

1. Pendahuluan

Clustering merupakan salah satu teknik dalam *unsupervised learning* yang banyak digunakan untuk mengidentifikasi pola dan struktur dalam data dengan cara mengelompokkan objek berdasarkan tingkat kemiripan tertentu [1]. Namun, metode *clustering* konvensional memiliki keterbatasan karena hanya mampu mengelompokkan data berdasarkan seluruh variabel secara keseluruhan, sehingga sering kali tidak mampu menangkap pola lokal yang hanya muncul pada sebagian variabel tertentu [2]. Untuk mengatasi keterbatasan tersebut, dikembangkan pendekatan *biclustering* yang memungkinkan pengelompokan secara simultan pada dimensi baris (objek) dan kolom (variabel), sehingga mampu mengidentifikasi pola tersembunyi dalam submatriks data [3].

Salah satu algoritma awal dalam *biclustering* adalah metode *Cheng & Church* (CC), yang berfokus pada pencarian *bicluster* dengan tingkat koherensi tinggi berdasarkan ukuran *Mean Squared Residue* (MSR) [4]. Metode ini dikenal efektif dalam menemukan pola yang homogen serta toleran terhadap noise dalam data [5]. Di sisi lain, metode *Iterative Signature Algorithm* (ISA) juga banyak digunakan karena kemampuannya dalam mengidentifikasi pola kompleks dan hubungan yang bersifat *nonlinier* melalui proses iteratif berbasis ambang batas [6]. Beberapa penelitian menunjukkan bahwa ISA unggul dalam menghasilkan *bicluster* yang stabil dan mampu menangkap hubungan yang tumpang tindih antar variabel [7], sedangkan metode CC lebih baik dalam menghasilkan pola yang koheren dan konsisten [8]. Perbedaan temuan tersebut menunjukkan bahwa belum terdapat kesimpulan yang konsisten mengenai metode *biclustering* yang paling optimal. Selain itu, belum diketahui apakah keunggulan masing-masing metode akan tetap memberikan hasil yang sama ketika diterapkan pada data kesehatan masyarakat yang memiliki karakteristik berbeda, khususnya data faktor-faktor yang memengaruhi penyebaran penyakit kusta di Pulau Sulawesi. Oleh karena itu, diperlukan penelitian yang membandingkan kinerja metode CC dan ISA untuk mengidentifikasi metode yang lebih optimal dalam menghasilkan *bicluster* yang koheren, terpisah dengan baik, dan homogen pada konteks data tersebut.

Permasalahan kesehatan masyarakat, khususnya penyakit kusta, merupakan salah satu isu yang masih signifikan di Indonesia. Kusta adalah penyakit menular yang disebabkan oleh *Mycobacterium leprae* dan dapat menimbulkan kecacatan permanen apabila tidak ditangani dengan baik [9]. Berdasarkan Profil Kesehatan Indonesia Tahun 2024, jumlah kasus baru kusta yang ditemukan mencapai 12.798 kasus. Tingginya jumlah kasus tersebut menempatkan Indonesia sebagai salah satu negara dengan beban kusta yang masih tinggi. Selain itu, beberapa provinsi di kawasan timur Indonesia masih menunjukkan angka prevalensi yang relatif tinggi, termasuk provinsi-provinsi di Pulau Sulawesi seperti Sulawesi Tengah, Sulawesi Utara, dan Gorontalo yang memiliki jumlah kasus kusta cukup signifikan [10]. Penyebaran penyakit kusta dipengaruhi oleh berbagai faktor sosial, ekonomi, dan lingkungan yang saling berinteraksi secara kompleks dan tidak selalu bersifat linier. Kondisi ini menyebabkan analisis konvensional kurang mampu menggambarkan pola hubungan antar wilayah dan faktor secara mendalam.

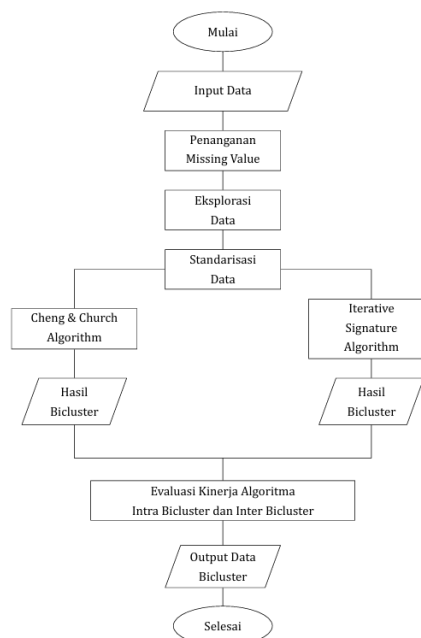
Meskipun pendekatan *biclustering* dinilai mampu mengidentifikasi pola hubungan antara wilayah dan faktor penyebab secara simultan, hasil penelitian terdahulu menunjukkan bahwa belum terdapat kesimpulan yang konsisten mengenai metode yang paling optimal. Beberapa penelitian melaporkan bahwa ISA lebih unggul dalam menghasilkan *bicluster* yang stabil dan mampu menangkap hubungan yang tumpang tindih, sedangkan penelitian lainnya menunjukkan bahwa CC lebih baik dalam menghasilkan *bicluster* yang koheren dan homogen. Perbedaan temuan tersebut menimbulkan pertanyaan mengenai metode yang lebih sesuai ketika diterapkan pada data kesehatan masyarakat yang memiliki karakteristik berbeda, khususnya data faktor-faktor yang memengaruhi penyebaran penyakit kusta di Pulau Sulawesi. Sejauh

penelusuran literatur yang dilakukan, penelitian yang secara khusus membandingkan metode *Cheng & Church* dan *Iterative Signature Algorithm* pada data faktor penyebaran penyakit kusta di Pulau Sulawesi masih terbatas. Oleh karena itu, kebaruan penelitian ini terletak pada penerapan dan perbandingan kedua metode *biclustering* tersebut dalam konteks data kesehatan masyarakat untuk mengidentifikasi pola keterkaitan antara wilayah dan faktor penyebab kusta. Berdasarkan uraian tersebut, rumusan masalah dalam penelitian ini adalah metode *biclustering* manakah yang lebih optimal antara *Cheng & Church* dan *Iterative Signature Algorithm* dalam mengelompokkan wilayah berdasarkan faktor-faktor yang memengaruhi penyebaran penyakit kusta di Pulau Sulawesi berdasarkan ukuran evaluasi MSR, *indeks Liu & Wang*, dan *Variance*.

