

Evaluasi Kinerja *Spectral Biclustering* dalam Identifikasi Potensi Produksi Komoditas Hortikultura di Indonesia

Merryanty Lestari P^{1*}, I Made Sumertajaya², Erfiani³

Jl. Raya Dramaga, Kabupaten Bogor, 16680

^{1,2,3}Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, IPB University

*e-mail: lestarip@apps.ipb.ac.id

Diajukan: 18 September 2024, Diperbaiki: 27 September 2024, Diterima: 1 Nopember 2024

Abstrak

Biclustering merupakan metode penggerombolan dua arah untuk menemukan *subset* baris dan kolom dari suatu matriks data. *Spectral biclustering* merupakan salah satu algoritma dari *biclustering*. Algoritma *spectral* mempunyai tiga metode normalisasi matriks antara lain *independent rescaling of rows and columns*, *bistochastization*, dan *log*. Penerapan *spectral biclustering* bertujuan untuk mengidentifikasi potensi produksi komoditas hortikultura jenis sayuran di Indonesia. Metode normalisasi *bistochastization* menghasilkan *bicluster* optimal dengan nilai rataan *mean squared residue* terkecil sebesar 0,079593. *Bicluster* yang dihasilkan sebanyak 5 *bicluster*. *Bicluster* 1 dan 2 terdiri dari wilayah Papua, Sulawesi Tenggara, Jakarta, dan Kep. Riau memiliki potensi produksi jenis tanaman sayuran mayoritas kategori rendah di antaranya kentang, bawang merah, bawang putih, dan bawang daun. *Bicluster* 3 dan 4 terdiri dari sebagian besar wilayah Kalimantan, Riau, Sumatera Selatan, Nusa Tenggara Timur, dan Maluku dengan potensi produksi mayoritas terkategori sedang di antaranya cabai rawit, tomat, buncis, labu siam, dan melinjo. *Bicluster* 5 merupakan wilayah Jawa, Bali, Nusa Tenggara Barat, sebagian besar wilayah Sumatera dan Sulawesi, serta Kalimantan Selatan. *Bicluster* 5 memiliki potensi produksi terkategori tinggi pada jenis sayuran sawi, kacang panjang, terung, ketimun, dan jengkol.

Kata Kunci: *Biclustering, bistochastization, hortikultura, mean square residue, spectral biclustering*

Abstract

Biclustering is a two-way clustering method used to identify subsets of rows and columns within a data matrix. *Spectral biclustering* is one of the algorithms in the *biclustering* analysis. The *spectral* algorithm includes three matrix normalization methods: *independent rescaling of rows and columns*, *bistochastization*, and *log* normalization. This study applies *spectral biclustering* to identify the production potential of horticultural vegetable commodities in Indonesia. The *bistochastization* normalization method resulted in the optimal *biclusters*, with the smallest *mean squared residue* value of 0.079593. A total of five *biclusters* were generated. *Biclusters* 1 and 2, which consist of the regions of Papua, Southeast Sulawesi, Jakarta, and Riau Islands, exhibit low production potential for vegetable types such as potatoes, shallots, garlic, and scallions. *Biclusters* 3 and 4, covering most of Kalimantan, Riau, South Sumatra, East Nusa Tenggara, and Maluku, show moderate production potential for vegetables like cayenne peppers, tomatoes, green beans, chayote, and melinjo. *Bicluster* 5 includes Java, Bali, West Nusa Tenggara, most of Sumatra and Sulawesi, and South Kalimantan, and has high production potential for vegetables such as mustard greens, long beans, eggplants, cucumbers, and jengkol.

Keywords: *Biclustering, bistochastization, horticultural, mean square residue, spectral biclustering*

1 Pendahuluan

Matriks data merupakan struktur tabular terdiri dari baris dan kolom untuk menyimpan berbagai informasi. Baris mewakili objek atau unit observasi, sedangkan kolom menunjukkan peubah atau fitur yang diukur. Nilai dalam setiap sel matriks data menghubungkan baris dan kolom, dan digunakan untuk analisis guna mengungkap pola atau hubungan dalam data. Aspek yang menarik untuk diamati dari suatu matriks data berupa pengelompokan objek berdasarkan kemiripan karakteristik. *Clustering* merupakan teknik analisis multivariat yang dapat mengelompokkan objek berdasarkan karakteristik tertentu, sehingga objek dalam satu gerombol memiliki kesamaan yang lebih tinggi dibandingkan dengan gerombol lainnya [1]. *Clustering* menerapkan teknik satu arah yang mengelompokkan objek atau peubah hanya berdasarkan baris atau kolom secara terpisah [2]. Metode *clustering* memiliki keterbatasan dalam melakukan pengelompokan dua arah antara baris dan kolom secara simultan.

Metode penggerombolan dua arah yang menghubungkan sebuah gerombol objek dengan sebuah gerombol peubah secara bersamaan disebut *biclustering*. Pada awal penerapannya, *biclustering* sering digunakan dalam bidang bioinformatika. Cheng dan Church [3] menerapkan metode penggerombolan dua arah pada data ekspresi gen, dengan *bicluster* yang dihasilkan berdasarkan kemiripan nilai pada *subset* baris dan kolom yang memiliki *Mean Squared Residue* (MSR) terkecil. Selain itu, Kluger dkk. [4] juga mengaplikasikan analisis *bicluster* pada data *microarray*. Saat ini, penggunaan *biclustering* semakin berkembang di berbagai bidang ilmu. Misalnya, algoritma *biclustering* digunakan untuk melihat pola ketahanan pangan di wilayah Indonesia [5], mendeteksi pola kesejahteraan rakyat di Indonesia [6], mengidentifikasi pola pengelompokan penyakit menular [7], mendeteksi pola indeks kerentanan ekonomi dan pandemi *coronavirus disease* [8], mengukur potensi perikanan tangkap di Indonesia [9], dan menganalisis pola konsumsi air bersih [10].

Pedoman khusus dalam pemilihan algoritma *biclustering* pada data tertentu belum ada. Algoritma *biclustering* dapat dipilih berdasarkan beberapa pertimbangan seperti keunggulan dari penelitian sebelumnya, karakteristik unik yang dihasilkan dari *bicluster* yang terbentuk, dan penanganan *bicluster* yang tumpang tindih [11]. Pontes dkk. [12] menyebutkan beberapa algoritma yang dapat digunakan untuk mendapatkan *bicluster* pada data matriks, di antaranya *Spectral Biclustering*. Algoritma *spectral* termasuk kelompok algoritma berbasis non metrik berdasarkan konsep aljabar linier. Kluger dkk. [4] menggunakan cara pendekatan aljabar linier seperti *Singular Value Decomposition* (SVD) untuk mendapatkan vektor ciri yang digunakan untuk menemukan pola *checkerboard-like* pada suatu matriks data. Standarisasi data diperlukan untuk membuat seluruh peubah memiliki skala dan rata-rata yang sama. Piscopo dan Resta [13] menerapkan

algoritma *spectral* untuk data *mortality* dengan menggunakan *log-normalization* pada tahapan normalisasi. Shahrudin dkk. [14] mengidentifikasi pola curah hujan menggunakan *spectral biclustering*.

Penelitian ini merupakan kajian awal penerapan algoritma *spectral biclustering* dalam bidang pertanian, yang sebelumnya belum pernah dilakukan. Algoritma ini digunakan untuk menganalisis data produksi berbagai jenis sayuran di Indonesia dengan menerapkan tiga pendekatan normalisasi matriks, yaitu *Independent Rescaling of Rows and Columns (IRRC)*, *bistochastization*, dan *log* [4]. Evaluasi kinerja algoritma ini dalam menghasilkan *bicluster* optimal dilakukan melalui penghitungan nilai rata-rata *Mean Squared Residue (MSR)*, Indeks Liu dan Wang, serta analisis *profiling bicluster*. Pendekatan evaluasi ini belum pernah diimplementasikan dalam penelitian *spectral biclustering* sebelumnya, sehingga dapat memberikan kontribusi baru dalam pengembangan metode analisis multivariat di bidang pertanian.

