DELTA: Jurnal Ilmiah Pendidikan Matematika
p.ISSN: 2303 -3983 e.ISSN:2548-3994
Vol. 14 No. 1 Bulan Januari Hal . 119 – 132
DOI: 10.31941/delta.v14i1.6722



PENDEKATAN TIGA FASE BICLUSTERING UNTUK MENGIDENTIFIKASI KORELASI BICLUSTER PADA DATA EKSPRESI GEN PENYAKIT DIABETES MELITUS

[1)]Soeganda Formalidin, [2)] Elise Natalia Manurung, [3)] Aulia Rahmawati, [4)] Muthia Arsyita Wardani, [5)] Rezi Prasetyo, [6)] Karolina Mulyati Edan
[1,2,3,4,5,6)] Universitas Pamulang
0wlmlkt@gmail.com

Received :
11/11/2025

Accepted :
25/12/2025

Published :
22/01/2026

Abstract

This study proposes the application of a three-phase biclustering method to identify patterns of association between genes and conditions in Diabetes Mellitus (DM) gene expression data, particularly in obese and lean samples. The proposed method consists of three main phases. The first phase applies Singular Value Decomposition (SVD) to reduce and transform gene expression data into gene- and condition-based matrix representations. The second phase employs Partition Around Medoids (PAM) with Euclidean distance to generate initial biclusters based on the similarity of expression patterns. The third phase evaluates and filters the biclusters using a Pearson correlation-based modified lift algorithm to detect more coherent additive-multiplicative patterns. The method was implemented on a DM microarray dataset comprising 1,331 selected genes. The SVD-PAM phase produced eight initial biclusters, while the filtering process using the modified lift algorithm resulted in three δ-corbiclusters. The δ-corbicluster values obtained were 0.097, 0.095, and 0.085, respectively. In the context of this method, smaller δ values indicate better pattern coherence between genes and conditions; therefore, values below 0.1 suggest that the resulting biclusters exhibit relatively strong and stable association patterns, although these values should not be directly interpreted as Pearson correlation coefficients. Thus, the three-phase biclustering method is able to refine initial biclusters into more coherent gene-condition groups and has the potential to support further medical research related to Diabetes Mellitus.

Keywords: *correlated bicluster; Diabetes mellitus; microarray data; modified lift algorithm; R*

Abstrak

Penelitian ini mengusulkan penerapan metode three-phase biclustering untuk mengidentifikasi pola keterkaitan antara gen dan kondisi pada data ekspresi gen Diabetes Melitus (DM), khususnya pada sampel obesitas dan kurus. Metode yang diusulkan terdiri atas tiga tahap utama. Tahap pertama menggunakan Singular Value Decomposition (SVD) untuk mereduksi dan mentransformasi data ekspresi gen ke dalam representasi matriks berbasis gen dan kondisi. Tahap kedua menerapkan Partition Around Medoids (PAM) dengan jarak Euclidean untuk membentuk bicluster awal berdasarkan kemiripan pola ekspresi. Tahap ketiga mengevaluasi dan menyaring bicluster menggunakan modified lift algorithm berbasis korelasi Pearson untuk mendeteksi pola additive-multiplicative yang lebih koheren. Implementasi dilakukan pada dataset microarray DM dengan 1.331 gen terpilih. Hasil tahap SVD-PAM menghasilkan delapan bicluster awal, sedangkan proses penyaringan menggunakan modified lift algorithm menghasilkan tiga δ-corbicluster. Nilai δ-corbicluster yang diperoleh masing-masing sebesar 0,097; 0,095; dan 0,085. Dalam konteks metode ini, nilai δ yang lebih kecil menunjukkan tingkat koherensi pola yang lebih baik antar gen dan kondisi, sehingga nilai di bawah 0,1 mengindikasikan bahwa bicluster yang terbentuk memiliki pola keterkaitan yang relatif kuat dan stabil. Dengan demikian, metode three-phase biclustering mampu menyaring bicluster awal menjadi kelompok gen-kondisi yang lebih koheren dan berpotensi mendukung analisis lanjutan dalam penelitian medis terkait Diabetes Melitus.

Kata Kunci: *Diabetes melitus*, korelasi *bicluster*, *microarray* data, modifikasi *lift algoritma*, R

To cite this article (in APA style):
Form alidin, S., Manurung, E. N., Rahmawati, A., Wardani, M. A., Prasetyo, R., & Edan, K. M. (2026). Pendekatan tiga fase biclustering untuk mengidentifikasi korelasi bicluster pada data ekspresi gen penyakit diabetes melitus. *Delta: Jurnal Ilmiah Pendidikan Matematika, 14*(1). https://doi.org/10.31941/delta.v14i1.6722



DELTA is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

1. Pendahuluan

Diabetes melitus (DM), atau dikenal sebagai penyakit kencing manis, merupakan kelainan metabolik kronis yang terjadi ketika pankreas tidak mampu memproduksi insulin secara cukup atau ketika tubuh tidak dapat menggunakan insulin secara efektif. Kondisi ini ditandai dengan hiperglikemia kronis serta gangguan metabolisme karbohidrat, lemak, dan protein yang dalam jangka panjang dapat menyebabkan komplikasi serius pada pembuluh darah, saraf, ginjal, jantung, dan mata (WHO, 2024). Diabetes melitus tipe 2 (DMT2) merupakan bentuk diabetes yang paling banyak ditemukan dan berkaitan erat dengan resistensi insulin, faktor genetik, pola hidup, serta obesitas. Secara global, beban penyakit diabetes terus meningkat. WHO melaporkan bahwa jumlah penderita diabetes meningkat dari sekitar 200 juta orang pada tahun 1990 menjadi sekitar 830 juta orang pada tahun 2022. Sementara itu, International Diabetes Federation melaporkan bahwa pada tahun 2025 terdapat sekitar 589 juta orang dewasa usia 20–79 tahun yang hidup dengan diabetes, dan jumlah ini diproyeksikan meningkat menjadi 853 juta pada tahun 2050 (IDF, 2025). Peningkatan tersebut menunjukkan bahwa diabetes, khususnya DMT2, masih menjadi masalah kesehatan global yang membutuhkan kajian lebih lanjut, termasuk dari aspek molekuler dan genetik.

