

PERBANDINGAN METODE ALGORITMA *K-MEANS* DAN *K-MEDOIDS* PADA PENGELOMPOKAN KABUPATEN/KOTA DI PROVINSI NUSA TENGGARA TIMUR BERDASARKAN DIMENSI INDEKS PEMBANGUNAN MANUSIA TAHUN 2020

Maria Khoncita Dasriana Bau¹, Yudi Setyawan², Maria Titah Jatipaningrum^{3*}

^{1,2,3}Jurusan Statistika, Fakultas Sains Terapan, IST AKPRIND Yogyakarta

Email : dasribau@gmail.com

*corresponding author

ABSTRACT. *Humans are the real wealth of the nation. It is appropriate if humans become the main goal in development. To find out the success of the development is to look at the of the human development index which can describe the development of human development in a measurable and representative manner. The growth of the development index in NTT in 2020 decreased by 0.04 points from the previous year, namely 2019. To find out the decline in the growth of the Human Development Index (HDI), research is needed on the similarity of characteristics in districts/cities in NTT to make it easier to see areas that affect the decline. growth in the Human Development Index (IPM). The clustering process can be done by implementing the clustering method, where the clustering method is grouping objects into a group that has a high similarity compared to objects in another group. Several methods in clustering include the K-Means and K-Medoids methods. The results of the analysis can be concluded that the best method in the case study of the Human Development Index indicator in NTT in 2020 is the K-Means clustering method with $k = 2$ with a Silhouette Index value of 0.18, the Davies-Bouldin Index value of 1.15 and the Dunn index value. of 0.24 where cluster 1 has 14 district/city members and cluster 2 has 8 district members.*

Keywords: *Human Development Index, K-Means, K-Medoids, Silhouette Index, Davies-Bouldin Index, Dunn Index*

ABSTRAK. Manusia merupakan kekayaan bangsa yang sesungguhnya. Sudah sepantasnya apabila manusia menjadi tujuan utama dalam pembangunan. Untuk mengetahui keberhasilan pembangunan tersebut adalah dengan melihat indeks pembangunan manusia yang dapat menggambarkan perkembangan pembangunan manusia secara terukur dan representatif. Pertumbuhan indeks pembangunan di NTT pada tahun 2020 mengalami penurunan sebesar 0,04 poin dari tahun sebelumnya yaitu tahun 2019. Untuk mengetahui penurunan pertumbuhan Indeks Pembangunan Manusia (IPM) diperlukan penelitian mengenai kesamaan karakteristik pada kabupaten/kota di NTT untuk mempermudah melihat wilayah yang mempengaruhi penurunan pertumbuhan Indeks Pembangunan Manusia (IPM) tersebut. Proses pengelompokan dapat dilakukan dengan mengimplementasikan metode *clustering*, yang dimana metode *clustering* merupakan pengelompokan objek kedalam suatu kelompok yang memiliki kesamaan tinggi dibandingkan dengan objek yang berada dalam satu kelompok lain. Beberapa metode yang ada dalam *clustering* antara lain metode *K-Means* dan *K-Medoids*. Hasil analisis dapat disimpulkan bahwa metode terbaik dalam studi kasus indikator Indeks Pembangunan Manusia di NTT tahun 2020 adalah metode *K-Means clustering* dengan $k = 2$ dengan nilai *Silhouette Index* sebesar 0,18, nilai *Davies-Bouldin Index* sebesar 1,15 dan nilai *Dunn index* sebesar 0,24 dimana *cluster 1* memiliki 14 anggota kabupaten/kota dan *cluster 2* memiliki 8 anggota kabupaten.

Kata Kunci : *Indeks Pembangunan Manusia, K-Means, K-Medoids, Silhouette Index, Davies-Bouldin Index, Dunn Index*

1. Pendahuluan

Manusia merupakan kekayaan bangsa yang sesungguhnya. Sudah sepantasnya apabila manusia menjadi tujuan utama dalam pembangunan. Pentingnya pembangunan manusia menjadi suatu kebutuhan dalam pembangunan sebuah wilayah. Menurut (UNDP, 2004) pembangunan manusia dirumuskan sebagai upaya perluasan pilihan bagi penduduk dan

sekaligus sebagai taraf yang dicapai dari upaya tersebut.