Indonesia merupakan salah satu negara agraris yang perekonomiannya didukung oleh sektor pertanian. Indonesia yang beriklim tropis dan memiliki tanah yang subur menjadi salah satu pendukung berkembangnya sektor pertanian. Sektor pertanian memberikan kontribusi sebesar 12,4% terhadap Produk Domestik Bruto (PDB) pada tahun 2022. Salah satu komoditas hortikultura yang menjadi unggulan adalah tanaman sayuran. Tanaman sayuran memiliki beberapa peranan strategis di antaranya sebagai sumber bahan makanan bergizi yang kaya akan vitamin, mineral dan serat, sebagai komoditas potensial ekspor yang dapat menjadi sumber devisa negara, dan memberi peluang kesempatan kerja bagi masyarakat melalui bercocok tanam. Permintaan terhadap komoditas sayuran segar terus meningkat pesat seiring terus meningkatnya jumlah penduduk di Indonesia. Kemampuan hasil produksi tanaman sayuran di berbagai provinsi belum cukup optimal dalam memenuhi kebutuhan nasional.

Menurut Badan Pusat Statistik (BPS) tahun 2022, Indonesia mengimpor sayuran sekitar satu juta ton, angka ini meningkat 3,17% dibandingkan tahun 2021 dan menjadi angka tertinggi dalam tiga tahun terakhir. Besaran nilai impor sayuran tahun 2022 mencapai 14,7 triliun rupiah. Sementara kemampuan ekspor sayuran sebesar 91 ribu ton dengan nilai 1,5 triliun rupiah. Terjadi defisit perdagangan sayuran sebesar 13,2 triliun rupiah. Salah satu upaya untuk mengetahui potensi produksi tanaman sayuran adalah dengan melakukan pemetaan berbagai jenis sayuran berdasarkan kesamaan potensi wilayah secara simultan. Setiap provinsi di Indonesia dengan berbagai karakteristik wilayah tentunya memiliki potensi produksi tanaman sayuran yang berbeda-beda. Berdasarkan permasalahan di atas, penelitian ini ingin mengetahui bagaimana pola pemetaan atau penggerombolan jenis tanaman sayuran pada seluruh provinsi di Indonesia, serta mengkaji

bagaimana hasil evaluasi kinerja algoritma *spectral biclustering* untuk menghasilkan *bicluster* optimal pada data hasil produksi tanaman sayuran di Indonesia.

2 Metode Penelitian

2.1 Data

Data yang digunakan merupakan data sekunder yang diperoleh dari Badan Pusat Statistik tahun 2022. Unit observasi atau objek (wilayah) terdiri dari 34 provinsi di Indonesia dan peubah kolom berupa komoditas hortikultura jenis sayuran sebanyak 23 peubah seperti pada Tabel 1. Data merupakan data numerik produksi tanaman sayuran.

Tabel 1. Matriks data komoditas hortikultura jenis sayuran

Provinsi (baris)	Peubah (kolom)	Satuan
34 Provinsi di Indonesia	Kentang, Bawang Merah, Bawang Putih, Bawang Daun, Kubis, Kembang Kol, Sawi, Wortel, Kacang Panjang, Cabai Besar, Cabai Rawit, Tomat, Terung, Buncis, Ketimun, Labu Siam, Kangkung, Bayam, Malinjo, Paprika, Jamur, Petai, Jengkol	Ton

2.2 Tahapan Penelitian

Analisis *biclustering* memiliki beberapa tahapan yaitu melakukan *preprocessing* data, melakukan analisis *biclustering* dengan algoritma *spectral biclustering*, dan melakukan evaluasi kinerja algoritma *spectral biclustering*. Diagram alir tahapan penelitian dapat dilihat pada Gambar 1.

- 1) *Preprocessing* data dilakukan dengan tiga tahapan. Pertama, membentuk matriks data yang terdiri dari 34 provinsi sebagai baris dan 23 jenis sayuran sebagai kolom. Kedua, melakukan standarisasi data (*scaling*) untuk membuat data mempunyai rentang nilai yang sama. Ketiga, melakukan eksplorasi data menggunakan *heatmap* dan *biplot* PCA untuk memberikan informasi gambaran awal matriks data yang akan diteliti.
- 2) Analisis *biclustering* dengan algoritma *spectral biclustering*
Spectral biclustering menggunakan pendekatan aljabar linier untuk mengidentifikasi struktur *bicluster* pada matriks data. Menurut Madeira dan Oliveira [15], *spectral* akan menghasilkan *bicluster* dengan struktur *checkerboard*. Struktur *checkerboard* akan terbentuk setelah dilakukan proses normalisasi matriks. Kluger dkk. [4] mengemukakan tiga metode pendekatan normalisasi matriks yang dapat digunakan di antaranya *independent rescaling of rows and columns* (IRRC), *bistochasticization*, dan *log*. Penjelasan detail

mengenai algoritma *spectral biclustering* dalam menemukan *bicluster* optimal dijelaskan sebagai berikut:

1. Misalkan terdapat matriks $\mathbf{A}_{34 \times 23}$. Melakukan normalisasi matriks data terlebih dahulu dengan metode IRRC, *bistochastization*, dan *log*. Penjelasan detail untuk masing-masing metode sebagai berikut: 1) IRRC, proses normalisasi diawali dengan menghitung nilai terkecil dari matriks \mathbf{A} . Didefinisikan matriks $\mathbf{B} = \mathbf{A} + |\min(\mathbf{A})|$. Normalisasi matriks \mathbf{B} dilakukan dengan bantuan matriks \mathbf{R} dan \mathbf{C} . Matriks \mathbf{R} adalah matriks diagonal yang elemennya merupakan jumlahan baris ke- i dari matriks \mathbf{B} . Matriks \mathbf{C} adalah matriks diagonal yang elemennya merupakan jumlahan kolom ke- j dari matriks \mathbf{B} . Maka, diperoleh matriks normalisasi $\mathbf{K} = \mathbf{R}^{-1/2} \mathbf{B} \mathbf{C}^{-1/2}$. Sun dan Das [16] mendefinisikan $\mathbf{R}^{-1/2}$ dan $\mathbf{C}^{-1/2}$ seperti pada Persamaan (1). 2) *Bistochastization* merupakan metode perulangan dari IRRC sampai diperoleh seluruh baris dan kolom memiliki nilai konstan tertentu. 3) *log*, diawali dengan menghitung nilai terkecil dari matriks \mathbf{A} . Jika $\min(\mathbf{A}) < 1$, maka matriks $\mathbf{B} = \mathbf{A} + (1 - \min(\mathbf{A}))$. Jika $\min(\mathbf{A}) \geq 1$, maka matriks $\mathbf{B} = \mathbf{A}$. Selanjutnya, menghitung matriks \mathbf{L} yang merupakan transformasi logaritma dari matriks \mathbf{B} sehingga diperoleh $\mathbf{L} = \log \mathbf{B}$. Matriks \mathbf{L}_i adalah matriks rata-rata baris ke- i dari matriks \mathbf{L} ; $i = 1, 2, \dots, n$. Matriks \mathbf{L}_j adalah matriks rata-rata kolom ke- j dari matriks \mathbf{L} ; $j = 1, 2, \dots, m$. Matriks \mathbf{L}_{ij} adalah matriks rata-rata seluruh data dari matriks \mathbf{L} . Jadi, didapatkan matriks normalisasi $\log \mathbf{K} = \mathbf{L} - \mathbf{L}_i - \mathbf{L}_j + \mathbf{L}_{ij}$.

