

- 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.

KAJIAN ALGORITMA *CHENG & CHURCH* DAN *SPECTRAL BICLUSTERING* PADA DATA BERKORELASI DAN STRUKTUR *BICLUSTER* TUMPANG TINDIH

SITI HAFSAH



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

**BOGOR
2025**



@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 “Kajian Algoritma *Cheng & Church* dan *Spectral Biclustering* pada Data Berkorelasi dan Stuktur *Bicluster* Tumpang Tindih” 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.

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

Bogor, Agustus 2025

Siti Hafsah
G1501231060

- 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.



RINGKASAN

SITI HAFSAH. Kajian Algoritma *Cheng & Church* dan *Spectral Biclustering* pada Data Berkorelasi dan Struktur *Bicluster* Tumpang Tindih. Dibimbing oleh INDAHWATI dan HARI WIJAYANTO.

Biclustering merupakan metode pengelompokan dua arah yang secara simultan menghubungkan subset baris dan subset kolom untuk membentuk submatriks koheren. Berbeda dengan *clustering* konvensional yang hanya bekerja pada satu dimensi (baris atau kolom), *biclustering* mampu mendeteksi pola lokal tersembunyi yang hanya muncul pada sebagian unit dan sebagian peubah secara bersamaan. Pendekatan *biclustering* telah berkembang dari bidang bioinformatika menuju aplikasi yang lebih luas, seperti sosial ekonomi, kesehatan dan lain sebagainya. Dalam penerapannya, *biclustering* menghadapi dua tantangan umum struktural utama: korelasi antarpeubah dan *overlap* keanggotaan. Korelasi menyebabkan informasi menjadi redundan sehingga menurunkan ketepatan deteksi pola, sedangkan *overlap* mengaburkan batas antar *bicluster* karena satu baris atau kolom dapat tergabung dalam lebih dari satu kelompok. Penelitian ini bertujuan mengkaji kinerja dua algoritma *biclustering* representatif, yaitu *Cheng & Church* (CC) yang berbasis minimisasi *Mean Squared Residue* (MSR) dan *Spectral Biclustering* yang berbasis *Singular Value Decomposition* (SVD) dalam mengidentifikasi pola *bicluster* pada data yang mengandung korelasi dan *overlap*.

Kajian dilakukan pada data simulasi dan data empiris. Data simulasi dibangkitkan dalam bentuk matriks 50×50 dengan 3 tingkat korelasi ($\rho = 0,3; 0,6; 0,9$) dan 3 tingkat *overlap* (tanpa, kecil, besar), serta dua *bicluster* konstan sebagai *ground truth*. Evaluasi dilakukan berdasarkan indeks Liu dan Wang (ILW) yang mengukur kesesuaian hasil *biclustering* terhadap struktur aktual, dan dianalisis menggunakan ANOVA tiga arah. Hasil menunjukkan bahwa algoritma CC optimal dalam kondisi korelasi rendah dan tanpa *overlap* (ILW mendekati 0,95), tetapi performanya menurun tajam ketika korelasi antar peubah tinggi dan *overlap* besar (ILW $\approx 0,50$). Sebaliknya, *Spectral Biclustering* mempertahankan performa stabil dengan ILW konsisten pada rentang 0,7 – 0,9 di hampir semua skenario. Analisis ANOVA menegaskan bahwa *overlap* merupakan sumber variasi dominan dalam penurunan performa algoritma ($\eta^2 = 42,85\%$), diikuti jenis algoritma dan korelasi. Interaksi algoritma dan korelasi juga signifikan, menunjukkan bahwa efektivitas algoritma sangat bergantung pada struktur data yang dihadapi.

Data empiris menggunakan IPD Provinsi Riau yang mencakup 172 kecamatan dan 21 peubah pembangunan desa hasil agregasi dari data PODES 2024. Struktur data menunjukkan korelasi antar peubah yang rendah (75,71% pasangan $r \leq 0,3$) dan distribusi nilai yang tidak homogen, menjadikannya relevan untuk pendekatan *biclustering*. Algoritma CC menghasilkan 16 *bicluster* non-*overlap* dengan nilai ASR sebesar 0,05, menggambarkan kekompakan internal dan segmentasi lokal yang tajam. Sementara itu, *Spectral Biclustering* menghasilkan 15 *bicluster* dengan pola *overlap* dan nilai ASR sebesar 0,43, mencerminkan struktur *checkerboard* global yang kompleks. Evaluasi nilai ILW menunjukkan kesamaan keanggotaan hanya sebesar 22%, yang menandakan bahwa masing-masing algoritma mengungkap pola dari perspektif yang berbeda. *Profiling bicluster* mengungkap bahwa CC mengelompokkan wilayah berdasarkan peubah penciri yang spesifik dan homogen, sedangkan *Spectral* mengidentifikasi keterkaitan



