

EVALUASI *BICLUSTERING* CHENG & CHURCH DAN BCBIMAX TERHADAP DATA MULTIDISTRIBUSI DAN *OVERLAP* DALAM PEUBAH SDGs

DESTRIANA AULIA RIFALDI



**PROGRAM STUDI MAGISTER STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



PERNYATAAN MENGENAI TESIS DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa tesis dengan judul “Evaluasi *Biclustering* Cheng & Church dan BCBimax terhadap Data Multidistribusi dan *Overlap* dalam Peubah SDGs” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir tesis ini. Dalam proses penyusunan tesis ini, saya memanfaatkan kecerdasan artifisial yaitu ChatGPT secara terbatas untuk membantu parafrase kalimat, meningkatkan keterbacaan naskah, serta membantu penelusuran referensi ilmiah. Seluruh analisis data, pengolahan hasil, interpretasi, penyusunan argumen ilmiah, verifikasi referensi, dan penarikan kesimpulan sepenuhnya merupakan tanggung jawab penulis.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Agustus 2026

Destriana Aulia Rifaldi



RINGKASAN

DESTRIANA AULIA RIFALDI. Evaluasi *Biclustering* Cheng & Church dan BCBimax terhadap Data Multidistribusi dan *Overlap* dalam Peubah SDGs. Dibimbing oleh INDAHAWATI dan MUHAMMAD NUR AIDI.

Biclustering merupakan metode pengelompokan dua dimensi yang mampu mengidentifikasi kelompok objek dan peubah secara simultan dalam suatu matriks data. Metode *clustering* konvensional umumnya melakukan pengelompokan objek atau peubah secara terpisah berdasarkan keseluruhan peubah atau objek yang tersedia. *Biclustering* mampu menemukan subkelompok objek yang memiliki pola serupa hanya pada sebagian peubah tertentu sehingga dapat mengungkap pola lokal yang tidak selalu terlihat melalui metode pengelompokan konvensional. Kemampuan tersebut menjadikan *biclustering* banyak digunakan pada data berdimensi tinggi dan heterogen, termasuk dalam bidang bioinformatika, ekonomi, kesehatan, dan ilmu sosial. Penerapan *biclustering* pada data *Sustainable Development Goals* (SDGs) dapat digunakan untuk mengidentifikasi kelompok negara yang memiliki pola capaian serupa pada peubah pembangunan tertentu.

Algoritme *biclustering* memiliki karakteristik dan mekanisme pembentukan *bicluster* yang berbeda. Algoritme Cheng & Church (CC) dirancang untuk data numerik kontinu dengan menggunakan *Mean Squared Residue* (MSR) sebagai ukuran koherensi internal *bicluster*. Algoritme tersebut mencari submatriks dengan pola yang relatif koheren berdasarkan nilai residue yang rendah. Algoritme BCBimax merupakan algoritme yang bekerja pada data biner dan dikembangkan dari algoritme Bimax untuk menemukan *bicluster* yang bersifat inklusi maksimal. Penerapan BCBimax memerlukan proses binarisasi sehingga hasil *biclustering* dipengaruhi oleh representasi biner dan pemilihan *threshold*. Perbedaan karakteristik tersebut dapat menyebabkan Cheng & Church dan BCBimax memberikan hasil yang berbeda ketika diterapkan pada kondisi data yang berbeda.

Evaluasi kinerja algoritme *biclustering* perlu mempertimbangkan karakteristik data yang digunakan. Penelitian sebelumnya banyak menggunakan data simulasi yang mengikuti distribusi normal dengan struktur *bicluster* tanpa *overlap*. Kondisi tersebut belum sepenuhnya merepresentasikan data nyata yang dapat memiliki distribusi tidak simetris, rentang nilai terbatas, serta struktur *overlap* antar-*bicluster*. Distribusi data *background* dapat memengaruhi kemampuan algoritme dalam membedakan pola *bicluster* dari pola yang terdapat pada data latar. Struktur *overlap* juga meningkatkan kompleksitas karena objek atau peubah tertentu dapat menjadi bagian dari lebih dari satu *bicluster*. Kondisi multidistribusi dan *overlap* menjadi aspek penting dalam mengevaluasi kemampuan algoritme *biclustering* secara lebih komprehensif.