Perkembangan teknologi microarray memungkinkan pengukuran ribuan ekspresi gen secara simultan, sehingga menghasilkan data ekspresi gen dalam bentuk matriks berdimensi tinggi. Data tersebut menyimpan informasi biologis yang penting, tetapi memiliki kompleksitas tinggi karena jumlah gen jauh lebih besar dibandingkan jumlah sampel. Oleh sebab itu, diperlukan pendekatan komputasi yang mampu mengidentifikasi pola lokal antara subset gen dan subset kondisi. Salah satu pendekatan yang banyak digunakan dalam analisis ekspresi gen adalah biclustering, yaitu metode pengelompokan simultan terhadap baris dan kolom matriks untuk menemukan submatriks yang memiliki pola ekspresi koheren. Berbeda dengan clustering konvensional yang mengelompokkan gen atau sampel secara terpisah pada seluruh dimensi, biclustering memungkinkan suatu gen berperan dalam beberapa proses biologis yang aktif hanya pada kondisi tertentu (Padilha & Campello, 2017; Castanho et al., 2024). Berbagai penelitian telah menunjukkan bahwa biclustering efektif untuk menganalisis data ekspresi gen, terutama karena mampu mengungkap pola tersembunyi yang tidak mudah ditemukan melalui metode clustering biasa. Padilha dan Campello (2017) menegaskan bahwa pemilihan algoritma biclustering sangat bergantung pada jenis pola biologis yang ingin dideteksi. Selanjutnya, Bustamam et al. (2020) mengembangkan pendekatan three-phase biclustering untuk menemukan korelasi kuat antara gen dan kondisi pada data ekspresi gen Diabetes Melitus dengan memanfaatkan modified lift algorithm berbasis korelasi Pearson. Perkembangan terbaru juga menunjukkan bahwa metode

biclustering semakin diarahkan untuk menangani data biologis berskala besar dan pola yang lebih kompleks. Liu et al. (2023), misalnya, memperkenalkan ARBic untuk mengidentifikasi bicluster signifikan dengan berbagai bentuk pada dataset ekspresi gen berskala besar, sedangkan Castanho et al. (2024) menyajikan taksonomi dan pipeline komprehensif untuk analisis biclustering pada data biologis dan biomedis. Selain itu, López-Fernández et al. (2025) menekankan bahwa pengembangan biclustering dalam bioinformatika semakin terkait dengan tantangan big data dan high-performance computing.

Meskipun demikian, sebagian besar pendekatan biclustering sebelumnya masih memiliki keterbatasan dalam mendeteksi pola korelasi yang kompleks, terutama ketika pola hubungan antara gen dan kondisi tidak hanya bersifat positif, tetapi juga negatif atau kombinasi keduanya. Padahal, korelasi negatif dalam data ekspresi gen juga dapat memiliki makna biologis penting karena dapat menunjukkan hubungan regulasi yang berlawanan antara kelompok gen dan kondisi tertentu. Dalam konteks DMT2 yang berkaitan dengan obesitas, identifikasi pola korelasi positif dan negatif secara simultan menjadi penting untuk memahami kemungkinan hubungan molekuler antara ekspresi gen, kondisi obesitas, dan diabetes. Oleh karena itu, penelitian ini mengusulkan penerapan metode three-phase biclustering untuk menemukan δ-corbicluster yang merepresentasikan pola keterkaitan koheren antara gen dan kondisi pada data ekspresi gen DMT2 dari sampel obesitas dan kurus.

Penelitian ini bertujuan untuk menemukan submatriks dengan pola korelasi yang kuat pada data ekspresi gen DMT2, khususnya pada sampel obesitas dan kurus, serta mengungkap informasi biologis yang berpotensi berkaitan dengan gen-gen yang berpengaruh terhadap diabetes. Hasil penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan metode komputasi untuk analisis data ekspresi gen serta mendukung eksplorasi biomarker atau kandidat gen yang relevan dalam kajian biomedis terkait Diabetes Melitus.

2. Metode Penelitian

Metode penelitian yang digunakan dalam studi ini terdiri dari beberapa tahapan utama untuk mengidentifikasi korelasi bicluster pada data ekspresi gen penyakit Diabetes Melitus dengan menerapkan metode three-phase biclustering. Penelitian ini menggunakan *microarray data* dari ekspresi gen *Diabetes Mellitus* yang diambil dari situs https://www.ncbi.nlm.nih.gov (*raw dataset*).

Tahap pertama dimulai dengan melakukan dekomposisi matriks data ekspresi gen menggunakan metode *Singular Value Decomposition* (SVD), yang bertujuan untuk mentransformasikan data menjadi dua matriks berbasis gen dan kondisi. Pada tahap kedua, dilakukan klasterisasi terhadap kedua matriks hasil SVD dengan menggunakan metode *Partition Around Medoids* (PAM) berbasis jarak Euclidean, sehingga diperoleh beberapa kandidat bicluster yang mewakili kelompok gen dan kondisi yang memiliki kemiripan pola ekspresi.

