

Pengelompokan Penyakit Menular di Pulau Jawa Tahun 2023 Menggunakan Algoritma *BCBimax*

Estu Widya Putri^{*}, Ilham Faishal Mahdy

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Islam Bandung, Indonesia.

estuwidya03@email.com, ilham.faishal@unisba.ac.id

Abstract. This study aims to identify simultaneous clustering patterns of infectious diseases and their distribution across three provinces in Java Island in 2023, using the biclustering method, specifically the BCBimax algorithm. Biclustering extends traditional clustering by grouping either binary data or binarized non-binary data to find submatrices consisting only of elements with value 1, according to clustering criteria. In this research, non-binary data were binarized using two scenarios: the median and the mean of each variable. The dataset includes the number of cases of HIV/AIDS, dengue fever (DBD), tuberculosis (TBC), leprosy, diarrhea, and measles in 2023, sourced from BPS portals and provincial open data. The analysis process involved visualization EDA, binarization, application of the BCBimax algorithm, and evaluation of selected biclusters. Results demonstrate that BCBimax effectively grouped regions across the three provinces into representative biclusters, reflecting the distribution patterns between diseases and their affected areas. This research contributes to the literature on biclustering applications and supports relevant authorities in making informed decisions for controlling infectious diseases and developing targeted public health strategies across these provinces in Java.

Keywords: *BCBimax, Biclustering, Infectious Diseases, Java Island.*

Abstrak. Penelitian ini bertujuan mengidentifikasi pola pengelompokan simultan penyakit menular dan sebarannya di tiga provinsi di Pulau Jawa pada tahun 2023 menggunakan metode biclustering, yaitu algoritma BCBimax. Biclustering merupakan pengembangan metode clustering standar yang dapat mengelompokkan data biner maupun data non-biner yang telah dibinarisasi, dengan menemukan submatriks yang hanya berisi elemen bernilai 1 sesuai kriteria pengelompokan. Dalam penelitian ini, data non-biner dibinarisasi berdasarkan dua skenario, yaitu median dan rata-rata masing-masing variabel. Data yang digunakan berupa jumlah kasus HIV/AIDS, demam berdarah dengue (DBD), tuberculosis (TBC), kusta, diare, dan campak tahun 2023 yang diperoleh dari portal BPS dan data terbuka dari masing-masing provinsi. Tahapan analisis meliputi visualisasi EDA, binarisasi data, penerapan algoritma BCBimax, serta evaluasi hasil bicluster yang terpilih. Hasil penelitian menunjukkan bahwa algoritma BCBimax mampu mengelompokkan wilayah ketiga provinsi menjadi beberapa bicluster representatif yang menggambarkan pola penyebaran antara penyakit dan daerah terdampak. Penelitian ini memberikan kontribusi pada literatur aplikasi metode biclustering serta mendukung pengambilan keputusan pihak terkait dalam pengendalian penyakit menular dan pengembangan strategi kesehatan masyarakat yang lebih tepat sasaran di wilayah tersebut.

Kata Kunci: *BCBimax, Biclustering, Penyakit Menular, Pulau Jawa.*

A. Pendahuluan

Pengelompokan standar (clustering) adalah salah satu metode pengelompokan yang paling sering atau umum dipakai dalam analisis data terutama diterapkan dalam analisis data kesehatan (Shuai et al., 2023). Namun, menurut, cara kerja metode pengelompokan standar yaitu subjek-subjek akan dikelompokkan berdasarkan kesamaan/kemiripan dari beberapa peubah saja atau peubah-peubah dikelompokkan berdasarkan kesamaan atau kemiripan perilaku subjek saja, yang dimana subjek atau perilaku yang memiliki kedekatan jarak akan terkelompokkan ke dalam satu cluster menggunakan beberapa metode kedekatan jarak seperti Euclidean, Manhattan, korelasi person, dll serta dalam memetakan informasi, berdasarkan baris atau kolomnya saja sehingga memungkinkan beberapa informasi dari dimensi lain tidak terkelompokkan dan saat melakukan analisis, tidak ada ketentuan pada tahap penelitian yang wajib dilakukan. Berbeda dengan metode biclustering (pengelompokan simultan), menurut beliau cara kerja metodenya lebih kompleks yaitu setiap subjek dalam bicluster (submatriks atau subkelompok) dipilih hanya menggunakan subset peubah dan setiap peubah dalam bicluster (submatriks atau subkelompok) dipilih hanya menggunakan subset subjek yang hasilnya akan terbentuk bicluster (subkelompok) atau matriks kombinasi minimum baris dan kolom yang hanya berisikan elemen 1 sehingga dapat mengelompokkan baris dan kolom yang memiliki perilaku serupa atau koheren serta dapat mengatasi keterbatasan pengelompokan standar. Dan juga pada metode biclustering ini memiliki ketentuan dalam tahapan penelitiannya yaitu harus menggunakan data biner atau data yang melalui proses binarisasi (menentukan ambang batas atau threshold) terlebih dahulu jika jenis data non-biner (Preli'c et al., 2006).