Menurut Badan Pusat Statistik (BPS, 2016) Indeks Pembangunan Manusia (IPM) merupakan indikator penting untuk mengukur keberhasilan dalam upaya membangun kualitas hidup manusia, IPM dapat menentukan peringkat atau level pembangunan suatu wilayah/Negara dan bagi Indonesia, IPM merupakan data strategis karena sebagai ukuran kinerja pemerintah. IPM mengukur capaian pembangunan manusia berbasis sejumlah komponen dasar kualitas hidup, pada dimensi pengetahuan, variabel harapan lama sekolah dan rata-rata lama sekolah merupakan refleksi dari kemampuan masyarakat untuk mengakses pendidikan, khususnya pendidikan berkualitas baik yang sangat diperlukan untuk kehidupan produktif dalam masyarakat modern. Dimensi ketiga dari pembangunan manusia adalah penguasaan sumber daya yang dibutuhkan untuk kehidupan yang layak. UNDP menggunakan data Pendapatan Nasional Bruto (PNB) sebagai indikator dimensi ini. Akan tetapi, mengingat tidak semua wilayah di Indonesia memiliki indikator tersebut, maka dipilih alternatif lain dengan memanfaatkan indikator pendapatan/pengeluaran (Nugroho & dkk, 2019).

Berdasarkan catatan (BPS N. , 2021) provinsi Nusa Tenggara Timur diketahui bahwa IPM di provinsi NTT pada tahun 2020 adalah sebesar 65,19 mengalami penurunan 0,04 poin dari capaian tahun sebelumnya. Penurunan pertumbuhan IPM tahun 2020 sangat dipengaruhi oleh turunnya rata-rata pengeluaran per kapita. Indikator ini turun dari 7.769.000 rupiah pada tahun 2019 menjadi 7.598.000 rupiah pada tahun 2020.

Penelitian ini menggunakan analisis multivariat dimana analisis multivariat merupakan metode statistik yang memungkinkan melakukan penelitian terhadap lebih dari dua variabel. Salah satu teknik analisis multivariat adalah analisis *cluster*. Analisis *cluster* merupakan cabang ilmu statistik yang digunakan untuk mengelompokkan obyek-obyek tertentu berdasarkan kesamaan karakter menjadi beberapa kelas.

Secara umum penelitian ini dilakukan untuk mengetahui penurunan pertumbuhan Indeks Pembangunan Manusia (IPM) diperlukan penelitian mengenai kesamaan karakteristik pada kabupaten/kota di NTT untuk mempermudah melihat wilayah yang mempengaruhi penurunan pertumbuhan Indeks Pembangunan Manusia (IPM) tersebut. Proses pengelompokan dapat dilakukan dengan mengimplementasikan metode *clustering*, yang dimana metode *clustering* merupakan pengelompokan objek ke dalam suatu kelompok yang memiliki kesamaan tinggi dibandingkan dengan objek yang berada dalam satu kelompok lain. Beberapa metode yang ada dalam *clustering* antara lain metode *K-Means* dan *K-Medoids*. Dari penelitian ini diharapkan dapat memberikan gambaran dan hasil pengelompokan kabupaten/kota berdasarkan dimensi Indeks Pembangunan Manusia agar pemerintah di provinsi NTT dapat melihat kabupaten/kota mana yang harus lebih diperhatikan.

2. Metode Penelitian

Data yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari publikasi Indeks Pembangunan Manusia pada website resmi BPS Provinsi NTT. Data <https://ntt.bps.go.id>. Data yang diambil merupakan dimensi Indeks Pembangunan Manusia untuk 22 kabupaten/kota di provinsi NTT tahun 2020.

1) Analisis Deskriptif

Statistik deskriptif adalah prosedur dan metode untuk pengumpulan, penyajian, peringkasan, dan penganalisisan data. Tujuan utama dari statistik deskriptif adalah mengumpulkan, menyajikan, dan meringkas sekumpulan data sehingga dapat memberikan informasi yang berguna. Statistik deskriptif hanya mengenai data yang dipunyai dan sama sekali tidak menarik kesimpulan yang lebih banyak dan lebih jauh dari data yang ada. Kegiatan memeriksa sifat-sifat penting dari data yang ada disebut analisis data secara (deskripsi).

2) Standarisasi Data

Standarisasi data dilakukan untuk mengurangi variasi data antar variabel karena setiap variabel memiliki satuan yang berbeda-beda dan akan mempengaruhi hasil keputusan analisis

karena perbedaan satuan membuat rentang nilai yang besar antar variabel. Oleh karena itu sebelum melakukan analisis dilakukan standarisasi data dengan persamaan sebagai berikut :

$$Z_{ir} = \frac{x_{ir} - \bar{x}_r}{s_r} \quad (1)$$

Dimana:

x_{ir} : Nilai dari data ke- i dan variabel ke- r , dengan $i = 1, 2, \dots, n$

s_r : standar deviasi dari variabel ke- r , dengan $r = 1, 2, \dots, p$

\bar{x}_r : Rata-rata variabel ke- r

Data dinyatakan *outlier* jika nilai $Z_{ir} > 2,5$ atau nilai $Z_{ir} < -2,5$. *Outliers* adalah data pengamatan yang nilainya “ekstrim” yang muncul karena memiliki karakteristik yang unik atau terlihat sangat jauh berbeda dari observasi-observasi lainnya.