Penelitian ini bertujuan mengevaluasi kinerja algoritme Cheng & Church dan BCBimax pada data simulasi dengan karakteristik distribusi *background* dan tingkat *overlap* yang berbeda. Penelitian ini juga bertujuan mengetahui pengaruh serta interaksi antara algoritme, distribusi data, dan tingkat *overlap* terhadap kinerja *biclustering*. Penerapan kedua algoritme pada data empiris SDGs dilakukan untuk mengidentifikasi pola kelompok negara dan peubah pembangunan berkelanjutan yang dihasilkan oleh masing-masing algoritme.

Data simulasi dibangkitkan menggunakan matriks *background* berukuran 50×50 dengan tiga karakteristik distribusi Beta, yaitu Beta(5,5), Beta(2,5), dan Beta(5,2). Distribusi Beta(5,5) digunakan untuk merepresentasikan distribusi yang relatif simetris, sedangkan Beta(2,5) dan Beta(5,2) merepresentasikan distribusi menjulur ke kanan dan menjulur ke kiri. *Bicluster* disisipkan ke dalam matriks *background* dan digunakan sebagai *ground truth* dalam proses evaluasi. Tingkat *overlap* yang digunakan terdiri atas 0%, 20%, dan 50% untuk merepresentasikan kondisi tanpa tumpang tindih hingga kondisi dengan area irisan yang lebih luas. Setiap kombinasi algoritme, distribusi, dan tingkat *overlap* diulang sebanyak 100 kali.

Kinerja algoritme pada data simulasi dievaluasi menggunakan *Average Squared Residue* (ASR) dan Indeks Jaccard. ASR digunakan untuk mengevaluasi tingkat koherensi internal *bicluster* yang dihasilkan oleh algoritme. Indeks Jaccard digunakan untuk mengukur kesesuaian antara *bicluster* hasil algoritme dan *ground truth*. Pengaruh algoritme, distribusi, dan tingkat *overlap* dianalisis menggunakan ANOVA tiga arah. Interaksi antara faktor-faktor tersebut digunakan untuk mengetahui apakah pengaruh suatu faktor terhadap kinerja *biclustering* bergantung pada faktor lainnya. Plot interaksi digunakan untuk memperjelas pola perubahan kinerja algoritme pada berbagai karakteristik distribusi dan tingkat *overlap*.

Hasil simulasi menunjukkan adanya interaksi antara algoritme, distribusi data *background*, dan tingkat *overlap* terhadap kinerja *biclustering*. Interaksi tiga faktor tersebut menunjukkan bahwa perubahan kinerja algoritme pada suatu tingkat *overlap* bergantung pada karakteristik distribusi *background* yang digunakan. Perubahan tingkat *overlap* tidak menghasilkan pola kinerja yang sama pada seluruh distribusi dan algoritme. Interaksi antara algoritme dan distribusi juga menunjukkan bahwa pola kinerja Cheng & Church dan BCBimax berbeda pada karakteristik distribusi *background* yang berbeda. Plot interaksi memperlihatkan perubahan nilai respons kedua algoritme pada setiap distribusi, dengan pola garis yang menunjukkan adanya perubahan perbedaan kinerja antaralgoritme. Garis yang tidak sejajar menunjukkan bahwa pengaruh algoritme terhadap kinerja tidak sama pada seluruh distribusi. Hasil tersebut menunjukkan bahwa tidak terdapat satu algoritme yang secara konsisten memberikan kinerja terbaik pada seluruh kombinasi kondisi simulasi. Evaluasi algoritme *biclustering* perlu mempertimbangkan interaksi antara algoritme, distribusi data, dan tingkat *overlap* secara bersamaan.

Data empiris yang digunakan merupakan data SDGs tahun 2024 yang terdiri atas 167 negara dan 17 peubah. Tahap praproses meliputi penanganan nilai hilang menggunakan *Multiple Imputation by Chained Equations* (MICE), standarisasi menggunakan transformasi *z-score*, serta proses binarisasi sesuai karakteristik distribusi data untuk penerapan BCBimax. Data hasil praproses digunakan untuk membentuk *bicluster* menggunakan algoritme Cheng & Church dan BCBimax. Hasil *biclustering* digunakan untuk mengidentifikasi kelompok negara yang memiliki pola capaian serupa pada peubah SDGs tertentu.