2. Bahan dan Metode

Penelitian ini menggunakan data sekunder tahun 2024 yang bersumber dari publikasi Badan Pusat Statistik (BPS), yaitu Provinsi dalam Angka 2025 dari enam provinsi di Pulau Sulawesi (Sulawesi Tengah, Sulawesi Tenggara, Sulawesi Utara, Sulawesi Barat, Sulawesi Selatan, dan Gorontalo) serta publikasi Statistik Kesejahteraan Rakyat 2024. Unit analisis dalam penelitian ini adalah seluruh kabupaten dan kota di Pulau Sulawesi. Variabel yang digunakan meliputi kepadatan penduduk, persentase penduduk miskin, dan jumlah tenaga medis yang diperoleh dari publikasi Provinsi dalam Angka 2024, sedangkan variabel persentase rumah tangga dengan sanitasi layak, akses air minum layak, penggunaan fasilitas buang air besar bersama, rumah tangga dengan lantai tanah, dan kepemilikan jaminan kesehatan diperoleh dari publikasi Statistik Kesejahteraan Rakyat 2024. Secara keseluruhan, penelitian ini menggunakan delapan variabel yang diduga berhubungan dengan penyebaran penyakit kusta.

Penelitian ini menggunakan pendekatan kuantitatif komparatif untuk membandingkan performa algoritma *Cheng & Church* (CC) dan *Iterative Signature Algorithm* (ISA) dalam mengelompokkan wilayah berdasarkan faktor penyebab kusta di Sulawesi. Evaluasi dilakukan menggunakan metrik *Mean Squared Residue* (MSR) dan *indeks Liu & Wang* untuk menentukan metode yang menghasilkan *bicluster* paling homogen dengan pemisahan yang jelas. Tahapan penelitian meliputi preprocessing data, penanganan *missing value* dengan mean imputation, standarisasi data, dan aplikasi kedua metode *biclustering*. Alur penelitian yang ditunjukkan pada **Gambar 1**.



Gambar 1. Alur Penelitian

Sebagaimana ditunjukkan pada Gambar 1, Tahapan penelitian dimulai dengan pengumpulan data, identifikasi variabel, dan pemeriksaan kualitas data. Selanjutnya dilakukan data *preprocessing* melalui mean imputation untuk menangani *missing value* serta standarisasi data. Selanjutnya dilakukan analisis dengan menerapkan metode *biclustering* CC dan ISA secara terpisah untuk mengidentifikasi pola hubungan antara wilayah dan faktor penyebab penyakit kusta. Tahap akhir dilakukan evaluasi hasil *bicluster* dilakukan menggunakan metrik *Mean Squared Residue* (MSR) untuk mengukur koherensi internal dan *indeks Liu & Wang* untuk mengukur pemisahan antar *bicluster*. Hasil dari kedua metode kemudian dibandingkan guna menentukan metode yang paling efektif dalam menghasilkan *bicluster* yang homogen dengan pemisahan yang jelas.

2.1. Preprocessing Data

Tahap *preprocessing* bertujuan memastikan kelengkapan dan konsistensi data. Penanganan *missing value* dilakukan menggunakan metode *mean imputation*, yaitu mengganti nilai hilang dengan rata-rata dari nilai yang tersedia pada masing-masing variabel. Metode ini dipilih karena sederhana, efektif, dan mampu mempertahankan karakteristik data ketika proporsi nilai hilang relatif kecil [11][12]. Proses penanganan *missing value* melibatkan persamaan berikut.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

dengan \bar{x} merupakan nilai rata-rata sampel (mean) yang diperoleh dari seluruh nilai observasi dalam suatu variabel. Selanjutnya, x_i adalah nilai observasi ke- i dari suatu variabel yang digunakan dalam perhitungan rata-rata tersebut. Adapun n merupakan jumlah sampel yang tidak hilang atau jumlah data valid yang digunakan dalam proses perhitungan rata-rata sampel. merupakan nilai rata-rata sampel (mean) yang diperoleh dari seluruh nilai observasi dalam suatu variabel.

2.2. Standarisasi Data

Standarisasi data adalah proses transformasi untuk menyamakan skala antar variabel sehingga memiliki rata-rata nol dan standar deviasi satu [13]. Tahap ini penting dilakukan ketika variabel memiliki satuan atau rentang nilai yang berbeda, karena perbedaan skala dapat menyebabkan bias dalam proses *biclustering*. Standarisasi membuat setiap variabel memiliki kontribusi yang setara sehingga hasil analisis lebih akurat. Proses standarisasi data melibatkan persamaan berikut.

$$Z = \frac{x_i - \bar{x}}{\sigma} \quad (2)$$

dengan Z adalah z -score dari suatu variabel, x nilai data dari suatu variabel, \bar{x} adalah rata-rata data (mean) dari suatu variabel, dan σ adalah standar deviasi dari suatu variabel.

2.3. Cheng & Church

Algoritma *Cheng & Church* (CC) adalah metode *biclustering* yang menemukan submatriks dengan pola nilai yang koheren pada baris dan kolom. Metode ini mengukur keseragaman pola menggunakan *Mean Squared Residue* (MSR), di mana nilai MSR yang lebih kecil menunjukkan tingkat koherensi yang lebih tinggi. Dalam implementasinya, algoritma CC memerlukan penentuan nilai ambang batas (*residue threshold*) δ yang digunakan untuk mengontrol tingkat koherensi *bicluster* yang terbentuk. Semakin kecil nilai δ , semakin ketat kriteria pembentukan *bicluster* sehingga pola yang dihasilkan cenderung lebih homogen. Sebaliknya, nilai δ yang lebih besar memungkinkan terbentuknya *bicluster* dengan ukuran lebih besar, namun tingkat koherensinya dapat menurun. Nilai δ ditentukan melalui pendekatan *trial and error* dengan mempertim-

bangkan jumlah *bicluster* yang terbentuk serta nilai rata-rata MSR yang dihasilkan [14]. Perhitungan msr pada metode ini melibatkan persamaan berikut.

$$H(I,J) = \frac{1}{|I||J|} \sum_{i \in I} \sum_{j \in J} (a_{ij} - a_{iJ} - a_{Ij} + a_{IJ})^2 \quad (3)$$

dengan $H(I,J)$ merupakan nilai *Mean Squared Residue* (MSR) dari sebuah *bicluster* yang dibentuk oleh himpunan baris I dan himpunan kolom J . Selanjutnya, a_{ij} adalah nilai elemen data yang berada pada baris ke- i dan kolom ke- j . a_{iJ} merupakan rata-rata nilai pada baris ke- i terhadap seluruh kolom yang terdapat dalam himpunan J , sedangkan a_{Ij} adalah rata-rata nilai pada kolom ke- j terhadap seluruh baris yang terdapat dalam himpunan I . Adapun a_{IJ} merupakan rata-rata keseluruhan dari semua elemen data yang berada di dalam *bicluster* (submatriks) yang dibentuk oleh himpunan baris I dan kolom J .

