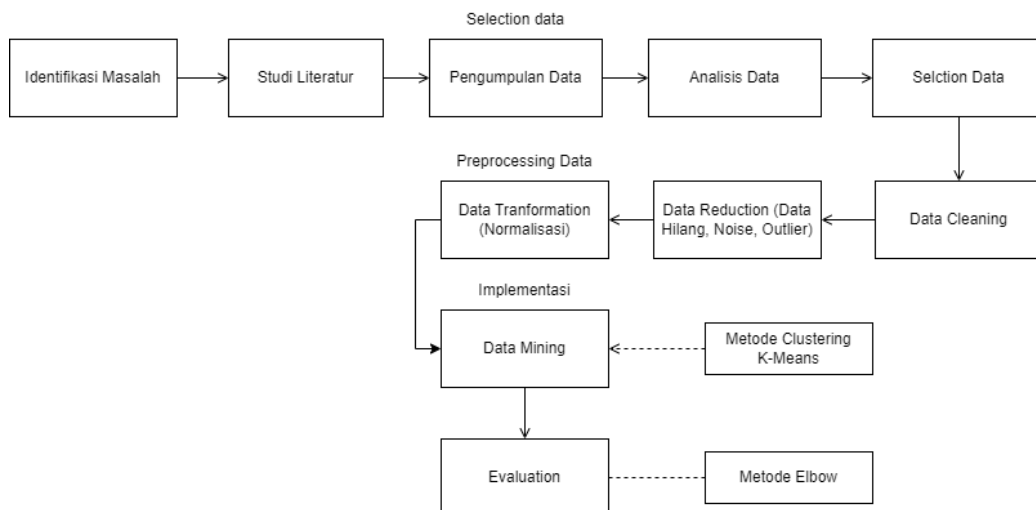
Pada penelitian ini [5] Pembangunan manusia yang bertujuan untuk menciptakan kesejahteraan masyarakat cukup dibutuhkan oleh suatu negara. Salah satu indikator penting yang dapat digunakan untuk melihat upaya pembangunan kualitas hidup manusia adalah Indeks Pembangunan Manusia (IPM). Tujuan dari penelitian ini adalah mengelompokkan Kabupaten/Kota di Jawa Tengah berdasarkan komponen IPM menggunakan metode k-means clustering. Penelitian ini menghasilkan 3 cluster dengan masing-masing karakteristik yang berbeda. Hasil dari penelitian ini diharapkan dapat dijadikan pertimbangan oleh pemerintah untuk lebih meningkatkan program-program yang mampu membangun nilai IPM pada beberapa daerah yang masih memiliki nilai IPM di bawah rata-rata.

Pada Penelitian ini [6] Menerapkan datamining menggunakan metode clustering dengan algoritma k-means. Penelitian ini bertujuan untuk mengklasifikasikan wilayah berdasarkan Indeks Pembangunan TIK ke dalam 3 cluster. Berdasarkan hasil analisis k-means diperoleh tingkat IP-TIK rendah (C1) terdiri atas 4 provinsi yaitu Nusa Tenggara Timur, Sulawesi Barat, Maluku Utara, dan Papua, cluster sedang (C2) terdiri atas 22 Provinsi dan cluster tinggi (C3) terdiri atas 8 Provinsi yaitu Kepulauan Riau, DKI Jakarta, Jawa Barat, DI Yogyakarta, Banten, Bali, Kalimantan Timur dan Kalimantan Utara. Dengan melakukan cluster Indeks Pembangunan TIK dapat memberikan gambaran wilayah-wilayah yang masih memerlukan perbaikan sehingga terjadi pemerataan pembangunan TIK di setiap provinsi di Indonesia.

Pada Penelitian Optimasi Performa Cluster K-Means Menggunakan Sum Of Squared Error (Sse) [4], K-Means merupakan suatu algoritma pengklasteran yang cukup sederhana yang mempunyai kemampuan mengelempokkan data dalam jumlah yang cukup besar, mempartisi dataset kedalam beberapa klaster k. Algoritmanya cukup mudah untuk diimplementasikan dan dijalankan, relative cepat, dan efisien. Disi lain K-Means masih memiliki beberapa kelemahan, yaitu dalam menentukan jumlah cluster. Hasil cluster yang terbentuk dari metode K-means ini sangatlah tergantung pada inisiasi nilai pusat awal cluster yang diberikan. Hal ini menyebabkan hasil clusternya berupa solusi yang sifatnya local optimal. Pada penelitian ini dilakukan untuk mengatasi kelemahan yang ada pada algoritma K-Means yaitu: perbaikan pada algoritma K-Means menghasilkan cluster yang lebih baik yaitu penerapan Sum Of Squared Error (SSE) untuk membatu K-Means Clustering dalam menentukan jumlah cluster yang paling optimum, dari proses modifikasi ini, diharapkan pusat cluster yang diperoleh nantinya akan menghasilkan cluster-cluster, dimana antar anggota cluster memiliki tingkat kemiripan yang tinggi. Perbaikan performa cluster K-Means akan diterapkan pada penentuan pusat cluster.