Penerapan algoritme Cheng & Church menghasilkan tujuh *bicluster* dengan cakupan negara yang relatif lebih luas. Struktur tersebut memungkinkan identifikasi pola capaian SDGs pada kelompok negara yang lebih besar berdasarkan nilai numerik kontinu. Algoritme BCBimax menghasilkan 38 *bicluster* dengan ukuran dan cakupan yang lebih spesifik serta relatif lebih homogen. Representasi data dalam bentuk biner memungkinkan BCBimax membentuk kelompok negara yang

memiliki pola kategori capaian yang lebih seragam pada peubah tertentu. Karakteristik tersebut membuat BCBimax mampu mengidentifikasi pola lokal yang lebih spesifik pada kelompok negara dan peubah tertentu. Perbedaan jumlah, cakupan, dan homogenitas *bicluster* menunjukkan bahwa kedua algoritme memberikan karakteristik hasil yang berbeda dalam mengungkap pola pada data SDGs.

Kesamaan hasil Cheng & Church dan BCBimax berdasarkan Indeks Jaccard sebesar 32,40% menunjukkan bahwa kedua algoritme mengidentifikasi sebagian pola yang sama, tetapi juga menghasilkan struktur *bicluster* yang berbeda. Perbedaan tersebut berkaitan dengan mekanisme pembentukan *bicluster* dan bentuk representasi data yang digunakan oleh masing-masing algoritme. Cheng & Church mempertahankan informasi numerik kontinu dalam proses pembentukan *bicluster*, sedangkan BCBimax menggunakan representasi biner sehingga pola yang terbentuk didasarkan pada kesamaan kategori hasil binarisasi. Hasil tersebut menunjukkan bahwa Cheng & Church lebih memberikan gambaran pola dengan cakupan yang lebih luas, sedangkan BCBimax lebih menonjolkan pola lokal yang spesifik dan relatif homogen.

Secara keseluruhan, penelitian ini menunjukkan bahwa karakteristik distribusi data dan interaksinya dengan algoritme merupakan aspek penting dalam evaluasi kinerja *biclustering*. Tingkat *overlap* juga perlu dipertimbangkan karena perubahan struktur tumpang tindih dapat memengaruhi kemampuan algoritme dalam mengidentifikasi *bicluster*. Penerapan pada data SDGs menunjukkan bahwa Cheng & Church dan BCBimax mampu mengungkap pola dengan karakteristik yang berbeda. Pemilihan algoritme *biclustering* perlu disesuaikan dengan karakteristik data, bentuk representasi data, serta tujuan analisis agar pola hubungan antara kelompok negara dan peubah pembangunan berkelanjutan dapat diidentifikasi dan diinterpretasikan secara lebih tepat.

Kata kunci: *Biclustering*, BCBimax, Cheng & Church, Multidistribusi, *Overlap*

SUMMARY

DESTRIANA AULIA RIFALDI. Evaluation of Cheng & Church and BCBimax *Biclustering* on Multidistributed Data and *Overlap* in SDGs Variables. Supervised by INDAH WATI and MUHAMMAD NUR AIDI.

Biclustering is a two-dimensional clustering method that simultaneously identifies groups of objects and variables within a data matrix. Conventional clustering methods generally group objects or variables separately based on the available variables or objects. Biclustering can identify subsets of objects that exhibit similar patterns across selected variables, enabling the detection of local patterns that may not be revealed by conventional clustering methods. This capability makes biclustering useful for high-dimensional and heterogeneous data and has led to its application in various fields, including bioinformatics, economics, health, and social sciences. The application of biclustering to Sustainable Development Goals (SDGs) data can be used to identify groups of countries with similar achievement patterns across selected development variables.

Biclustering algorithms differ in their characteristics and mechanisms for constructing biclusters. The Cheng & Church (CC) algorithm is designed for continuous numerical data and uses the Mean Squared Residue (MSR) as a measure of internal bicluster coherence. The algorithm searches for submatrices with relatively coherent patterns based on low residue values. The BCBimax algorithm operates on binary data and was developed from the Bimax algorithm to identify inclusion-maximal biclusters. The application of BCBimax requires a binarization process, making the resulting biclustering structure dependent on the binary representation and threshold selection. These differences in algorithm characteristics may result in different performance when Cheng & Church and BCBimax are applied to different data conditions.