Secara formal, dengan matriks data A dan ambang batas $\delta > 0$, algoritma CC bertujuan untuk menemukan δ -*bicluster*, yaitu subset baris I dan subset kolom J yang memiliki nilai MSR tidak melebihi δ [15]. Proses pencarian dilakukan secara iteratif dengan meminimalkan nilai MSR sehingga diperoleh *bicluster* yang homogen dan memiliki pola yang konsisten. Tahapan dari algoritma CC [5] yaitu :

1. Inisialisasi *bicluster*, menetapkan matriks awal data input (A) dan ambang batas δ .
2. Tahap penghapusan beberapa simpul, menghapus baris dan kolom berdasarkan rata-rata residu kuadrat baris ($d(i)$) dan kuadrat kolom ($d(j)$) yang lebih besar dari $1,5 \times$ rata-rata residu kuadrat seluruh matriks ($a(H(I,J))$), asalkan memenuhi kondisi residu kuadrat rata-rata ($MSR \leq \delta$).
3. Fase penghapusan simpul tunggal, menghapus baris atau kolom berdasarkan kondisi $d(i)$ atau $d(j)$ yang memiliki nilai maksimum, asalkan masih memenuhi kondisi $MSR \leq \delta$.
4. Tahap penambahan node, menambahkan baris dan kolom berdasarkan kondisi $d(i) \leq H(I,J)$ dan $d(j) \leq H(I,J)$, asalkan penambahan node menghasilkan $H(I,J)$ yang tetap $\leq \delta$.
5. Tahap substitusi, mengganti elemen matriks hasil *bicluster* dengan bilangan acak untuk mencegah tumpang tindih antar *bicluster*.
6. Iterasi, ulangi langkah 1 sampai 5 sebanyak k kali, yaitu sebanyak jumlah *bicluster* yang ingin ditemukan.

2.4. Iterative Signature Algorithm

Metode *Iterative Signature Algorithm* (ISA) adalah metode *biclustering* yang menghasilkan *bicluster* berupa *Transcription Module* (TM), yaitu subset baris dan kolom dengan tingkat kesamaan tertentu berdasarkan ambang batas [16]. Metode ini menggunakan pendekatan iteratif dengan ruang vektor dan pemetaan linear untuk mengidentifikasi submatriks yang memiliki korelasi tinggi antar baris dan kolom [5]. Proses pencarian *bicluster* dilakukan melalui pembaruan skor baris dan kolom secara bergantian hingga konvergen. Skor baris dihitung berdasarkan rata-rata nilai pada kolom yang memenuhi ambang batas, sedangkan skor kolom dihitung berdasarkan rata-rata nilai pada baris terpilih. Hal ini memungkinkan ISA mengidentifikasi pola kompleks dan hubungan saling terkait antar variabel.

Dalam implementasinya, metode ISA memerlukan penentuan beberapa parameter, yaitu ambang baris (*row threshold*, (t_c)), ambang kolom (*column threshold*, (t_g)), dan jumlah *seed* awal. Parameter (t_c) digunakan untuk menentukan objek yang dipertahankan dalam *bicluster* berdasarkan tingkat keterkaitannya terhadap pola yang terbentuk, sedangkan parameter (t_g) digunakan untuk menyeleksi variabel yang memiliki kontribusi signifikan dalam *bicluster*. Nilai ambang yang lebih tinggi akan

menghasilkan bicluster yang lebih selektif, sementara nilai yang lebih rendah cenderung menghasilkan bicluster yang lebih besar tetapi kurang spesifik.

Nilai (t_c) dan (t_g) ditentukan melalui pendekatan *trial and error* dengan mempertimbangkan jumlah bicluster yang terbentuk serta kualitas bicluster berdasarkan ukuran evaluasi yang digunakan. Selain itu, jumlah *seed* awal ditetapkan. Penggunaan beberapa *seed* bertujuan untuk meningkatkan peluang menemukan pola bicluster yang stabil dan mengurangi ketergantungan hasil terhadap kondisi awal algoritma. Parameter yang menghasilkan bicluster dengan kualitas terbaik selanjutnya digunakan pada tahap analisis utama. Tahapan dari algoritma ISA [17] yaitu :

1. Menetapkan ambang batas baris dan kolom (t_c, t_g) , nilai benih, dan jumlah benih (n)
2. Membuat matriks baris ternormalisasi (A^g) dan matriks kolom ternormalisasi (A^c) dari matriks $A_{|U| \times |V|} \left(\begin{bmatrix} U \\ V \end{bmatrix} \right)$.
3. Memilih secara acak beberapa vector kolom (sampel kolom) $\left(\begin{bmatrix} U \\ V' \end{bmatrix} \right)$.
4. Menghitung rata-rata setiap baris dari sampel kolom $(a_{u'v}^c)$ menggunakan A^c .
5. Memilih sampel baris yang memenuhi kondisi $(a_{u'v}^c > t_c \sigma_c)$ dan nilai rata-ratanya menjadi "skor baris" $\left(\begin{bmatrix} U \\ V' \end{bmatrix} \right)$.
6. Menghitung rata-rata setiap kolom dari sampel baris $(a_{u'v}^g)$ menggunakan A^g .
7. Memilih sampel kolom yang memenuhi kondisi $a_{u'v}^g > t_g \sigma_g$ dan nilai rata-ratanya menjadi "skor kolom" $\left(\begin{bmatrix} U \\ V' \end{bmatrix} \right)$.
8. Mengulangi tahap 4 hingga 7 hingga jumlah benih (n) tercapai atau kondisi konvergen terpenuhi.
9. Ketika kondisi konvergen terpenuhi, yaitu $\frac{|V' \Delta V''|}{|V' \cup V''|} < \varepsilon$ maka baris dan kolom (bicluster) dipilih
10. Mengulangi tahap 3 hingga 9 sebanyak jumlah bicluster yang ingin dibentuk