multidimensi antar kecamatan dan peubah, dengan kecenderungan *overlap* keanggotaan pada beberapa *bicluster* sekaligus.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa algoritma *Cheng & Church* lebih sesuai untuk data dengan batas kelompok tegas dan korelasi rendah, sedangkan *Spectral Biclustering* unggul dalam menangani struktur data yang *overlap* dan multidimensi. Temuan ini menegaskan pentingnya mempertimbangkan karakteristik struktural data dalam memilih algoritma *biclustering*, agar hasil segmentasi wilayah lebih adaptif dan dapat dijadikan dasar dalam pengambilan keputusan pembangunan yang berbasis data.

Kata kunci: *biclustering*, *Cheng & Church*, korelasi, *overlap*, *Spectral Biclustering*.

- 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.



SUMMARY

SITI HAFSAH. A Study of The Cheng & Church Algorithm and Spectral Biclustering on Correlated and Overlapping Bicluster Structure. Supervised by INDAHWATI dan HARI WIJAYANTO.

Biclustering is a bidirectional clustering method that simultaneously links a subset of rows (objects) and a subset of columns (variables) to form a coherent submatrix. Unlike clustering that only works in one dimension, biclustering is able to detect hidden local patterns that only appear across multiple units and multiple variables at a time. This approach has expanded from the field of bioinformatics to broader applications, such as socioeconomics, healthcare, and so on. In its application, biclustering faces two main structural common challenges: collinearity between variables and membership overlap. Collinearity causes information to be redundant, which decreases the accuracy of pattern detection, while overlap blurs the boundaries between biclusters as a single row or column can belong to more than one cluster. This study aims to evaluate the performance of two representative biclustering algorithms, namely Cheng & Church (CC) based on Mean Squared Residue (MSR) minimization and Spectral Biclustering based on singular value decomposition (SVD) in identifying bicluster patterns on data containing collinearity and overlap.

The study was conducted on simulated and empirical data. Simulated data were generated in the form of a 50×50 matrix with 3 levels of collinearity ($\rho = 0,3; 0,6; 0,9$) and 3 levels of overlap (no, small, large), and two constant biclusters as ground truth. Evaluation was performed based on the Liu & Wang (ILW) index which measures the conformity of the biclustering results to the actual structure, and analyzed using three-way ANOVA. The results show that the CC algorithm is optimal under the condition of low collinearity without overlap (ILW close to 0,95), but its performance drops sharply when the correlation between variables is high and overlap is large ($ILW \approx 0,50$). In contrast, Spectral Biclustering maintained stable performance with ILW consistent in the range of 0,7 – 0,9 in almost all scenarios. ANOVA analysis confirmed that overlap was the dominant source of variation in algorithm performance degradation ($\eta^2 = 42,85\%$), followed by algorithm type and collinearity. The interaction of algorithm and collinearity was also significant, suggesting that the effectiveness of the algorithm is highly dependent on the structure of the data at hand.

The empirical data uses Riau Province's IPD covering 172 sub-districts and 21 village development variables aggregated from PODES 2024 data. The structure of the data shows low correlation between variables (75,71% pairs $r \leq 0,3$) and inhomogeneous distribution of values, making it relevant for a biclustering approach. The CC algorithm generated 16 non-overlapping biclusters with an ASR value of 0,05, illustrating internal compactness and sharp local segmentation. Meanwhile, Spectral Biclustering produced 15 biclusters with an overlap pattern and an ASR value of 0,43, reflecting a complex global checkerboard structure. Evaluation of the ILW values showed a membership similarity of only 22%, indicating that each algorithm revealed patterns from different perspectives. Bicluster profiling revealed that CC clustered areas based on specific and homogeneous characterizing variables, while Spectral identified multidimensional



linkages between sub-districts and variables, with a tendency to overlap membership in several biclusters at once.

Overall, the results show that Cheng & Church algorithm is more suitable for data with strict group boundaries and low correlation, while Spectral Biclustering excels in handling overlapping and multidimensional data structures. These findings emphasize the importance of considering the structural characteristics of the data in selecting a biclustering algorithm, so that the results of area segmentation are more adaptive and can be used as a basis for data-driven development decision-making.

Keywords: biclustering, Cheng & Church, correlated, overlapped, Spectral Biclustering.