$$\mathbf{R}^{-1/2} = \text{diag} \left(\frac{1}{\sqrt{r_{11}}}, \frac{1}{\sqrt{r_{22}}}, \dots, \frac{1}{\sqrt{r_{nn}}} \right); \mathbf{C}^{-1/2} = \text{diag} \left(\frac{1}{\sqrt{c_{11}}}, \frac{1}{\sqrt{c_{22}}}, \dots, \frac{1}{\sqrt{c_{mm}}} \right) \quad (1)$$

2. Menerapkan *Singular Value Decomposition* (SVD) pada matriks \mathbf{K} sehingga diperoleh submatriks, yaitu matriks \mathbf{U} dan \mathbf{V} , matriks \mathbf{U} berisi informasi provinsi dan matriks \mathbf{V} berisi informasi produksi jenis sayuran.
3. Menentukan kombinasi parameter banyak vektor ciri (E) dan vektor ciri proyeksi (e) yang akan digunakan pada proses *biclustering*, terdapat 42 kombinasi pada Tabel 2. Menurut Kluger dkk [4], penentuan parameter E adalah minimal 2 karena menggunakan 1 vektor ciri dapat memberikan informasi yang kurang relevan, dan maksimal sebanyak $n - 1$ peubah, serta merekomendasikan penggunaan parameter e sebesar 2 atau 3 untuk mendapatkan hasil yang optimal.

Tabel 2. Kombinasi parameter yang digunakan

Parameter	Vektor ciri
Banyak vektor ciri (E)	2,3,4, ..., 22
Banyak vektor ciri proyeksi (e)	2,3

4. Melakukan proses *biclustering* dan pengurutan. Jika menggunakan normalisasi IRRC dan *bistochastization* dilakukan penghapusan vektor ciri pertama dari matriks U dan V . Selanjutnya, memilih vektor ciri sebanyak nilai parameter E untuk masing-masing matriks U dan V yang dinotasikan dengan matriks u dan v . Penggerombolan dilakukan pada masing-masing matriks u dan v dengan metode *iterative kmeans* yaitu algoritma *kmeans* dikenakan pada masing-masing vektor ciri yang dilakukan secara berulang hingga diperoleh gerombol optimal berdasarkan *Within Sum of Square* (WSS) terkecil. Bentuk *piecewise* terbaik didapatkan dengan mengurutkan vektor ciri u dan v berdasarkan nilai WSS terkecil ke terbesar. Selanjutnya, mengurutkan banyak gerombol optimal pada provinsi dan jenis sayuran berdasarkan nilai WSS terkecil ke terbesar.
5. Melakukan proses proyeksi dan partisi. Vektor ciri proyeksi digunakan untuk mendapatkan partisi terbaik bagi provinsi dan produksi jenis tanaman sayuran. Pertama, menentukan vektor ciri proyeksi untuk provinsi, $\bar{u} = K^T u$, dengan banyak vektor ciri u sebanyak e dan vektor ciri proyeksi untuk jenis tanaman sayuran, $\bar{v} = K v$, dengan banyak vektor ciri v sebanyak e . Kedua, partisi provinsi diperoleh berdasarkan *iterative kmeans* pada \bar{v} . Ketiga, partisi produksi jenis tanaman sayuran berdasarkan *iterative kmeans* pada \bar{u} .
6. Langkah 4 dan 5 diulangi sebanyak kombinasi parameter yang telah ditentukan.
7. Pemilihan parameter yang menghasilkan *bicluster* optimal berdasarkan nilai Indeks Calinski-Harabasz (CH). Pertama, menghitung nilai indeks CH dari gerombol provinsi dan jenis tanaman sayuran. Kedua, menghitung rata-rata dari Indeks CH provinsi dan jenis tanaman sayuran. Ketiga, parameter dipilih berdasarkan nilai rata-rata indeks CH terbesar.
8. Selanjutnya, pada hasil *bicluster* dari parameter terpilih, dilakukan pemilihan *bicluster* berdasarkan banyak keanggotaan dalam *bicluster* minimal dua objek dan dua peubah, serta nilai *within variances* sesuai dengan nilai *threshold* yang telah ditentukan, serta melakukan interpretasi hasil *bicluster*.

3) Evaluasi kinerja algoritma *spectral biclustering*

Kinerja algoritma *biclustering* dapat dievaluasi menggunakan *Average Residu* (ASR), Indeks Liu dan Wang, dan *profiling bicluster*. ASR digunakan untuk mengukur koherensi dalam *bicluster*. Lee dkk. [17] memperkenalkan ASR untuk mengevaluasi seluruh k *bicluster* yang dihasilkan, didefinisikan seperti Persamaan (2). ASR merupakan nilai rata-rata dari MSR yang didefinisikan seperti Persamaan (3),

$$ASR = \frac{1}{k} \sum_{i=1}^k MSR_i(I, J) \quad (2)$$

$$MSR_{(I, J)} = \frac{\sum_{i=1}^{|I|} \sum_{j=1}^{|J|} (a_{ij} - a_{i\cdot} - a_{\cdot j} + a_{\cdot\cdot})^2}{|I| \times |J|} = \frac{\sum_{i=1}^{|I|} \sum_{j=1}^{|J|} (e_{ij})^2}{|I| \times |J|} \quad (3)$$