Salah satu metode biclustering (pengelompokan simultan) yang dapat menghasilkan pola pengelompokan dengan baik yaitu metode algoritma Bimax. Tetapi, menurut (Preli'c et al., 2006) metode tersebut memiliki keterbatasan dalam hal efisiensi kecepatan komputasi dan penanganan tumpang tindih dalam hasil submatriks atau subkelompoknya. Sehingga, sebagai solusi atas keterbatasan metode algoritma Bimax tersebut, dikembangkan metode Algoritma BCBimax oleh (Dolnicar et al., 2012) yang lebih efektif dalam kecepatan komputasi atau tergolong cepat dalam menemukan submatriks atau subkelompok yang berisi elemen 1, serta menghasilkan pola pengelompokan hasil submatriks atau subkelompok yang lebih minim tumpang tindih. Dalam penelitian terdahulu, menurut (Cynthia Wulandari, 2023) dengan algoritma BCBimax berhasil menemukan 11 submatriks atau bicluster optimal wilayah Indonesia berdasarkan provinsi dibandingkan metode Algoritma lain yang diteliti. Selain itu, penelitian oleh (Nur Hikmah, 2024) dengan algoritma ini berhasil mengidentifikasi pola pengelompokan aspek ketahanan pangan di wilayah Indonesia yang lebih baik dibanding metode lain yang diteliti. Dan juga penelitian oleh Aidi et al., (2023), berdasarkan provinsi di Indonesia menemukan 8 bicluster (submatriks) optimal dengan ada 1 peubah yang masuk kedalam 6 bicluster dari 8 bicluster optimal yang terbentuk, yang berarti bahwa peubah tersebut harus ditinjau lebih lanjut.

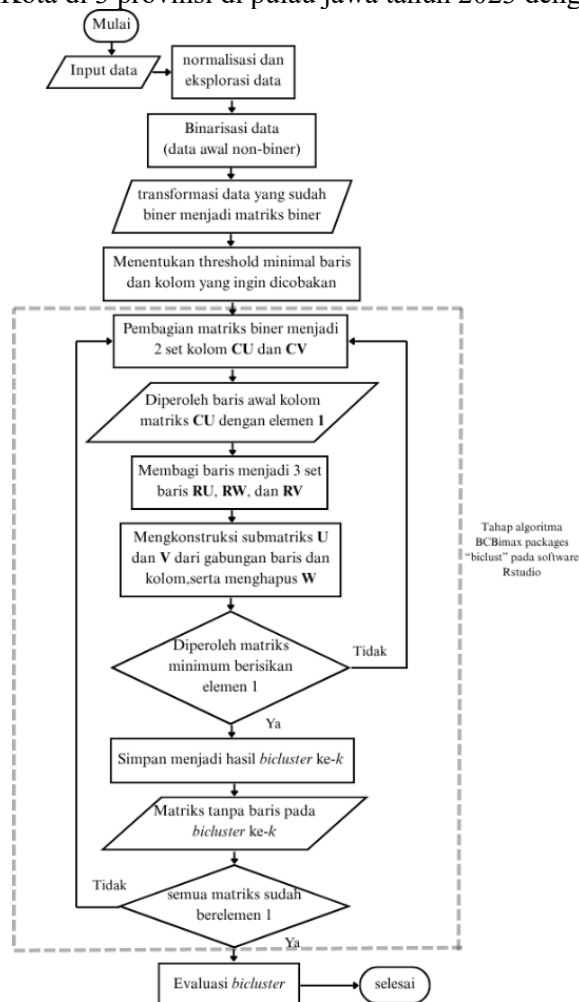
Pada penelitian ini, penerapan algoritma BCBimax difokuskan pada aspek tingkat Kabupaten atau Kota pada 3 provinsi di pulau jawa yaitu Jawa Barat, Jawa Tengah, dan Jawa Timur. Dilansir dari berita detik.com berdasarkan data BPS tahun 2023 yang diterbitkan tahun 2024, ketiga provinsi ini masih menjadi 3 peringkat teratas daerah yang memiliki tingkat penduduk terpadat pada tahun 2023 sehingga sangat berisiko penyakit menular mewabah secara cepat di masyarakat. Penyakit menular yang dijadikan peubah atau variabel di penelitian ini yaitu jumlah kasus HIV/AIDS, Demam Berdarah Dengue (DBD), Tuberkulosis (TBC), Kusta, Diare, dan Campak dengan jenis data numerik, sehingga dilakukan proses binarisasi terlebih dahulu dengan threshold median dan rata-rata masing-masing peubah untuk memenuhi syarat metode biclustering yang jenis datanya harus biner. Peubah-peubah ini dipilih berdasarkan pertimbangan penelitian terdahulu dan data BPS dari masing-masing provinsi yang menunjukkan adanya kenaikan jumlah kasus pada tahun 2023 yang melonjak naik tajam pada beberapa Kabupaten/Kotanya atau bahkan keseluruhan Kabupaten/Kotanya. Dampak penyakit ini tidak hanya memengaruhi kesehatan masyarakat saja, tetapi juga berkontribusi pada masalah-masalah lainnya pada kehidupan sosial bermasyarakat. Kondisi ini menuntut pendekatan analisis yang mampu mengidentifikasi pola penyebaran penyakit menular di Pulau Jawa untuk mendukung perencanaan intervensi kesehatan yang lebih efektif. Maka dari hal tersebut, penelitian ini memiliki rumusan masalah yaitu bagaimana penyebaran pengelompokan penyakit menular di Pulau Jawa tahun 2023 menggunakan algoritma *BCBimax* dengan *threshold median* dan *rata-rata* dalam binarisasi data, mengingat urgensi penanganan berbasis data di wilayah padat penduduk dengan risiko penyebaran