3) Analisis Cluster

Analisis *cluster* adalah suatu teknik statistik yang bertujuan untuk mengelompokkan objek ke dalam suatu kelompok sedemikian sehingga objek yang berada dalam satu akan memiliki kesamaan tinggi dibandingkan dengan objek yang berada dalam satu kelompok lain.

Terdapat dua asumsi dari analisis *cluster* menurut (Kurniawan M, 2018), yaitu:

1. Sampel Representatif
Sampel yang diambil benar-benar dapat mewakili populasi yang ada atau bagian dari populasi yang berupaya untuk secara akurat mencerminkan karakteristik kelompok yang lebih besar.
2. Tidak Terjadi Multikolinieritas
Uji multikolinieritas dilakukan untuk mengetahui adanya hubungan linear yang sempurna atau pasti diantara beberapa atau semua variabel (Arniva & dkk, 2016). Koefisien korelasi biasanya dilambangkan dengan huruf r yang dimana nilai r bervariasi dari -1 hingga +1, dan nilai r mendekati nilai 0 yang mengidentifikasi lemahnya hubungan antar variabel. Jika nilai r positif (+) maka hubungan antar variabel berbanding lurus, tapi jika nilai r negatif (-) yang artinya hubungan antar variabel berbanding terbalik.

4) Jarak Euclidean

Dalam menjelaskan atau mengukur kemiripan/kesamaan antar objek digunakan pendekatan ukuran jarak. Menurut (Ulinuh & dkk, 2020), jarak *Euclidean* adalah besar suatu jarak pada garis lurus yang menghubungkan antar objek.

$$d_{i,j} = \sqrt{\sum_{r=1}^p (x_{ir} - x_{jr})^2} \quad (2)$$

Dimana:

$d_{i,j}$: Jarak antara data i dengan data j , dengan $i, j = 1, 2, \dots, n$

x_{ir} : Data ke - i pada variabel ke - r

x_{jr} : Data ke - j pada variabel ke - r , dengan $r = 1, 2, \dots, p$

n : Jumlah data

p : Jumlah variabel

5) Metode Elbow

Dalam menentukan jumlah *cluster* optimal, dapat digunakan metode *Elbow*. Menurut (Rahman, 2017) metode *Elbow* merupakan suatu metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah *cluster* terbaik dengan cara melihat persentase hasil perbandingan antara jumlah *cluster* yang akan membentuk siku pada suatu titik. Pada metode *Elbow* nilai *cluster* terbaik yang akan diambil dari nilai *Sum of Square Within (SSW)* yang mengalami penurunan yang signifikan. Berikut merupakan perhitungan nilai *SSW* seperti dibawah ini :

$$SSW = \sum_{k=1}^K \sum_{x_i \in S_k} \|x_i - c_k\|^2 \quad (3)$$

Dimana :

K : Jumlah *Cluster*

x_i : Data ke- i , dengan $i = 1, 2, \dots, n$

S_k : Kelompok data dalam *cluster* ke- k dengan $k = 1, 2, \dots, K$

c_k : Pusat *cluster* ke- k

6) Metode K-Means

Algoritma *K-means* merupakan salah satu algoritma dengan *partitional*, karena *K-Means* didasarkan pada penentuan jumlah awal kelompok dengan mendefinisikan nilai *centroid* awalnya. Adapun tujuan pengklasteran data ini adalah untuk meminimalkan fungsi objektif yang diset dalam proses pengklasteran, yang pada umumnya berusaha meminimalkan variasi di dalam suatu klaster dan memaksimalkan variasi antar klaster. Adapun langkah-langkah untuk *K-Means Clustering* adalah sebagai berikut (Prasetyo, 2012) :

1. Inisialisasi : tentukan nilai k sebagai jumlah klaster yang diinginkan dan matriks jarak yang diinginkan.
2. Pilih k data dari set data X sebagai *centroid*.
3. Alokasikan semua data ke *centroid* terdekat dengan matriks jarak yang sudah ditetapkan.
4. Hitung kembali *centroid* berdasarkan data yang mengikuti klaster masing-masing.
5. Ulangi langkah 3 dan 4 hingga kondisi konvergen tercapai, yaitu tidak ada data yang berpindah klaster.