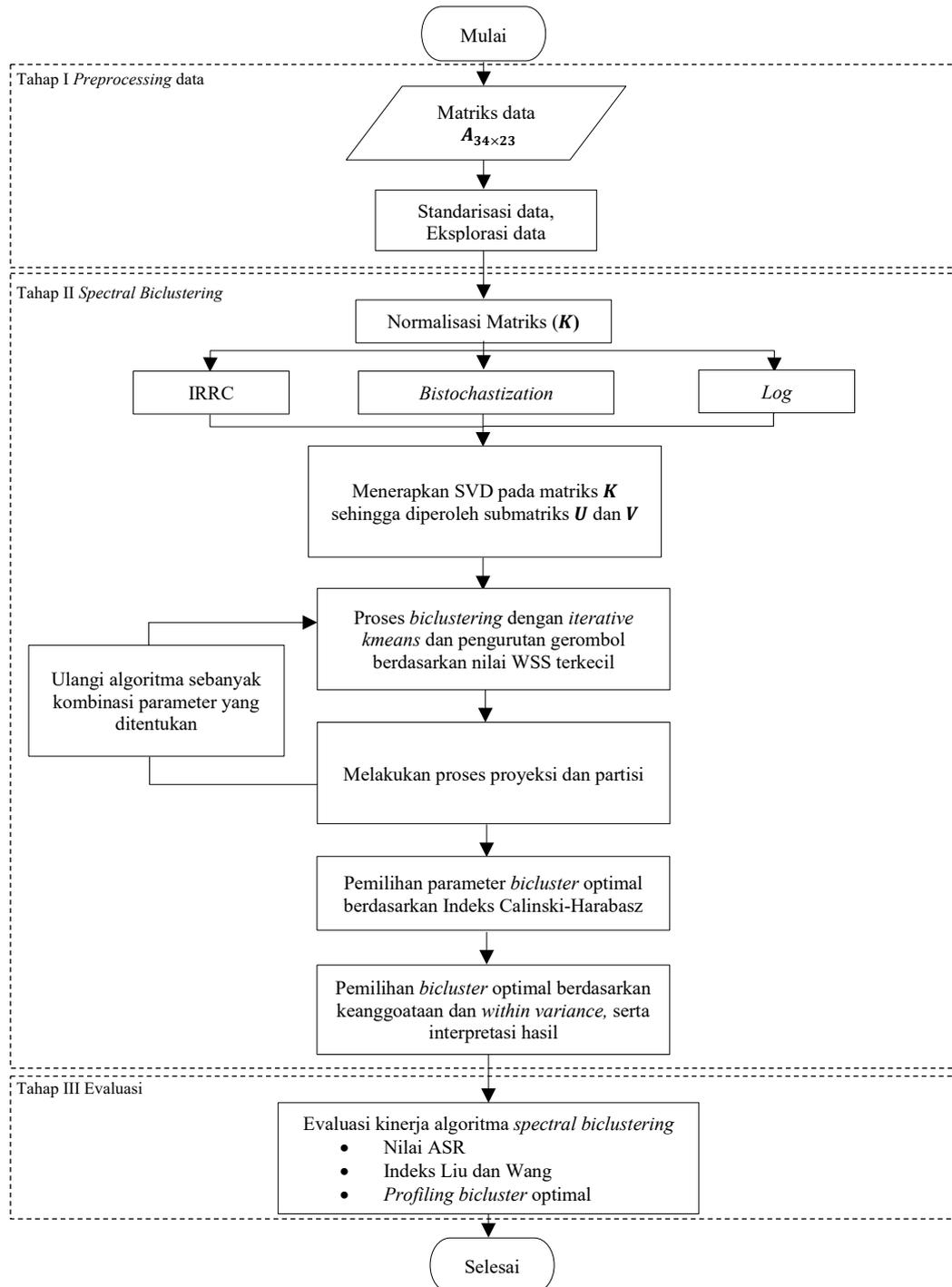
dengan a_{ij} adalah elemen *bicluster* pada baris ke- i dan kolom ke- j , $a_{i\cdot}$ adalah rata-rata kolom ke- j , $a_{\cdot j}$ adalah rata-rata baris ke- i , $a_{\cdot\cdot}$ adalah rata-rata seluruh elemen *bicluster*, dan $|I| \times |J|$ adalah dimensi *bicluster* (volume), yakni jumlah baris dalam *bicluster* ($|I|$) dikali jumlah kolom dalam *bicluster* ($|J|$). Kualitas *bicluster* semakin baik apabila nilai ASR yang dihasilkan semakin kecil atau mendekati nol [17].

Indeks Liu dan Wang digunakan untuk mengukur tingkat kemiripan dari *bicluster* yang dihasilkan, semakin kecil nilai Indeks Liu dan Wang, maka kemiripan karakteristik anggota hasil *bicluster* semakin berbeda. Didefinisikan seperti pada Persamaan (4),

$$I_{Liu\&Wang}(M_{opt}, M) = \frac{1}{k_{opt}} \sum_{i=1}^{k_{opt}} \max \left(\frac{|G_i \cap G_j| + |C_i \cap C_j|}{|G_i \cup G_j| + |C_i \cup C_j|} \right) \quad (4)$$

dengan M_{opt} adalah kelompok *bicluster* yang memiliki nilai rata-rata MSR paling kecil dan M adalah kelompok *bicluster* yang lain, k_{opt} adalah banyak *bicluster* pada M_{opt} , $|G_i \cap G_j|$ adalah banyak baris (G) pada M_{opt} yang beririsan dengan baris pada M , $|C_i \cap C_j|$ adalah banyak kolom (C) pada M_{opt} yang beririsan dengan kolom pada M , $|G_i \cup G_j|$ adalah banyaknya gabungan baris pada M_{opt} dan M , $|C_i \cup C_j|$ adalah banyak gabungan kolom pada M_{opt} dan M .

Profiling bicluster merupakan plot garis antara *subset* provinsi dan *subset* jenis tanaman sayuran. *Profiling bicluster* bertujuan untuk melihat kemiripan karakteristik keanggotaan dari suatu *bicluster*. Pola *profiling* yang diharapkan merupakan pola berhimpit dan sejajar. Semakin garis yang dihasilkan berhimpit dan sejajar, maka *bicluster* yang dihasilkan semakin homogen.



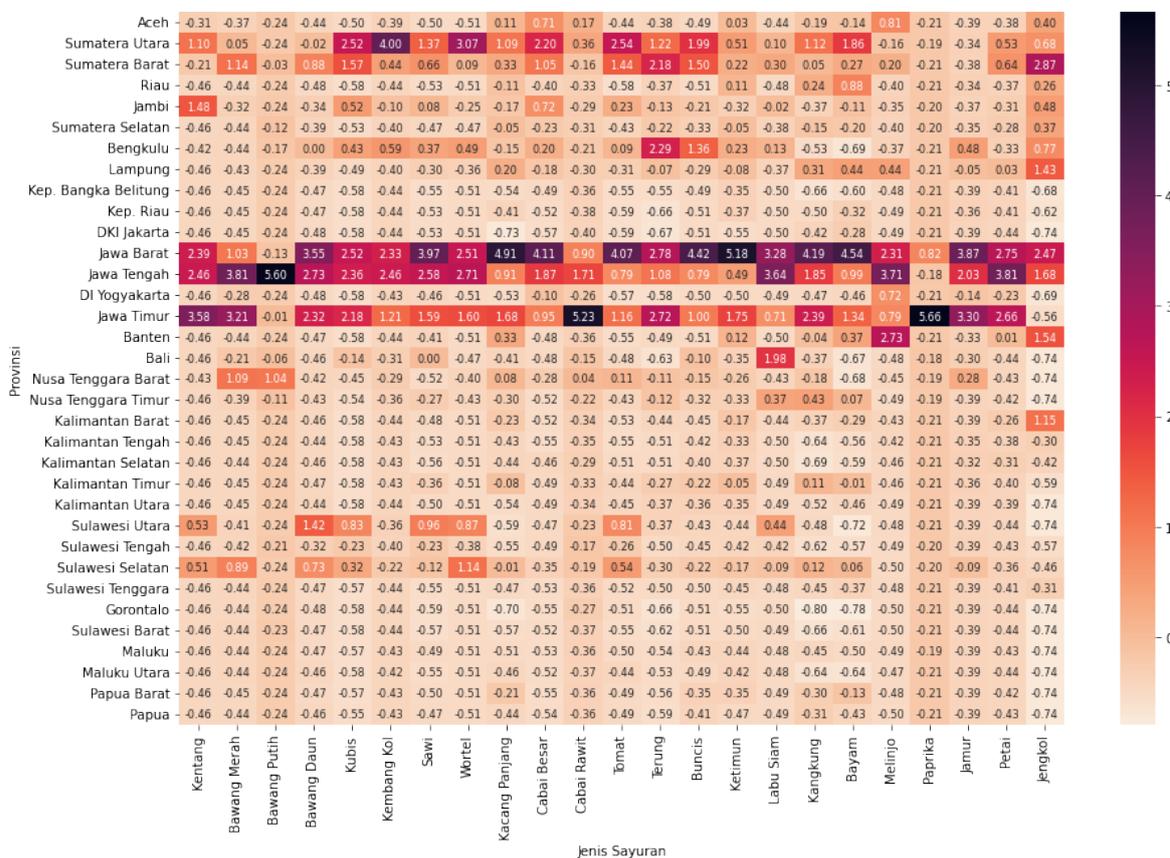
Gambar 1. Diagram alir tahapan penelitian

3 Hasil dan Pembahasan

3.1 Eksplorasi data

Gambaran awal data dapat dilihat melalui *heatmap* pada Gambar 2. *Heatmap* digambarkan menggunakan representasi warna. Semakin gelap warna yang dihasilkan maka menunjukkan data bernilai semakin tinggi. *Heatmap* menunjukkan jumlah produksi tanaman sayuran terbagi dalam