BAB III METODE

Pada penelitian ini, dibuat sebuah kerangka kerja berupa skema untuk mempermudah penelitian. Kerangka kerja dibuat berupa suatu tahapan-tahapan sebagai panduan yang harus dilakukan untuk mendapatkan hasil akhir dari penelitian. Berikut adalah tahapan penelitian.



Gambar 1. Flowchart Metode

Gambar 1 menjelaskan tahapan penelitian yang akan dilaksanakan, mulai dari tahap identifikasi masalah hingga tahap evaluasi hasil penelitian seperti penjelasan sebagai berikut:

A. Identifikasi Masalah

Permasalahn dalam penelitian ini yaitu masih belum merata indeks pembangunan manusia dan belum adanya klasterisasi daerah berdasarkan indeks pembangunan manusia di indonesia. Dengan adanya identifikasi masalah berfungsi agar peneliti fokus pada titik permasalahan yaitu bagaimana implementasi algoritma k-means dalam menentukan Klastering daerah berpotensi maju berdasar

indeks pembangunan manusia di Indonesia. Dikarenakan luasnya permasalahan yang ada, diperlukan adanya pembatasan masalah diantaranya metode yang digunakan clustering k-means, fokus penelitian terhadap klastering daerah berpotensi maju berdasar indeks pembangunan manusia di Indonesia, dataset yang digunakan bersumber dari BPS terdiri dari 548 kabupaten/kota dan provinsi di Indonesia, periode 2021, dan metode evaluasi yang digunakan yaitu Elbow Method. Metode analisis menggunakan studi literatur dilakukan dengan mencari sumber pustaka relevan dengan masalah yang dapat dijadikan topik sehingga dapat dianalisis untuk mendapatkan pemahaman mengenai penerapan metode k-means dalam Klastering data indeks pembangunan manusia 2021.

B. Selection Data

Data yang digunakan diambil dari dokumentasi Badan Pusat Statistik (BPS) pada url <https://www.bps.go.id> dataset terdiri dari 548 kabupaten/kota dan provinsi di Indonesia dengan atribut Daerah, PPK, RLS, UHH, HLS, dan IPM tahun 2021.

Tabel 1. Informasi Dataset

Atribut	Type	Information
Daerah	Object	Daerah Menurut Kabupaten/Kota dan Provinsi di Indonesia
PPK 2021	Float	Pendapatan Per Kapita (Standar hidup layak)
RLS 2021	Float	Rata-rata Lama Sekolah (Pendidikan)
HLS 2021	Float	Harapan Lama Sekolah (Pendidikan)
AHH 2021	Float	Angka Harapan Hidup Saat Lahir (Kesehatan)
IPM 2021	Float	Indeks Pembangunan Manusia

C. Data Cleaning

Tahap data cleaning berfungsi untuk memeriksa apakah dalam data tersebut terdapat data noise, data hilang. Pada tahap ini menghilangkan missing values (data hilang pada dataset).

D. Data Reduction

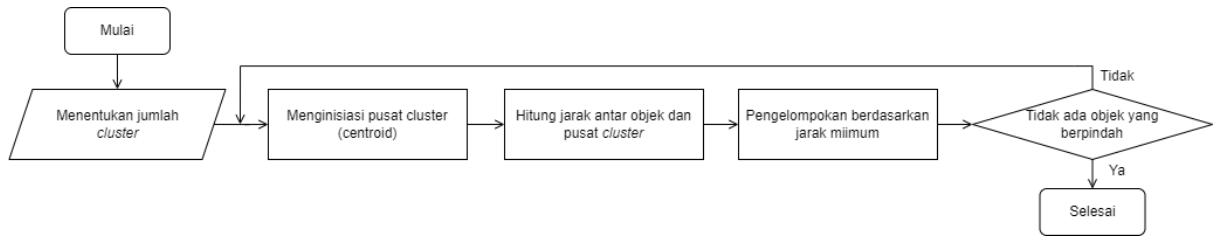
Data Reduction merupakan pemilihan, pemusatan pada penyederhanaan data kompleks tanpa mengurangi integritas data asli serta tidak mengurangi kualitas informasi yang dihasilkan. Data reduction berfungsi untuk mengatasi keterbatasan penyimpanan dan mengatasi lamanya waktu dalam pemrosesan data. Penerapan data reduction yaitu dengan menghilangkan atribut 'IPM'. Untuk proses klastering menggunakan atribut parameter penentuan Indeks Pembangunan Manusia (IPM).