penyakit menular yang tinggi dan bertujuan untuk menganalisis dan memvisualisasikan penyebaran pengelompokan penyakit menular di Pulau Jawa tahun 2023 menggunakan algoritma *BCBimax* dengan binarisasi *threshold median* dan *rata-rata* di wilayah padat penduduk dengan risiko penyebaran penyakit menular yang tinggi. Serta manfaat yang dapat diambil dari dilakukannya penelitian ini adalah diharapkan penelitian ini dapat memberikan kontribusi dalam memahami pola pemetaan pengelompokan penyakit menular di Pulau Jawa tahun 2023 dan dapat menambah wawasan memperkaya literatur tentang penerapan metode *BCBimax* dalam analisis kesehatan di Pulau Jawa.

B. Metode

Penelitian ini menggunakan data sekunder yang bersumber dari portal Badan Pusat Statistik dan Open data Provinsi tahun 2023 yang diterbitkan pada tahun 2024 dengan data berjumlah 100 Kabupaten atau Kota pada 3 Provinsi di Pulau Jawa yaitu Jawa Barat, Jawa Tengah, dan Jawa Timur. Dengan penyakit menular sebagai peubah yaitu HIV/AIDS, TBC, DBD, Kusta, Diare, dan Campak. Karena data yang digunakan pada penelitian adalah data yang bersifat non-biner, maka dilakukan satu tahap sebelum melanjutkan ke analisis yaitu binarisasi data terlebih dahulu atau dengan kata lain mengubah data numerik atau non-biner ini menjadi data biner, gunanya untuk memenuhi syarat metode biclustering yang harus memakai data biner dalam tahapan analisisnya.

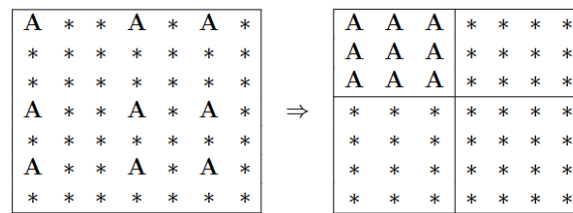
Berikut bagan yang berisi mengenai tahapan penelitian yang dilakukan dengan menggunakan data 100 Kabupaten atau Kota di 3 provinsi di pulau Jawa tahun 2023 dengan Algoritma *BCBimax*.



Gambar 1 tahapan analisis penelitian dengan Algoritma *BCBimax*

Konsep biclustering pertama kali diperkenalkan Hartigan (1972) dengan nama direct clustering, kemudian diaplikasikan pada data ekspresi gen (microarray DNA) oleh Cheng & Church (2000). Secara esensial, biclustering bertujuan untuk mengidentifikasi subset atau submatriks objek dan

variabel yang menunjukkan pola serupa namun berbeda dengan subset atau submatriks lainnya Kaiser (2011).



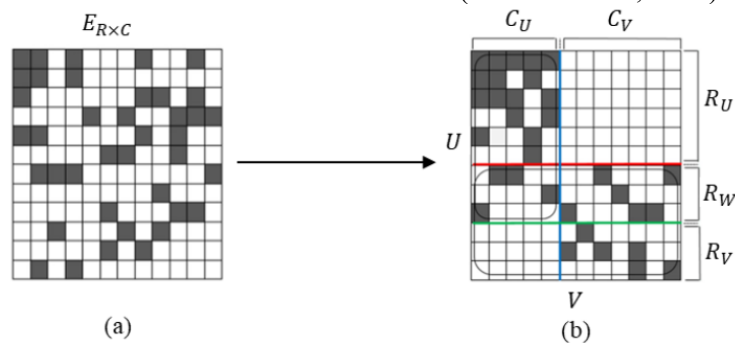
Gambar 2. ilustrasi biclustering oleh kaiser (2011)

Tahap pertama sebelum melakukan analisis menggunakan Algoritma BCBimax, penelitian ini melalui tahapan normalisasi data terlebih dahulu menggunakan Z-Score rata-rata, dengan rumus sebagai berikut:

$$Z_{rata-rata} = \frac{x_i - \text{rerata}}{\text{simp.baku}} \quad \dots(1)$$

dengan $Z_{rata-rata}$ adalah data yang telah dinormalisasi atau data scalling, x_i adalah nilai data awal, rerata adalah rata-rata peubah, dan simp.baku adalah simpangan baku peubah. Setelah itu, data scalling ini akan digunakan dalam membinarisasi data dengan threshold atau ambang batas rata-rata dan median dari setiap variabel peubah. Dimana jika data scalling masing-masing peubah lebih dari angka threshold atau ambang batas median dan rata-rata masing-masing peubah maka dimisalkan dengan satu (1) dan jika data scalling masing-masing peubah kurang dari angka ambang batas median dan rata-rata masing-masing peubah maka dimisalkan dengan nol (0).