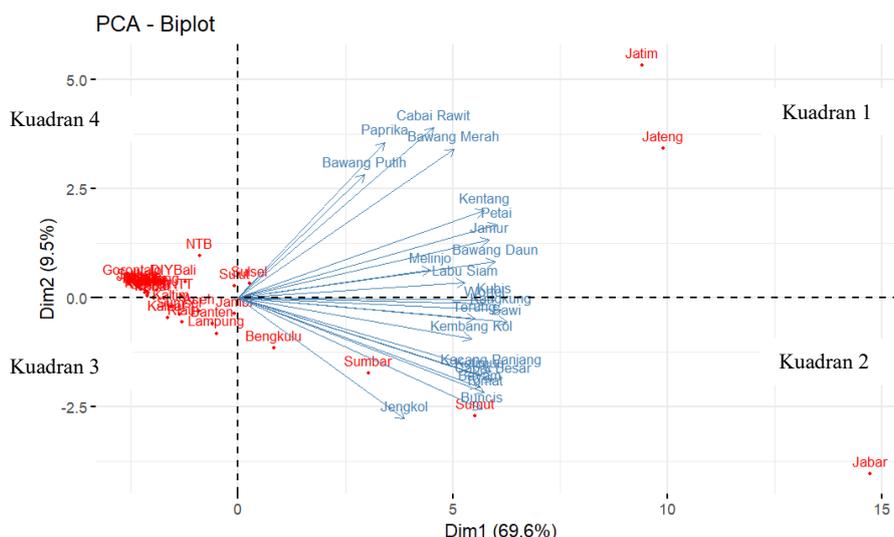
tiga kategori warna yaitu produksi tinggi berwarna hitam, sedang berwarna merah tua, dan rendah berwarna oranye. Provinsi Jawa Barat mempunyai warna hitam dan merah tua menandakan jumlah produksi yang tinggi seperti kacang panjang, cabai besar, ketimun, kangkung dan bayam. Provinsi Jawa Tengah, Jawa Timur, Sumatera Utara, dan Sumatera Barat merupakan provinsi yang memiliki jumlah produksi tanaman sayuran yang relatif tinggi. Provinsi Sumatera Utara memiliki jumlah produksi tinggi pada kembang kol, kubis, wortel, cabai besar, dan tomat. Provinsi Jawa Tengah mempunyai jumlah produksi tinggi pada bawang putih, bawang merah, labu siam, melinjo, dan petai. Provinsi Jawa Timur mempunyai jumlah produksi tinggi pada cabai rawit, paprika, kentang, bawang merah, dan jamur. Secara keseluruhan, jumlah produksi tanaman sayuran mayoritas berwarna oranye. Hal ini menandakan produksi jenis sayuran di Indonesia cenderung memiliki kategori sedang ke rendah.



Gambar 2. Heatmap matriks data scaling

Metode lain yang dapat digunakan untuk melihat karakteristik awal data adalah *biplot* PCA. Gambaran awal data terkait pengelompokan provinsi dan karakteristik peubah jenis sayuran dapat dilihat pada Gambar 3. PCA mereduksi peubah jenis sayuran menjadi dua komponen utama dengan total keragaman yang dapat dijelaskan sebesar 79,1%. Kelompok provinsi terbagi ke dalam 4 kelompok yaitu kuadran 1, 2, 3, dan 4. Provinsi yang berada pada kuadran yang sama dapat dikatakan memiliki kesamaan karakteristik jumlah produksi sayuran yang sama. Sedangkan,

provinsi yang berada pada arah berlawanan dari vektor jenis sayuran maka memiliki jumlah produksi yang bertolak belakang atau bernilai lebih rendah dari rata-rata. Contohnya pada kuadran 1, terdapat Provinsi Jawa Timur, Jawa Tengah dan Sulawesi Selatan dengan peubah penciri potensi tinggi yaitu bawang putih, paprika, cabai rawit, bawang merah, kentang, petai, jamur, bawang daun, labu siam, dan kubis. Maka, provinsi pada kuadran 1 memiliki nilai peubah penciri yang relatif lebih tinggi dibandingkan dengan provinsi pada kuadran 3. Kuadran 2, Provinsi Jawa Barat, Bengkulu, Sumatera Barat, dan Sumatera Utara mempunyai peubah penciri dengan potensi tinggi yaitu wortel, kangkung, terung, sawi, kembang kol, kacang panjang, cabai besar, ketimun, cabai besar, bayam, tomat, buncis, dan jengkol. Provinsi kuadran 2 memiliki nilai peubah penciri yang relatif lebih tinggi dibandingkan dengan provinsi kuadran 4. Keragaman peubah juga dapat dilihat melalui panjang vektor jenis sayuran. Kuadran 1, bawang putih dan labu siam memiliki keragaman yang lebih kecil dibandingkan peubah lainnya. Kuadran 2, keragaman peubah jenis sayuran relatif hampir sama.



Gambar 3. Biplot PCA

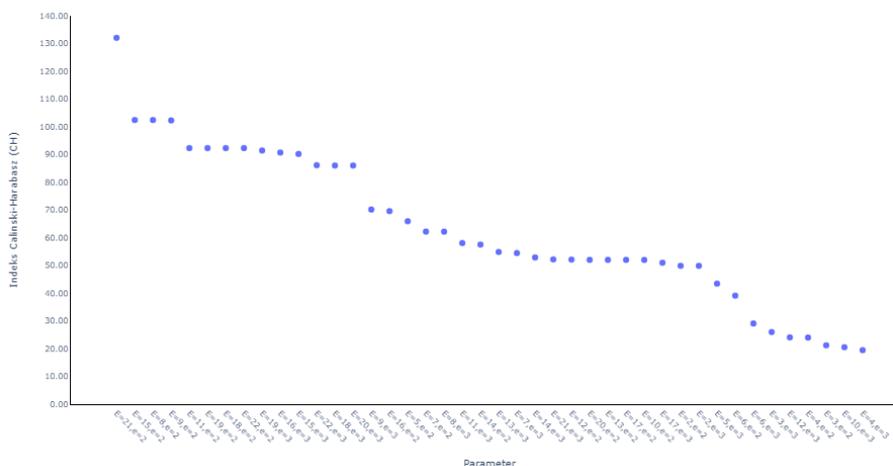
3.2 Spectral biclustering

Algoritma *spectral biclustering* melakukan proses normalisasi matriks data terlebih dahulu. *Spectral* mempunyai *tuning* parameter untuk mendapatkan *bicluster* optimal berupa banyak vektor ciri (E) dan banyak vektor ciri proyeksi (e). Penentuan parameter optimal berdasarkan nilai rata-rata Indeks Calinski Harabasz (CH) terbesar. Berdasarkan Tabel 3, metode *bistochasticization* memiliki Indeks CH terbesar. Selanjutnya, akan diperlihatkan perbandingan nilai Indeks CH pada metode *bistochasticization* untuk masing-masing kombinasi parameter pada Gambar 4.