Tahap ketiga dilakukan dengan modified lift algorithm berbasis korelasi Pearson dan new residue score. Tahap ini bertujuan untuk mengevaluasi dan memperbaiki original bicluster yang telah terbentuk pada tahap sebelumnya. Proses evaluasi dilakukan melalui dua prosedur utama, yaitu single-node deletion dan single-node addition. Pada tahap single-node deletion, baris atau kolom yang menyebabkan nilai residue bicluster melebihi ambang batas δ dihapus secara bertahap. Sebaliknya, pada tahap single-node addition, baris atau kolom di luar bicluster diuji kembali dan dapat ditambahkan apabila mampu menurunkan nilai residue. Bicluster yang memenuhi kriteria nilai residue lebih kecil atau sama dengan ambang batas δ serta memenuhi ukuran minimum gen dan kondisi dinyatakan sebagai δ-corbicluster.

Evaluasi hasil dilakukan dengan meninjau jumlah bicluster awal, jumlah δ-corbicluster yang terbentuk, nilai residue masing-masing δ-corbicluster, serta pola keterkaitan antara gen dan kondisi. Nilai δ yang lebih kecil menunjukkan tingkat koherensi pola yang lebih baik, sehingga bicluster dengan nilai residue rendah dianggap memiliki pola hubungan gen-kondisi yang lebih stabil. Hasil akhir dari penelitian ini berupa δ-corbicluster yang merepresentasikan kelompok gen dan kondisi dengan pola korelasi kuat, baik positif maupun negatif, sehingga dapat digunakan sebagai dasar untuk eksplorasi informasi biologis terkait Diabetes Melitus tipe 2.

3. Hasil dan Pembahasan

3.1 Data Sintetik dengan Dua Embedded Overlapping Biclusters

Sintetik data adalah data yang dibuat untuk mengukur seberapa bagus implementasi algoritma yang dibuat. Sintetik data yang digunakan adalah matriks berukuran 8×7 dengan memasukan dua overlapping bicluster $corB_1 = \{r_1, r_2, r_3, r_4\} \times \{c_1, c_2, c_3, c_4, c_5\}$ dan $corB_2 = \{r_4, r_5, r_6, r_7\} \times \{c_4, c_5, c_6, c_7\}$ dengan

background angka acak U(0,2). Bicluster yang pertama memilki nilai korelasi yang mendekati nol sedangkan yang kedua merupakan korelasi bicluster yang sempurna.

Algoritma three phase biclustering diimplementasikan pada matriks ini dengan parameter $\delta = 0.1$, nr = 3 dan nc = 3. Data ini ditampilkan dalam bentuk matriks seperti berikut:

	c1	c2	c3	c4	c5	c6	c7
r1	1	10	21	6	17	1.36	0.2
r2	1.56	20	41	11	32	0.79	0.52
r3	1.43	14	29	8	23	0.73	0.67
r4	1.81	24	49	13	38	22	32
r5	1.78	0.52	0.59	21	46	30	40
r6	0.67	0.82	0.64	15	40	24	34
r7	1.4	1.19	0.85	27	52	36	46
r8	0.4	0.52	1.02	1.16	1.25	1.59	1.31

Pertama – tama tahapan *clustering* dengan menggunakan PAM-IDB pada matriks $R^{(l)}$ dan $C^{((l)}$ yang mana dihasilkan dari SVD, menghasilkan empat *clusters* $\{r_1, r_2, r_3\}, \{r_4\}, \{r_5, r_6, r_7\}$ dan $\{r_8\}$ pada $R^{(l)}$ dan tiga *cluster* $\{c_1, c_2, c_4\}, \{c_3\}$ dan $\{c_5, c_6, c_7\}$ pada $C^{(l)}$. *Node* r_4, r_8 dan c_3 dihapus karena ketika digabung hanya berbentuk vektor. Sehingga *original bicluster* yang didapat:

$$B_1 = \{r_1, r_2, r_3\} \times \{c_1, c_2, c_4\}$$

$$B_2 = \{r_1, r_2, r_3\} \times \{c_5, c_6, c_7\}$$

$$B_3 = \{r_5, r_6, r_7\} \times \{c_1, c_2, c_4\}$$

$$B_3 = \{r_5, r_6, r_7\} \times \{c_5, c_6, c_7\}$$

Selanjutnya keempat *bicluster* ini dilakukan proses *lift algorithm* pada aplikasi R. Nilai dari masing – masing residue terdapat pada Tabel 2.

Tabel 1. Nilai Residu dari *Original Bicluster*

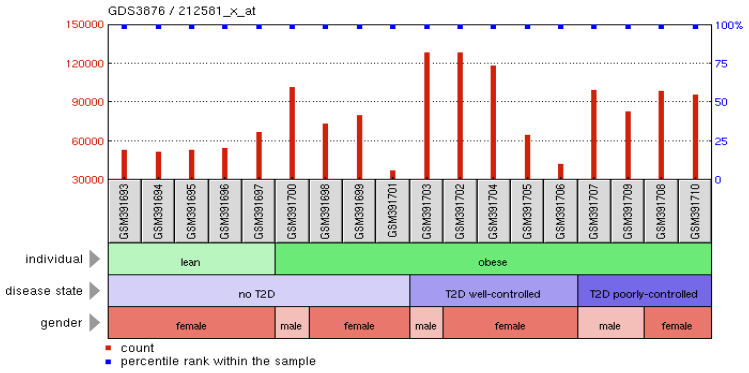
Bicluster	Nilai Residue
1	0.08
2	0.34
3	0.58
4	0

Pada *bicluster* pertama karena memiliki nilai residu 0.08 yang mana kurang dari *threshold* δ maka langsung dilakukan fase kedua *single-node addition* dengan

menambahkan *node* r_4 dan c_3 sehingga menghasilkan δ-*corbicluster* $B_1 = \{r_1, r_2, r_3, r_4\} \times \{c_1, c_2, c_3, c_4\}$. Untuk *bicluster* kedua proses algoritma menghasilkan status 'kurang dari batas parameter nr dan nc'. Untuk *bicluster* ketiga nilai residu lebih dari delta sehingga dilakukan fase *single-node deletion* dan menghasilkan status 'tidak bisa di*lift*'. Sedangkan untuk *bicluster* keempat menghasilkan δ-*corbicluster* $B_4 = \{r_4, r_5, r_6, r_7\} \times \{c_4, c_5, c_6, c_7\}$ dengan menambahkan *node* r_4 dan c_4. Hasil evaluasi *bicluster* pada algoritma ini meng-*cover* semua *embedded bicluster* pada data sintetik yang digunakan yang mana menunjukan *robustness* dari algoritma *three phase biclustering* yang digunakan pada penelitian ini.