Untuk konsep Algoritma BCBimax itu sendiri pertama kali diperkenalkan oleh (Prelić et al., 2006) yang merupakan metode yang dikembangkan dari Algoritma Bimax yang gunanya untuk mengidentifikasi submatriks bernilai 1 dalam data biner. Algoritma ini termasuk kedalam kategori divide and conquer, yang dimana metode ini bekerja secara berulang dengan mempartisi matriks utama menjadi submatriks hingga menemukan dengan nilai seragam satu (1) dan melakukan penghapusan pada elemen yang bernilai nol atau yang tidak bisa masuk kriteria pembentukan bicluster. Metode BCBimax dan Bimax memiliki perbedaan yang terletak pada kemampuan meminimalkan overlap antar bicluster melalui penghapusan data yang tidak termasuk kriteria bicluster yang dimana metode BCBimax lebih bisa meminimalkan overlap dibanding Bimax. Proses kerja algoritma BCBimax dalam mengidentifikasi bicluster bernilai 1 divisualisasikan oleh (Dolnicar et al., 2012) sebagai berikut:



Gambar 3. ilustrasi algoritma BCBimax (Dolnicar et al., 2012)

Algoritma BCBimax mengandalkan data scalling menjadi matriks biner sebagai tahap pokok yang menentukan jumlah bicluster. Proses binarisasi ini menjadi sangat penting karena memengaruhi efektivitas partisi matriks menjadi tiga submatriks berbeda, dimana submatriks berisi elemen nol seluruhnya akan dieliminasi. Mekanisme kerja algoritma ini dapat diamati melalui Gambar 3 yang mengilustrasikan pembagian matriks biner (dengan R merepresentasikan kumpulan baris dan C sebagai kolom, dimana nilai 1 dan 0 masing-masing ditunjukkan oleh kotak hitam dan putih) menjadi tiga bagian sebelum akhirnya dipartisi lebih lanjut menjadi dua submatriks utama U dan V.

Proses penilaian performa algoritma biclustering dilakukan melalui dua pendekatan utama yaitu evaluasi intrabicluster dan interbicluster. Menurut Aidi et al., (2023), evaluasi intrabicluster

menggunakan Average Squared Residue (ASR) yang berasal dari perhitungan Mean Squared Residue (MSR) bertujuan untuk mengukur kualitas suatu bicluster. Indikator kualitas ini berbanding terbalik dengan nilai ASR dan MSR, dimana semakin kecil nilai ASR (mendekati nol), semakin baik kualitas bicluster yang dihasilkan. Penelitian Nur Hikmah (2024) memperkuat pendapat ini dengan menyatakan bahwa nilai MSR yang rendah menunjukkan tingkat homogenitas yang tinggi dalam bicluster. Adapun persamaan matematis untuk menghitung MSR telah diusulkan oleh Cheng & Church (2000) sebagai berikut:

$$MSR(R, C) = \frac{1}{|R||C|} \sum_{r \in R, c \in C} (x_{rc} - \bar{x}_{ri} - \bar{x}_{cj} + \bar{x}_{RC})^2 \quad \dots(2)$$

Dan untuk mengevaluasi kualitas gabungan dari n bicluster yang dihasilkan memakai ASR dengan rumus sebagai berikut:

$$ASR = \frac{1}{n} (\sum_{i=1}^n MSR(R, C)) \quad \dots(3)$$

Indeks (Liu & Wang, 2007) digunakan sebagai metrik evaluasi interbicluster untuk mengukur tingkat kesamaan masing-masing gerombol berdasarkan kombinasi maximum baris dan kolom yang terbentuk. Nilai indeks ini berkisar antara 0 hingga 1, di mana nilai yang lebih tinggi menunjukkan kualitas pengelompokan yang lebih baik. Untuk dua skenario hasil biclustering, $Mopt$ dan M , persamaannya dapat dirumuskan sebagai berikut:

$$I_{liu\&wang}(Mopt, M) = \frac{1}{Kopt} \sum_{i=1}^{Kopt} \left(\frac{[R_i \cap R_j] + [C_i \cap C_j]}{[R_i \cup R_j] + [C_i \cup C_j]} \right) \quad \dots(4).$$

C. Hasil Penelitian dan Pembahasan

Bagian Normalisasi data (Data scalling)