- 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 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 2025
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.

KAJIAN ALGORITMA *CHENG & CHURCH* DAN *SPECTRAL BICLUSTERING* PADA DATA BERKORELASI DAN STRUKTUR *BICLUSTER* TUMPANG TINDIH

SITI HAFSAH

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
2025**



@Hak cipta milik IPB University

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:
Prof. Dr. Ir. Muhammad Nur Aidi, M.S.



Digitally signed by:
Muhammad Nur Aidi

Date: 13 Agu 2025 10.06.18 WIB
Verify at disign.ipb.ac.id

Judul Tesis : Kajian Algoritma *Cheng & Church* dan *Spectral Biclustering*
pada Data Berkorelasi dan Struktur *Bicluster* Tumpang Tindih
Nama : Siti Hafsa
NIM : G1501231060

Disetujui oleh

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

Pembimbing 2:
Prof. Dr. Ir. Hari Wijayanto, M.Si



Diketahui oleh

Ketua Program Studi:
Dr. Agus Mohamad Soleh, S.Si., M.T.
NIP. 19750315 199903 1 004

Dekan Sekolah Sains Data, Matematika, dan Informatika
Prof. Dr. Ir. Agus Buono, M.Si., M.Kom.
NIP. 19660702 199302 1 001



Tanggal Ujian:
06 Agustus 2025

Tanggal Lulus:



PRAKATA

Puji dan Syukur alhamdulillah penulis panjatkan kepada Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga karya ilmiah dengan judul “Kajian Algoritma *Cheng & Church* dan *Spectral Biclustering* pada Data Berkorelasi dan Stuktur *Bicluster* Tumpang Tindih” berhasil diselesaikan. Penulisan karya ilmiah ini penulis susun sebagai salah satu syarat untuk menyelesaikan Studi Magister Sains pada Program Studi Statistika dan Sains Data. Penulisan karya ilmiah ini dapat terselesaikan tidak terlepas dari dukungan, bimbingan, dan bantuan berbagai pihak sehingga penulis menyampaikan ucapan terima kasih kepada:

1. Ibu Dr. Ir. Indahwati, M.Si. dan Bapak Prof. Dr. Ir. Hari Wijayanto, M.Si. selaku Komisi Pembimbing yang senantiasa meluangkan waktu untuk memberikan bimbingan, masukan, dukungan, dan arahan selama proses penyusunan penelitian ini.
2. Bapak Prof. Dr. Ir. Muhammad Nur Aidi, M.S. selaku penguji luar komisi yang telah memberikan saran, kritik, dan masukan yang sangat berharga demi penyempurnaan tesis ini.
3. Kedua orang tua tercinta, Bapak Abdurrahman dan Ibu Maryeni, serta adik-adik Ahamad Husein Aljabar dan Rohadhotul Aisy, atas doa, dukungan, dan motivasi yang tiada henti.
4. Seluruh dosen pengajar Program Studi Statistika dan Sains atas ilmu yang telah diberikan;
5. Seluruh staf terutama staf tata usaha Program Studi Statistika dan Sains;
6. Rekan-rekan mahasiswa Pascasarjana Program Studi Statistika dan Sains Data angkatan 2023 atas dukungan, kebersamaan, dan kerja sama yang terjalin dengan baik.

Semoga penelitian ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Bogor, Agustus 2025

Siti Hafsa

DAFTAR ISI

DAFTAR TABEL	xii
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xiii
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan	3
1.4 Manfaat	3
II TINJAUAN PUSTAKA	4
2.1 <i>Exploratory Data Analysis</i> (EDA)	4
2.2 <i>Biclustering</i>	4
2.3 Algoritma <i>Cheng & Church</i> (CC)	8
2.4 Algoritma <i>Spectral Biclustering</i>	11
2.5 Evaluasi Kinerja Algoritma <i>Biclustering</i>	13
2.6 Pembangunan Wilayah	14
III METODOLOGI PENELITIAN	15
3.1 Data	15
3.2 Tahapan Analisis Data	17
IV HASIL DAN PEMBAHASAN	21
4.1 Eksplorasi Skenario Simulasi	21
4.2 Hasil Kajian Simulasi	23
4.3 Eksplorasi Data Empiris	27
4.4 Hasil Algoritma CC <i>Biclustering</i>	30
4.5 Hasil Algoritma <i>Spectral Biclustering</i>	35
4.6 Evaluasi Kinerja Algoritma <i>Biclustering</i>	40
V SIMPULAN DAN SARAN	42
5.1 Simpulan	42
5.2 Saran	42
DAFTAR PUSTAKA	43
LAMPIRAN	46
RIWAYAT HIDUP	57



