

PENGELOMPOKAN JENIS PENDERITA PENYAKIT MENULAR BERDASARKAN KECAMATAN DI KABUPATEN TUBAN MENGUNAKAN ALGORITMA K-MEANS

Kurniawan Indra Jaya^{1*}, Kresna Oktafianto²

^{1,2} Program Studi Matematika, Universitas PGRI Ronggolawe

*Email: kurniawanindrajaya12317@gmail.com

ABSTRAK

Penyakit menular dapat ditularkan dari penderita penyakit menular ke orang yang sehat. Pada umumnya, penyakit menular penyebarannya tidak merata pada suatu wilayah akan menjadi permasalahan yang menuntut suatu solusi dan kerja sama untuk mengatasinya. Hal ini perlu untuk menjadi perhatian bersama supaya masyarakat pada Kabupaten Tuban dapat mencapai tingkat kesehatan yang lebih baik. Solusi yang dapat digunakan adalah dengan mengumpulkan data penyakit menular pada kabupaten tuban berdasarkan kecamatan dengan data yang diambil adalah data dokumentasi penyakit HIV, IMS, DBD, Diare, dan TBC pada tahun 2021 dan dikelompokkan menggunakan algoritma k-means menjadi 3 bagian, yaitu 2 kelompok, 3 kelompok, dan 4 kelompok, untuk mengetahui hasil akhir yaitu kelompok yang mempunyai kemiripan objek. Dari proses pengelompokan tersebut terpilih hasil kelompok dengan pengelompokan yang paling baik menggunakan evaluasi pengelompokan Indeks Davies Bouldin sebesar 0.322 dengan pengelompokan 2 kelompok, yaitu kelompok kasus tinggi, dan kelompok kasus rendah. Dari 20 kecamatan pada Kabupaten Tuban, diketahui 3 kecamatan termasuk kelompok kasus tinggi, dan 17 kecamatan termasuk kelompok kasus rendah. Faktor penyakit Diare dan TBC sangat mempengaruhi pada kelompok kasus tinggi, dan jenis penyakit menular cukup jarang terjadi pada kelompok kasus rendah dengan faktor jenis penyakit menular yang paling mempengaruhi adalah Diare.

Kata Kunci: Pengelompokan; K-Means; Penyakit Menular; IDB

PENDAHULUAN

Penyakit Menular adalah penyakit yang dapat ditularkan dari orang penderita penyakit menular ke orang yang sehat. Pada umumnya, penyakit menular penyebarannya tidak merata pada wilayah tertentu dan akan menjadi permasalahan yang membutuhkan suatu solusi dan kerja sama untuk mengatasinya [1]. Menurut data dari Badan Pusat Statistik Kabupaten Tuban, penduduk pada Kabupaten Tuban yang terinfeksi penyakit menular cukup banyak dan tidak merata pada wilayah-wilayah Kabupaten Tuban. Penyakit menular dapat mengganggu aktivitas, bahkan meningkatkan resiko kematian masyarakat pada Kabupaten Tuban. Untuk itu, penyakit menular perlu untuk jadi perhatian bersama agar masyarakat di Kabupaten Tuban dapat mencapai tingkat Kesehatan yang lebih baik. Solusi ditawarkan adalah dengan mengumpulkan data dari penyakit menular di pada Kabupaten Tuban dan dikelompokkan untuk mencari kelompok dengan local optimal. Pengetahuan tentang pengelompokan penyakit menular yang sedang menjangkit suatu wilayah diharapkan dapat

membantu upaya dalam mencapai derajat kesehatan masyarakat yang lebih baik dengan upaya promotif seperti edukasi dan sosialisasi penyakit menular dan preventif tentang penyakit menular untuk mencapai derajat kesehatan masyarakat yang baik tentang penyakit menular di Kabupaten Tuban.

Algoritma k-means digunakan dalam penelitian ini untuk melakukan pengelompokan non-hierarki atau klasterisasi data jumlah orang yang menderita penyakit menular pada Kabupaten Tuban. Data tersebut akan dikelompokkan dengan algoritma k-means dan dikelompokkan pada kelompok yang mempunyai kemiripan pada anggotanya dan mempunyai perbedaan antar kelompoknya [2]. Keunggulan algoritma k-means dari algoritma yang lainnya adalah algoritma k-means sederhana, mudah diimplementasikan, tidak lambat, mudah disesuaikan, serta sering digunakan dalam proses data mining [3]–[5]. Pengelompokan dengan algoritma k-means akan menghasilkan masing-masing kelompok atau klaster yang mempunyai karakteristik berbeda-beda antar

kelompok dan mempunyai kelompok yang homogen.

Beberapa kajian ilmiah tentang algoritma K-Means untuk pengelompokan penyakit menular adalah, "Penerapan Metode K-Means Clustering Untuk Pemetaan Penyebaran Penyakit Demam Berdarah (DBD)", hasil dari kajian ilmiah tersebut ditemukan bahwa desa di Kecamatan Setia janji yang paling sering dan paling banyak warganya yang terjangkit penyakit DBD yaitu desa 2 dan 4 [6]. "Pengelompokan Jumlah Kasus Penyakit Aids Berdasarkan Provinsi Menggunakan Metode K-Means", kesimpulan dari kajian ilmiah tersebut menghasilkan beberapa cluster, yaitu cluster tinggi ada 4 provinsi, dan cluster rendah ada 30 provinsi [7]. "Algoritma K-Means Untuk Pengelompokan Penyakit Pasien Pada Puskesmas Cigugur Tengah", kesimpulan dari kajian ilmiah ini dapat mengelompokkan penyakit pasien pada penyakit akut dan tidak akut, maka jumlah cluster yang digunakan dalam kajian ini sebanyak 2 cluster, yaitu cluster_0 dan cluster_1, dimana pada cluster_0 untuk kelompok penyakit akut sebanyak 241 orang, sedangkan dari cluster_1 untuk kelompok penyakit tidak akut sebanyak 9 orang pada algoritma k-means. [8], "Klasterisasi Persebaran Virus Corona (Covid-19) Di Dki Jakarta Menggunakan Metode K-Means", kesimpulan dari kajian ilmiah ini menunjukkan bahwa metode k-means dapat digunakan dengan baik [9].