E. Data Transformation

Data yang sudah melalui data selection dan data cleaning akan di transformasi, bertujuan untuk meratakan persebaran data sehingga data menjadi data normal dan lebih mudah untuk dikelompokkan. Dalam penelitian ini menggunakan feature untuk normalisasi dataset.

F. Data Mining

Tahap datamining dilakukan dengan pemodelan menggunakan algoritma K-Means. Proses clustering menggunakan algoritma K-Means ditunjukkan oleh flowchart pada Gambar 2.



Gambar 2. Flowchart algoritma K-Means

Berdasarkan konsep yang sudah dijelaskan, proses komputasi untuk K-Means dijelaskan sebagai berikut [9].

- 1) Tentukan K sebagai jumlah cluster yang akan dibentuk
- 2) Buat K centroid secara random
- 3) Hitung jarak setiap data ke masing – masing centroid
- 4) Setiap data memilih centroid yang terdekat
- 5) Tentukan posisi centroid yang baru dengan cara menghitung nilai rata-rata dari data-data yang terletak pada centroid yang sama
- 6) Jika posisi centroid baru dengan centroid yang lama masih belum memiliki kesamaan ma diulang lagi pada langkah ke 3 sampai menemukan kesamaan.
- 7) Tentukan jumlah klaster yang akan dibentuk dan k sebagai nilainya h. Hitung jarak setiap inputan terhadap centroid – centroid menggunakan persamaan Euclidean, dengan rumus i

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

G. Evaluasi

Tahap evaluasi merupakan tahap interpretasi terhadap pemodelan data mining berdasarkan domain pengetahuan pada pemahaman bisnis. Tahap evaluasi menggunakan *Elbow Method*. *Elbow Method* merupakan salah satu metode yang digunakan untuk menguji kualitas klaster yang dihasilkan dari proses clustering. Berikut merupakan persamaan yang digunakan pada *Elbow Method* [10].

$$SSE = \sum_{K=1}^K \sum_{X_i} |X_i - C_k|^2 \quad (2)$$

Dimana: k = cluster ke-c, X_i = jarak data objek ke-I dan C_k = pusat cluster ke-i

BAB IV HASIL DAN PEMBAHASAN

A. Pengumpulan data (*Selection Data*)

Dataset yang digunakan yaitu data parameter perhitungan IPM di indonesia tahun 2021. Dataset yang digunakan untuk clustering terdiri dari 6 atribut yaitu Daerah, PPK, RLS, UHH, HLS, dan IPM, untuk dataset sampel dapat dilihat pada tabel 2.

Tabel 2. Sample Dataset

No	Daerah	PPK	RLS	UHH	HLS	IPM
1	ACEH	9572.0	9.37	69.96	14.36	72.18
2	Simeulue	7148.0	9.48	65.28	13.9	66.41
3	Aceh Singkil	8776.0	8.68	67.43	14.32	69.22

.....						
546	Deiyai	4673.0	3.25	65.36	9.82	49.96
547	Kota Jayapura	14937.0	11.57	70.52	15.02	80.11
548	Indonesia	111156.0	8.54	71.57	13.08	72.29

Pada tabel 2 merupakan sampel dataset, kemudian diubah menjadi dataframe. Setelah dataset diubah ke dalam bentuk dataframe maka data akan dianalisis mengenai strktur dari isi data tersebut dapat dilihat pada tabel 3.

Tabel 3. Struktur Dataset

Column	Non-Null Count	Dtype
Daerah	549 non-null	Object
PPK_2021	549 non-null	Float64
RLS_2021	549 non-null	Float64
UHH_2021	549 non-null	Float64
HLS_2021	549 non-null	Float64
IPM_2021	549 non-null	Float64

B. Preprocessing Data

a. Data Cleaning

Data cleaning berfungsi untuk memeriksa apakah dalam data tersebut terdapat data noise, data hilang. Pada tahap ini menghilangkan missing values (data hilang) pada dataset dapat dilihat pada tabel 4.