3.2 Microarray Data Ekspresi Gen Diabetes Melitus Type 2 pada Organ Hati

Dataset ini terdiri dari 22283 baris gen, 18 sampel yang terdiri dari 5 wanita kurus yang tidak terkena penyakit DM, 1 laki – laki gemuk tidak terkena penyakit DM, 3 wanita gemuk tidak terkena penyakit DM, 1 laki – laki gemuk terkontrol terkena penyakit DM, 4 wanita gemuk terkontrol terkena penyakit DM, dan 2 wanita dan 2 laki – laki gemuk kurang terkontrol yang terkena penyakit.



Gambar 1. Kondisi Sampel Penyakit *Diabetes Melitus*
Sumber:https://www.ncbi.nlm.nih.gov/geo/tools/profileGraph.cgi?ID=GDS3876:212581_x_at

3.2.1 Seleksi Gen pada Banyak *Cluster*

Tabel 2. Hasil Pemilihan *Threshold* pada Banyak Gen Terseleksi dan Banyak *Cluster*

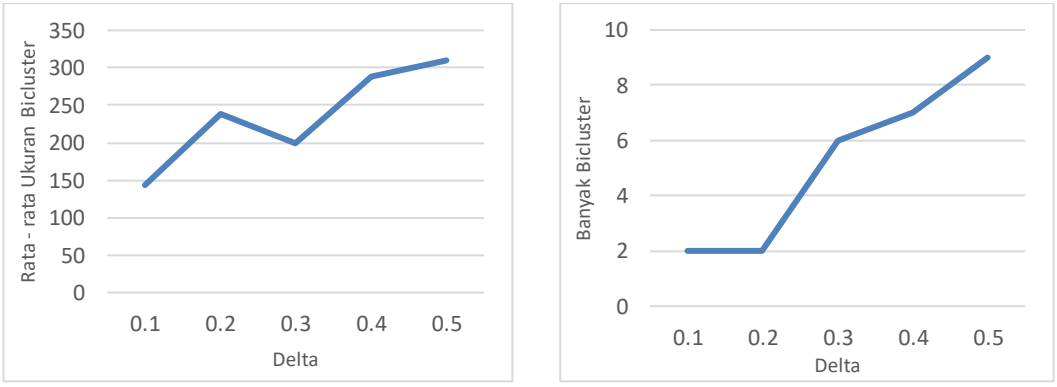
Threshold	Banyak Gen Terseleksi	R		C		Banyak *Bicluster*
		IDB	*Cluster*	IDB	*Cluster*	
200000/150	41	0,194	2	0,337	5	10
120000/130	99	0,2905	2	0,419	6	12
90000/110	173	0,3506	2	0,3199	5	10
60000/70	527	0,3686	2	0,4708	2	4
20000/50	1331	0,4673	2	0,2575	4	8
10000/30	3126	0,5321	2	0,2808	3	6

5000/20	5968	0,6163	2	0,1416	5	10

Pada Tabel 2. dapat dilihat bahwa semakin kecil pemilihan *threshold* maka gen yang terseleksi semakin besar. Banyak gen yang terseleksi ini juga berpengaruh pada banyak *cluster* sampel, namun tidak berpengaruh pada banyak *cluster* gen yang dihasilkan (2 *cluster*). Oleh karena itu, untuk analisa lebih lanjut dipilih *threshold* 20000/50 mengingat nilai IDB yang dihasilkan relatif lebih kecil dibandingkan dengan *threshold* yang lain.

3.2.2. Analisis Dataset DM dengan 99 Gen Terseleksi

Tujuan dari pengambilan 99 gen terseleksi dari dataset DM ini adalah untuk mengetahui pengaruh parameter δ (*threshold* nilai korelasi) pada banyak *bicluster* dan rata – rata ukuran *bicluster*. Algoritma *three-phase biclustering* diimplementasikan pada 99 gen terseleksi ini dengan mengambil perameter *l* terbaik dan memvariasikan parameter δ.



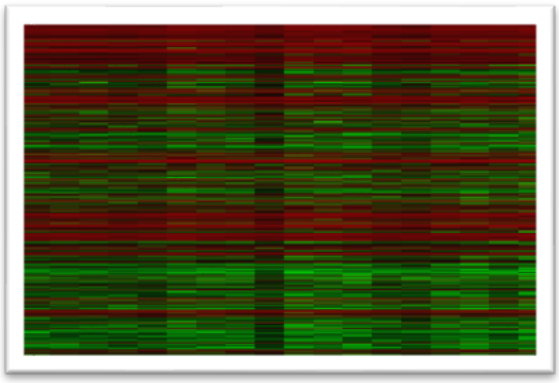
(a) (b)

Gambar 2. (a) Grafik Perbandingan Rata – Rata Ukuran *Bicluster* vs Delta (b) Grafik Perbandingan Banyak *Bicluster* vs Delta

3.2.3 Analisis Dataset DM dengan 1331 Gen Terseleksi

Sebelum masuk ke dalam proses algoritma pertama kali data diseleksi dengan menggunakan *absolute* dan *relative deviation* dengan masing – masing *threshold* 20.000 dan 50, menyisakan 1331 gen dan 18 sampel. Setelah itu gen yang terseleksi dinormalisasikan dengan *Log – Mean Centering* menghasilkan angka diantara -7,73 sampai 5.22 yang data ini sebagai gambaran ditampilkan dalam bentuk *heatmap* pada gambar 3.