Pengelompokan dengan Algoritma K-Means di kajian ilmiah ini akan mengelompokkan data dari penyakit menular di Kabupaten Tuban, yaitu HIV, IMS, DBD, Diare, dan TBC pada tahun 2021. Data ini dikelompokkan beberapa kelompok, yaitu menjadi 2, 3, atau 4 kelompok bagian dan dievaluasi dengan mengecek kelompok yang baik dari 3 pilihan kelompok tersebut. Maka, rumusan masalah dari penelitian ini. Pertama adalah bagaimana hasil pengelompokan data jenis penderita penyakit menular berdasarkan wilayah kecamatan di kabupaten Tuban menggunakan algoritma K-Means, kedua adalah bagaimana hasil dari evaluasi pengelompokan berdasarkan Indeks Davies Bouldin, dan ketiga adalah apa saja faktor-faktor penyakit menular yang mempengaruhi hasil dari pengelompokan berdasarkan kecamatan di Kabupaten Tuban menggunakan Algoritma K-Means. Maka, salah satu tujuan dari penelitian ini adalah memperoleh hasil dari faktor-faktor penyakit

menular yang mempengaruhi hasil dari pengelompokan menggunakan k-means pada penderita penyakit menular di Kabupaten Tuban.

METODE PENELITIAN

Pengumpulan Data

Data yang digunakan untuk penelitian pengelompokan penyakit menular ini adalah data sekunder yang peneliti dapatkan dari BPS Kabupaten Tuban dan penelitian ini termasuk ke penelitian kuantitatif. Data penyakit menular meliputi jenis penyakit HIV, IMS, DBD, Diare, TBC, dan Malaria. Peneliti mendapatkan data penyakit menular tersebut pada tahun 2021 dengan sumber data dari Dinas Kesehatan Kabupaten Tuban. Data yang telah diperoleh diolah menggunakan aplikasi Microsoft Excel, Rapidminer, dan SPSS 25.

Algoritma K-Means

Algoritma K-Means adalah salah satu Algoritma non-hirarki untuk pengelompokan data yang mempunyai output berupa kelompok berjumlah satu atau lebih, dengan objek yang mempunyai sifat sama akan menjadi satu kelompok dengan karakter yang sama dan berbeda dengan satu kelompok yang lain [10]–[12]. Algoritma K-Means dapat berkerja pada atribut angka [13]. Algoritma K-means memproses data atau objek secara berulang-ulang kali untuk mendapatkan data-data atau objek klaster pada kelompok dan akan dipilih dengan jarak terkecil untuk pengelompokan dalam kelompoknya [14][15]. Untuk itu, dibutuhkanlah jumlah klaster atau kelompok yang diperlukan sebagai suatu masukan titik pusat awal untuk menghasilkan titik pusat akhir sebagai output [16]. Pada implementasi menyelesaikan masalah, algoritma K-Means adalah algoritma yang cukup sederhana serta fleksibel. Artinya perhitungan tidak terlalu rumit, algoritma ini dapat diimplementasikan pada banyak bidang, tidak lambat, mudah disesuaikan, serta seringkali digunakan dalam proses data mining [3]–[5], [17], [18].

Berikut ini adalah tahapan-tahapan yang dibutuhkan pada algoritma K-Means:

1. Tentukan k sebagai jumlah kelompok yang diperlukan untuk dibentuk.
2. Tentukan k-titik pusat awal kelompok (*Centroid*) awal secara acak.

3. Hitung jarak pada setiap objek kepada titik pusat kelompok dari masing-masing kelompok. Untuk menghitung jarak antara objek dengan titik pusat kelompok bisa menggunakan rumus *Euclidian Distance*.

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

Keterangan:

x_i : objek atau data x ke i.

y_i : titik pusat kelompok ke i.

n : banyaknya objek.

i : 1,2,3,...,n .

4. Alokasikan data atau objek pada jarak titik pusat kelompok yang paling kecil atau terdekat.
5. Setelah melakukan perhitungan berulang dan alokasi objek, kemudian tentukan posisi titik pusat kelompok yang baru dengan menggunakan persamaan berikut:

$$v_x = \frac{\sum_{i=1}^n x_i}{n}, n = 1,2,3, \dots, n \quad (2)$$

Keterangan:

v_x : *centroid* atau titik pusat pada kelompok.

x_i : objek atau data ke-i.

n : banyaknya objek atau data yang menjadi anggota dari suatu kelompok.

6. Jika terdapat perubahan pada titik pusat kelompok baru dan titik pusat awal kelompok, maka ulangi langkah 3 dengan menggunakan titik pusat kelompok dari Langkah 5. Jika tidak terdapat perubahan pada titik pusat kelompok awal dengan titik pusat kelompok baru, maka proses pengelompokan pada – k telah selesai.