7) Metode K-Medoids

K-Medoids Clustering dikenal sebagai *Partitioning Around Medoids (PAM)*, adalah varian dari metode *K-Means*. Hal ini didasarkan pada penggunaan *medoids* bukan dari pengamatan mean yang dimiliki oleh setiap klaster, dengan tujuan mengurangi sensitivitas dari partisi sehubungan dengan nilai ekstrim yang ada dalam dataset. Menurut (Han & Kamber, 2006) tahapan *K-Medoids Clustering* adalah sebagai berikut:

1. Secara acak pilih k objek pada sekumpulan n objek sebagai *medoids*.
2. Tempatkan objek-objek *non medoids* ke dalam klaster yang paling dekat dengan *medoids*.
3. Hitung total cost yang merupakan total jarak semua objek ke pusat klasternya.
4. Melakukan percobaan perhitungan pada beberapa O_{random} (k objek *non medoids* sebagai *medoids* baru).
5. Hitung total *cost* yang baru dari pertukaran *medoids* O_j dengan O_{random} .
6. Hitung nilai S yang merupakan selisih total cost baru-total cost lama.
7. Jika $S < 0$ maka tukar O_j dengan O_{random} , untuk membentuk sekumpulan k objek baru sebagai *medoids*.
8. Ulangi langkah 4 sampai 7 hingga tidak ada perubahan *cluster*.

8) Silhouette Index

Menurut (Trayasiwi, 2015) untuk mengetahui seberapa baik suatu objek ditempatkan dalam suatu *cluster* yang terbentuk dengan melihat nilai *silhouette indeks* maka dilakukan validasi. Rentang nilai SI adalah -1 hingga +1. Hasil *cluster* yang baik ditunjukkan dengan nilai *silhouette* yang mendekati rentang nilai 1. Berikut merupakan perhitungan *silhouette index*:

$$SI = \frac{\sum_{i=1}^n SC_i^j}{n} \quad (4)$$

Dimana :

SC_i^j : *silhouette coefficient*

n : jumlah data

9) Davies Bouldin Index

Davies-Bouldin Index merupakan salah satu metode evaluasi internal yang mengukur evaluasi *cluster* pada suatu metode pengelompokan yang didasarkan pada nilai kohesi dan separasi (Muhammad, 2015).

Menurut (Kartikasari, 2021), *Davies-Bouldin Index* adalah rasio dari jumlah jarak *intra-cluster* dan jarak *inter-cluster* harus rendah. Oleh karena itu, nilai DBI yang lebih rendah diperlukan agar menunjukkan hasil *clustering* yang baik. Berikut merupakan perhitungan

Davies-Bouldin Index :

$$DBI = \frac{1}{K} \sum_{e=1}^K D_e \quad (5)$$

Dimana :

$$D_e: \max_{f \neq e} R_{e,f} , e, f=1, 2, \dots, K$$

Dengan :

$$R_{e,f} = \frac{SSW_e + SSW_f}{SSB_{ef}} \quad (6)$$

Dimana :

SSW_e : Sum of Square Within Cluster ke- e

SSW_f : Sum of Square Within Cluster ke- f

$SSB_{e,f}$: Sum of Square Between Cluster ke- e dan f

10) Dunn Index

Dunn Index (DI) menghitung validitas *cluster* menggunakan diameter *cluster* (kohesi) dan jarak antara dua *cluster* (separasi). Untuk mendapatkan diameter sebuah *cluster* ke- k dilakukan dengan menghitung jarak pasangan dua data dalam sebuah *cluster*, kemudian diambil yang terbesar. Jika nilai *Dunn* semakin besar maka hasil *clustering* semakin bagus. Perhitungan *DI* dinyatakan oleh persamaan berikut :

$$DI = \min_{1 \leq e \leq K} \left\{ \min_{\substack{1 \leq f \leq K \\ f \neq e}} \left\{ \frac{\delta(x_e, x_f)}{\max_{1 \leq k \leq K} \{\Delta(x_k)\}} \right\} \right\} \quad (7)$$

Dimana :