2.5. Evaluasi Kinerja Metode

Kinerja algoritma biclustering dievaluasi melalui dua fungsi, yaitu evaluasi intra-bicluster dan inter-bicluster. Evaluasi intra-bicluster menilai tingkat koherensi atau keseragaman nilai dalam suatu bicluster. Semakin tinggi koherensi, semakin baik kualitas bicluster [18]. Salah satu ukuran yang umum digunakan adalah Mean Squared Residue (MSR), yang mengukur penyimpangan nilai dalam bicluster terhadap pola rata-ratanya.

$$MSR_{average \text{ per volume}} = \frac{1}{b} \sum_{i=1}^b \frac{MSR_i}{Volume_i} \quad (4)$$

Evaluasi inter-bicluster digunakan untuk menilai kualitas keseluruhan bicluster dengan melihat kemampuan algoritma dalam mengidentifikasi struktur data yang sebenarnya [19]. Dalam penelitian ini, indeks Liu & Wang digunakan sebagai ukuran evaluasi inter-bicluster untuk menilai tingkat pemisahan dan kesesuaian antar bicluster yang dihasilkan.

$$I_{Liu \& Wang}(M_{opt}, M) = \frac{1}{K_{opt}} \sum_{i=1}^{K_{opt}} Max \left(\frac{|G_i \cap G_j| + |G_i \cap G_j|}{|G_i \cup G_j| + |G_i \cup G_j|} \right) \tag{5}$$

Evaluasi homogenitas bicluster digunakan untuk menilai tingkat keseragaman elemen-elemen di dalam setiap bicluster yang terbentuk. Dalam penelitian ini, *Variance* (VAR) digunakan sebagai ukuran evaluasi homogenitas untuk mengukur tingkat penyebaran data terhadap rata-rata elemen pada setiap bicluster. Nilai VAR yang lebih kecil menunjukkan bahwa elemen-elemen dalam bicluster semakin homogen sehingga pola yang dihasilkan menjadi lebih konsisten dan stabil.

$$VAR = \frac{1}{|I||J|} \sum_{i=1}^{|I|} \sum_{j=1}^{|J|} (bij - bIJ)^2 \tag{6}$$

3. Hasil

Bagian ini menyajikan hasil analisis biclustering pada data faktor penyebab penyakit kusta di Pulau Sulawesi. Analisis mencakup tahapan *preprocessing* data, penerapan metode *Cheng & Church* (CC) dan *Iterative Signature Algorithm* (ISA), serta evaluasi performa menggunakan *Mean Squared Residue* (MSR) dan *indeks Liu & Wang*. Hasil disajikan dalam bentuk tabel dan visualisasi sebagai dasar pembahasan lebih lanjut.

3.1. Hasil Preprocessing Data

Berikut adalah hasil uji preprocessing data yaitu pada penanganan missing value dengan menggunakan mean imputation.

Tabel 1. Penanganan Missing Value

Kabupaten/Kota	Persentase Rumah Tangga Dengan Alas Lantai Tanah
Tojo Una-Una	1,239859
Palu	1,239859
Kepulauan Selayar	1,239859
Maros	1,239859
Barru	1,239859
Bone	1,239859
Wajo	1,239859
Kota Parepare	1,239859
Kabupaten Pohuwato	1,239859
Kolaka Utara	1,239859

Berdasarkan **Tabel 1**, terdapat 10 daerah yang memiliki missing value pada variabel Persentase Rumah Tangga dengan Alas Lantai Tanah. Untuk menjaga kelengkapan dan konsistensi data, nilai yang hilang ditangani menggunakan metode imputasi rata-rata (mean). Metode ini dipilih agar jumlah observasi tetap sama tanpa perlu menghapus data.

3.2. Hasil Standarisasi Data

Sebelum dilakukan proses biclustering, data terlebih dahulu distandarisasi menggunakan metode Z-Score. Standarisasi dilakukan karena variabel penelitian memiliki satuan dan rentang nilai yang berbeda, sehingga diperlukan penyetaraan skala agar setiap variabel memberikan kontribusi yang seimbang dalam proses analisis. Data yang digunakan terdiri atas 81 kabupaten/kota di Pulau Sulawesi sebagai unit amatan (baris) dan 8 variabel penelitian sebagai unit pengamatan (kolom).

Adapun kedelapan variabel tersebut meliputi: (X1) yaitu kepadatan penduduk, (X2) persentase sanitasi layak, (X3) persentase air minum layak, (X4) persentase penduduk miskin, (X5) persentase penduduk yang memiliki jaminan kesehatan, (X_6) persentase penggunaan tempat buang air bersama, (X7) persentase rumah tangga dengan alas lantai

tanah, dan (X8) jumlah tenaga medis. Hasil standarisasi menghasilkan data dengan rata-rata mendekati nol dan simpangan baku sebesar satu, sehingga perbedaan skala antar-variabel tidak memengaruhi proses pembentukan bicluster. Berikut adalah hasil standarisasi data dengan menggunakan perhitungan Z.Score.

Tabel 2. Standarisasi Data

Index	X1	X2	...	X7	X8
1	-0,342	-0,621	...	-0,375	-0,342
2	-0,348	-0,271	...	-0,375	-0,331
3	-0,355	-1,037	...	-0,375	-0,331
.
.
.
79	-0,333	-0,159	...	-0,375	0,164
80	-0,324	0,119	...	-0,375	-0,177
81	-0,340	-0,434	...	-0,375	-0,318

Berdasarkan **Tabel 2**, hasil standarisasi data menggunakan metode Z-score menunjukkan bahwa seluruh variabel telah berada pada skala yang sama dengan nilai berpusat di sekitar nol. Nilai positif mengindikasikan bahwa suatu daerah memiliki nilai di atas rata-rata, sedangkan nilai negatif menunjukkan nilai di bawah rata-rata. Selain itu, terlihat adanya beberapa nilai ekstrem (outlier) yang ditunjukkan oleh nilai Z-score yang cukup besar secara absolut, seperti pada variabel tertentu. Proses standarisasi ini penting untuk menghilangkan perbedaan skala antar variabel sehingga analisis lanjutan, seperti clustering, dapat dilakukan secara lebih akurat dan tidak bias.

3.3. Analisis Cheng & Church

Analisis dengan menggunakan metode Cheng and Church (CC) bertujuan untuk mengidentifikasi submatriks data yang memiliki tingkat koherensi tinggi antara baris dan kolom berdasarkan nilai Mean Squared Residue (MSR). Suatu bicluster dianggap baik apabila memiliki nilai MSR yang rendah, yang menunjukkan bahwa pola hubungan antara wilayah dan variabel di dalam bicluster tersebut relatif homogen. Analisis diawali dengan penentuan nilai ambang batas (δ) sebagai parameter utama dalam algoritma Cheng & Church. Nilai δ ditentukan melalui trial and error dengan menguji rentang nilai 0,1–1,0 untuk memperoleh bicluster yang optimal. Hasil pengujian tersebut disajikan pada tabel berikut.