Indeks Davies Bouldin

Indeks Davies Bouldin (IDB) adalah suatu metode untuk mencari hasil pengelompokan yang baik dari melihat nilai dari hasil IDB, metode IDB ini berfungsi untuk memaksimalkan jarak antar kelompok serta meminimalkan jarak antar objek pada sebuah kelompok [19], [20]. Berikut adalah tahapan-tahapan untuk mencari IDB yaitu dengan persamaan berikut [21], [22]:

1. Tentukan Nilai SSW

Sum Squared Within (SSW) kelompok atau ukuran nilai dispersi pada kelompok yang didapatkan dan diformulasikan dengan rumus:

$$SSW = \frac{1}{m} \sum_{j=1}^{m_j} d(x_j, c_i) \quad (3)$$

Keterangan:

m_j : objek atau data dalam kelompok.

$d(x_j, c_i)$: jarak antara objek atau data pada titik pusat kelompok.

m : banyaknya objek yang terdapat dalam kelompok.

2. Tentukan Nilai SSB

Sum Squared Between (SSB) kelompok merupakan persamaan yang digunakan untuk mengetahui separasi antar kelompok yang didapatkan dan diformulasikan dengan rumus:

$$SSB = d(c_j, c_i) \quad (4)$$

Keterangan:

$d(x_j, c_i)$: jarak antara titik pusat kelompok j dengan titik pusat kelompok i.

c_j : titik pusat kelompok j

c_i : titik pusat kelompok i

3. Tentukan Nilai Rasio

Nilai SSB yang didapatkan dan diformulasikan dengan rumus:

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{ij}} \quad (5)$$

Keterangan:

R_{ij} adalah nilai Rasio

SSW_i pada kelompok i.

SSW_j pada kelompok j.

SSB_{ij} pada kelompok ij.

4. Tentukan Nilai IDB

Nilai IDB didapatkan dan diformulasikan dengan rumus:

$$IDB = \frac{1}{k} \sum_{n=1}^k \max_{i \neq j} (R_{ij}) \quad (6)$$

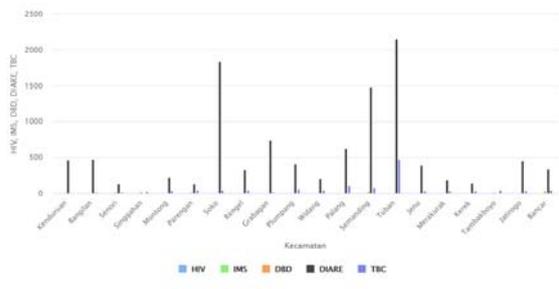
Keterangan:

k adalah jumlah total kelompok

Dari perhitungan IDB, maka dapat diambil suatu kesimpulan bahwa, semakin kecil nilai indeks davies bouldin atau mendekati nol, maka skema pengelompokan tersebut akan semakin baik [23]–[25].

HASIL DAN PEMBAHASAN

Pada penelitian ini, peneliti menggunakan data sekunder berupa data penderita penyakit menular berdasarkan kecamatan di Kabupaten Tuban pada tahun 2021, berikut adalah rincian data yang akan digunakan untuk penelitian penyakit menular, yaitu:



Gambar 1. Data penyakit menular HIV, IMS, DBD, Diare, TBC

Pada Gambar 1, diketahui data yang akan digunakan untuk pengelompokan adalah penyakit menular dengan jenis HIV, IMS, DBD, Diare dan TBC yang terdapat di 20 kecamatan pada Kabupaten Tuban.

Proses Pengelompokan

Proses pengelompokan pada penelitian ini menggunakan algoritma k-means, Langkah pertama untuk menggunakan algoritma k-means adalah tentukan jumlah kelompok untuk menjadi hasil akhir. Pada penelitian ini mengambil jumlah kelompok yaitu, $k=2$, $k=3$, dan $k=4$. Selanjutnya, tentukan titik pusat kelompok secara random. Berikut adalah proses pengelompokan untuk tiap kelompok, yaitu:

Pengelompokan dengan 2 kelompok

Pengelompokan dengan jumlah 2 kelompok dipilih untuk pengelompokan pada tahap ini. Langkah selanjutnya adalah tentukan titik pusat kelompok untuk pengelompokan pada 2 kelompok, titik pusat awal kelompok pada perhitungan tahap kedua bisa dilihat pada Tabel 1.

Tabel 1. Titik pusat awal kelompok 2 pada tahap perhitungan berulang kedua

Data	Titik Pusat Kelompok	
	1	2
HIV	3	4.333333
IMS	0.23529412	0.333333
DBD	10.3529412	15.66667
Diare	307	2825
TBC	39.41176	194

Pada Tabel 1, dapat diketahui titik pusat awal untuk perhitungan berulang pada tahap 2, hasil pada Tabel 1 diketahui dari menghitung titik pusat yang baru menggunakan persamaan (2) pada perhitungan berulang pertama. Berikut adalah hasil alokasi objek kecamatan pada titik pusat kelompok yang terkecil:

Tabel 2. Hasil alokasi objek 2 kelompok pada perhitungan berulang kedua

Kecamatan	1	2	AI
Kenduruan	158.32	1374.0	1
Bangilan	165.29	1365.2	1
Senori	182.67	1708.3	1
Singgahan	307.70	1832.6	1
Montong	86.850	1612.2	1
Parengan	183.24	1707.6	1
Soko	1539.0	153.69	2
Rengel	21.793	2153.8	1
Grabagan	432.57	1100.1	1
Plumpang	107.72	1418.2	1
Widang	105.45	1629.4	1
Palang	322.16	1205.5	1
Semanding	1170.5	369.64	2
Tuban	1894.5	427.93	2
Jenu	87.327	1439.5	1
Merakurak	129.28	1654.3	1
Kerek	168.32	1693.1	1
Tambakboyo	307.20	1831.0	1
Jatirogo	153.07	1373.9	1
Bancar	41.962	1491.1	1