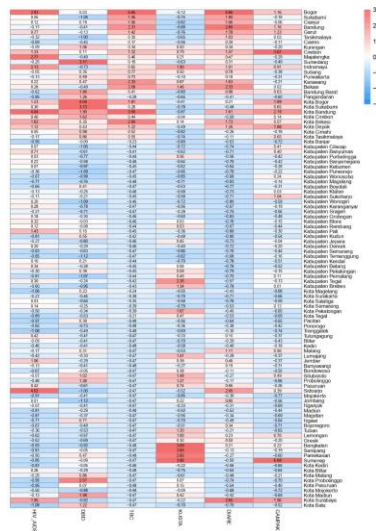
Langkah pertama yang perlu dilakukan yaitu normalisasi data dengan menggunakan Z-Score pada penelitian ini memakai rata-rata setiap peubah. Dengan hasil tabel sebagai berikut:

Tabel 1. hasil data perhitungan Z-score rata-rata setiap peubah

Kab/kota	HIV/AIDS	DBD	TBC	Kusta	Diare	Campak
Bogor	2.9331	0.0296	5.6462	-0.1244	4.6597	1.1583
Sukabumi	0.0541	-1.0435	1.9550	-0.7445	1.9008	-0.1801
Cianjur	0.1158	0.1886	1.5551	-0.8196	1.6623	-0.0580
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Kota Batu	-1.0770	1.2220	-0.4734	-0.6995	-0.7020	-0.5525

Eksplorasi Data Scalling dan binarisasi data

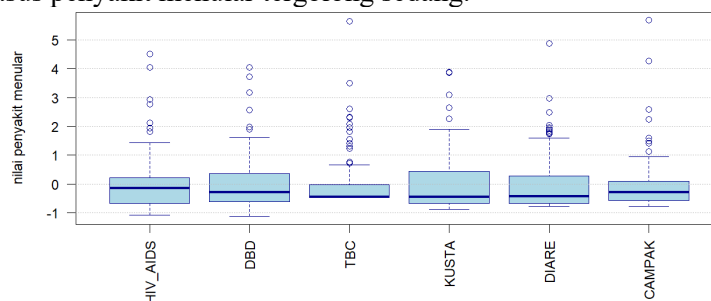
Eksplorasi awal terhadap matriks data scalling ini divisualisasikan dengan heatmap untuk mendapatkan karakteristik dasar data yang sekiranya mungkin dapat menjadi informasi bahwa ada wilayah dan penyakit mana saja yang bisa termasuk ke dalam satu bicluster yang sama.



Gambar 4. Heatmap matriks data scalling

Berdasarkan *heatmap* pada gambar 4, terlihat bahwa jumlah kasus penyakit menular di 100 Kabupaten/Kota di 3 provinsi pulau jawa tahun 2023 terbagi menjadi 3 kategori warna. Warna oren menandakan provinsi dengan tingkat penyakit menular yang tinggi, sementara warna *peach* ke putih menandakan tingkat penyakit menular yang sedang, dan Kabupaten/Kota yang memiliki tingkat penyakit menular rendah ditandai dengan warna biru. Daerah Kabupaten Bogor memiliki jumlah kasus penyakit HIV/AIDS, TBC, dan Diare paling tinggi di 3 Provinsi pulau jawa yang ditandai dengan warna oren, dan juga 25 Kabupaten/Kota teratas lainnya memiliki jumlah kasus penyakit menular yang tinggi dan cenderung sedang mengacu pada warna oren pekat dan *peach*.

Melalui visualisasi *boxplot* pada gambar 5, diketahui terdapat enam peubah dalam penelitian yang memiliki penciran. Dimana rata-rata pada datanya hampir seluruh peubahnya bernilai nol yang berarti bahwa jumlah kasus penyakit menular tergolong sedang.



Gambar 5. Boxplot matriks data scalling

Selanjutnya dari data scalling tersebut diubah menjadi data biner atau binarisasi data. Jika data lebih dari angka rata-rata dan mediannya maka dimisalkan dengan 1 dan jika angkanya kurang dari rata-rata dan mediannya maka dimisalkan dengan 0. Seperti tabel 2 dan tabel 3 berikut:

Tabel 2. data binarisasi dengan threshold median setiap peubah

Kab/kota	HIV/AIDS	DBD	TBC	Kusta	Diare	Campak
Bogor	1	1	1	1	1	1
sukabumi	1	0	1	0	1	1
Cianjur	1	1	1	0	1	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Kota Batu	0	1	0	0	0	0

Tabel 3. data binarisasi dengan threshold rata-rata setiap peubah

Kab/kota	HIV/AIDS	DBD	TBC	Kusta	Diare	Campak
Bogor	1	1	1	0	1	1
sukabumi	1	0	1	0	1	1
Cianjur	1	1	1	0	1	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Kota Batu	0	1	0	0	0	0

Analisis dan Pembahasan Algoritma BCBimax

Pada algoritma BCBimax diperlukan tahapan binarisasi data terlebih dahulu jika datanya bersifat non-biner seperti yang sudah dijelaskan sebelumnya, dikarenakan dalam tahapannya data harus berbentuk biner. Dalam penelitian ini, proses binarisasi data dilakukan dengan 2 skenario yaitu dengan mengambil ambang batas *rata-rata* dari masing-masing peubah dan *median* dari masing-masing peubah, data dimisalkan elemen 1 jika melebihi angka *rata-rata* dan *median* setiap peubah dan data dimisalkan dengan 0 jika kurang dari angka *rata-rata* dan *median* setiap peubah. Hal ini bertujuan untuk membandingkan proporsi elemen nol (0) atau seberapa besar angka 0 yang terbentuk pada proses binarisasi.