$\delta(x_e, x_f)$: Jarak data antar *cluster*

$\Delta(x_k)$: Jarak data di dalam *cluster* yang sama

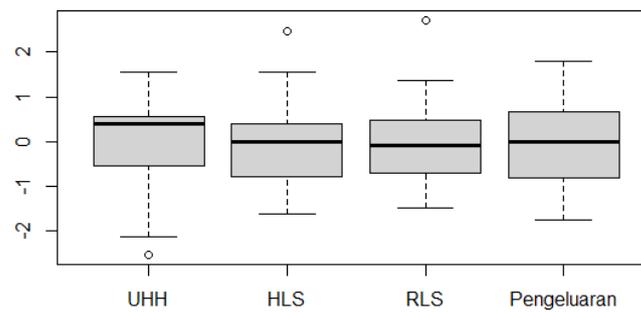
3. Analisis Dan Pembahasan

Analisis Deskriptif

Berdasarkan studi kasus dimensi Indeks Pembangunan Manusia (IPM) di NTT tahun 2020 memiliki karakteristik yang disajikan dalam analisis deskriptif. Pada variabel Umur Harapan Hidup dapat diketahui bahwa dari 22 kabupaten/kota di NTT nilai terendah untuk umur harapan hidup sebesar 60,64 tahun terdapat pada kabupaten Sabu Raijua. Nilai tertinggi untuk usia harapan hidup terletak pada kabupaten Kota Kupang sebesar 69,55 tahun. Pada variabel Harapan Lama Sekolah diketahui nilai terendah untuk harapan lama sekolah sebesar 11,99 tahun terdapat pada kabupaten Manggarai Timur. Jumlah tertinggi untuk harapan lama sekolah terletak pada kabupaten Kota Kupang sebesar 14,4 tahun. Pada variabel Rata-rata Lama Sekolah diketahui jumlah terendah untuk harapan lama sekolah sebesar 6,25 tahun terdapat pada kabupaten Sumba Tengah. Jumlah tertinggi untuk rata-rata lama sekolah terletak pada kabupaten Kota Kupang sebesar 9,58 tahun. Sedangkan Pada variabel Pengeluaran diketahui NTT jumlah terendah pengeluaran sebesar Rp5.265.000 terdapat pada kabupaten Sabu Raijua. Jumlah tertinggi untuk pengeluaran terletak pada kabupaten Sumba Timur sebesar Rp9.406.000.

Standarisasi Data

Dimensi Indeks Pembangunan Manusia (IPM) memiliki nilai skala yang berbeda-beda, sehingga perlu dilakukan standarisasi untuk mengurangi variasi data antar variabel. Berdasarkan hasil analisis menggunakan bantuan *software R-studio* diketahui bahwa terdapat beberapa data yang memiliki karakteristik berbeda dengan data yang lainnya dibuktikan dengan melihat boxplot pada Gambar 1 berikut :



Gambar 1 Boxplot Untuk Mendeteksi *Outlier*

Dengan: UHH : Usia Harapan Hidup
 HLS : Harapan Lama Sekolah
 RLS : Rata-rata Lama Sekolah

Dari Gambar 4.5 dapat diketahui bahwa terdapat data *outlier* pada variabel Umur Harapan Hidup, variabel Harapan Lama Sekolah dan variabel Rata-rata Lama Sekolah. Pada penelitian ini data *outlier* tetap dipertahankan agar tidak mempengaruhi hasil analisis.

Asumsi Cluster

Terdapat dua asumsi dalam analisis *cluster* sebagai suatu *cluster* yang baik menurut (Kurniawan M, 2018), yaitu:

- 1) Sampel Representatif :
 Pada penelitian ini tidak dilakukan uji asumsi sampel representatif karena pada asumsi dilakukan jika penelitian mengambil data sampel yang mewakili populasi, sedangkan pada penelitian ini data yang digunakan adalah data populasi.
- 2) Tidak Terjadi Multikolinearitas :
 Data dimensi Indeks Pembangunan Manusia di NTT memiliki empat variabel, sehingga perlu dilakukan uji multikolinearitas untuk mengetahui adanya hubungan linear yang sempurna atau pasti diantara beberapa atau semua variabel. Adanya hubungan linear antara variabel dapat dilihat dari nilai koefisien korelasi antar variabel. Nilai koefisien Pearson yang melebihi 0,8 menandakan bahwa terjadi multikolinieritas (Alwi & dkk, 2018). Berdasarkan hasil analisis menggunakan *software R-Studio* pada disajikan Tabel 1 dan Lampiran 3 mengenai nilai korelasi antar variabel :

Tabel 1 Nilai Korelasi Antar Variabel

Variabel	UHH	HLS	RLS	Pengeluaran
UHH	1,000	0,127	0,091	0,284
HLS	0,127	1,000	0,233	0,318
RLS	0,091	0,233	1,000	0,559
Pengeluaran	0,284	0,318	0,559	1,000

Data pada Tabel 1 menunjukkan bahwa tidak ada nilai korelasi antar variabel yang melebihi 0,8. Hal ini menunjukkan bahwa tidak terdapat multikolinearitas antar variabel-variabel indikator Indeks Pembangunan Manusia yang diteliti, sehingga dapat melanjutkan proses berikutnya.

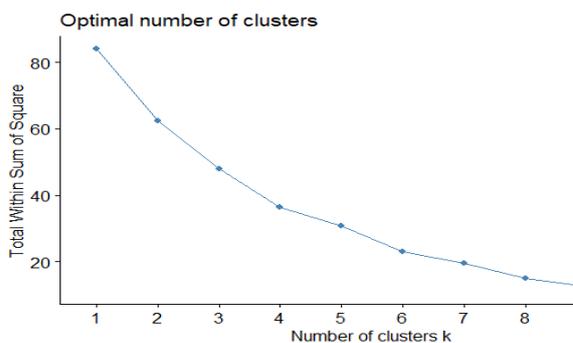
Jarak Euclidean

Berdasarkan data dimensi Indeks Pembangunan Manusia pada tahun 2020 maka akan dihitung jarak terdekat menggunakan nilai *Euclidean* untuk menjelaskan atau mengukur kemiripan/kesamaan antar objek. Berdasarkan hasil output *R-Studio* diperoleh bahwa jarak