Dari Tabel 2, dapat diketahui alokasi kecamatan berdasarkan jarak ke titik pusat yang terkecil. Terdapat 17 kecamatan yang termasuk kelompok 1 dan terdapat 3 kecamatan yang termasuk kelompok 2 dan dapat dilihat di Tabel 2. Berikut adalah hasil dari titik pusat akhir pada perhitungan berulang kedua:

Tabel 3. Hasil perhitungan dari titik pusat 2 kelompok yang baru pada perhitungan berulang kedua

Data	Titik Pusat Kelompok	
	1	2
HIV	3	4.333333
IMS	0.23529412	0.333333
DBD	10.3529412	15.66667
Diare	307	2825
TBC	39.41176	194

Dari Tabel 3, dapat diketahui titik pusat baru untuk perhitungan kedua dengan hasil perhitungan titik pusat awal berulang kedua ini adalah sama. Maka, perhitungan berulang berhenti pada perhitungan berulang kedua. Dapat diketahui dari titik pusat kelompok pada Tabel 3, kelompok 1 merupakan kelompok kasus rendah, dan kelompok 2 merupakan kelompok kasus tinggi.

Pengelompokan dengan 3 Kelompok

Pengelompokan dengan jumlah 3 kelompok dipilih untuk pengelompokan pada tahap ini, Langkah selanjutnya adalah tentukan titik pusat kelompok untuk pengelompokan pada 3 kelompok, titik pusat awal kelompok pada perhitungan tahap keenam bisa dilihat pada Tabel 4.

Tabel 4. Titik pusat awal 3 kelompok pada perhitungan berulang keenam

Data	Titik Pusat Kelompok		
	1	2	3
HIV	2.625	3.333333	4.333333
IMS	0.25	0.222222	0.333333
DBD	12.25	8.666667	15.66667
Diare	123.75	469.8889	1825
TBC	36.375	42.11111	194

Pada Tabel 4, dapat diketahui titik pusat awal untuk perhitungan berulang pada tahap 6, hasil pada Tabel 4 diketahui dari menghitung titik pusat yang baru menggunakan persamaan (2) pada perhitungan berulang kelima. Berikut adalah hasil alokasi objek kecamatan pada titik pusat kelompok yang terkecil:

Tabel 5. Hasil alokasi objek 3 kelompok pada perhitungan berulang keenam

Kecamatan	1	2	3	Al
Kenduruan	340.14	30.185	1374.0	2
Bangilan	347.68	23.656	1365.2	2
Senori	21.353	345.01	1708.3	1
Singgahan	124.98	470.48	1832.6	1
Montong	98.012	249.22	1612.2	1
Parengan	9.9105	346.03	1707.6	1
Soko	1722.2	1376.1	153.69	3
Rengel	200.98	146.26	1507.9	2
Grabagan	615.58	270.20	1100.1	2
Plumpang	290.69	56.683	1418.2	2
Widang	79.274	267.99	1629.4	1
Palang	503.57	164.42	1205.5	2
Semanding	1353.7	1007.6	369.64	3
Tuban	2074.0	1735.6	427.93	3
Jenu	270.41	76.223	1439.5	2
Merakurak	55.323	391.98	1654.3	1
Kerek	19.753	331.01	1693.1	1
Tambakboyo	124.29	470.03	1831.0	1
Jatirogo	336.31	11.163	1373.9	2
Bancar	218.60	131.34	1491.1	2

Dari Tabel 5, dapat diketahui alokasi kecamatan berdasarkan jarak ke titik pusat yang terkecil. Terdapat 8 objek kecamatan yang termasuk kedalam kelompok 1, 9 objek kecamatan yang termasuk kedalam kelompok 2, dan 3 objek kecamatan yang termasuk kedalam kelompok 3, dan bisa dilihat pada Tabel 5. Berikut adalah hasil dari titik pusat akhir pada perhitungan berulang kedua:

Tabel 6. Hasil perhitungan dari titik pusat 3 kelompok yang baru pada perhitungan berulang ke 6

Data	Titik Pusat Kelompok		
	1	2	3
HIV	2.625	3.333333	4.333333
IMS	0.25	0.222222	0.333333
DBD	12.25	8.666667	15.66667
Diare	123.75	469.8889	1825
TBC	36.375	42.11111	194

Dari Tabel 6, dapat diketahui titik pusat awal untuk perhitungan keenam dengan hasil perhitungan titik pusat baru berulang keenam ini adalah sama. Maka, perhitungan berulang berhenti pada perhitungan berulang keenam. Dapat diketahui dari titik pusat kelompok pada Tabel 6, kelompok 1 merupakan kelompok kasus rendah, dan kelompok 2 merupakan kelompok kasus sedang, dan kelompok 3 merupakan kelompok kasus tinggi.

Pengelompokan dengan 4 Kelompok

Pengelompokan dengan jumlah 4 kelompok dipilih untuk pengelompokan pada tahap ini, Langkah selanjutnya adalah tentukan titik pusat kelompok untuk pengelompokan pada 4 kelompok, titik pusat awal kelompok pada tahap kedua bisa dilihat pada Tabel 7.