Tabel 4. perbandingan skenario kriteria binarisasi dalam pembentukan biclustering

No	Skenario binarisasi	Proporsi elemen "0"
1.	Median setiap peubah	50.33%
2.	Rata -rata setiap peubah	66.83%

Pada tabel 2 ini, didapatkan kisaran elemen nol (0) pada kedua skenario binarisasi yaitu untuk skenario *median* setiap peubah memiliki proporsi elemen nol (0) sebesar 50.33% dan untuk skenario *rata-rata* setiap peubah memiliki proporsi elemen nol (0) lebih besar dari skenario *median* yaitu sebesar 66.83%. Kedua skenario ini akan dilihat pengaruhnya terhadap jumlah *bicluster* yang terbentuk dan evaluasinya.

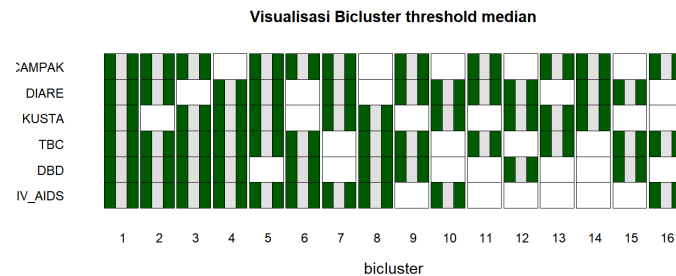
Setelah proses binarisasi, tahapan dalam algoritma BCBimax memerlukan pemilihan ambang batas minimum kombinasi baris (*minr*) dan kolom (*minc*) untuk membentuk hasil *biclustering*. Algoritma BCBimax akan melakukan pengelompokan seluruh submatriks/subkelompok dengan elemen bernilai 1 dan memaksimalkan hasil pengelompokan di awal. Berdasarkan tabel 4.4, didapatkan hasil bahwa semakin besar penentuan kombinasi ambang batas baris dan kolomnya, maka akan semakin kecil *bicluster* yang terbentuk, dan hal tersebut akan mengurangi informasi penyebaran keanggotaan yang termasuk dalam *bicluster*. Maka dari itu, perlu pemilihan ambang batas minimum baris dan minimum kolom berdasarkan kombinasi yang dicobakan.

Tabel 5. jumlah bicluster terbentuk

Kombinasi batas minimum baris dan kolom	Jumlah bicluster (data biner median)	Jumlah bicluster (data biner rata-rata)
2,2	16	11
3,3	7	2
4,4	2	1
5,5	1	0

Ambang batas minimum baris dan kolom terpilih berdasarkan banyaknya informasi atau submatriks (*bicluster*) yang terbentuk. Diketahui dari tabel 4.2 diatas, didapatkan pada kombinasi batas minimum baris dan kolom 2,2 yang menghasilkan informasi atau submatriks (*bicluster*) berjumlah 16 *bicluster* pada skenario binarisasi *median* dan 11 *bicluster* pada skenario binarisasi *rata-rata*. Maka hasil *bicluster* dari kombinasi minimum baris dan kolom 2,2 dipilih sebagai jumlah *bicluster* yang akan lanjut diteliti dan dilihat hasil evaluasinya.

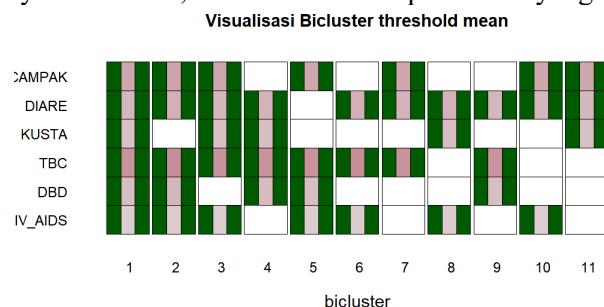
Karakteristik keanggotaan hasil *bicluster* dengan skenario *threshold median* dan kombinasi minimum baris dan kolom 2,2 yang terdapat pada gambar, diketahui bahwa dari 100 Kabupaten/kota dan 6 penyakit menular, terbentuk 54 Kabupaten/kota yang tersebar dalam 16 *bicluster*.