Tabel 4. Data Cleaning

Sebelum Data Cleaning		Sesudah Data Cleaning	
Atribut	Jumlah noise	Atribut	Jumlah noise
Daerah	1	Daerah	0
PPK	4	PPK	0
RLS	4	RLS	0
UHH	4	UHH	0
HLS	4	HLS	0
IPM	4	IPM	0

b. Data Reduction

Data reduction adalah transformasi informasi numeric atau alphabet berdasarkan penelitian yang dirubah, dikoreksi, dan disederhanakan. Karena data yang digunakan hanya 5 atribut parameter perhitungan IPM maka dilakukan data *reduction* dengan menghapus atribut yang tidak digunakan yaitu “IPM” dan menghapus data “Indonesia” karna bukan merupakan bagian dari daerah kabupaten/kota dan provinsi di Indonesia.

c. Data Transformation

Tujuan data transformation berfungsi untuk meratakan persebaran data sehingga data menjadi data normal dan lebihh mudah untuk dikelompokan. Pada data *transformation* dilakukan *normalisasi* hasilnya dapat dilihat pada tabel 5.

Tabel 5. Hasil *Normalisasi*

No	Daerah	PPK	RLS	UHH	HLS
1	ACEH	0.28103656	0.69675723	0.65156951	0.75305097
2	Simeulue	0.15930092	0.7063979	0.44170404	0.72002872
3	Aceh Singkil	0.24106067	0.63628396	0.53811659	0.75017947
.....					
546	Deiyai	0.03500402	0.16038563	0.44529148	0.42713568
547	Kota Jayapura	0.55047208	0.88957055	0.67668161	0.80043073

C. Implementasi Data Mining (Metode Clustering K-Means)

Setelah data dilakukan proses *preprocessing* selanjutnya dilakukan implementasi data mining. Tahap implementasi data mining menggunakan Metode Clustering dengan algoritma K-Means.

a. Menentukan Jumlah Cluster

Jumlah *cluster* yang digunakan terdiri dari 3 *cluster* yaitu: cluster daerah berpotensi maju rendah (cluster 0), cluster daerah berpotensi maju sedang (cluster 1), dan cluster daerah berpotensi maju tinggi (cluster 2).

b. Menginisiasi Pusat Cluster (Centroid)

Menginisiasi atau menentukan pusat cluster (centroid) awal dengan menggunakan atribut random.

c. Hitung Jarak Antar Objek dan Pusat Cluster

Hasil perhitungan jarak antar objek dan pusat cluster (centroid) dapat dilihat pada tabel 5.

Tabel 6. Perhitungan Jarak Antar Objek dan Pusat *Cluster Iterasi 1*

No	Daerah	Cluster 0	Cluster 1	Cluster 2
1	Aceh	0,148986	0,319018	0,288029
2	Simeulue	0,310029	0,228486	0,502731
3	Aceh Singkil	0,191152	0,21218	0,404517
.....				
546	Deiyai	0,611772	0,426368	0,934519
547	Kota Jayapura	0,403037	0,596114	0,146803

d. Klastering Berdasarkan Jarak Minimum

Hasil klastering berdasarkan jarak minimum dari perhitungan jarak antar objek dan pusat cluster dapat dilihat pada tabel 6.

Tabel 6. Perhitungan Jarak dan Hasil *Clustering Iterasi 1*

No	Daerah	Cluster 0	Cluster 1	Cluster 2	Cluster
1	Aceh	0,148986	0,319018	0,288029	0
2	Simeulue	0,310029	0,228486	0,502731	1
3	Aceh Singkil	0,191152	0,21218	0,404517	0
.....					
546	Deiyai	0,611772	0,426368	0,934519	1
547	Kota Jayapura	0,403037	0,596114	0,146803	2

e. Jika Tidak Ada Perubahan Centroid Selesai

Setelah semua data dimasukkan kedalam kelas yang sama berdasarkan *cluster* terdekat, lalu hitung kembali pusat *cluster* (centroid) untuk menentukan iterasi kedua hasil perhitungan pada centroid baru dapat dilihat pada tabel 7.

Tabel 7. Centroid dari Perhitungan *Iterasi 1*

Centroid	PPK	RLS	UHH	HLS
Cluster 0	0.31506629	0.60018391	0.67700213	0.64785755
Cluster 1	0.19622077	0.51127007	0.4539664	0.60770149
Cluster 2	0.50852772	0.81170628	0.78564583	0.75735118

Setelah *centroid* baru dari setiap *cluster* didapatkan, lakukan *iterasi* kembali hingga setiap *cluster* tidak berubah lagi dengan dengan melihat jumlah *iterasi* atau perubahan posisi *centroid* dalam *iterasi* berurutan dan tidak ada data yang berpindah.