Tabel 7. Titik pusat awal 4 kelompok pada perhitungan berulang kedua

Data	Titik Pusat Kelompok			
	1	2	3	4
HIV	6	3.5	3.333333	2.625
IMS	0	0.5	0.222222	0.25
DBD	11	18	8.666667	12.25
Diare	2152	1661.5	469.8889	123.75
TBC	470	56	42.11111	36.375

Pada Tabel 7, dapat diketahui titik pusat awal untuk perhitungan berulang pada tahap kedua, hasil pada Tabel 7 diketahui dari menghitung titik pusat yang baru menggunakan persamaan (2) pada perhitungan berulang kedua. Berikut adalah hasil alokasi objek kecamatan pada titik pusat kelompok yang terkecil:

Tabel 8. Hasil alokasi objek kelompok pada perhitungan berulang kedua

Kecamatan	1	2	3	4	AI
Kenduruan	1749	1199	30.18	340.1	3
Bangilan	1740	1191	23.65	347.6	3
Senori	2076	1536	345.0	21.35	4
Singgahan	2197	1661	470.4	124.9	4
Montong	1980	1440	249.2	98.01	4
Parengan	2072	1537	346.0	9.910	4
Soko	526.1	185.3	1376	1722	2
Rengel	1875	1337	146.2	200.9	3
Grabagan	1483	923.3	270.2	615.5	3
Plumpang	1787	1247	56.68	290.6	3
Widang	1994	1459	267.9	79.27	4
Palang	1572	1039	164.4	503.5	3
Semanding	784.8	185.3	1017	1353	2
Tuban	0	641.9	1735	2074	1
Jenu	1810	1267	76.22	270.4	3
Merakurak	2020	1483	291.9	55.32	4
Kerek	2058	1522	331.0	19.75	4
Tambakboyo	2193	1661	470.0	124.2	4
Jatirogo	1746	1201	11.16	336.3	3
Bancar	1859	1320	131.3	218.6	3

Dari Tabel 8, dapat diketahui alokasi kecamatan berdasarkan jarak ke titik pusat yang terkecil. Terdapat 1 objek kecamatan yang termasuk kedalam kelompok 1, 2 objek kecamatan yang termasuk kedalam kelompok 2, 9 objek kecamatan yang termasuk kedalam kelompok 3, dan 8 objek kecamatan yang termasuk kedalam kelompok 4. Berikut adalah hasil dari titik pusat akhir pada perhitungan berulang kedua:

Tabel 9. Hasil perhitungan dari titik pusat 4 kelompok yang baru pada perhitungan berulang kedua

Data	Titik Pusat Kelompok			
	1	2	3	4
HIV	6	3.5	3.333333	2.625
IMS	0	0.5	0.222222	0.25
DBD	11	18	8.666667	12.25
Diare	2152	1661.5	469.8889	123.75
TBC	470	56	42.11111	36.375

Dari Tabel 9, dapat diketahui titik pusat awal untuk perhitungan kedua dengan hasil perhitungan titik pusat baru berulang kedua ini adalah sama. Maka, perhitungan berulang berhenti pada perhitungan berulang kedua. Dapat diketahui, kelompok 1 merupakan kelompok kasus tertinggi, kelompok 2 merupakan kasus cukup tinggi, kelompok 3 merupakan kelompok kasus cukup rendah, dan kelompok 4 merupakan kelompok kasus terendah.

Perhitungan Indeks Davies Bouldin

Perhitungan Indeks Davies Bouldin untuk mengetahui jumlah kelompok yang baik dari hasil pengelompokan sebelumnya. Berikut adalah proses perhitungan Indeks Davies Bouldin, yaitu:

Perhitungan IDB Pada 2 Kelompok

Perhitungan IDB pada 2 kelompok ini digunakan untuk mengetahui nilai IDB dari hasil pengelompokan menggunakan algoritma k-means. Berikut adalah perhitungannya:

- Mencari nilai SSW
 Indikator SSW adalah indikator *Sum Square Within*. Nilai SSW ini dicari untuk mengetahui nilai kohesi didalam kelompok hasil pengelompokan. Berikut adalah hasil dari pencarian nilai SSW pada 2 kelompok, yaitu:

Tabel 10. Hasil Nilai SSW

Nilai SSW	Jumlah SSW
SSW ₁	174.1754
SSW ₂	317.0937

Dari Tabel 10, dapat diketahui Nilai SSW untuk kelompok 1 dan kelompok 2.

- Mencari Nilai SSB
 Indikator SSB adalah indikator *Sum Square Between*. Nilai SSB ini dicari untuk mengetahui nilai separasi antara kelompok dari hasil pengelompokan. Berikut adalah hasil dari pencarian nilai SSB pada 2 kelompok, yaitu:

Tabel 11. Hasil Nilai SSB

Nilai SSB	Jumlah SSB
SSB _{1,2}	1525.8609
SSB _{2,1}	1525.8609

Dari Tabel 11, dapat diketahui Nilai SSB antara kelompok 1 dan kelompok 2, dan kelompok 2 dan kelompok 1.

- Mencari Nilai Rasio
 Indikator Rasio adalah indikator untuk mengetahui seberapa baik nilai rasio antara nilai SSW dan nilai SSB, berikut adalah nilai Rasio dari perhitungan 2 kelompok:

Tabel 12. Hasil Nilai Rasio

Nilai Rasio	Jumlah Rasio
Rasio _{1,2}	0.322
Rasio _{2,1}	0.322

Dari Tabel 12, dapat diketahui Nilai Rasio dari kelompok 1 dan kelompok 2 dan Rasio dari kelompok 2 dan kelompok 1.

- Mencari Nilai IDB
 Indikator nilai IDB memerlukan hasil dari perhitungan nilai Rasio, dan akan diambil hasil maksimal dari perhitungan rasio, dan akan dibagi dengan jumlah banyaknya pengelompokan. Berikut adalah nilai IDB:

Tabel 13. Nilai Rasio Maksimal

R	1	2	R _{max}
1	0	0.322	0.322
2	0.322	0	0.322

Dari Tabel 13, dapat diketahui R_{max} dari rasio-rasio yang telah dihitung nilainya.