DAFTAR TABEL

1	Posisi baris dan kolom <i>bicluster</i> simulasi	15
2	Daftar peubah IPD	16
3	Hasil ANOVA tiga arah terhadap nilai ILW	25
4	Ringkasan Nilai ILW pada Berbagai Skenario Simulasi	26
5	Evaluasi nilai <i>threshold</i> delta	31
6	Karakteristik <i>bicluster</i> optimal hasil algoritma CC	33
7	Evaluasi normalisasi <i>Spectral Biclustering</i>	37
8	Karakteristik <i>bicluster</i> optimal hasil algoritma <i>Spectral Biclustering</i>	37
9	Evaluasi kinerja algoritma CC dan <i>Spectral Biclustering</i>	40

DAFTAR GAMBAR

1	Ilustrasi perbedaan <i>clustering</i> dan <i>biclustering</i> (Maâtouk <i>et al.</i> 2021)	5
2	Ilustrasi tipe <i>Bicluster</i> (Silva <i>et al.</i> 2022)	7
3	Struktur <i>bicluster</i> (Castanho <i>et al.</i> 2024)	7
4	Diagram alir CC <i>biclustering</i>	10
5	Diagram alir <i>Spectral Biclustering</i>	13
6	Diagram alir tahapan penelitian untuk data simulasi	18
7	Diagram alir tahapan penelitian untuk data empiris	20
8	Visualisasi data simulasi	21
9	Ilustrasi pengacakan matriks data simulasi	22
10	Rata-rata korelasi antarpeubah pada setiap tingkat korelasi	22
11	<i>Boxplot</i> nilai indeks Liu dan Wang pada algoritma CC dan <i>Spectral Biclustering</i>	23
12	Plot Interaksi rata-rata indeks Liu dan Wang berdasarkan (a) No <i>overlap</i> , (b) <i>Small overlap</i> , (c) <i>Large overlap</i>	26
13	<i>Heatmap scaling</i> peubah IPD	27
14	<i>Boxplot scaling</i> peubah IPD	28
15	<i>Biplot</i> PCA	29
16	Korelasi antarpeubah	30
17	Evaluasi <i>threshold</i> delta algoritma CC berdasarkan (a) Jumlah <i>Bicluster</i> , (b) Rataan MSR per volume, dan (c) ASR	31
18	Indeks Liu dan Wang terhadap nilai delta	32
19	<i>Radar chart</i> CC berdasarkan (a) profil <i>bicluster</i> 1, (b) profil <i>bicluster</i> 5, (c) profil <i>bicluster</i> 8, dan (d) profil <i>bicluster</i> 14	34
20	Peta sebaran proporsi peubah potensi tinggi per kecamatan CC <i>biclustering</i>	35
21	Plot keragaman kumulatif menurut vektor ciri	36
22	<i>Elbow plot</i> penentuan jumlah <i>cluster</i> optimal untuk normalisasi (a) IRRC, (b) <i>Bistochasticization</i> , dan (c) <i>Log</i>	36
23	<i>Radar chart</i> <i>Spectral Biclustering</i> berdasarkan (a) profil <i>bicluster</i> 1, (b) profil <i>bicluster</i> 3, dan (c) profil <i>bicluster</i> 7	39
24	Peta sebaran proporsi peubah potensi tinggi per kecamatan <i>Spectral Biclustering</i>	40
25	Plot <i>profiling bicluster</i>	41

DAFTAR LAMPIRAN

1	Tuning parameter delta CC	47
2	Karakteristik isi <i>bicluster</i> berdasarkan potensi pembangunan wilayah pada algoritma CC	49
3	Karakteristik isi <i>bicluster</i> berdasarkan potensi pembangunan wilayah pada algoritma <i>Spectral Biclustering</i>	50
4	Karakteristik masing-masing kecamatan berdasarkan potensi pembangunan wilayah pada algoritma CC	51
5	Karakteristik masing-masing kecamatan berdasarkan potensi pembangunan wilayah pada algoritma <i>Spectral Biclustering</i>	52
6	<i>Radar Chart bicluster</i> algoritma CC	53
7	<i>Radar Chart bicluster</i> algoritma <i>Spectral Biclustering</i>	54
8	Profil masing-masing <i>bicluster</i> algoritma CC	55
9	Profil masing-masing <i>bicluster</i> algoritma <i>Spectral Biclustering</i>	56