Tabel 8. Centroid dari Perhitungan *Iterasi 2*

Centroid	PPK	RLS	UHH	HLS
<i>Cluster 0</i>	0.50852772	0.81170628	0.67700213	0.75735118
<i>Cluster 1</i>	0.19622077	0.51127007	0.4539664	0.60770149
<i>Cluster 2</i>	0.31506629	0.60018391	0.67700213	0.64785755

Tabel 9. Centroid dari Perhitungan *Iterasi 64*

Centroid	PPK	RLS	UHH	HLS
<i>Cluster 0</i>	0.19622077	0.51127007	0.4539664	0.60770149
<i>Cluster 1</i>	0.31506629	0.60018391	0.67700213	0.64785755
<i>Cluster 2</i>	0.50852772	0.81170628	0.78564583	0.75735118

Tabel 10. Centroid dari Perhitungan *Iterasi 65*

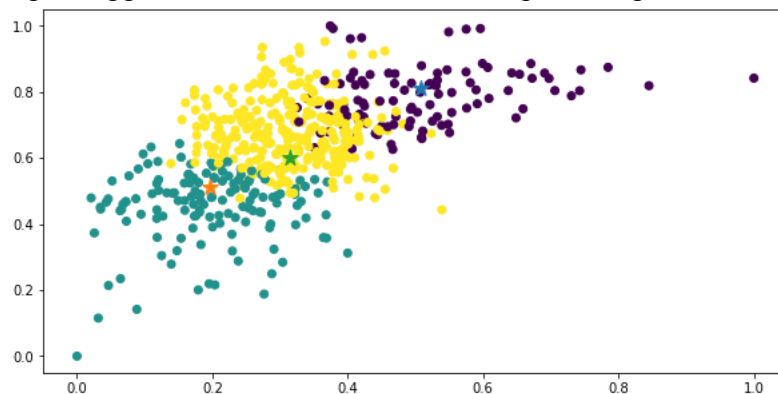
Centroid	PPK	RLS	UHH	HLS
<i>Cluster 0</i>	0.19622077	0.51127007	0.4539664	0.60770149
<i>Cluster 1</i>	0.31506629	0.60018391	0.67700213	0.64785755
<i>Cluster 2</i>	0.50852772	0.81170628	0.78564583	0.75735118

Berdasarkan hasil perhitungan diatas, dapat dilihat titik *centroid* tidak mengalami perubah pada *iterasi* ke-65, Dari hasil tersebut dapat disimpulkan pemodelan menggunakan K-Means selesai dengan 65 *iterasi*, hasil perhitungan jarak dan hasil clustering pada *iterasi* ke-65 dapat dilihat pada tabel 11.

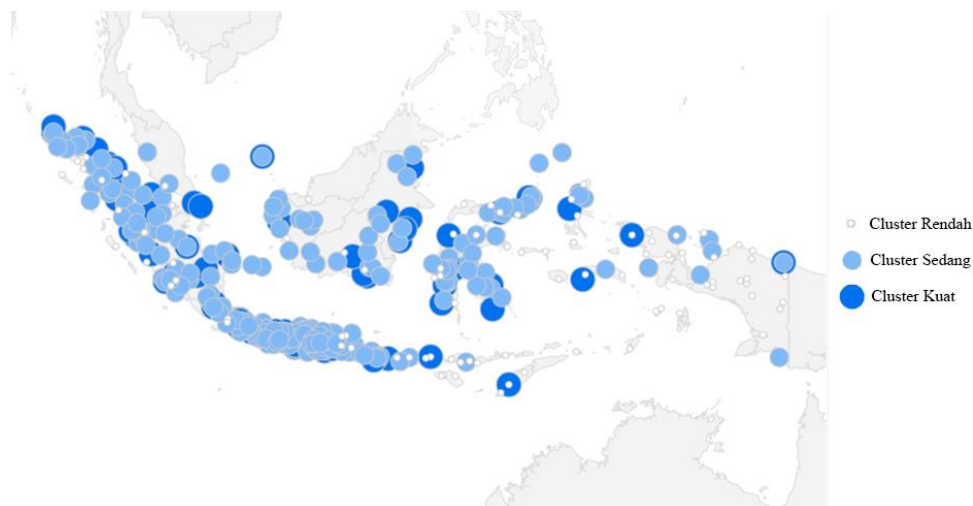
Tabel 11. Perhitungan Jarak dan Hasil Clustering *Iterasi 65*

No	Daerah	Cluster 0	Cluster 1	Cluster 2	Cluster
1	Aceh	0,319018	0,148986	0,288029	1
2	Simeulue	0,228486	0,310029	0,502731	0
3	Aceh Singkil	0,21218	0,191152	0,404517	1
.....					
546	Deiyai	0,426368	0,611772	0,934519	0
547	Kota Jayapura	0,596114	0,403037	0,146803	2

Visualisasi hasil klastering dari implementasi data mining menggunakan algoritma K-Means dapat dilihat pada gambar 3. Pada gambar 3 menunjukkan titik persebaran setiap data yang dihasilkan dari proses clustering dengan 3 cluster. Dimana titik berwarna hijau menunjukkan cluster rendah, titik berwarna kuning menunjukkan cluster sedang dan titik berwarna ungun menunjukkan cluster tinggi sedangkan untuk bentuk bintang menggambarkan centroid dari masing- masing cluster.