Berikutnya adalah menghitung Nilai IDB, yaitu:

$$IDB = \frac{1}{2}(0.322 + 0.322)$$

$$IDB = 0.322$$

Perhitungan IDB Pada 3 Kelompok

Perhitungan IDB pada 3 kelompok ini digunakan untuk mengetahui nilai IDB dari hasil pengelompokan. Berikut adalah hasil perhitungan IDB pada 3 kelompok:

- Mencari nilai SSW
 Indikator SSW adalah indikator *Sum Square Within* untuk mengetahui nilai kohesi didalam kelompok. Berikut adalah hasil dari pencarian nilai SSW pada kelompok 3, yaitu:

Tabel 14. Hasil Nilai SSW

Nilai SSW	Jumlah SSW
SSW ₁	66.6134
SSW ₂	101.128
SSW ₃	317.093

Dari Tabel 14, dapat diketahui Nilai SSW untuk kelompok 1, kelompok 2 dan kelompok 3.

- Mencari Nilai SSB
 Indikator SSB adalah indikator *Sum Square Between*. Nilai SSB ini dicari untuk mengetahui nilai separasi antara kelompok dari hasil pengelompokan. Berikut adalah hasil dari pencarian nilai SSB pada 3 kelompok, yaitu:

Tabel 15. Hasil Nilai SSB

SSB	1	2	3
1	0	346.20	1708.5
2	346.20	0	1363.6
3	1708.5	1363.6	0

Dari Tabel 15, dapat diketahui Nilai SSB antara kelompok 1 dan kelompok 2, kelompok 2 dan kelompok 3.

- Mencari Nilai Rasio
 Indikator Rasio adalah indikator untuk mengetahui seberapa baik nilai rasio antara nilai SSW dan nilai SSB, berikut adalah nilai Rasio:

Tabel 16. Hasil Nilai Rasio

R	1	2	3
1	0	0.484	0.224
2	0.484	0	0.306
3	0.224	0.306	0

Dari Tabel 16, dapat diketahui Nilai Rasio dari kelompok 1 dan kelompok 2, Rasio dari kelompok 2 dan kelompok 3, dan seterusnya.

- Mencari Nilai IDB
 Indikator nilai IDB memerlukan hasil dari perhitungan nilai Rasio, dan akan diambil hasil maksimal dari perhitungan rasio, dan akan dibagi dengan jumlah banyaknya pengelompokan. Berikut adalah nilai IDB:

Tabel 17. Nilai Rasio Maksimal

R	1	2	3	R _{max}
1	0	0.484	0.224	0.484
2	0.484	0	0.306	0.484
3	0.224	0.306	0	0.306

Dari Tabel 17, dapat diketahui R_{max} dari rasio-rasio yang telah dihitung nilainya. Berikutnya adalah menghitung Nilai IDB, yaitu:

$$IDB = \frac{1}{3}(0.484 + 0.484 + 0.306)$$

$$IDB = 0.425$$

Perhitungan IDB Pada 4 Kelompok

Perhitungan IDB ini digunakan untuk mengetahui nilai IDB dari hasil pengelompokan dengan 4 kelompok. Berikut adalah perhitungannya:

- Mencari nilai SSW
 Indikator SSW adalah indikator *Sum Square Within*. Nilai SSW ini dicari untuk mengetahui nilai kohesi didalam kelompok. Berikut adalah hasil dari pencarian nilai SSW pada 4 kelompok, yaitu:

Tabel 18. Hasil Nilai SSW

Nilai SSW	Jumlah SSW
SSW ₁	0
SSW ₂	185.334
SSW ₃	101.128
SSW ₄	66.6134

Dari Tabel 18, dapat diketahui Nilai SSW dari pengelompokan dengan 4 kelompok.

- Mencari Nilai SSB
 Indikator SSB adalah indikator *Sum Square Between*. Nilai SSB ini dicari untuk

mengetahui nilai separasi antara kelompok. Berikut adalah hasil dari pencarian nilai SSB pada 3 kelompok, yaitu:

Tabel 19. Hasil Nilai SSB

SSB	1	2	3	4
1	0	641.90	1735.6	2074.1
2	641.90	0	1191.7	1537.8
3	1735.6	1191.7	0	346.20
4	2074.1	1537.8	346.20	0

Dari Tabel 19, dapat diketahui Nilai SSB antara kelompok 1 dan kelompok 2, kelompok 2 dan kelompok 1, dan seterusnya dari hasil pengelompokan menggunakan algoritma k-means.

- Mencari Nilai Rasio
 Indikator Rasio adalah indikator untuk mengetahui seberapa baik nilai rasio antara nilai SSW dan nilai SSB, berikut adalah nilai Rasio dari perhitungan 3 kelompok dengan menggunakan algoritma k-means:

Tabel 20. Hasil Nilai Rasio

R	1	2	3	4
1	0	0.288	0.058	0.032
2	0.288	0	0.240	0.163
3	0.058	0.240	0	0.484
4	0.032	0.163	0.484	0

Dari Tabel 20, dapat diketahui Nilai Rasio dari pengelompokan dengan 4 kelompok.