DELTA: **Jurnal Ilmiah Pendidikan Matematika**
Vol. 14 No. 1 Bulan Januari hal. 119 – 132



Gambar 3. *Heatmap* Hasil Normalisasi

Hasil Parameter Reduksi Dimensi Berdasarkan Nilai IDB

Setelah data dinormalisasi dengan menggunakan normalisasi *log – mean centering,* data normalisasi ini melalui fase SVD dan PAM membentuk *original bicluster* $(\hat{R}_i^{(l)}, \hat{C}_j^{(l)})$ dengan $i = 1,2,3,\dots,m$ dan $j = 1,2,3,\dots,n$ sebanyak $m \times n$ *bicluster.* Parameter l menunjukan reduksi dimensi kolom matriks $\hat{R}_i^{(l)}$ dan $\hat{C}_j^{(l)}$ adalah nilai IDB di setiap rasio parameter l di setiap *cluster* $\hat{R}_i$ dan $\hat{C}_j$ pada dataset penyakit *diabetes militus*.

$list_parameter_R

	10%	20%	30%	40%	50%	60%	70%	80%	90%	Full Dim
2	0.4673035	0.6033778	0.7372878	0.7696339	0.8002185	1.002196	1.169751	1.230161	1.136357	1.136357
3	0.7479437	1.0899353	1.5458799	1.5104444	1.5827453	1.775309	1.803696	1.750485	1.850228	1.850228
4	0.8177596	1.5068612	1.7853417	1.9472021	2.0353072	2.055658	1.825315	1.819937	1.978164	1.978164
5	0.7617159	1.3993459	1.6537341	2.0663166	1.9861734	2.188257	1.853678	1.916464	2.018162	2.018162
6	0.7745830	1.3962316	1.8174668	2.0472834	2.1110918	2.124234	1.864496	1.963160	2.209949	2.209949
7	0.7497022	1.4342297	1.8802796	1.9940435	2.0599636	2.117181	2.120519	2.010709	2.253733	2.253733
8	0.7493308	1.3305204	1.8488580	1.9372872	2.0025175	2.097279	2.157701	1.915076	2.497650	2.497650
9	0.7116267	1.3084125	1.7277609	1.9560932	2.0422688	2.148247	2.125805	2.136038	2.370256	2.370256
10	0.6651511	1.3022146	1.5530102	1.8717656	1.9950336	2.052643	2.018037	2.278642	2.306640	2.306640
11	0.6240832	1.3573884	1.6068703	1.9154295	1.9844788	1.942746	2.086025	2.142321	2.219789	2.219789
12	0.6460973	1.4640123	1.6184069	1.8295709	1.8125617	2.079932	2.141031	2.051078	2.127416	2.127416
13	0.6358904	1.3702529	1.5881530	1.7426936	1.8152878	2.072133	2.144625	2.033408	2.243780	2.243780
14	0.6373888	1.3367527	1.7114095	1.7916157	1.8673581	2.181972	2.163343	2.122410	2.192234	2.192234
15	0.6375262	1.2829587	1.6387875	1.7967215	1.7427682	2.098649	2.203608	2.116515	2.364250	2.364250
16	0.6040148	1.2853184	1.6798340	1.8188755	2.0847439	2.152962	2.300137	2.263439	2.333382	2.333382
17	0.5978978	1.2445664	1.6680418	1.8994938	2.1262219	2.119203	2.415787	2.249293	2.296910	2.296910
18	0.6070530	1.2304874	1.6024346	1.9611624	1.9959118	2.132786	2.370271	2.235865	2.317312	2.317312
19	0.6112460	1.3013809	1.5924079	1.9109834	1.9574521	2.217392	2.501360	2.197427	2.283164	2.283164
20	0.6064549	1.2658697	1.7532180	1.8852532	1.9525443	2.188277	2.400865	2.177689	2.302830	2.302830
21	0.6121542	1.2769174	1.6509809	1.9177369	1.9352823	2.127205	2.480836	2.143355	2.294690	2.294690
22	0.6122174	1.2998870	1.6572041	1.8883098	1.9807606	2.127664	2.470582	2.256267	2.328616	2.328616
23	0.6110331	1.3387934	1.6509044	1.8679610	1.9865898	2.152872	2.395759	2.292171	2.270871	2.270871
24	0.6111189	1.3155547	1.6433414	1.8522799	1.9852822	2.149092	2.238151	2.258905	2.248416	2.248416
25	0.6056298	1.3046141	1.6105998	1.9473900	1.9457450	2.205184	2.354713	2.305778	2.390188	2.390188
26	0.6216497	1.2865822	1.7158519	1.9252302	2.0096467	2.142818	2.327298	2.288907	2.418138	2.418138
27	0.6174345	1.2882714	1.6339685	1.9105021	1.9197760	2.181545	2.356823	2.213676	2.298869	2.298869
28	0.6216263	1.3161752	1.6409360	1.9084733	1.9227409	2.148999	2.262230	2.165866	2.312383	2.312383
29	0.6272710	1.3189493	1.6062467	1.8854034	1.9340647	2.136482	2.375940	2.169141	2.279601	2.279601
30	0.6305770	1.3458874	1.6846178	1.8151987	1.9483909	2.113334	2.305379	2.196693	2.252673	2.252673