Tabel 3. Trial and Error Parameter

δ	Jumlah BC	Rata-Rata Min MSR
0,1	15	0,0802110
0,2	15	0,05116104
0,3	15	0,0555519
0,4	15	0,591065
0,5	15	0,0501123
0,6	15	0,0672087
0,7	15	0,1538672
0,8	15	0,1511137
0,9	15	0,1473789
1,0	15	0,1471639

Berdasarkan hasil pada **Tabel 3**, seluruh nilai ambang batas menghasilkan jumlah bicluster yang sama, yaitu 15 bicluster. Oleh karena itu, pemilihan nilai δ didasarkan pada

rata-rata Mean Squared Residue (MSR). Nilai $\delta = 0,5$ menghasilkan rata-rata MSR terkecil sebesar 0,0501123, sehingga dipilih sebagai nilai ambang batas karena menunjukkan tingkat homogenitas bicluster yang lebih baik dibandingkan nilai δ lainnya.

Analisis selanjutnya dilakukan dengan menetapkan nilai ambang batas δ dan jumlah bicluster (n) sebesar 15. Parameter δ digunakan untuk mengontrol tingkat keseragaman dalam bicluster melalui batas maksimum nilai MSR, sedangkan parameter n digunakan untuk menentukan jumlah bicluster yang akan dibentuk pada proses iterasi algoritma.

Proses pembentukan bicluster dilakukan melalui tahapan multiple deletion node, single deletion node, dan node addition untuk meminimalkan nilai MSR

Tabel 4. Nilai MSR

Bicluster	Hapus Beberapa Simpul		Hapus Simpul Tunggal		Penambahan Simpul	
	Sebelum	Sesudah	Sebelum	Sesudah	Sebelum	Sesudah
1	0,870035	0.178975	0.178975	0.178975	0.178975	0.178975
2	0,870035	0,625953	0,625953	0,0000195	0,0000195	0,0000195
3	0,870035	0,659100	0,659100	0,0000138	0,0000138	0,0000138
4	0,870035	0,723913	0,723913	0,0000667	0,0000667	0,0000667
5	0,870035	0,655957	0,655957	0,0000191	0,0000191	0,0000191

Berdasarkan **Tabel 4**, proses pembentukan bicluster dilakukan melalui tiga tahap, yaitu multiple node deletion, single node deletion, dan node addition dengan nilai ambang (δ) sebesar 0,5. Pada tahap multiple node deletion, nilai MSR awal sebesar 0,870035 mengalami penurunan pada seluruh bicluster. Bicluster 1 telah mencapai nilai MSR sebesar 0,178975 yang berada di bawah nilai ambang, sedangkan bicluster 2 hingga bicluster 5 masih memiliki nilai MSR di atas 0,5 sehingga memerlukan proses lanjutan. Pada tahap single node deletion, nilai MSR pada bicluster 2 hingga bicluster 5 menurun secara signifikan menjadi 0,0000195; 0,0000138; 0,0000667; dan 0,0000191. Selanjutnya, tahap node addition tidak mengubah nilai MSR yang telah diperoleh, sehingga kualitas koherensi bicluster tetap terjaga. Hasil tersebut menunjukkan bahwa kombinasi ketiga tahapan dalam algoritma Cheng & Church mampu menghasilkan bicluster dengan nilai MSR yang jauh di bawah ambang batas yang ditetapkan, sehingga pola yang terbentuk memiliki tingkat homogenitas yang baik.

Nilai MSR yang tetap pada tahap node addition menunjukkan bahwa bicluster yang terbentuk telah optimal dan tidak memerlukan penambahan baris maupun kolom. Seluruh nilai MSR akhir juga berada di bawah ambang $\delta = 0,5$, yang mengindikasikan bahwa bicluster yang dihasilkan bersifat homogen dan koheren.

Tabel 5. Hasil Bicluster CC

Bc	Kab/Kota	Variabel	MSR
1	Kepulauan Selayar, Maros, Pangkajene dan Kepulauan, Barru, Minahasa Utara, Buton, Muna, Konawe Selatan, Kolaka Utara, Konawe Utara	Persentase rumah tangga dengan sanitasi layak, persentase rumah tangga dengan akses air minum layak, persentase penduduk miskin, persentase penduduk yang memiliki jaminan kesehatan, persentase penggunaan tempat buang air besar bersama, dan persentase rumah tangga dengan alas lantai tanah.	0.178975
2	Bolaang Mongondow, Bolaang Mongondow Timur, Bone Bolango	Kepadatan penduduk (jiwa/km ²) dan persentase rumah tangga dengan sanitasi layak.	0,0000195

Bc	Kab/Kota	Variabel	MSR
3	Soppeng, Wajo	Kepadatan penduduk (jiwa/km ²), persentase rumah tangga dengan sanitasi layak, persentase rumah tangga dengan akses air minum layak, dan persentase penduduk miskin.	0,0000138
4	Konawe, Kolaka	Persentase rumah tangga dengan sanitasi layak dan persentase rumah tangga dengan akses air minum layak	0,0000667
5	Jeneponto, Takalar	epadatan penduduk (jiwa/km ²), persentase rumah tangga dengan sanitasi layak, dan persentase rumah tangga dengan akses air minum layak.	0,0000191

Berdasarkan **Tabel 5** Bicluster pertama menunjukkan keterkaitan antara sanitasi, akses air minum, kemiskinan, jaminan kesehatan, BAB bersama, dan kondisi lantai rumah. Karakteristik ini menggambarkan wilayah dengan pola serupa pada aspek kesehatan lingkungan dan sosial ekonomi. Bicluster kedua dicirikan oleh keterkaitan antara kepadatan penduduk dan persentase sanitasi layak, menunjukkan kesamaan pola antara kondisi kependudukan dan sanitasi. Bicluster ketiga terbentuk dari kombinasi kepadatan penduduk, sanitasi layak, air minum layak, dan persentase penduduk miskin, menunjukkan bahwa keempat aspek tersebut saling berkaitan. Bicluster keempat didominasi oleh variabel sanitasi layak dan air minum layak, mengindikasikan bahwa karakteristik wilayah dalam bicluster ini terutama ditentukan oleh layanan dasar kesehatan lingkungan. Bicluster kelima dicirikan oleh keterkaitan antara kepadatan penduduk, sanitasi layak, dan air minum layak, menunjukkan kesamaan pola pada aspek kependudukan dan kualitas layanan dasar. Secara keseluruhan, faktor sanitasi layak dan akses air minum layak merupakan variabel yang paling sering muncul dalam pembentukan bicluster. Temuan ini mengindikasikan bahwa kedua variabel tersebut memiliki peran dominan dalam membedakan karakteristik wilayah terkait faktor-faktor yang memengaruhi penyebaran penyakit kusta di Pulau Sulawesi.