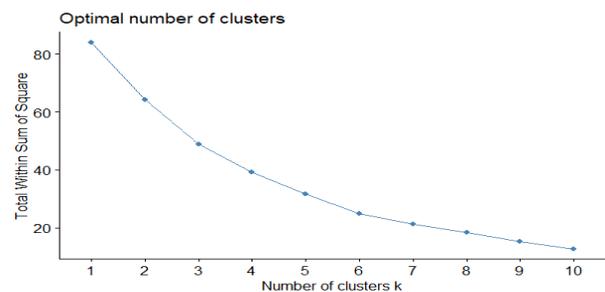
terdekat sebesar 0,312 antara kabupaten Sumba Barat Daya dan kabuapten Sumba Tengah sedangkan jarak terjauh sebesar 5,864 anatar kabupaten Kota Kupang dan kabupaten Alor.

Metode *Elbow*

Pada penelitian ini, penentuan jumlah *cluster* menggunakan metode *elbow* yang dimana yang dimana penentuan banyaknya *cluster* yang terbentuk bertujuan untuk mencari nilai *cluster* optimal dengan cara melihat hasil grafik yang membentuk siku. Berdasarkan hasil output *software R-Studio*, berikut adalah grafik metode *elbow* untuk menentukan jumlah *cluster* optimal untuk metode *K-Means* dan *K-Medoids* yang disajikan pada Gambar 2 dan 3:



Gambar 2 Grafik Metode *K-Means*



Gambar 3 Grafik Metode *K-Medoids*

Untuk penentuan nilai *k* dengan metode *elbow* dengan grafik yang membentuk siku. Dari Gambar 2 dapat dilihat grafik membentuk siku pada angka 2, jadi nilai *k* optimum pada metode *elbow* untuk metode *K-Means* adalah 2 dan juga untuk metode *K-Medoids* pada Gambar 3 dapat dilihat grafik membentuk siku pada angka 2.

Metode *K-Means*

Berdasarkan data penelitian berikut adalah perhitungan algoritma *K-Means Clustering* dengan bantuan *software Microsoft Excel* dan *R-Studio* dan dikelompokkan berdasarkan jarak terdekat terdapat pada Tabel 2 :

Tabel 2 Pengelompokan Berdasarkan Jarak Terdekat

C_1	Data : 1,4,5,6,7,8,9, 14,15,16,17,19,20,21
C_2	Data : 2,3,10,11,12,13, 18,22

Pada Tabel 2 dapat diketahui bahwa pada *cluster* 1 terdiri dari 14 anggota *cluster* yaitu kabupaten Sumba Barat, Timor Tengah Selatan, Timor Tengah Utara, Belu, Alor, Lembata, Flores Timur, Rote Ndao, Manggarai Barat, Sumba Tengah, Sumba Barat Daya, Manggarai Timur, Sabu Raijua dan Malaka. Sedangkan pada *cluster* 2 terdiri dari 8 anggota *cluster* yaitu kabupaten Sumba Timur, Kupang, Sikka, Ende, Ngada, Manggarai, Nagekeo, dan Kota Kupang.

Metode *K-Medoids*

Berdasarkan data penelitian berikut adalah perhitungan algoritma *K-Medoids Clustering* dengan bantuan *software Microsoft Excel* dan *R-Studio* dan dikelompokkan berdasarkan jarak terdekat terdapat pada Tabel 3 :