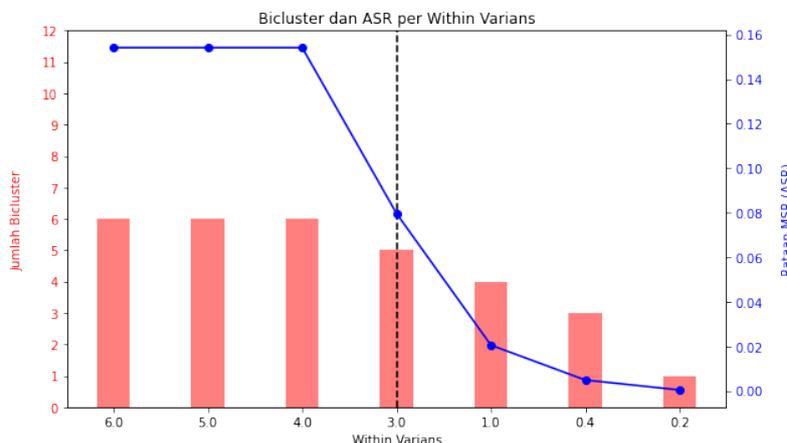
Tabel 3. Parameter *bicluster* optimal untuk metode normalisasi

Metode	Parameter optimal	Indeks CH
IRRC	$E = 22, e = 2$	72,61
<i>Bistochastization</i>	$E = 21, e = 2$	132,07
<i>log</i>	$E = 20, e = 2$	84,09

Parameter optimal untuk normalisasi matriks *bistochastization* adalah $E = 21$, dan $e = 2$ karena memiliki nilai Indeks CH terbesar dapat dilihat pada Gambar 4. Selanjutnya, dipilih beberapa nilai ambang batas *within variance* secara berurutan. Nilai yang dicobakan merupakan kelipatan 0,1 yakni mulai dari 0,1 sampai 1, dan kelipatan 1 yakni mulai 1 sampai 10. Berdasarkan Gambar 5, garis vertikal putus-putus pada *within variance* sama dengan 3 menunjukkan titik di mana perubahan paling signifikan terjadi pada rata-ran MSR. Hal ini merupakan indikasi bahwa pada titik tersebut, terdapat keseimbangan terbaik antara jumlah *bicluster* yang cukup banyak dan rata-ran MSR yang rendah. Oleh karena itu, ambang batas yang dipilih adalah *within variance* = 3. Jumlah *bicluster* optimal diperoleh sebanyak 5 *bicluster* dengan rata-ran MSR sebesar 0,07959.



Gambar 4. Tuning parameter *bistochastization* berdasarkan Indeks CH



Gambar 5. Nilai ASR dan Jumlah *Bicluster* per *Within Variance*

Karakteristik keanggotaan hasil *bicluster* dengan metode normalisasi *bistochastization* ditampilkan pada Tabel 4. *Bicluster* 1 dan 2 terdiri dari Provinsi Kepulauan Riau, Jakarta, Sulawesi Tenggara, Papua Barat, Papua dengan peubah penciri yang berbeda. *Bicluster* 1 potensi produksi sedang ke rendah pada jenis sayuran sawi, kacang panjang, terung, ketimun, dan jengkol. *Bicluster* 2 mayoritas memiliki potensi produksi rendah, hanya terdapat dua jenis sayuran dengan potensi produksi sedang yaitu Kubis dan Petai. *Bicluster* 3 dan 4 terdiri dari Provinsi Riau, Sumatera Selatan, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Maluku dengan potensi produksi didominasi kategori sedang. *Bicluster* 5 memiliki keanggotaan provinsi yang berbeda dari *bicluster* lainnya. *Bicluster* 5 terdiri dari 20 provinsi dengan potensi produksi tinggi pada jenis sayuran sawi, kacang panjang, terung, ketimun dan jengkol.

Tabel 4. Karakteristik dan keanggotaan hasil *bicluster* optimal normalisasi *bistochastization*

<i>Bicluster</i>	Ukuran	Provinsi	Potensi Produksi Tanaman Sayuran		
			Rendah	Sedang	Tinggi
1	5 × 5	Kep. Riau, Jakarta, Sultra, Pabar, Papua	Sawi, Terung, Ketimun	Kacang Panjang, Jengkol	-
2	5 × 16	Kep. Riau, Jakarta, Sultra, Pabar, Papua	Kentang, Bawang Merah, Bawang Putih, Bawang Daun, Kembang Kol, Wortel, Cabai Besar, Cabai Rawit, Tomat, Buncis, Labu Siam, Melinjo, Paprika, Jamur	Kubis, Petai	-
3	8 × 5	Riau, Sumsel, NTT, Kalbar, Kalteng, Kaltim, Kaltara, Maluku	Jengkol	Sawi, Terung, Ketimun, Kacang Panjang	-
4	8 × 16	Riau, Sumsel, NTT, Kalbar, Kalteng, Kaltim, Kaltara, Maluku	Kentang, Bawang Merah, Bawang Putih, Kubis, Wortel, Cabai Besar, Paprika	Bawang Daun, Kembang Kol, Cabai Rawit, Tomat, Buncis, Labu Siam, Melinjo, Jamur, Petai	-

<i>Bicluster</i>	Ukuran	Provinsi	Potensi Produksi Tanaman Sayuran		
			Rendah	Sedang	Tinggi
5	20 × 5	Aceh, Sumut, Sumbar, Jambi, Bengkulu, Lampung, Babel, Jabar, Jateng, Yogyakarta, Jatim, Banten, Bali, NTB, Kalsel, Sulut, Sulteng, Sulsel, Sulbar, Malut	-	-	Sawi, Kacang Panjang, Terung, Ketimun, Jengkol

Keterangan:

Potensi Rendah \leq *Percentil*_{0,33}

*Percentil*_{0,33} < Potensi Sedang < *Percentil*_{0,67}

Potensi Tinggi \geq *Percentil*_{0,67}

3.3 Evaluasi Kinerja Algoritma *Spectral Biclustering*

Hasil *bicluster* dari tiga metode normalisasi dievaluasi menggunakan nilai rata-ran MSR, Indeks Liu dan Wang, dan *profiling bicluster*. Semakin kecil nilai rata-ran MSR, maka kualitas *bicluster* semakin baik [17]. Perbandingan hasil nilai rata-ran MSR untuk masing-masing metode normalisasi dapat dilihat pada Tabel 5.

Tabel 5. Nilai rata-ran MSR berdasarkan normalisasi matriks

IRRC		<i>Bistochastization</i>		<i>Log</i>	
<i>Bicluster</i>	MSR	<i>Bicluster</i>	MSR	<i>Bicluster</i>	MSR
1	0,3779	1	0,008459	1	0,2416
2	0,3693	2	0,000356	2	0,3034
3	0,6205	3	0,067526	3	0,0794
4	0,0236	4	0,005646	4	0,0414
5	0,0701	5	0,315976	-	-
6	0,0277	-	-	-	-
Rataan	0,2482	Rataan	0,079593	Rataan	0,1665

Metode normalisasi *bistochastization* memiliki nilai MSR terkecil sebesar 0,079593. Hal ini mengindikasikan bahwa metode *bistochastization* mempunyai tingkat kemiripan antar keanggotaan *bicluster* yang lebih baik atau semakin homogen *bicluster* yang dihasilkan. Berdasarkan Tabel 4 banyaknya provinsi yang tergerombolkan sebesar 31 provinsi dan peubah sebanyak 21 peubah. Selanjutnya, dilakukan evaluasi dengan Indeks Liu dan Wang untuk mengukur kemiripan keanggotaan antar metode normalisasi. Berdasarkan Tabel 6, metode normalisasi *bistochastization* memiliki persentase kemiripan rendah terhadap metode IRRC dan *log* artinya masing-masing metode normalisasi menghasilkan keanggotaan *bicluster* yang berbeda.

Persentase kemiripan antara *bistochastization* terhadap IRRC sebesar 31,76% artinya terdapat beberapa keanggotaan dalam *bicluster* IRRC telah digambarkan oleh *bistochastization*.