$list_parameter_C

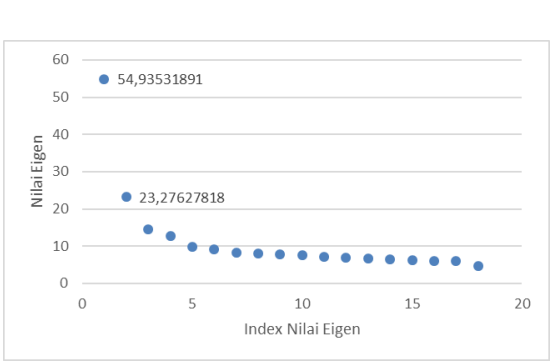
	10%	20%	30%	40%	50%	60%	70%	80%	90%	Full Dim
2	0.4789922	1.1396151	1.8251895	2.2390296	1.9940768	2.252233	0.7027167	0.7515782	2.628491	2.628491
3	0.7877645	0.7725286	1.2431199	1.5846847	1.4479825	1.656489	0.7286899	0.7758337	1.985251	1.985251
4	0.2575505	0.6285064	1.0171268	1.2779452	1.1641563	1.356770	0.7171518	0.7657074	1.660655	1.660655
5	0.3686982	0.5024773	0.7181184	0.8743130	0.9788260	1.240123	1.1785912	0.7531278	1.483208	1.483208
6	0.3024442	0.3940958	0.5926589	0.7565049	0.8289913	1.074949	1.0718172	0.7362386	1.360103	1.360103

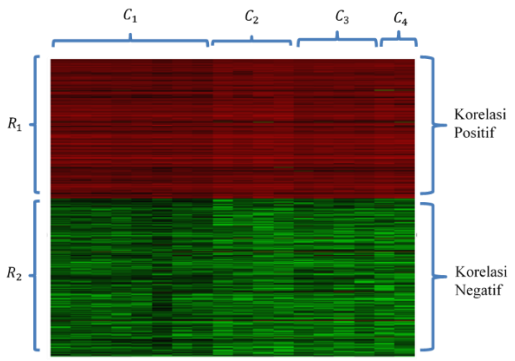
Gambar 4. Nilai IDB Di Setiap Rasio Parameter l Dan *Cluster*

Nilai minimum IDB pada Gambar 4. terletak pada *cluster* ke – 2 dengan rasio reduksi dimensi l 10% dengan nilai 0,467 pada matriks $\hat{R}$ sedangkan untuk $\hat{C}$ terletak pada *cluster* ke – 4 dan 10% dari kolom dengan nilai IDB 0,257 dengan mengambil nilai maximun 30 cluster dan 6 cluster untuk masing – masing $\hat{R}$ dan $\hat{C}$ dan *relative significant* dari nilai eigen mencapai 0.89. Gambar 5(a). menunjukan grafik

penyebaran nilai eigen yang dihasilkan dari SVD. Selanjutnya cluster terbaik ini akan dipilih untuk membentuk sebanyak $2 \times 4 = 8$ *original bicluster* yang masing – masing *bicluster* nya dievaluasi dengan modifikasi *lift algorithm.* Gambar 5(b) menunjukan *heatmap* dari hasil penggabungan *bicluster* ini. Hasil *Heatmap* Dari Pembentukan *Original Biclusters*.

Heatmap pada Gambar 5(b) menunjukkan ekspresi gen dari 8 *original bicluster* yang terkelompok berdasarkan gen dan kondisi. Hasilnya mengungkap korelasi positif pada 624 gen pertama (berwarna merah), yang memiliki aktivitas lebih tinggi dibandingkan 707 gen lainnya di semua sampel. Selain itu, *heatmap* juga memperlihatkan bahwa pengelompokan gen dan kondisi masih dalam korelasi positif atau negatif tertentu, sesuai hasil residu tiap *bicluster*. Oleh karena itu, untuk memperkuat analisis korelasi (positif dan negatif) pada gen dan sampel, diperlukan tahap ketiga, yaitu modifikasi *lift algorithm.*





5(a) Grafik Penyebaran Nilai Eigen 5(b) Hasil *Heatmap* Dari Pembentukan *Original Biclusters*

Hasil Evaluasi *Original Bicluster* Penyakit DM

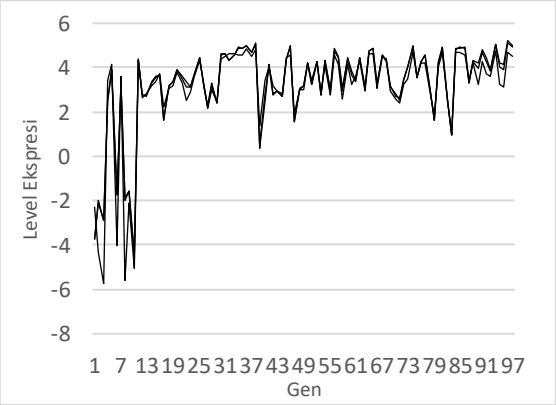
Sebanyak 8 *bicluster* didapat dari tahapan kedua dievaluasi secara sekuensial dengan menggunakan program R dengan menetapkan nilai *threshold residue* $\delta = 0.1$ dan minimum banyak baris nr = 5 dan minimum banyak kolom nc = 3. Alasan penetapan nilai *threshold* $\delta = 0.1$ adalah untuk mendapatkan hasil korelasi yang bagus mengingat semakin nilai residu mendekati nol maka korelasi akan semakin bagus. Hasil lengkapnya dapat dilihat pada Tabel 5.