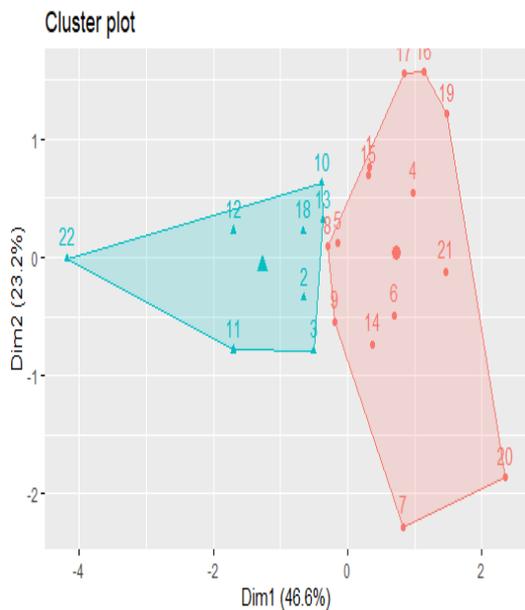
Tabel 3 Anggota *cluster*

C_1	Data : 1,4,15,16,17,19,20,21
C_2	Data : 2,3,5,6,7,8, 9,10,11,12,13,14,18,22

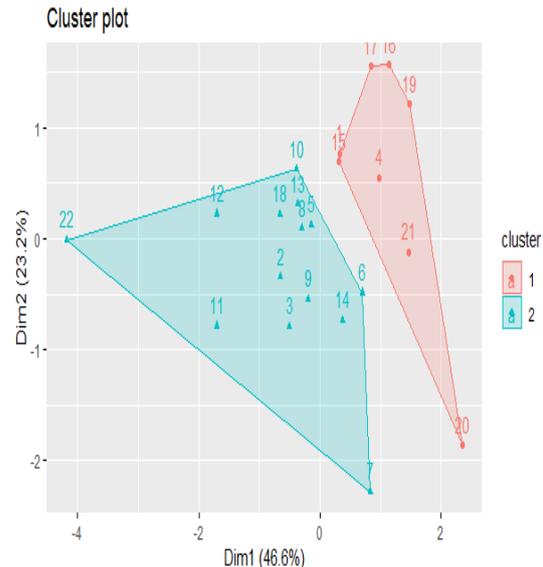
Tabel 3 menunjukkan bahwa hasil pengelompokan 22 kabupaten/kota di NTT kedalam 2 *cluster* adalah sebagai berikut: *Cluster 1* : Sumba Barat, Timor Tengah Selatan, Manggarai Barat, Sumba Tengah, Sumba Barat Daya, Manggarai Timur, Sabu Raijua dan Malaka. *Cluster 2* : Sumba Timur, Kupang, Timor Tengah Utara, Belu, Alor, Lembata, Flores Timur, Sikka, Ende, Ngada, Manggarai, Rote Ndao, Nagekeo dan Kota Kupang.

Plot K-Means dan K-Medoids

Pola sebaran setiap anggota yang terdapat pada setiap *cluster* dapat dilihat pada plot di Gambar 4 dan 5 :



Gambar 4 Plot *K-Means*



Gambar 5 Plot *K-Medoids*

Pada Gambar 4 warna merah untuk *cluster* pertama dengan jumlah anggota didalamnya ada 14 anggota, pada *cluster* kedua ditandai dengan warna biru ada 8 anggota. Pada Gambar 5 warna merah untuk *cluster* pertama dengan jumlah anggota didalamnya ada 8 anggota, pada *cluster* kedua ditandai dengan warna biru ada 14 anggota. Selain itu pada plot terdapat sumbu X dan sumbu Y yang merupakan komponen-komponen utama. Masing-masing komponen yaitu Dim1 dan Dim2 yang menjelaskan variansi data sebesar 46,6% dan 23,2%.

Penentuan Metode Terbaik

Dari proses *clustering* kedua metode yaitu metode *K-Means Clustering* dan metode *K-Medoids Clustering* kemudian dilakukan penilaian metode mana yang terbaik untuk studi kasus dimensi Indeks Pembangunan Manusia (IPM) tahun 2020 dengan menggunakan *silhouette index*, *Davies-Bouldin index* dan *Dunn index*. Berikut adalah perhitungan *silhouette index*, *Davies-Bouldin index* dan *Dunn Index* dengan bantuan *software R-Studio* pada Tabel 4:

Tabel 4 Penentuan Metode Terbaik

	Metode	
	<i>K-Means</i>	<i>K-Medoids</i>
	k = 2	k = 2
<i>Silhouette Index</i>	0,18	0,17
<i>Davies-Bouldin index</i>	1,15	2,14
<i>Dunn Index</i>	0,24	0,13

Pada penelitian ini, kriteria nilai yang digunakan dalam menentukan metode terbaik adalah dengan menggunakan nilai *Silhouette Index*, *Davies Bouldin Index* dan *Dunn Index*. Berdasarkan Tabel 4 dapat diketahui bahwa pada metode *Silhouette Index* metode terbaik yang digunakan adalah metode Algoritma *K-Means* dengan nilai sebesar 0,18 dan memiliki jumlah *cluster* sebanyak 2 *cluster*. Pada metode *Davies Bouldin Index* untuk penentuan metode terbaik adalah metode *K-Means* dengan nilai sebesar 1,15 dengan jumlah *cluster* sebanyak 2 *cluster*.

Pada metode *Dunn Index* metode terbaik yang digunakan adalah metode Algoritma *K-Means* dengan nilai 0,24 dengan banyaknya *cluster* 2 *cluster*.