Gambar 6. grafik hasil bicluster threshold median

Dari Gambar 6, dapat diinterpretasikan bahwa, *bicluster* 1 terdiri dari 6 kabupaten/kota yaitu Bogor, Kuningan, Cirebon, Karawang, Kota Bogor, dan Kota Bekasi dengan 6 penyakit menular yaitu HIV/AIDS, DBD, TBC, Kusta, Diare, dan Campak. Dalam *bicluster* 1 tersebut memperlihatkan bahwa kabupaten/kota yang terkelompokan memiliki kemiripan yang tinggi di seluruh kasus penyakit menular sehingga butuh penanganan yang lebih serius untuk menangani kasus penyakit tersebut. dan bicluster seterusnya sampai dengan bicluster 16 yang terdiri dari Kabupaten Cilacap dan Kabupaten Banyumas dengan kasus penyakit menular HIV/AIDS, TBC, Campak. Dan untuk 46 baris wilayah yang tidak masuk kedalam bicluster merupakan noise atau tidak memenuhi kriteria untuk masuk terkelompokan kedalam bicluster yang terbentuk berdasarkan minimum baris dan kolom yang dicobakan.

Selanjutnya, karakteristik keanggotaan hasil bicluster dengan skenario threshold rata-rata dan kombinasi minimum baris dan kolom 2×2 yang terdapat pada tabel 4.6, diketahui bahwa dari 100 Kabupaten/kota dan 6 penyakit menular, terbentuk 29 Kabupaten/kota yang tersebar dalam 11 *bicluster*.



Gambar 7. visualisasi hasil bicluster threshold rata-rata

Dari gambar 7, dapat diinterpretasikan bahwa bicluster 1 terdiri dari Cirebon dan Kota Bekasi dengan 6 kasus penyakit yaitu HIV/AIDS, DBD, TBC, Kusta, Diare, dan Campak, sehingga dalam bicluster ini membutuhkan penanganan yang lebih serius karena kabupaten/kota yang terkelompokan memiliki kemiripan yang tinggi di seluruh kasus penyakit menular. Dan seterusnya sampai bicluster 11 terdiri dari Lamongan dan Bangkalan dengan kasus penyakit menular Kusta, Diare, dan Campak. Dan 71 baris wilayah yang tidak masuk kedalam bicluster merupakan noise atau tidak memenuhi kriteria untuk masuk terkelompokan kedalam bicluster yang terbentuk berdasarkan minimum baris dan kolom yang dicobakan.

Evaluasi Kinerja Algoritma BCBimax

Pada tabel 6, didapatkan hasil evaluasi *intrabicluster* dengan *average residu* (ASR) dari *Mean Square Residue* (MSR) yang dapat dilihat pada lampiran 9, sebesar 0 pada kedua skenario binarisasi yaitu median dan rata-rata serta pada kombinasi minimum baris dan kolom 2,2, dikarenakan nilai dari ASR nya sama, maka dilihat pada hasil evaluasi *interbicluster* nya untuk menyimpulkan. Dan untuk hasil evaluasi *interbicluster* dengan indeks Liu and Wang pada lampiran 10, didapatkan hasil untuk skenario binarisasi median setiap peubah jika diasumsikan optimal terhadap skenario rata-rata setiap peubah sebesar 55.59%, dan hasil untuk skenario rata-rata setiap peubah jika diasumsikan optimal

terhadap skenario median setiap peubah sebesar 63.38%.

Tabel 6. perbandingan nilai evaluasi bicluster

Skenario	Liu & wang	Jumlah bicluster	ASR	Keanggotaan baris
Median setiap peubah diasumsikan optimal	55.59%	16	0	54
Rata-rata setiap peubah diasumsikan optimal	63.38%	11	0	29

Berdasarkan nilai perbandingan tersebut, dapat diinterpretasikan bahwa skenario binarisasi dengan rata-rata setiap peubah diasumsikan optimal lebih baik. Dan pada kedua skenario bicluster yang terbentuk, tidak ditemukan adanya overlap/tumpang tindih antara kabupaten/kota dan penyakit, sehingga membuktikan bahwa tahap binarisasi yang diterapkan pada metode algoritma BCBimax berhasil menghindari tumpang tindih.

D. Kesimpulan

Hasil penelitian menunjukkan bahwa metode biclustering BCBimax dapat mengelompokkan kabupaten atau kota di pulau jawa beserta kasus penyakit menular secara simultan. Algoritma BCBimax berhasil mengelompokkan kabupaten atau kota di 3 provinsi pulau jawa dengan 2 skenario yaitu binarisasi median setiap peubah dan binarisasi rata-rata setiap peubah serta pemilihan ambang batas minimum baris dan kolom sebesar (2,2). Untuk skenario dengan data binarisasi median setiap peubah menghasilkan 16 buah bicluster yang bersifat tidak tumpang tindih dengan nilai ASR sebesar 0. Hasil bicluster 1 pada skenario data binarisasi median setiap peubah memaksimalkan pengelompokan 6 kabupaten/kota yaitu Bogor, Kuningan, Cirebon, Karawang, Kota Bogor, dan Kota Bekasi dengan 6 penyakit menular yaitu HIV/AIDS, DBD, TBC, Kusta, Diare, dan Campak. Sedangkan skenario data binarisasi rata-rata setiap peubah menghasilkan 11 buah bicluster yang juga bersifat tidak tumpang tindih dengan ASR sebesar 0. Hasil bicluster 1 pada skenario binarisasi rata-rata setiap peubah memaksimalkan pengelompokan pada 2 Kab/kota yaitu Cirebon dan Kota Bekasi dengan 6 kasus penyakit yaitu HIV/AIDS, DBD, TBC, Kusta, Diare, dan Campak. Pada skenario binarisasi median setiap peubah, penyakit TBC merupakan penyakit yang terkelompok paling banyak yaitu 12 bicluster. Dan pada skenario rata-rata setiap peubah, penyakit Diare paling banyak terkelompok yaitu 10 bicluster. Perbandingan persentase kemiripan anggota berdasarkan nilai indeks Liu and Wang dari kedua skenario Algoritma BCBimax menunjukan angka yang cukup jauh yaitu 55.59% dan 63.38%, sehingga dapat disimpulkan dalam minimum baris dan kolom (2,2), skenario binarisasi rata-rata setiap peubah lebih baik atau optimal dibandingkan skenario binarisasi median setiap peubah.