Evaluation of biclustering performance needs to consider the characteristics of the underlying data. Previous studies have frequently used simulation data following a normal distribution with non-overlapping bicluster structures. Such conditions do not fully represent real-world data, which may exhibit asymmetric distributions, bounded ranges, and overlapping structures among biclusters. Background data distributions may affect the ability of an algorithm to distinguish bicluster patterns from structures present in the background data. Overlapping structures may also increase data complexity because particular objects or variables may belong to more than one bicluster. Multidistribution and overlap conditions are therefore important aspects in evaluating biclustering algorithms more comprehensively.

This study aims to evaluate the performance of the Cheng & Church and BCBimax algorithms on simulated data with different background distribution characteristics and overlap levels. The study also examines the effects and interactions among algorithm, data distribution, and overlap level on biclustering performance. The application of both algorithms to empirical SDGs data is conducted to identify patterns among groups of countries and sustainable development variables generated by each algorithm.

The simulation data were generated using a 50×50 background matrix with three Beta distribution characteristics, namely Beta(5,5), Beta(2,5), and Beta(5,2).



Beta(5,5) represents a relatively symmetric distribution, whereas Beta(2,5) and Beta(5,2) represent right-skewed and left-skewed distributions, respectively. Biclusters were embedded into the background matrix and used as the ground truth for evaluation. The overlap levels were 0%, 20%, and 50%, representing conditions ranging from no overlap to conditions with larger intersection areas. Each combination of algorithm, distribution, and overlap level was replicated 100 times.

Algorithm performance on the simulation data was evaluated using the Average Squared Residue (ASR) and the Jaccard Index. ASR was used to evaluate the internal coherence of the biclusters generated by the algorithms. The Jaccard index was used to measure the agreement between the biclusters generated by the algorithms and the ground truth. The effects of algorithm, distribution, and overlap were analyzed using three-way ANOVA. Interactions among the factors were examined to determine whether the effect of one factor on biclustering performance depended on another factor. Interaction plots were used to further illustrate changes in algorithm performance across different distribution characteristics and overlap levels.

The simulation results showed an interaction among algorithm, background distribution, and overlap level in determining biclustering performance. The three-way interaction indicates that changes in algorithm performance at a particular overlap level depend on the characteristics of the background distribution. Changes in overlap levels did not produce the same performance pattern across all distributions and algorithms. The interaction between algorithm and distribution also demonstrated that the performance patterns of Cheng & Church and BCBimax differed across background distribution characteristics. The interaction plots illustrated changes in the response values of the two algorithms across the distributions, with line patterns indicating changes in the performance difference between the algorithms. Non-parallel lines indicate that the effect of the algorithm on performance is not consistent across all distributions. The findings demonstrate that no single algorithm consistently provides the best performance across all combinations of simulation conditions. Evaluation of biclustering algorithms therefore needs to consider the interaction among algorithm, data distribution, and overlap simultaneously.

The empirical data consisted of the 2024 SDGs dataset containing 167 countries and 17 variables. Data preprocessing included missing-value handling using Multiple Imputation by Chained Equations (MICE), standardization using the z-score transformation, and binarization according to the distribution characteristics for the BCBimax algorithm. The preprocessed data were analyzed using the Cheng & Church and BCBimax algorithms. The resulting biclusters were used to identify groups of countries exhibiting similar achievement patterns across selected SDGs variables.

The Cheng & Church algorithm produced seven biclusters with relatively broader country coverage. This structure enabled the identification of SDGs achievement patterns across larger groups of countries based on continuous numerical values. The BCBimax algorithm produced 38 biclusters with more specific sizes and coverage and relatively higher homogeneity. Binary data representation enabled BCBimax to form groups of countries with more homogeneous achievement categories across particular variables. This characteristic allowed BCBimax to identify more specific local patterns among



particular groups of countries and variables. Differences in the number, coverage, and homogeneity of the resulting biclusters demonstrate that the two algorithms produce different characteristics in revealing patterns within the SDGs data.

The similarity between the Cheng & Church and BCBimax results based on the Jaccard Index was 32.40%, indicating that the two algorithms identified some common patterns while also producing different bicluster structures. These differences are associated with the mechanisms used to construct biclusters and the form of data representation used by each algorithm. Cheng & Church retains continuous numerical information during bicluster construction, whereas BCBimax uses binary representation, resulting in patterns based on similarities in the categories obtained from binarization. The findings indicate that Cheng & Church provides broader patterns with wider coverage, whereas BCBimax emphasizes more specific and relatively homogeneous local patterns.