3.4. Analisis Iterative Signature Algorithm

Analisis menggunakan metode Iterative Signature Algorithm (ISA) bertujuan untuk mengidentifikasi bicluster berupa submatriks data yang menunjukkan keterkaitan kuat antara baris dan kolom melalui proses iteratif. Metode ini dimulai dengan pemilihan ambang batas baris (tC) dan kolom (tG). Penentuan nilai ambang batas baris (tC) dan kolom (tG) dilakukan melalui parameter tuning dengan menguji beberapa kombinasi nilai parameter. Nilai ambang yang lebih rendah menghasilkan seleksi yang lebih longgar sehingga lebih banyak objek dan variabel yang masuk ke dalam bicluster, sedangkan nilai yang lebih tinggi menghasilkan seleksi yang lebih ketat. Kombinasi parameter yang memberikan hasil bicluster terbaik selanjutnya digunakan dalam analisis. Hasil pengujian beberapa kombinasi nilai tC dan tG disajikan pada tabel berikut.

Tabel 6. Hasil Percobaan Nilai Thresholds

Threshold (tG = tC)	Jumlah Baris	Jumlah Kolom	Jumlah BC
0,1	75	8	32
0,2	56	8	29
0,3	56	8	27
0,4	52	8	24
0,5	48	8	20
0,6	37	5	15
0,7	24	2	8
0,8	38	2	6
0,9	15	1	2
1,0	18	1	1

Berdasarkan **Tabel 6**, peningkatan nilai threshold menyebabkan jumlah baris, kolom, dan bicluster yang terbentuk semakin berkurang. Pada threshold 0,1–0,5, proses seleksi masih relatif longgar sehingga sebagian besar baris dan kolom tetap dipertahankan. Sebaliknya, pada threshold 0,7–1,0, proses seleksi menjadi lebih ketat yang ditunjukkan oleh berkurangnya jumlah kolom dan bicluster secara signifikan. Nilai threshold 0,6 dipilih karena menghasilkan keseimbangan antara jumlah baris, kolom, dan bicluster, sehingga struktur bicluster yang terbentuk dinilai cukup representatif.

Selanjutnya, jumlah seed ditetapkan sebanyak 50 sebagai titik awal pembentukan bicluster. Penggunaan beberapa seed bertujuan untuk meningkatkan peluang menemukan pola yang stabil melalui proses iterasi dari berbagai kondisi awal

Berdasarkan hasil parameter tuning, Adapun hasil bicluster yang terbentuk disajikan pada tabel berikut.

Tabel 7. Hasil Pembentukan Bicluster ISA

Bicluster	Jumlah Wilayah	Variabel Terlibat
1	23	Sanitasi Layak, Penduduk Miskin, BAB Bersama
2	29	Sanitasi Layak, Penduduk Miskin, Jaminan Kesehatan, Alas Lantai Tanah
3	19	Sanitasi Layak, Air Minum Layak
4	19	Penduduk Miskin, Alas Lantai Tanah
5	25	Sanitasi Layak, Air Minum Layak, BAB Bersama
6	16	Sanitasi Layak, Penduduk Miskin, Jaminan Kesehatan
7	16	Air Minum Layak, Penduduk Miskin, Jaminan Kesehatan
8	16	Penduduk Miskin, Alas Lantai Tanah
9	23	Sanitasi Layak, Air Minum Layak, Alas Lantai Tanah
10	21	Penduduk Miskin, Jaminan Kesehatan, Alas Lantai Tanah
11	14	Air Minum Layak, Jaminan Kesehatan, Alas Lantai Tanah
12	10	Penduduk Miskin, Jaminan Kesehatan, Alas Lantai Tanah
13	13	Sanitasi Layak, Penduduk Miskin
14	19	Sanitasi Layak, Jaminan Kesehatan
15	34	Sanitasi Layak, Jaminan Kesehatan, Alas Lantai Tanah

Berdasarkan **Tabel 7**, Metode ISA menghasilkan 15 bicluster yang terbentuk dari kombinasi variabel: persentase rumah tangga dengan sanitasi layak, akses air minum layak, penduduk miskin, kepemilikan jaminan kesehatan, penggunaan fasilitas BAB bersama, dan rumah tangga dengan alas lantai tanah. Hasil tersebut menunjukkan adanya pola keterkaitan antar variabel pada kelompok wilayah tertentu.

Variabel penduduk miskin muncul pada sebagian besar bicluster dan sering berasosiasi dengan sanitasi layak, akses air minum layak, jaminan kesehatan, serta alas lantai tanah. Temuan ini menunjukkan bahwa kondisi sosial ekonomi memiliki keterkaitan dengan kualitas lingkungan permukiman dan akses layanan kesehatan. Selain itu, sanitasi layak dan akses air minum layak juga sering muncul bersamaan, mengindikasikan kesamaan karakteristik infrastruktur dasar antarwilayah. Kombinasi variabel penduduk miskin, jaminan kesehatan, dan alas lantai tanah menggambarkan kelompok wilayah dengan kerentanan sosial ekonomi yang relatif lebih tinggi.

Secara keseluruhan, hasil biclustering ISA menunjukkan bahwa faktor kemiskinan, sanitasi, akses air minum, dan kondisi tempat tinggal merupakan variabel paling dominan dalam membentuk pola karakteristik wilayah terkait faktor-faktor yang memengaruhi penyebaran penyakit kusta di Pulau Sulawesi.

3.5. Evaluasi Kinerja Metode

Evaluasi ini bertujuan untuk menilai sejauh mana pola hubungan antara wilayah dan variabel dalam setiap *bicluster* bersifat homogen dan stabil

3.5.1. Evaluasi Intra-Bicluster

Evaluasi *intra bicluster* dilakukan menggunakan *Mean Squared Residue* (MSR) untuk mengukur kualitas dan koherensi internal *bicluster*. Semakin kecil nilai MSR, semakin tinggi tingkat keseragaman pola dalam *bicluster*, sehingga kualitas *bicluster* yang terbentuk semakin baik.