Gambar 3. Visualisasi Hasil *Clustering* Algoritma K-Means

Dari hasil pemodelan k-means untuk analisis clustering daerah berpotensi maju dapat divisualisasikan dalam bentuk diagram geografis dapat dilihat pada gambar 4.



Gambar 4. Visualisasi Hasil *Clustering* pada Diagram Geografis Indonesia

Pada kelompok cluster pada gambar diatas, terlihat bahwa cluster rendah disimbolkan dengan lingkaran putih, cluster cluster sedang lingkaran biru muda, serta cluster kuat lingkaran biru tua. Hasil Clustering dari 547 kabupaten/kota dan provinsi dapat dilihat pada tabel 12.

Tabel 12. Hasil *Clustering*

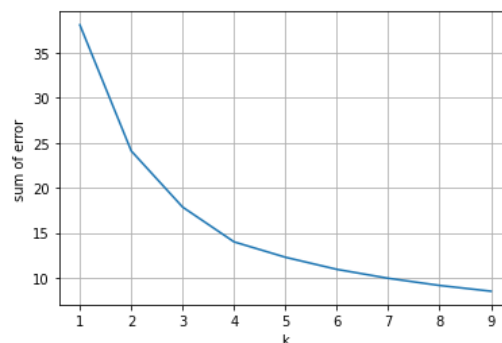
Kelompok Cluster	Data Cluster	Jumlah Data
Rendah	Simeulue, Aceh Selatan, Aceh Barat Daya, Gayo Lues, Kota Subulussalam, Nias, Mandailing Natal, Tapanuli Selatan, Nias Selatan, Pakpak Bharat, Nias Utara, Nias Barat, Kota Tanjung Balai, Kepulauan Mentawai, Sijunjung, Pasaman, Indragiri Hilir, Kepulauan Meranti, Tanjung Jabung Timur, Lahat, Ogan Komering Ulu Selatan, Ogan Ilir, Empat Lawang, Penukal, Abab Lematang Ilir, Musi Rawas Utara, Kaur, Seluma, Mukomuko, Lebong, Mesuji, Pesisir Barat, Lingga, Bondowoso, Probolinggo, Sampang, Pamekasan, Pandeglang, Lebak, Serang, Nusa Tenggara Barat, Lombok Barat, Lombok Tengah, Lombok Timur, Dompu, Bima, Lombok Utara, Nusa Tenggara Timur, Sumba Barat, Sumba Timur, Kupang, Timor Tengah Selatan, Timor Tengah Utara, Belu, Alor, Lembata, Flores Timur, Sikka, Ende, Manggarai, Rote Ndao, Manggarai Barat, Sumba Tengah, Sumba Barat Daya, Nagekeo, Manggarai Timur, Sabu Raijua, Malaka, Kayong Utara, Katingan, Barito Kuala, Hulu Sungai Utara, Bolaang Mongondow Selatan, Bolaang Mongondow Timur, Banggai Kepulauan, Donggala, Toli-Toli, Parigi Moutong, Tojo Una-Una, Banggai Laut, Jeneponto, Takalar, Sinjai, Bone, Buton, Buton Tengah, Buton Selatan, Gorontalo, Pohuwato, Gorontalo Utara, Sulawesi Barat, Majene, Polewali Mandar, Mamuju Utara, Mamuju Tengah, Maluku Tenggara Barat, Maluku Tenggara, Kepulauan Aru, Seram Bagian Barat, Seram Bagian Timur, Maluku Barat Daya, Buru Selatan, Kota Tual, Halmahera Barat, Halmahera Tengah, Kepulauan Sula, Halmahera Selatan, Pulau Morotai, Pulau Taliabu, Papua Barat, Kaimana, Teluk Wondama, Teluk Bintuni, Sorong Selatan, Sorong, Raja Ampat, Tambora, Maybrat, Manokwari Selatan, Pegunungan Arfak, Papua, Jayawijaya, Paniai, Puncak Jaya, Boven Digoel, Mappi, Asmat, Yahukimo, Pegunungan Bintang, Tolikara, Sarmi, Keerom, Waropen, Supiori, Mamberamo Raya, Nduga, Lanny Jaya, Mamberamo Tengah, Yalimo, Puncak, Dogiyai, Intan Jaya, Deiyai	142
Sedang	...	303
Kuat	Kota Banda Aceh, Kota Sabang, Kota Langsa, Kota Lhokseumawe, Toba Samosir, Karo, Deli Serdang, Kota Pematang Siantar, Kota Tebing Tinggi, Kota Medan, Kota Binjai, Kota Padangsidimpuan, Kota Padang, Kota Solok, Kota Padang Panjang, Kota Bukittinggi, Kota Payakumbuh, Kota Pariaman, Kota Pekanbaru, Kota Dumai, Kota Jambi, Kota Sungai Penuh, Kota Palembang, Kota Prabumulih, Kota Lubuklinggau, Kota Bengkulu, Kota Bandar Lampung, Kota Metro, Kota Pangkal Pinang, Kepulauan Riau, Kota Batam, Kota Tanjung Pinang, DKI Jakarta, Kota Jakarta Selatan, Kota Jakarta Timur, Kota Jakarta Pusat, Kota Jakarta Barat, Kota Jakarta Utara, Kota Bogor, Kota Bandung, Kota Cirebon, Kota Bekasi, Kota Depok, Kota Cimahi, Klaten, Sukoharjo, Karanganyar, Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Semarang, Kota Pekalongan, Kota Tegal, D I Yogyakarta, Bantul, Sleman, Kota Yogyakarta, Sidoarjo, Gresik, Kota Kediri, Kota Blitar, Kota Malang, Kota Pasuruan, Kota Mojokerto, Kota Madiun, Kota Surabaya, Kota Batu, Kota Tangerang, Kota, Tangerang Selatan, Bali, Tabanan, Badung, Gianyar, Kota Denpasar, Kota Mataram, Kota Bima, Kota Kupang, Kota Pontianak, Kota Palangka Raya, Kota Banjarmasin, Kota Banjar Baru, Kalimantan Timur, Berau, Kota Balikpapan, Kota Samarinda, Kota Bontang, Kota Tarakan, Minahasa, Kota Manado, Kota Tomohon, Kota Palu, Kota Makassar, Kota Parepare, Kota Palopo, Kota Kendari, Kota Baubau, Kota Gorontalo, Kota Ambon, Kota Ternate, Kota Sorong, Mimika, Kota Jayapura	102