Ucapan Terimakasih

Penulis mengucapkan terima kasih kepada pembimbing Bapak Ilham Faishal Mahdy S.Stat., M.Stat yang telah memberikan bimbingan, motivasi, dan saran selama proses penelitian ini. Terima kasih terbesar juga penulis sampaikan kepada orang tua yang selalu memberikan dukungan dan doa tanpa henti. Selain itu, penulis mengapresiasi portal pengumpulan data penelitian BPS dan Open data Provinsi yang menyediakan sumber informasi yang sangat membantu kelancaran penelitian ini. Semoga hasil penelitian ini bermanfaat bagi pihak yang membutuhkan.

Daftar Pustaka

Aidi, M. N., Wulandari, C., Oktarina, S. D., Aditra, T. R., Ernawati, F., Efriwati, Nurjanah, N., Rachmawati, R., Julianti, E. D., Sundari, D., Retiaty, F., Arifin, A. Y., Dewi, R. M., Nazaruddin, N., Salimar, Fuada, N., Widodo, Y., Setyawati, B., Nurhidayati, N., ...

- Widoretno. (2023). Province clustering based on the percentage of communicable disease using the BCBimax biclustering algorithm. *Geospatial Health*, 18(2). <https://doi.org/10.4081/gh.2023.1202>
- Ajeng Mega Pratiwi, & Mutaqin, A. K. (2021). Penerapan Algoritma Naïve Bayes Classifier dalam Memprediksi Status Keberlanjutan Polis Nasabah Asuransi PT.X. *Jurnal Riset Statistika*, 1(2), 117–126. <https://doi.org/10.29313/jrs.v1i2.435>
- Cheng Y, C. G. (2000). *biclustering of expression data*.
- Cynthia Wulandari. (2023). *EVALUASI HASIL ANALISIS BICLUSTER PROVINSI DAN POTENSI PERIKANAN TANGKAP DI INDONESIA, ANTARA ALGORITME BCBIMAX DAN ISA*.
- Dolnicar, S., Kaiser, S., Lazarevski, K., & Leisch, F. (2012). Biclustering: Overcoming data dimensionality problems in market segmentation. *Journal of Travel Research*, 51(1), 41–49. <https://doi.org/10.1177/0047287510394192>
- J. A. Hartigan. (1972). *direct clustering of a data matrix*.
- Kaiser, S. (2011). *Biclustering: Methods, Software and Application*.
- Liu, X., & Wang, L. (2007). Computing the maximum similarity bi-clusters of gene expression data. *Bioinformatics*, 23(1), 50–56. <https://doi.org/10.1093/bioinformatics/btl560>
- Nur Hikmah. (2024). *EVALUASI KINERJA BICLUSTERING ALGORITME BCBIMAX DAN MODEL PLAID DALAM MELIHAT POLA KETAHANAN PANGAN WILAYAH INDONESIA*.
- Preli'c, A. P., Bleuler, S., Zimmermann, P., Wille, A., Bühlmann, P., Gruissem, W., Hennig, L., Thiele, L., & Zitzler, E. (2006). *Comparison of Biclustering Methods: A Systematic Comparison and Evaluation of Biclustering Methods for Gene Expression Data*. <http://www.tik.ee.ethz.ch/sop/bimax>
- Shafira Agnia Latfalia, & Reny Rian Marlina. (2024). Klasifikasi Status Indeks Desa Membangun Jawa Barat Menggunakan Algoritma XGBoost. *Jurnal Riset Statistika*, 75–82. <https://doi.org/10.29313/jrs.v4i2.5011>
- Shuai, Y., Dai, X., Wang, X., Liang, Z., Yan, G., & Chen, M. (2023). *Methods for Clinical Disease Classification and Clustering: A Review*. 116–124. <https://doi.org/10.5220/0012014700003633>
- Suhartina, N., & Sukarsih, I. (2024). Model SVEIR Penyebaran Penyakit Rabies Terhadap Anjing dengan Vaksinasi. *DataMath: Journal of Statistics and Mathematics*, 2(1), 25–32.