Setelah mengetahui pengelompokan terbaik maka akan dilakukan *profiling cluster* untuk melihat karakteristik *cluster* yang terbentuk. Berikut adalah *profiling cluster* dimensi IPM tahun 2020 di NTT yang disajikan pada Tabel 5 :

Tabel 5 Profiling Cluster

	UHH	HLS	RLS	Pengeluaran	IPM
C_1	65,8	12,7	7,21	6658429	62,16
C_2	66,8	13,3	7,83	8477625	67,49

Dari Tabel 5 dapat diketahui pengelompokan dari masing-masing *cluster* yang dikategorikan berdasarkan nilai rata-rata setiap variabel pada *cluster* yang terbentuk. Dilihat dari nilai rata-rata setiap variabel pada *cluster* 1 dan *cluster* 2. Dari dua *cluster* tersebut dikelompokkan menjadi wilayah yang tinggi mempengaruhi dan rendah mempengaruhi penurunan Indeks Pembangunan Manusia pada tahun 2020 di NTT. Daerah yang tinggi mempengaruhi penurunan IPM terdapat pada *cluster* 1. Hal ini dikarenakan nilai rata-rata pada untuk variabel umur harapan hidup, harapan lama sekolah, rata-rata lama sekolah dan pengeluaran pada *cluster* 2 lebih besar dibandingkan *cluster* 1 yang menunjukkan bahwa pada daerah *cluster* 2 untuk keempat variabel yang diteliti mengalami peningkatan IPM tahun 2020 di NTT dan juga untuk variabel Indeks Pembangunan Manusia sendiri rata-rata pada *cluster* 2 lebih besar dari pada rata-rata IPM untuk kabupaten/kota pada *cluster* 1.

4. Kesimpulan

Berdasarkan hasil analisis dapat disimpulkan bahwa dari ketiga metode validasi yang digunakan, metode terbaik dalam pengelompokan kabupaten/kota di provinsi NTT yang berpengaruh dalam penurunan Indeks Pembangunan Manusia tahun 2020 adalah metode *K-Means Clustering* dengan banyaknya *cluster* 2. *Cluster* 1 memiliki 14 anggota *cluster* yaitu kabupaten Sumba Barat, Timor Tengah Selatan, Timor Tengah Utara, Belu, Alor, Lembata, Flores Timur, Rote Ndao, Manggarai Barat, Sumba Tengah, Sumba Barat Daya, Manggarai Timur, Sabu Raijua dan Malaka. Sedangkan pada *cluster* 2 terdiri dari 8 anggota *cluster* yaitu kabupaten Sumba Timur, Kupang, Sikka, Ende, Ngada, Manggarai, Nagekeo, dan Kota Kupang.

Daftar Pustaka

- Alwi, W., & dkk. (2018). Analisis Klaster Untuk Pengelompokan Kabupaten/Kota Di Provinsi Sulawesi Selatan Berdasarkan Indikator Kesejahteraan Rakyat . *Jurnal MSA* , 6.
- Arniva, N. S., & dkk. (2016). Pemodelan Dan Pemetaan Kasus Demam Berdarah Dengue Di Provinsi Jawa Timur Tahun 2014 Dengan Generalize Poisson Regression, Regresi Binomial Negatif Dan Flexibly Shaped Spatisal Scan Statistic. (I. T. Statistika, Ed.) *Jurnal Sains dan Seni ITS* , 5.
- BPS. (2016). *Indeks Pembangunan Manusia*.
- BPS, N. (2021). *Indeks Pembangunan Manusia*.
- Han, J., & Kamber, M. (2006). *Data Mining : Concept and Techniques*. Waltham: Morgan Kauffman Publisher.
- Kartikasari, M. D. (2021). Self-Organizing Map Menggunakan Davies-Bouldin Index Dalam Pengelompokan Wilayah Indonesia Berdasarkan Konsumsi Pangan. *Jurnal Jambura J Math* , Volume 3.
- Muhammad, A. F. (2015). *Klasterisasi Proses Seleksi Pemain Menggunakan Algoritma K-Means (Studi Kasus : Tim Hockey Kabupaten Kendal)*. Semarang: Fakultas Ilmu Komputer, Teknik Informatika, Universitas Dian Nuswantoro.
- Nugroho, A., & dkk. (2019). *Indeks Pembangunan Manusia* . Jakarta: Badan Pusat Statistik.
- Prasetyo, E. (2012). *Data Mining Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi Offset.
- Rahman, A. T. (2017). Coal Trade Data Clustering Using K-Means (Case Study Pt, Globab

- Bangkit Utama). *Jurnal Ilmiah Teknologi dan Informasi* , Volume 6.
- Trayasiwi, G. P. (2015). *Penerapan Metode Klustering Dengan Algoritma K-Means Untuk Prediksi Kelulusan Mahasiswa Pada Program Studi Teknik Informatika Strata Satu*. Semarang: Universitas Dian Nuswantoro .
- Ulinuh, N., & dkk. (2020). Analisis Cluster Dalam Pengelompokan Provinsi Di Indonesia Berdasarkan Variabel Penyakit Menular Menggunakan Metode Complete Linkage, Average Linkage, Dan Ward . *Jurnal Nasional Informatika dan Teknologi Jaringan* , 5.
- UNDP. (2004). *The Economics Democracyn: Financing Human Development in Indonesia. Indonesia Human Development Report* .