Overall, this study demonstrates that data distribution characteristics and their interactions with the algorithm are important aspects of biclustering performance evaluation. The level of overlap also needs to be considered because changes in overlapping structures may affect the ability of algorithms to identify biclusters. The application to SDGs data demonstrates that Cheng & Church and BCBimax can reveal patterns with different characteristics. Selection of a biclustering algorithm should consider the characteristics of the data, the form of data representation, and the analytical objective so that patterns between groups of countries and sustainable development variables can be identified and interpreted appropriately.

Keywords: *Biclustering*, BCBimax, Cheng & Church, Multidistribution, *Overlap*.



Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2026
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.



EVALUASI *BICLUSTERING* CHENG & CHURCH DAN BCBIMAX TERHADAP DATA MULTIDISTRIBUSI DAN *OVERLAP* DALAM PEUBAH SDGs

DESTRIANA AULIA RIFALDI

Tesis
sebagai salah satu syarat untuk memperoleh gelar
Magister pada
Program Studi Statistika dan Sains Data

**PROGRAM STUDI MAGISTER STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Penguji Luar Komisi Pembimbing pada Ujian Tesis:
Dr. Ir. Erfiani, M.Si.



Judul Tesis : Evaluasi *Biclustering* Cheng & Church dan BCBimax terhadap Data Multidistribusi dan *Overlap* dalam Peubah SDGs
Nama : Destriana Aulia Rifaldi
NIM : M0501241081

Disetujui oleh

Pembimbing 1:
Dr. Ir. Indahwati, M. Si.

Pembimbing 2:
Prof. Dr. Ir. Muhammad Nur Aidi M.S.

Diketahui oleh

Ketua Program Studi:
Dr. Agus Mohamad Soleh, S. Si., MT.
NIP. 197503151999031000

Dekan Sekolah Sains Data, Matematika, dan Informatika:
Prof. Dr. Ir. Agus Buono, M.Si., M.Kom.
NIP 196607021993021001

Tanggal Ujian :
4 Agustus 2026

Tanggal Lulus :



PRAKATA

Puji dan syukur penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat, karunia, dan hidayah-Nya, sehingga tesis yang berjudul “Evaluasi *Biclustering* Cheng & Church dan BCBimax terhadap Data Multidistribusi dan *Overlap* dalam Peubah SDGs” dapat disusun dan diselesaikan dengan baik. Penulis menyadari bahwa tesis ini masih memiliki keterbatasan, baik dari segi penyajian maupun substansi. Oleh karena itu, kritik dan saran yang bersifat konstruktif sangat diharapkan sebagai bahan perbaikan dan penyempurnaan penelitian ini di masa mendatang.

Penyusunan tesis ini tidak terlepas dari bantuan, bimbingan, serta dukungan dari berbagai pihak yang telah memberikan kontribusi secara langsung maupun tidak langsung. Pada kesempatan ini, penulis ingin menyampaikan rasa terima kasih dan penghargaan yang setulus-tulusnya kepada:

1. Ibu Dr. Ir. Indahwati, M.Si. dan Bapak Prof. Dr. Ir. Muhammad Nur Aidi, M.S. selaku komisi pembimbing yang telah meluangkan waktu, tenaga, dan pikiran untuk memberikan bimbingan, arahan, masukan, serta motivasi yang sangat berharga selama proses penyusunan tesis ini. Saran dan pandangan yang diberikan telah banyak membantu penulis dalam memperjelas arah penelitian dan memperdalam pemahaman metodologis.
2. Bapak Dr. Agus Mohamad Soleh, S.Si., M.T. selaku moderator kolokium dan Prof. Dr. Ir. Jarwadi Budi Hernowo, M.Sc selaku moderator seminar hasil yang telah memberikan arahan, masukan, dan saran yang konstruktif sehingga bermanfaat dalam penyempurnaan tesis ini
3. Keluarga besar Departemen Statistika dan Sains Data IPB University atas lingkungan akademik yang kondusif, ilmu pengetahuan, serta berbagai fasilitas dan dukungan yang diberikan, sehingga penulis dapat menjalani proses penyusunan tesis ini dengan baik.
4. Orang tua dan keluarga yang senantiasa memberikan doa, dukungan moral, perhatian, serta bantuan material yang tak ternilai, sehingga penulis dapat menyelesaikan tesis ini dengan penuh semangat dan ketekunan.
5. Rekan-rekan mahasiswa dan sahabat yang telah memberikan dukungan, motivasi, serta diskusi yang membangun selama proses penyusunan tesis ini.