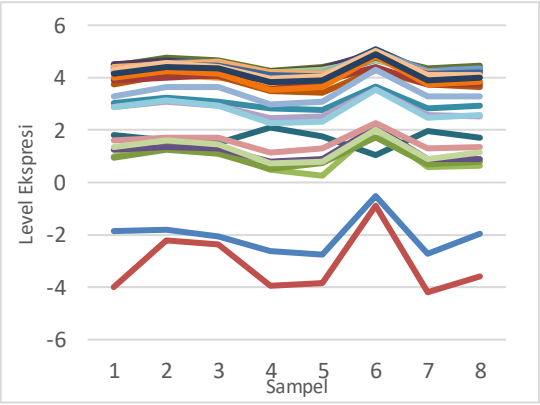
DELTA: **Jurnal Ilmiah Pendidikan Matematika**
Vol. 14 No. 1 Bulan Januari hal. 119 – 132

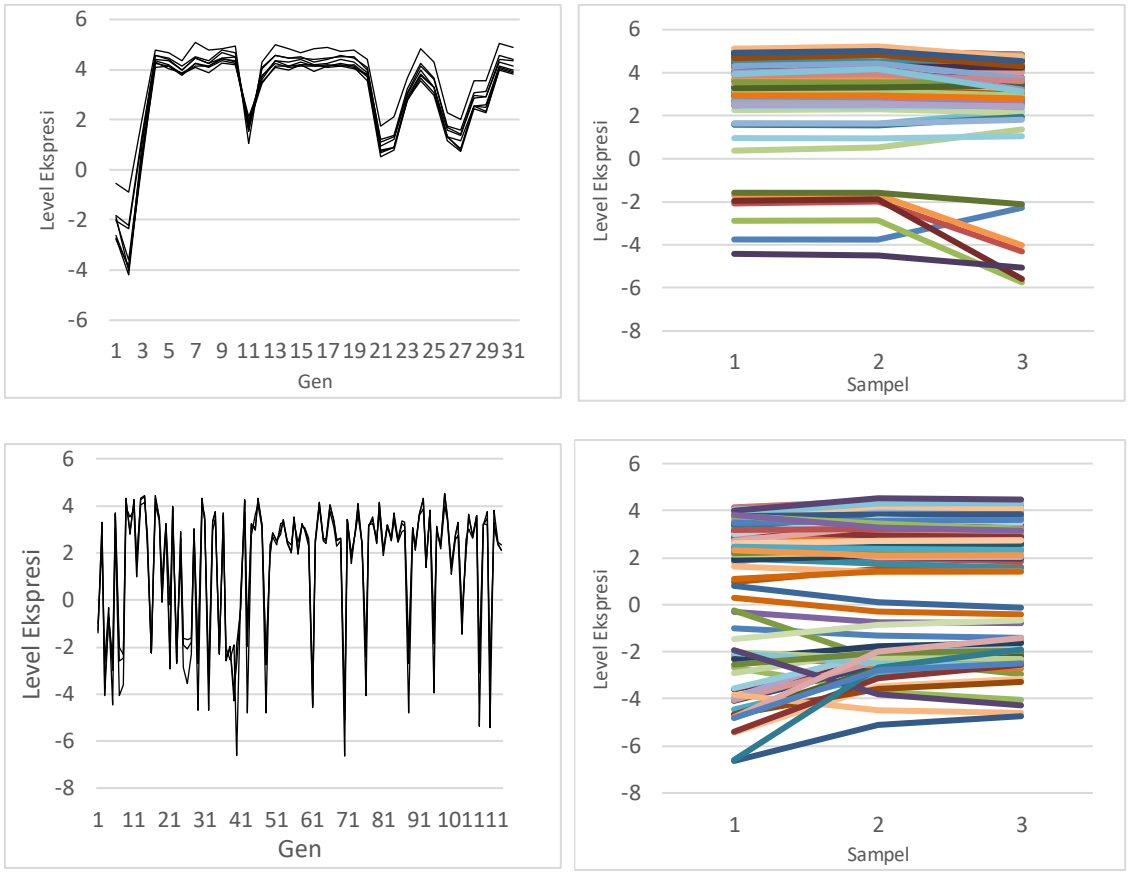
Tabel 5. Hasil Evaluasi Modifikasi *Lift Algorithm* Pada Dataset DM

Bicluster	Status	Residu Sebelum Dievaluasi	Residu Setelah Dievaluasi	Dimensi Sebelum Dievaluasi	Dimensi Setelah Dievaluasi
1	Berhasil	0.778	0.097	624×8	31×8
2	Berhasil	0.638	0.095	624×4	97×3
3	Berhasil	0.558	0.085	624×4	114×3
4	Tidak bisa di*lift*	0.113	-	624×2	624×2
5	Tidak bisa di*lift*	0.737	-	707×8	85×2
6	Tidak bisa di*lift*	0.529	-	707×4	680×2
7	Kurang dari batas parameter nr dan nc	0.57	-	707×4	98×2
8	Tidak bisa di*lift*	0.451	-	707×2	707×2

Untuk lebih jelas tentang pola korelasi ini dapat dilihat pada Gambar 6. yang mana dapat disimpulkan bahwa pada dataset dengan 1331 gen terseleksi ini efek dari korelasi negatif terlihat signifikan secara visual pada gen tetapi tidak tampak terlihat pada sampel. Kesimpulan ini tidak bertentangan dengan hasil kesimpulan penelitian yang dilakukan oleh Yang dengan menggunakan algoritma yang sama (*Lift Algorithm)* pada dataset organ manusia. Selain itu pada δ-corbicluster terbaik (ketiga), GO (Gen Ontology) Function pada 114 gen, diantaranya 70 gen menunjukan fungsi yang sama yaitu *protein binding* sedangkan 12 gen tanpa keterangan. Untuk keterangan lengkap tentang GO function.







Gambar 6. Grafik Level Ekspresi pada Tiga δ – *CorBicluster*; Sisi Kiri Grafik Level Ekspresi Sampel pada Gen, Sisi Kanan Gen pada Sampel

Analisis Sampel Hasil Evaluasi *Bicluster* Penyakit DM

Dataset DM terdiri dari 16 sampel dengan kriteria seperti dijelaskan pada subbab 3.2. Pada *bicluster* pertama, korelasi positif dominan terlihat pada 8 sampel dalam *δ-corbicluster*, terdiri dari 5 wanita kurus tanpa DM, 2 wanita gemuk tanpa DM, dan 1 wanita gemuk dengan DM yang rutin kontrol. Sebanyak 31 gen memengaruhi 7 sampel non-DM, mengindikasikan potensi kerentanan terhadap DM. Di *bicluster* kedua, 97 gen memengaruhi 2 dari 3 sampel obesitas dengan DM, menunjukkan kecenderungan serupa pada sampel kontrol. Pada *bicluster* ketiga, 113 gen memengaruhi 2 sampel obesitas dengan DM, memperkuat dugaan kerentanan DM pada kelompok obesitas. Hasil ketiga *bicluster* ini menyimpulkan bahwa obesitas merupakan faktor risiko DM (lihat Tabel 6).