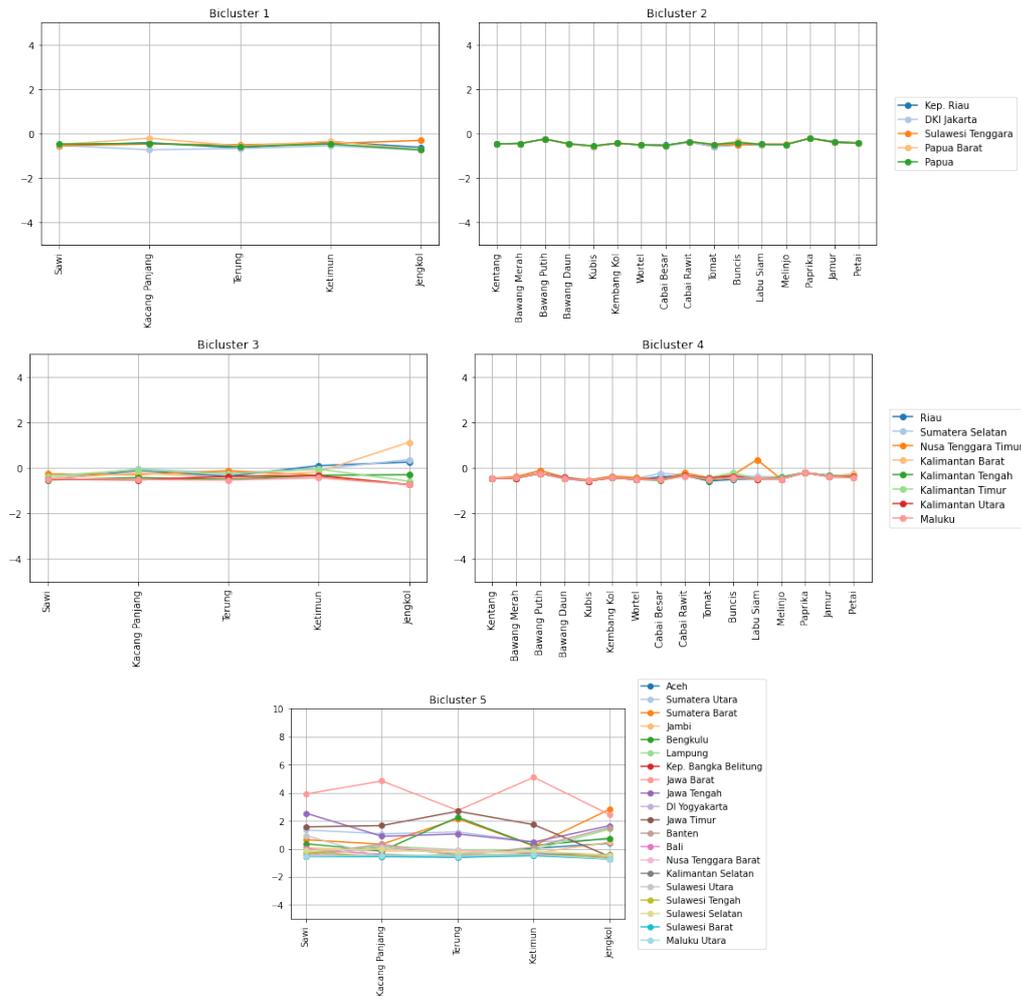
Tabel 6. Perbandingan evaluasi kinerja Indeks Liu dan Wang

	Normalisasi	Indeks Liu dan Wang
<i>Bistochastization</i> *)	IRRC	31,76%
<i>Bistochastization</i> *)	Log	24,20%

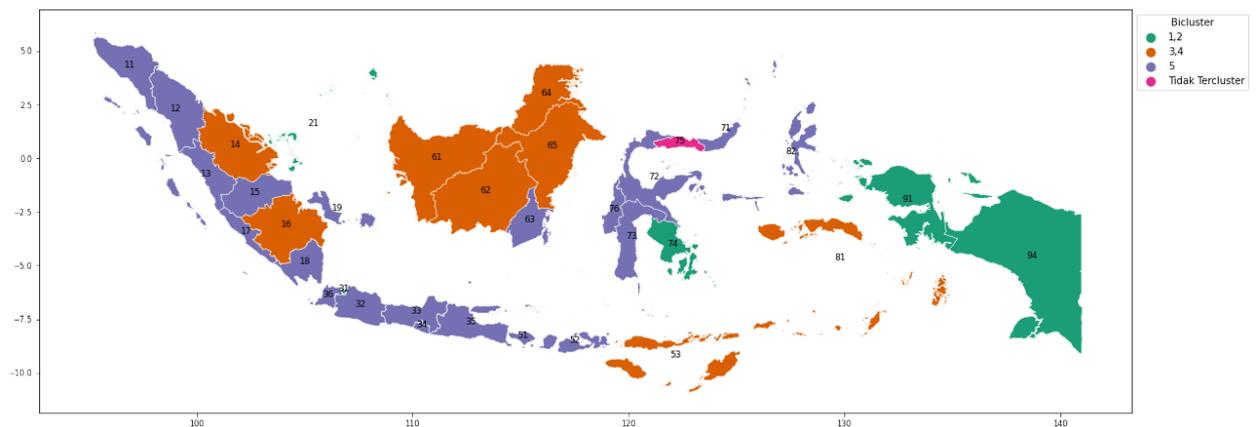
*) metode dengan *bicluster* optimal

Profiling keanggotaan terhadap *subset* provinsi terhadap *subset* peubah menggunakan metode normalisasi *bistochastization* dapat dilihat pada Gambar 6. *Bicluster* dikatakan semakin homogen apabila garis profil yang dimiliki semakin berhimpit dan sejajar. *Bicluster* 1, 2, 3, dan 4 mempunyai garis profil yang sejajar dan cenderung berhimpit sehingga dapat dikatakan bahwa *bicluster* ini memiliki tingkat kemiripan yang sama. *Bicluster* 5 mempunyai garis profil yang cenderung tidak berhimpit dan tidak sejajar menandakan terdapat perbedaan nilai produksi tanaman sayuran di beberapa provinsi. Provinsi Jawa Barat memiliki potensi produksi sayuran sawi, kacang panjang, dan ketimun yang lebih tinggi dibandingkan provinsi lainnya. Berdasarkan nilai ASR dan *profiling bicluster* maka metode normalisasi *bistochastization* memiliki kinerja yang lebih baik dalam menghasilkan *bicluster* optimal dibandingkan metode IRRC dan *log*.

Peta sebaran hasil *bicluster* dapat digunakan untuk melihat sebaran provinsi berdasarkan kelompok *bicluster*. Peta sebaran hasil *bicluster* optimal dengan metode normalisasi *bistochastization* disajikan dalam Gambar 7. *Bicluster* 1 dan 2 terdiri dari wilayah Papua dan Sulawesi Tenggara dan memiliki potensi produksi tanaman sayuran mayoritas kategori rendah. Hanya beberapa jenis sayuran yang memiliki potensi produksi sedang yaitu kacang panjang, jengkol, kubis, dan petai. *Bicluster* 3 dan 4 sebagian besar merupakan wilayah Kalimantan, Riau, Sumatera Selatan, Nusa Tenggara Timur dan Maluku. Wilayah ini memiliki potensi produksi tanaman sayuran mayoritas kategori sedang. Terdapat beberapa jenis sayuran yang terkategori rendah di antaranya jengkol, kentang, bawang merah, bawang putih, kubis, wortel, cabai besar, dan paprika. *Bicluster* 5 terdiri dari wilayah Jawa, Bali, Nusa Tenggara Barat, sebagian besar wilayah Sumatera dan Sulawesi, serta Kalimantan Selatan. *Bicluster* 5 memiliki peubah penciri berupa sawi, kacang panjang, terung, ketimun, jengkol dengan potensi produksi terkategori tinggi.