D. Evaluation (Metode Elbow)

Pada evaluasi model menggunakan metode evaluasi *Elbow* dengan hasil seperti pada tabel 13 dan gambar 5.

Tabel 13. Hasil Evaluasi *Elbow*

K	Nilai Tiap K	Selisih Antar Jarak K
1	38,10363258125401	38,10363258
2	24,10023592078211	14,00339666
3	17,86192617719919	6,238309744
4	14,00184541062832	3,860080767
5	12,29075380879772	1,711091602
6	10,94723368350315	1,343520125
7	9,95590851216456	0,991325171
8	9,157788103832317	0,798120408
9	8,518608810741629	0,639179293



Gambar 5. Grafik Hasil Evaluasi Metode *Elbow*

Dari hasil analisis evaluasi nilai K dari range 1-10, didapat bahwa dari nilai K 1-3 terdapat selisih yang cukup drastis, namun dari nilai K 3-9 selisih cenderung landai sebagaimana yang bisa dilihat pada visualisasi diatas. hal tersebut dapat diartikan bahwa nilai K = 3 merupakan nilai K yang paling optimal untuk dijadikan *cluster* karena cenderung membentuk siku dan Nilai *Sum of Square Error* sebesar 6,238309744. Maka Implementasi Data Mining menggunakan metode Klastering dengan algoritma K-Means dapat dikatakan optimal.

BAB V PENUTUP

Kesimpulan

Berdasarkan hasil analisis clustering daerah berpotensi maju di indonesia berdasarkan parameter penilaian Indeks Pembangunan Manusia (IPM) menggunakan algoritma K-Means didapatkan hasil sebagai berikut:

1. *Clustering* daerah berpotensi maju di Indonesia berhasil dilakukan dengan menerapkan data mining menggunakan clustering dengan algoritma K-Means
2. Hasil *clustering* daerah berpotensi maju di Indonesia dibagi menjadi 3 macam cluster yaitu *cluster* 0 untuk daerah berpotensi maju rendah, *cluster* 1 untuk daerah berpotensi maju sedang dan *cluster* 2 untuk daerah berpotensi maju rendah.
3. Terdapat 547 daerah kabupaten/kota dan provinsi di indonesia yang dilakukan proses *clustering*, hasil dari *clustering* yaitu pada *cluster* rendah terdiri atas 142 daerah atau sebesar 25,96%, *cluster* sedang terdiri dari 303 daerah atau sebesar 55,39%, dan *cluster* tinggi terdiri 102 daerah atau sebesar 18,65%.
4. Evaluasi model menggunakan metode *Elbow*, didapatkan hasil bahwa pemodelan *clustering* menggunakan algoritma K-Means sudah optimal.
5. Hasil analisis *clustering* untuk daerah berpotensi maju di indonesia didominasi oleh *cluster* daerah berpotensi sedang namun masih ada *cluster* daerah berpotensi rendah yang jumlahnya masih lebih banyak dibandingkan *cluster* daerah berpotensi tinggi. Hal ini dapat dijadikan masukan bagi pemerintahan yang berwenang sebagai bahan informasi untuk memperhatikan daerah-daerah yang masih dikategorikan berpotensi rendah sehingga terjadi pemerataan Indeks Pembangunan Manusia di setiap kabupaten/kota dan provinsi di Indonesia.

DAFTAR PUSTAKA

- [1] B. P. Statistik, *Indeks Pembangunan Manusia 2015*. Badan Pusat Statistik, 2016. [Online]. Available: <https://www.bps.go.id/publication.html?Publikasi%5BtahunJudul%5D=2015&Publikasi%5BkataKunci%5D=IPM&Publikasi%5BcekJudul%5D=0&yt0=Tampilkan>
- [2] N. Cahyana and A. Aribowo, "Metode Data Mining K-Means Untuk Klasterisasi Data Penanganan Dan Pelayanan Kesehatan Masyarakat," *Semin. Nas. Inform. Medis*, no. 5, pp. 24–31, 2018.
- [3] W. Yustanti, N. Rahmawati, and Y. Yamasari, "Klastering Wilayah Kota/Kabupaten Berdasarkan Data Persebaran Covid-19 Di Propinsi Jawa Timur dengan Metode K-Means," *J. Inf. Eng. Educ. Technol.*, vol. 4, no. 1, pp. 1–9, 2020, doi: 10.26740/jieet.v4n1.p1-9.
- [4] R. Nainggolan and G. Lumbantoruan, "Optimasi Performa Cluster K-Means Menggunakan Sum of Squared Error (Sse)," *METHOMIKA J. Manaj. Inform. Komputerisasi Akunt.*, vol. 2, no. 2, pp. 103–108, 2018, [Online]. Available: <https://ejurnal.methodist.ac.id/index.php/methomika/article/view/138>
- [5] G. Z. SHAFa, "ANALISIS K-MEANS CLUSTERING KOMPONEN INDEKS PEMBANGUNAN MANUSIA 2021 BERDASARKAN KABUPATEN/KOTA DI JAWA TENGAH," 2016. [Online]. Available: <https://dspace.uui.ac.id/handle/123456789/38880>
- [6] S. Suharni, "Data Mining Analisis Cluster K-Means Pada Indeks Pembangunan Teknologi, Informasi dan Telekomunikasi," *J. Technopreneursh. Inf. Syst.*, vol. 3, no. 3, pp. 7–11, 2021, doi: 10.36085/jtis.v3i3.1228.
- [7] P. N. Akbar *et al.*, "Klasterisasi Daerah Pelaku Pencurian Menggunakan Algoritma K-Means," vol. 2, no. 9, pp. 1–13, 2022.
- [8] A. Charolina and D. Ruswanti, "Klastering Kota Dan Kabupaten Di Indonesia Berdasarkan Umur Harapan Hidup Saat Lahir Dengan K-Medoids," *Gaung Inform.*, vol. 12, p. 11, 2019.
- [9] W. Saputro, M. Reza Pahlevi, and A. Wibowo, "Analisis Algoritma K-Means Untuk Klasterisasi Tindak Pidana Korupsi Di Wilayah Hukum Indonesia," *JIKO (Jurnal Inform. dan Komputer)*, vol. 3, no. 3, pp. 137–142, 2020, doi: 10.33387/jiko.v3i3.1960.
- [10] E. Virantika, Kusnawi, and J. Ipmawati, "Evaluasi Hasil Pengujian Tingkat Clusterisasi Penerapan Metode K-Means Dalam Menentukan Tingkat Penyebaran Covid-19 di Indonesia," *J. Media Inform. Budidarma*, vol. 6, no. 3, pp. 1657–1666, 2022, doi: 10.30865/mib.v6i3.4325.