Tabel 8. Hasil Evaluasi Intra-Biclustering

Metode	MSR
Cheng & Church	0,0006015
Iterative Signature Algorithm	0,015006

3.5.2. Evaluasi Inter-Bicluster

Evaluasi *inter-bicluster* dilakukan menggunakan indeks *Liu & Wang* untuk mengukur tingkat keterpisahan antar *bicluster*. Nilai indeks yang lebih rendah menunjukkan tingkat kemiripan yang lebih kecil, sehingga *bicluster* yang terbentuk memiliki karakteristik yang lebih spesifik dan tidak saling tumpang tindih.

Tabel 9. Hasil Evaluasi Inter-Biclustering

Metode	Indeks Liu & Wang
Cheng & Church	0,2905
Iterative Signature Algorithm	0,7356

3.5.3. Homogenitas Bicluster

Evaluasi homogenitas *bicluster* dilakukan menggunakan *Variance* (VAR) untuk mengukur tingkat keseragaman data dalam setiap *bicluster*. Nilai *variance* yang lebih kecil menunjukkan bahwa elemen-elemen dalam *bicluster* semakin homogen, sehingga pola yang terbentuk lebih konsisten.

Tabel 10. Hasil Evaluasi Homogenitas Biclustering

Metode	VAR
Cheng & Church	0,7737
Iterative Signature Algorithm	2,7859

4. Pembahasan

Berdasarkan evaluasi intra-bicluster, metode *Cheng & Church* (CC) menghasilkan nilai MSR Average per Volume sebesar 0,0006015, lebih rendah dibandingkan metode *Iterative Signature Algorithm* (ISA) sebesar 0,015006. Hasil ini menunjukkan bahwa *bicluster* yang dihasilkan oleh CC memiliki tingkat koherensi internal dan keseragaman pola yang lebih baik dibandingkan ISA.

Pada evaluasi *inter-bicluster*, metode CC memperoleh nilai indeks *Liu & Wang* sebesar 0,2905, sedangkan ISA sebesar 0,7356. Nilai yang lebih rendah pada metode CC menunjukkan bahwa *bicluster* yang terbentuk memiliki tingkat kemiripan yang lebih kecil, sehingga karakteristik antar *bicluster* lebih spesifik dan memiliki tingkat *overlap* yang lebih rendah dibandingkan metode ISA.

Sementara itu, evaluasi homogenitas menggunakan *Variance* (VAR) menghasilkan nilai 0,7737 untuk metode CC dan 2,7859 untuk metode ISA. Nilai VAR yang lebih rendah pada metode CC menunjukkan bahwa elemen-elemen dalam *bicluster* memiliki tingkat penyebaran yang lebih kecil, sehingga pola yang terbentuk lebih homogen dan konsisten.

Secara keseluruhan, metode *Cheng & Church* memberikan hasil yang lebih baik dibandingkan *Iterative Signature Algorithm* pada ketiga ukuran evaluasi yang digunakan, yaitu koherensi internal (*MSR*), keterpisahan antar *bicluster* (indeks *Liu & Wang*), dan homogenitas (*Variance*). Hasil ini menunjukkan bahwa metode *Cheng & Church* lebih efektif dalam mengelompokkan wilayah berdasarkan faktor-faktor yang memengaruhi penyebaran penyakit kusta di Pulau Sulawesi.

Selain menunjukkan kinerja yang lebih baik berdasarkan nilai evaluasi, metode *Cheng & Church* juga menghasilkan *bicluster* yang lebih koheren dalam mengelompokkan wilayah berdasarkan faktor-faktor yang berkaitan dengan penyebaran kusta. Hasil *biclustering* menunjukkan bahwa beberapa wilayah seperti Kepulauan Selayar, Maros, Pangkajene dan Kepulauan, Barru, Minahasa Utara, Buton, Muna, Konawe Selatan, Kolaka Utara, dan Konawe Utara tergabung dalam kelompok yang dicirikan oleh variabel sanitasi layak, akses air minum layak, kemiskinan, jaminan kesehatan, penggunaan fasilitas BAB bersama, dan kondisi lantai rumah. Temuan ini mengindikasikan bahwa faktor lingkungan dan sosial ekonomi masih berperan secara bersamaan dalam membentuk karakteristik wilayah yang berpotensi memengaruhi penyebaran kusta.

Selain itu, kemunculan variabel penduduk miskin, sanitasi layak, dan akses air minum layak pada banyak *bicluster* menunjukkan bahwa ketiga faktor tersebut merupakan karakteristik yang dominan pada wilayah-wilayah di Pulau Sulawesi. Dari perspektif epidemiologis, hasil ini menunjukkan bahwa upaya pengendalian kusta tidak hanya perlu difokuskan pada aspek pelayanan kesehatan, tetapi juga pada peningkatan kualitas sanitasi lingkungan, akses air bersih, dan kondisi sosial ekonomi masyarakat. Oleh karena itu, intervensi yang terintegrasi antara sektor kesehatan, sanitasi, dan pengentasan kemiskinan berpotensi memberikan dampak yang lebih efektif dalam menekan penyebaran penyakit kusta.

Hasil penelitian ini juga sejalan dengan penelitian terdahulu yang menyatakan bahwa metode *Cheng & Church* lebih efektif dalam menghasilkan *bicluster* yang koheren dan konsisten, sedangkan metode ISA cenderung menghasilkan *bicluster* yang lebih fleksibel namun memiliki tingkat homogenitas yang lebih rendah. Pada data penelitian ini, keunggulan metode *Cheng & Church* terlihat dari nilai *MSR* yang jauh lebih kecil serta kemampuan membentuk kelompok wilayah dengan pola yang lebih seragam.

Meskipun demikian, penelitian ini memiliki beberapa keterbatasan. Pertama, analisis hanya menggunakan data *cross-section* tahun 2024 sehingga belum mampu menggambarkan perubahan pola faktor penyebaran kusta dari waktu ke waktu. Kedua, variabel yang digunakan terbatas pada delapan indikator sosial, ekonomi, dan lingkungan yang tersedia pada publikasi BPS. Ketiga, hasil *biclustering* dipengaruhi oleh pemilihan parameter pada masing-masing metode, sehingga kemungkinan diperoleh hasil yang berbeda apabila menggunakan kombinasi parameter atau algoritma *biclustering* lainnya.