- Mencari Nilai IDB
 Indikator nilai IDB memerlukan hasil dari perhitungan nilai Rasio, dan akan diambil hasil maksimal dari perhitungan rasio, dan akan dibagi dengan jumlah banyaknya pengelompokan. Berikut adalah nilai IDB:

Tabel 21. Nilai Rasio Maksimal

R	1	2	3	4	R _{max}
1	0	0.288	0.058	0.032	0.288
2	0.288	0	0.240	0.163	0.288
3	0.058	0.240	0	0.484	0.484
4	0.032	0.163	0.484	0	0.484

Dari Tabel 21, dapat diketahui R_{max} dari rasio-rasio yang telah dihitung nilainya. Berikutnya adalah menghitung Nilai IDB, yaitu:

$$IDB = \frac{1}{4}(0.288 + 0.288 + 0.484 + 0.484)$$

$$IDB = 0.387$$

Perbandingan Hasil IDB

Dari perhitungan IDB pada penelitian ini, berikut rangkuman hasil dari nilai IDB dapat dilihat pada Tabel 22, yaitu:

Tabel 22. Rangkuman Hasil Nilai IDB

Kelompok	IDB
2 Kelompok	0.322
3 Kelompok	0.425
4 Kelompok	0.387

Dari Tabel 22, dapat diketahui nilai IDB yang terkecil pada kelompok 2, maka pengelompokan dengan 2 kelompok dipilih karena mempunyai nilai IDB yang baik. Pada pengelompokan dengan 2 kelompok, terdapat 2 jenis kelompok, yaitu: kelompok kasus tinggi dan kelompok kasus rendah. Pada kelompok kasus rendah, penyakit menular yang terjadi cukup jarang terjadi, dengan penyakit menular jenis diare sangat mempengaruhi. Pada kelompok kasus tinggi, penyakit menular cukup sering terjadi, dengan faktor penyakit menular jenis diare dan TBC adalah yang paling mempengaruhi pada kelompok kasus tinggi.

Kesimpulan

Berdasarkan penelitian yang telah dilakukan untuk pengelompokan jenis penderita penyakit menular di Kabupaten Tuban menggunakan algoritma k-means, dapat disimpulkan algoritma k-means dapat mengelompokkan data penyakit menular menjadi 2 kelompok, 3 kelompok, dan 4 kelompok. Selanjutnya, setelah dicek dengan nilai IDB, peneliti mendapatkan nilai hasil IDB untuk 2 kelompok sebesar 0.322, nilai untuk 3 kelompok sebesar 0.425, nilai untuk 4 kelompok sebesar 0.387. Peneliti mendapatkan hasil pengelompokan dengan kategori 2 kelompok merupakan kelompok yang baik dalam banyaknya kelompok pada penelitian ini. Pada kategori 2 kelompok, terdapat 2 jenis kelompok, yaitu kelompok 1 merupakan kelompok kasus rendah, dan kelompok 2 merupakan kelompok kasus tinggi. Terdapat 17 kecamatan yang termasuk kelompok kasus rendah, yaitu Terdapat 3 kecamatan yang termasuk kelompok kasus tinggi dan wilayah kecamatan untuk alokasi pengelompokan dengan 2 kelompok dan untuk objek Kecamatan dapat dilihat pada Tabel 3. Pada kelompok kasus tinggi, jenis penyakit menular cukup banyak terjadi, dengan faktor jenis penyakit Diare, dan TBC adalah yang

paling mempengaruhi pada kelompok kasus tinggi. Pada kelompok kasus rendah, jenis penyakit menular cukup jarang terjadi, dengan faktor jenis penyakit menular Diare yang paling mempengaruhi pada kelompok kasus rendah.

DAFTAR PUSTAKA

- [1] A. Bainus and J. Budi Rachman, "Editorial: Pandemi Penyakit Menular (Covid-19) Hubungan Internasional," *Intermestic J. Int. Stud.*, vol. 4, no. 2, p. 111, 2020, doi: 10.24198/intermestic.v4n2.1.
- [2] S. F. Susilo, A. Jamaludin, and I. Purnamasari, "Pengelompokan Desa Menggunakan K-Means Untuk Penyelenggaraan Penanggulangan Bencana Banjir," *JOINS (Journal Inf. Syst.*, vol. 5, no. 2, pp. 156–167, 2020, doi: 10.33633/joins.v5i2.3709.
- [3] L. P. Refialy, H. Maitimu, and M. S. Pesulima, "Perbaikan Kinerja Clustering K-Means Pada Data Ekonomi Nelayan Dengan Perhitungan Sum Of Square Error (SSE) Dan Optimasi Nilai K Cluster," *Techno.Com*, vol. 20, no. 2, pp. 321–329, 2021, doi: 10.33633/tc.v20i2.4572.
- [4] Y. P. Sari, A. Primajaya, and A. S. Y. Irawan, "Implementasi Algoritma K-Means Untuk Clustering Penyebaran Tuberkulosis Di Kabupaten Karawang," *INOVTEK Polbeng - Seri Inform.*, vol. 5, no. 2, p. 229, 2020, doi: 10.35314/isi.v5i2.1457.
- [5] J. Wandana, S. Defit, and S. Sumijan, "Klasterisasi Data Rekam Medis Pasien Pengguna Layanan BPJS Kesehatan Menggunakan Metode K-Means," *J. Inf. dan Teknol.*, vol. 2, pp. 4–9, 2020, doi: 10.37034/jidt.v2i4.73.
- [6] M. A. Sembiring, "Penerapan Metode Algoritma K-Means Clustering Untuk Pemetaan Penyebaran Penyakit Demam Berdarah Dengue (DBD)," *J. Sci. Soc. Res.*, vol. 4, no. 3, p. 336, 2021, doi: 10.54314/jssr.v4i3.712.
- [7] R. I. L. Sinaga, W. Saputra, and H. Qurniawan, "Pengelompokan Jumlah Kasus Penyakit Aids Berdasarkan Provinsi Menggunakan Metode K-Means," *KESATRIA Jurnal Penerapan Sist. Inf. (Komputer Manajemen)*, vol. 2, no. 2, pp. 99–107, 2021.
- [8] C. A. Sugianto, A. H. Rahayu, and A. Gusman, "Algoritma K-Means Untuk Pengelompokkan Penyakit Pasien Pada