Tabel 6. Penjelasan Sampel dan Kesimpulan pada Evaluasi *Bicluster*

Bicluster	Kode Sampel	Penjelasan Sampel	Kesimpulan
1	GSM391693	Wanita, Kurus, Tidak Berpenyakit	Wanita gemuk berpenyakit dengan kontrol yang baik berpotensi sembuh
	GSM391694	Wanita, Kurus, Tidak Berpenyakit	
	GSM391695	Wanita, Kurus, Tidak Berpenyakit	
	GSM391696	Wanita, Kurus, Tidak Berpenyakit	
	GSM391697	Wanita, Kurus, Tidak Berpenyakit	
	GSM391701	Wanita, Gemuk, Tidak Berpenyakit	
	GSM391705	Wanita, Gemuk, Berpenyakit (Kontrol Baik)	
2	GSM391700	Pria, Gemuk, Tidak Berpenyakit	Pria gemuk berpotensi terkena penyakit
	GSM391702	Wanita, Gemuk, Berpenyakit	
	GSM391710	Wanita, Gemuk, Berpenyakit	
3	GSM391699	Wanita, Gemuk, Tidak Berpenyakit	Wanita gemuk sangat berpotensi terkena penyakit
	GSM391707	Pria, Gemuk, Berpenyakit (Kontrol Jelek)	
	GSM391708	Wanita, Gemuk, Berpenyakit (Kontrol Jelek)	

4. Kesimpulan dan Saran

Berdasarkan hasil penelitian, implementasi metode three-phase biclustering pada data ekspresi gen penyakit Diabetes Melitus berhasil mengidentifikasi korelasi yang kuat antar gen dan kondisi pada sampel obesitas dan kurus. Metode ini secara efektif membentuk delapan bicluster utama dari 1331 gen terseleksi, yang kemudian dievaluasi lebih lanjut menggunakan modifikasi algoritma lift berbasis korelasi Pearson. Dari delapan bicluster tersebut, tiga δ-corbicluster dengan nilai korelasi tinggi (0,097; 0,095; dan 0,085) berhasil ditemukan, menunjukkan adanya kelompok gen dan sampel yang memiliki keterkaitan erat dalam konteks patofisiologi Diabetes Melitus. Temuan ini membuktikan bahwa pendekatan three-phase biclustering mampu memberikan hasil yang signifikan dalam mengungkap pola korelasi tersembunyi pada data ekspresi gen, serta berpotensi besar untuk diaplikasikan dalam penelitian biomedis lanjutan, khususnya dalam pengembangan diagnosis dan terapi berbasis genetik untuk Diabetes Melitus.

Sebagai saran, penelitian selanjutnya dapat mengembangkan metode ini dengan memperluas jumlah data dan variasi sampel, serta mengintegrasikan data ekspresi gen dengan data klinis pasien untuk memperoleh pemahaman yang lebih komprehensif mengenai mekanisme molekuler Diabetes Melitus. Selain itu, penggunaan algoritma biclustering lain atau penggabungan dengan metode machine learning dapat dieksplorasi guna meningkatkan akurasi dan interpretasi hasil bicluster yang diperoleh.

Pustaka

Castanho, E. N., et al. (2024). *Biclustering data analysis: A comprehensive survey*. Briefings in Bioinformatics, 25(4), bbae342.

Diabetes Melitus Data: Obese patients with and without type 2 diabetes: liver. (2009). https://www.ncbi.nlm.nih.gov/sites/GDSbrowser

Walston, J. Silver, K. (2010). *Insulin Response to Glucose Is Lower in Individuals Homozygous for the Arg 64 Variant of the ß-3-Adrenergic Receptor*. Johns Hopkins University School of Medicine. https://www.news-medical.net/health/Obesity-and-Diabetes.aspx

International Diabetes Federation. (2025). *IDF Diabetes Atlas*.

Yang, W.-H. Dai, D.-Q. Yan, H. (2011). *Finding Correlated Biclusters from Gene Expression Data*. IEEE Trans. Knowl. Data Eng. 23 568–584.

Yang, W.-H. Dai, D.-Q. Yan, H. (2007). *Biclustering of microarray data based on singular value decomposition*. Conference: Emerging Technologies in Knowledge Discovery and Data Mining. PAKDD, International Workshops, Nanjing, China, Revised Selected Papers.

Zhao, Y.H., Yu, J.X, Wang, G.R., Chen, L., Wang, B., and Yu, Gu. (2008). *Maximal Subspace Coregulated Gene Clustering*. IEEE Trans. Knowledge and Data Eng. vol. 20. no. 1, pp. 83-98.

Liu, X., et al. (2023). *ARBic: An all-round biclusterin*

López-Fernández, A., et al. (2025). *Biclustering in bioinformatics using big data and High Performance Computing applications: Challenges and perspectives, a review*. The Journal of Supercomputing.

Padilha, V. A., & Campello, R. J. G. B. (2017). *A systematic comparative evaluation of biclustering techniques*. BMC Bioinformatics, 18, 55.

World Health Organization. (2024). *Diabetes*.

DELTA: **Jurnal Ilmiah Pendidikan Matematika**
Vol. 14 No. 1 Bulan Januari hal. 119 – 132