Akhir kata, penulis berharap semoga tesis ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan, khususnya dalam bidang statistika dan analisis data, serta menjadi kontribusi yang berguna bagi pihak-pihak yang membutuhkan.

Bogor, Agustus 2026

Destriana Aulia Rifaldi

DAFTAR ISI

DAFTAR ISI	v
DAFTAR GAMBAR	vi
DAFTAR TABEL	vi
DAFTAR LAMPIRAN	vii
I. PENDAHULUAN	8
1.1 Latar Belakang	8
1.2 Rumusan Masalah	11
1.3 Tujuan Penelitian	11
1.4 Manfaat Penelitian	11
II. TINJAUAN PUSTAKA	12
2.1 <i>Biclustering</i>	12
2.2 <i>Binary Inclusion-Maximal (BCBimax) Biclustering</i>	13
2.3 Cheng and Church (CC) <i>Bicluster</i>	14
2.4 Evaluasi Hasil Algoritme <i>Biclustering</i>	15
2.5 <i>Analysis of Variance</i> (ANOVA) Tiga Arah	17
2.6 <i>Sustainable Development Goals</i> (SDGs)	18
III. METODE	21
3.1 Sumber Data dan Peubah Penelitian	21
3.2 Langkah Analisis	22
3.2.1. Langkah Analisis Data Simulasi	22
3.2.2. Langkah Analisis Data Empiris	24
IV. PEMBAHASAN	26
4.1 Eksplorasi Data Simulasi	26
4.2 Kajian Simulasi	27
4.3 Eksplorasi Data Empiris	35
4.4 Hasil Algoritme CC <i>Biclustering</i>	41
4.5 Hasil Algoritme BCBimax <i>Biclustering</i>	49
4.6 Evaluasi Kinerja Algoritme <i>Biclustering</i>	58
V. SIMPULAN DAN SARAN	61
DAFTAR PUSTAKA	63
LAMPIRAN	66



DAFTAR GAMBAR

1	Perbandingan <i>outcome cluster</i> dan <i>bicluster</i>	12
2	Alur penelitian data simulasi	24
3	Alur penelitian data empiris	25
4	Ilustrasi skenario data simulasi	26
5	Ilustrasi pengacakan matriks	27
6	<i>Boxplot</i> sebaran nilai indeks jaccard dan ASR CC dan BCBimax	28
7	Plot interaksi respon ASR	31
8	Plot interaksi respon indeks jaccard	34
9	Distribusi peubah data SDGs	36
10	<i>Heatmap</i> scaling peubah SDGs	37
11	<i>Boxplot</i> scaling peubah SDGs	38
12	<i>Biplot</i> PCA SDGs	39
13	Evaluasi <i>threshold</i> dengan jumlah <i>bicluster</i> (a), evaluasi <i>threshold</i> dengan rata-rata MSR per volume (b), evaluasi <i>threshold</i> dengan rata-rata MSR (c), evaluasi <i>threshold</i> dengan indeks jaccard (d)	42
14	Grafik profilisasi hasil algoritme CC	47
15	Peta sebaran negara hasil algoritme CC	49
16	Jumlah <i>bicluster</i> berdasarkan kombinasi baris dan kolom	50
17	Nilai ASR kombinasi baris dan kolom	51
18	Nilai indeks jaccard kombinasi baris dan kolom	52
19	Peta sebaran negara hasil algoritme BCBimax	58

DAFTAR TABEL

1	Peubah penelitian data empiris	21
2	Skenario <i>overlap</i>	23
3	Ringkasan nilai rata – rata ASR dan indeks jaccard data simulasi	29
4	Ringkasan anova tiga arah respon ASR	30
5	Ringkasan anova tiga arah respon indeks jaccard	32
6	Hasil evaluasi <i>threshold</i>	42
7	Karakteristik <i>bicluster</i> optimal algoritme CC	43
8	Pendekatan binerisasi BCBimax <i>biclustering</i>	49
9	Karakteristik <i>bicluster</i> optimal algoritme BCBimax	52
10	Evaluasi nilai ASR algoritme CC dan BCBimax	59
11	Ringkasan evaluasi algoritme CC dan BCBimax	60



DAFTAR LAMPIRAN

1	Hasil <i>tunning</i> delta algoritma CC	67
2	Profil masing – masing <i>bicluster</i> hasil algoritma BCBimax	69