- Puskesmas Cigugur Tengah,” *J. Inf. Technol.*, vol. 2, no. 2, pp. 39–44, 2020, doi: 10.47292/joint.v2i2.30.
- [9] A. Solichin and K. Khairunnisa, “Klasterisasi Persebaran Virus Corona (Covid-19) Di DKI Jakarta Menggunakan Metode K-Means,” *Fountain Informatics J.*, vol. 5, no. 2, p. 52, 2020, doi: 10.21111/fij.v5i2.4905.
- [10] W. A. Suputra, “Klasterisasi Hasil Ujian Nasional SMA/MA Dengan Algoritma K-Means,” *Wahana Mat. dan Sains J. Mat. Sains, dan Pembelajarannya*, vol. 15, no. 1, pp. 22–30, 2021, [Online]. Available: <https://ejournal.undiksha.ac.id/index.php/JPM/article/view/25380>
- [11] I. N. M. Adiputra, “Clustering Penyakit Dbd Pada Rumah Sakit Dharma Kerti Menggunakan Algoritma K-Means,” *Inser. Inf. Syst. Emerg. Technol. J.*, vol. 2, no. 2, p. 99, 2022, doi: 10.23887/insert.v2i2.41673.
- [12] A. Septianingsih, “Analisis K-Means Clustering Pada Pemetaan Provinsi Indonesia Berdasarkan Indikator Rumah Layak Huni,” *J. Ilm. Pendidik. Mat. Mat. dan Stat.*, vol. 3, no. 1, pp. 224–241, 2022.
- [13] M. W. Talakua, Z. A. Leleury, and A. W. Talluta, “Analisis Cluster Dengan Menggunakan Metode Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014,” *J. Ilmu Mat. dan Terap.*, vol. 11, no. 2, pp. 119–128, 2017.
- [14] R. A. Indraputra and R. Fitriana, “K-Means Clustering Data Covid-19,” *J. Tek. Ind.*, vol. 10, no. 3, p. 3, 2020.
- [15] D. N. Solikhun, Verdi Yasin, “Optimization Of The Number Of Clusters Of The K-Means Method In Grouping Egg Production Data In Indonesia,” *Int. J. Artif. Intell. Robot.*, vol. 4, no. 1, pp. 39–47, 2022.
- [16] Sani Askia, Nurul Hidayat, and Ratih Kartika Dewi, “Kelompok Penyakit Tanaman Apel Menggunakan Metode K-Means Berbasis Web,” *Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, No.8, no. e-ISSN: 2548-964X, pp. 2435–2441, 2018.
- [17] M. Nurhidayati and N. Khasanah, “Penggunaan Metode K-Means Cluster Untuk Mengklasifikasikan Kemampuan 4C Mahasiswa,” *J. Ilm. Mat. Dan Terap.*, vol. 18, no. 2, pp. 160–169, 2021, doi: 10.22487/2540766x.2021.v18.i2.15615.
- [18] Q. Wang, X. Mao, X. Jiang, D. Pei, and X. Shao, “Digital Image Processing Technology Under Backpropagation Neural Network And Kmeans Clustering Algorithm On Nitrogen Utilization Rate Of Chinese Cabbages,” *PLoS One*, vol. 16, no. 3 March 2021, pp. 1–24, 2021, doi: 10.1371/journal.pone.0248923.
- [19] W. Gie and D. Jollyta, “Perbandingan Euclidean Dan Manhattan Untuk Optimasi Cluster Menggunakan Davies Bouldin Index : Status Covid-19 Wilayah Riau,” *Pros. Semin. Nas. Ris. Dan Inf. Sci. 2020*, vol. 2, no. April, pp. 187–191, 2020.
- [20] A. Badruttamam, S. Sudarno, and D. A. I. Maruddani, “Penerapan Analisis Klaster K-Modes Dengan Validasi Davies Bouldin Index Dalam Menentukan Karakteristik Kanal Youtube Di Indonesia (Studi Kasus: 250 Kanal Youtube Indonesia Teratas Menurut Socialblade),” *J. Gaussian*, vol. 9, no. 3, pp. 263–272, 2020, doi: 10.14710/j.gauss.v9i3.28907.
- [21] Z. Nabila, “Analisis Data Mining Untuk Clustering Kasus Covid-19,” *J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, pp. 100–108, 2021.
- [22] A. A. Az-zahra, A. F. Marsaoly, I. P. Lestyani, R. Salsabila, and W. O. Z. Madjida, “Penerapan Algoritma K-Modes Clustering Dengan Validasi Davies Bouldin Index Pada Pengelompokan Tingkat Minat Belanja Online Di Provinsi Daerah Istimewa Yogyakarta,” *J. MSA (Mat. dan Stat. serta Apl.)*, vol. 9, no. 1, p. 24, 2021, doi: 10.24252/msa.v9i1.18555.
- [23] S. Nawrin, M. Rahatur, and S. Akhter, “Exploreing K-Means With Internal Validity Indexes For Data Clustering In Traffic Management System,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 3, 2017, doi: 10.14569/ijacsa.2017.080337.
- [24] A. E. Haryati, S. Sugiyarto, and R. D. A. Putri, “Comparison Of Fuzzy Subtractive Clustering And Fuzzy C-Means,” *J. Ilm. Kursor*, vol. 11, no. 1, pp. 1–8, 2021, doi: 10.21107/kursor.v11i1.254.
- [25] D. Jollyta, S. Efendi, M. Zarlis, and H. Mawengkang, “Optimasi Cluster Pada Data Stunting: Teknik Evaluasi Cluster Sum Of Square Error Dan Davies Bouldin Index,” *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 918, 2019, doi: 10.30645/senaris.v1i0.100.