5. Kesimpulan

Dengan demikian, metode *Cheng & Church* menunjukkan kinerja yang lebih baik dibandingkan *Iterative Signature Algorithm* berdasarkan ketiga ukuran evaluasi yang digunakan. Metode *Cheng & Church* menghasilkan nilai *MSR Average per Volume* yang lebih rendah (0,0006015) dibandingkan *Iterative Signature Algorithm* (0,015006), yang menunjukkan tingkat koherensi internal yang lebih baik. Selain itu, nilai indeks *Liu & Wang* yang lebih rendah pada metode *Cheng & Church* (0,2905) dibandingkan *Iterative Signature Algorithm* (0,7356) menunjukkan keterpisahan antar *bicluster* yang lebih baik. Dari sisi homogenitas, metode *Cheng & Church* juga menghasilkan nilai *Variance* yang lebih rendah (0,7737) dibandingkan *Iterative Signature Algorithm* (2,7859). Oleh karena itu,

metode *Cheng & Church* lebih optimal dalam mengelompokkan wilayah di Pulau Sulawesi berdasarkan faktor-faktor yang memengaruhi penyebaran penyakit kusta.

Referensi

- [1] D. N. Alfiansyah, V. R. S. Nastiti, and N. Hayatin, "Penerapan metode K-means pada data penduduk miskin per kecamatan Kabupaten Blitar," *Jurnal Repository*, vol. 4, no. 1, 2022. <https://ejournal.umm.ac.id/index.php/repositor/article/view/31083>
- [2] R. N. Puspita, "Analisis K-means cluster pada kabupaten/kota di Provinsi Banten berdasarkan indikator indeks pembangunan manusia," *Jurnal Lebesgue*, vol. 2, no. 3, pp. 267–281, 2021. <https://lebesgue.lppmbinabangsa.id/index.php/home/article/download/85/59>
- [3] S. C. Madeira and A. L. Oliveira, "Biclustering algorithms for biological data analysis: A survey," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 1, no. 1, pp. 24–45, 2004. <https://doi.org/10.1109/TCBB.2004.2>.
- [4] Y. Cheng and G. M. Church, "Biclustering of expression data," in *Proc. 8th Int. Conf. Intelligent Systems for Molecular Biology*, 2000, pp. 93–103. <https://staff.fmi.uvt.ro/.../ChengChurch.pdf>
- [5] B. Pontes, R. Giráldez, and J. S. Aguilar-Ruiz, "Biclustering on expression data: A review," *Journal of Biomedical Informatics*, vol. 57, pp. 163–180, 2015. <https://doi.org/10.1016/j.jbi.2015.06.028>.
- [6] S. Bergmann, J. Ihmels, and N. Barkai, "Iterative signature algorithm for the analysis of large-scale gene expression data," *Physical Review E*, vol. 67, no. 3, p. 031902, 2003. <https://arxiv.org/pdf/physics/0210038>
- [7] K. Eren, M. Deveci, O. Küçükünç, and Ü. V. Çatalyürek, "A comparative analysis of biclustering algorithms for gene expression data," *Briefings in Bioinformatics*, vol. 14, no. 3, pp. 279–292, 2013. <https://doi.org/10.1093/bib/bbs032>.
- [8] S. Baehera, U. D. Syafitri, and A. M. Soleh, "Evaluasi perbandingan kinerja algoritma Cheng & Church biclustering terhadap algoritma clustering klasik K-means," *Jurnal Statistika dan Aplikasinya*, vol. 7, no. 2, pp. 149–161, 2023. <https://doi.org/10.21009/JSA.07204>.
- [9] A. Alrehaili, "Leprosy classification, clinical features, epidemiology, and host immunological responses," *Cureus*, vol. 15, no. 9, 2023. <https://doi.org/10.7759/cureus.44767>.
- [10] Kementerian Kesehatan RI, "Indonesia percepat eliminasi kusta dan filariasis, target bebas NTDs pada 2030," 2025.
- [11] J. C. Ortiz-Bayliss, A. K. Gárate-Escamilla, and H. Terashima-Marín, "Missing Data and their Effect on Algorithm Selection for the Bin Packing Problem," in *Proceedings of the Mexican Conference on Pattern Recognition (MCP)*, 2024.
- [12] M. Shadbahr *et al.*, "A comparative study of imputation methods for missing data," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 5, pp. 456–463, 2023. <https://doi.org/10.1155/acs/2766701>.
- [13] S. Spyridonidis, J. Dykxhoorn, A.-C. Hollander, C. Dalman, and J. B. Kirkbride, "Neighborhood-level predictors of age-at-first-diagnosis of psychotic disorders: A Swedish register-based cohort study," *Schizophrenia Bulletin Open*, vol. 3, no. 1, 2022. <https://doi.org/10.1093/schizbullopen/sgac045>.
- [14] Hafsah *et al.*, "Analisis biclustering menggunakan algoritma Cheng & Church untuk identifikasi pola data," *Jurnal Statistika dan Aplikasinya*, vol. 9, no. 1, pp. 45–55, 2025. <https://journal.unnes.ac.id/journals/sji/article/view/26413>
- [15] Di Iorio, A. Frigieri, and V. Vittorini, "Biclustering techniques based on Cheng and Church algorithm: A survey and extensions," *Information Sciences*, vol. 520, pp. 41–59, 2020. <https://pmc.ncbi.nlm.nih.gov/articles/PMC7812681/>
- [16] S. Balamurugan *et al.*, "Iterative signature algorithm for biclustering gene expression data," *International Journal of Computer Applications*, vol. 120, no. 5, pp. 1–6, 2015. <https://www.inderscienceonline.com/doi/abs/10.1504/IJCBDD.2023.133849>
- [17] W. A. L. Ningsih, I. M. Sumertajaya, and A. Saefuddin, "Biclustering application in Indonesian economic and pandemic vulnerability," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 16, no. 4, pp. 1453–1462, 2022. <https://doi.org/10.30598/barekengvol16iss4pp1453-1464>.
- [18] N. K. Sri and R. Porkodi, "An Extensive Survey on Biclustering Approaches and Algorithms for Gene Expression Data," *International Journal of Scientific & Technology Research*, vol. 8, no. 9, pp. 2228–2236, 2019..
- [19] R. Henriques and S. C. Madeira, "BSig: Evaluating the statistical significance of biclustering solutions," *Data Mining and Knowledge Discovery*, vol. 32, no. 1, pp. 124–161, 2018. <https://doi.org/10.1007/s10618-017-0521-3>.