Gambar 6. Plot *profiling bicluster*



Gambar 7. Peta sebaran hasil *bicluster* optimal menurut provinsi

4 Simpulan

Spectral biclustering dengan metode normalisasi *bistochastization* menghasilkan *bicluster* optimal berdasarkan nilai ASR terkecil sebesar 0,079593. *Bicluster* optimal didapatkan dengan

tuning parameter $E = 21$, dan $e = 2$, serta nilai *threshold within variance* sebesar 3. *Bicluster* optimal yang dihasilkan sebanyak 5 *bicluster*. *Spectral biclustering* menghasilkan tumpang tindih wilayah di antaranya *bicluster* 1 dan 2, *bicluster* 3 dan 4. Wilayah *bicluster* 1 dan 2 merupakan sebagian besar wilayah Indonesia Timur yaitu wilayah Papua, Sulawesi Tenggara, Jakarta, dan Kep. Riau. *Bicluster* 1 dan 2 memiliki peubah penciri potensi produksi jenis tanaman sayuran mayoritas terkategori rendah di antaranya sawi, terung, ketimun, kentang, bawang merah, bawang putih, bawang daun, kembang kol, wortel, cabai besar, cabai rawit, tomat, buncis, labu siam, melinjo, paprika, dan jamur. Wilayah *bicluster* 3 dan 4 merupakan wilayah Kalimantan, Riau, Sumatera Selatan, Nusa Tenggara Timur dan Maluku. *Bicluster* 3 dan 4 dicirikan oleh jenis tanaman sayuran dengan potensi produksi mayoritas kategori sedang di antaranya sawi, kacang panjang, terung, ketimun, bawang daun, kembang kol, cabai rawit, tomat, buncis, labu siam, melinjo, jamur, dan petai. *Bicluster* 5 merupakan wilayah Jawa, Bali, Nusa Tenggara Barat, sebagian besar wilayah Sumatera dan Sulawesi, serta Kalimantan Selatan. *Bicluster* 5 memiliki peubah penciri berupa sawi, kacang panjang, terung, ketimun, jengkol dengan potensi produksi terkategori tinggi.

Hasil penelitian ini dapat dijadikan sebagai bahan masukan dalam menyusun kebijakan pemerintah dalam meningkatkan jumlah produksi jenis tanaman sayuran. Pemerintah dapat memberikan dukungan untuk wilayah dengan produksi sayuran terkategori rendah dan sedang melalui pelatihan, penyuluhan, dan penyediaan sarana produksi seperti penyediaan bibit unggul dan pupuk. Wilayah dengan potensi produksi jenis sayuran terkategori tinggi dapat mendorong atau menjadi pendukung rantai pasok ke wilayah dengan kategori produksi rendah. Selain itu, pemerintah juga perlu mendorong diverifikasi tanaman sayuran, dengan pengenalan varietas unggul dan teknik budidaya yang efisien sehingga produksi tanaman sayuran terkategori rendah dapat meningkat, serta dapat mengadakan program penyuluhan intensif untuk petani tentang teknik pertanian yang efisien.

5 Daftar Pustaka

- [1] N. Trianasari, I. M. Sumertajaya, Erfiani, and I. W. Mangku, "Application of beta mixture distribution in data on gpa proportion and course scores at the mbti telkom university," *Commun. Math. Biol. Neurosci.*, vol. 2021, p. Article-ID, 2021, doi: 10.28919/cmbn/5391.
- [2] A. Tanay, R. Sharan, and R. Shamir, "Biclustering algorithms: A survey," *Handb. Comput. Mol. Biol.*, vol. 9, no. 1–20, pp. 122–124, 2005.
- [3] Y. Cheng and G. M. Church, "Biclustering of Expression Data," in *Proceedings of the 8th*

-
- International Conference on Intelligent Systems for Molecular Biology, ISMB 2000*, 2000, pp. 93–103.
- [4] Y. Kluger, R. Basri, J. T. Chang, and M. Gerstein, “Spectral biclustering of microarray data: Coclustering genes and conditions,” *Genome Res.*, vol. 13, no. 4, pp. 703–716, 2003, doi: 10.1101/gr.648603.
- [5] N. Hikmah, I. M. Sumertajaya, and F. M. Afendi, “Pattern Recognition of Food Security in Indonesia Using Biclustering Plaid Model,” *JTAM (Jurnal Teor. dan Apl. Mat.)*, vol. 7, no. 4, pp. 1178–1192, 2023.
- [6] L. Marifni, I. M. Sumertajaya, and U. D. Syafitri, “Bicluster Analysis of Cheng and Church’s Algorithm to Identify Patterns of People’s Welfare in Indonesia,” *JUITA J. Inform.*, vol. 11, no. 2, p. 261, 2023, doi: 10.30595/juita.v11i2.17446.
- [7] M. N. Aidi *et al.*, “Province clustering based on the percentage of communicable disease using the BCBimax biclustering algorithm,” *Geospat. Health*, vol. 18, no. 2, 2023, doi: 10.4081/gh.2023.1202.
- [8] W. A. L. Ningsih, I. M. Sumertajaya, and A. Saefuddin, “Pattern Detection of Economic and Pandemic Vulnerability Index in Indonesia Using Bi-Cluster Analysis,” *JUITA J. Inform.*, vol. 10, no. 2, pp. 273–282, 2022.
- [9] C. Wulandari, I. M. Sumertajaya, and M. N. Aidi, “Evaluation of Bicluster Analysis Results in Capture Fisheries Using the BCBimax Algorithm,” *JUITA J. Inform.*, vol. 11, no. 1, pp. 57–66, 2023.
- [10] M. G. Silva, S. C. Madeira, and R. Henriques, “Water Consumption Pattern Analysis Using Biclustering: When, Why and How,” *Water (Switzerland)*, vol. 14, no. 12, p. 1954, 2022, doi: 10.3390/w14121954.
- [11] E. N. Castanho, H. Aidos, and S. C. Madeira, “Biclustering fMRI time series: a comparative study,” *BMC Bioinformatics*, vol. 23, no. 1, pp. 1–30, 2022.
- [12] B. Pontes, R. Giráldez, and J. S. Aguilar-Ruiz, “Biclustering on expression data: A review,” *J. Biomed. Inform.*, vol. 57, pp. 163–180, 2015, doi: 10.1016/j.jbi.2015.06.028.
- [13] G. Piscopo and M. Resta, “Applying spectral biclustering to mortality data,” *Risks*, vol. 5, no. 2, 2017, doi: 10.3390/risks5020024.
- [14] S. Shazlyn Milleana, S. Ismail, C. M. N. Siti Mariana, and A. Norhaiza, “An efficient method to improve the clustering performance using hybrid robust principal component analysis-spectral biclustering in rainfall patterns identification,” *IAES Int. J. Artif. Intell.*, vol. 8, no. 3, pp. 237–243, 2019, doi: 10.11591/ijai.v8.i3.pp237-243.
- [15] S. C. Madeira and A. L. Oliveira, “Biclustering algorithms for biological data analysis: a

- survey,” *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 1, no. 1, pp. 24–45, 2004.
- [16] S. Sun and K. C. Das, “On the second largest normalized Laplacian eigenvalue of graphs,” *Appl. Math. Comput.*, vol. 348, pp. 531–541, 2019.
- [17] Y.-R. Lee, J.-H. Lee, and C.-H. Jun, “Validation measures of bicluster solutions,” *Ind. Eng. Manag. Syst.*, vol. 8, no. 2, pp. 101–108, 2009.