



Clustering Microarray Adenoma Menggunakan Spectral Clustering dengan Algoritma Partitioning Around Medoid (PAM)

Septian Wulandari

Fakultas Teknik dan Ilmu Komputer, Program Studi Informatika, Universitas Indraprasta PGRI

E-mail: septian.pmb09@rocketmail.com

Info Artikel

Kata kunci:

Adenoma, *Spectral Clustering*,
Partitioning Around Medoid (PAM)

Abstrak

Adenoma merupakan tumor jinak yang ditemukan di dalam sel-sel mirip kelenjar di dalam jaringan epitel. Adenoma dapat dianalisis dengan mengamati tingkat ekspresi gen yang dimiliki. Menganalisis ekspresi gen dapat dilakukan dengan melakukan pengelompokan atau *clustering* yang membutuhkan waktu cukup lama karena menggunakan dimensi yang besar. Oleh karena itu, salah satu teknik dalam *clustering* yang dapat digunakan adalah mengelompokkan *microarray* menjadi kelompok-kelompok yang memiliki kesamaan sifat dan fungsi. Metode *clustering* yang digunakan dalam penelitian ini adalah *spectral clustering* dengan algoritma *Partitioning Around Medoid (PAM)*. Metode *spectral clustering* merupakan metode yang memperlakukan setiap titik data sebagai simpul grafik. Setelah dimensi data dikurangi, kemudian data akan dipartisi menggunakan algoritma PAM. Algoritma ini meminimalkan fungsi objektif dengan menukar semua titik non-medoid menjadi titik medoid secara iteratif hingga bertemu. Tujuan dari penelitian ini adalah mengelompokkan *microarray* adenoma menggunakan *spectral clustering* dengan algoritma PAM. Data yang digunakan pada penelitian ini adalah 7.087 *microarray* adenoma berdasarkan nilai kesamaan. Hasil yang diperoleh adalah terdapat tiga *cluster* adenoma yaitu 4.349 *microarray* adenoma pada *cluster* ke-1, 2.115 *microarray* adenoma pada *cluster* ke- 2, dan 623 *microarray* adenoma pada *cluster* ke- 3 dengan nilai *average silhouette* 0,5289522.

How to Cite: Wulandari, S. (2020). *Clustering Microarray Adenoma Menggunakan Spectral Clustering dengan Algoritma Partitioning Around Medoid (PAM)*. *Prosiding Seminar Nasional Sains 2020*, 1 (1): 345-351.

PENDAHULUAN

Adenoma merupakan tumor atau neoplasma jinak yang timbul di dalam jaringan epitel. Namun, dalam beberapa kasus adenoma dapat berkembang menjadi tumor ganas. Terdapat beberapa macam adenoma yang dinamakan sesuai dengan tempat tumbuhnya adenoma diantaranya adenoma basofilik, adenoma kromofob, adenoma sebaceous, dan adenoma kolorektal. Menurut Arifah, pada tahun 2012 terdapat 2800 kasus tumor sistem saraf pusat (SSP) yang di diagnosa di Kanada, dan 1850 kematian dari tumor SSP yang akan terjadi dalam periode yang sama (Hidayat, 2015). Ekspresi gen mempunyai banyak sekali manfaat. Salah satunya adalah untuk mengetahui perkembangan penyakit di dalam tubuh manusia. Berdasarkan kasus tersebut, maka perlu untuk menganalisis atau mengamati adenoma melalui tingkat ekspresi gen yang dimiliki. Ekspresi gen atau DNA (*deoxyribonucleic acid*) *microarray* memanifestasi gen fenotip dari proses transkripsi dan translasi genetik dalam adenoma. Menganalisis ekspresi gen dapat dilakukan dengan melakukan pengelompokan atau *clustering*.

Clustering data merupakan salah satu metode dalam data mining yang dapat digunakan untuk memetakan data kedalam kelompok-kelompok yang lebih kecil berdasarkan kesamaan karakteristik yang dimilikinya (Taslim & Fajrizal, 2016). Proses *clustering* dalam menganalisis

ekspresi gen membutuhkan waktu yang cukup lama, karena data yang digunakan menggunakan dimensi yang cukup besar. Oleh karena itu, salah satu teknik dalam *clustering* yang dapat digunakan adalah menggolongkan DNA *microarray* menjadi kelompok-kelompok yang memiliki kesamaan sifat dan fungsi. *Spectral clustering* merupakan salah satu metode yang menjanjikan untuk mengidentifikasi ekspresi gen pada tipe dataset gen tumor (Cahyaningrum, Bustamam, & Siswantining, 2017). Metode ini mampu mempartisi data dengan struktur yang lebih rumit dibandingkan dengan metode *cluster* tradisional seperti *k-means*, *fuzzy c-means*, SOM, dan lain sebagainya. Disamping itu, metode ini menggunakan informasi yang diperoleh dari nilai eigen dan vektor eigen dari kedekatan matriks untuk mempartisi grafik (Sharma, Kumar, & Mansotra, 2016).

Secara umum, metode *clustering* dibagi menjadi dua kelompok yaitu metode hirarki *clustering* dan metode partisi *clustering*. Partitioning Around Medoid (PAM) merupakan salah satu metode atau algoritma partisi *clustering* yang banyak digunakan untuk menggolongkan data menjadi bagian-bagian kecil. Algoritma ini memiliki kelebihan yaitu hasil proses *clustering* tidak bergantung pada urutan masuk dataset sehingga mudah untuk diaplikasikan atau diimplementasikan pada dataset apapun (Rizby, Marji, & Muflikhah, 2018). Disamping itu, algoritma ini mengatasi *sensitive* terhadap *noise* dan *outlier*, dimana objek dengan nilai yang besar yang memungkinkan menyimpang pada dari distribusi data. Umumnya, dalam metode *spectral clustering*, algoritma partisi yang digunakan adalah algoritma *k-means*. Namun, penelitian yang dilakukan oleh Hengki Muradi algoritma PAM juga dapat digunakan dalam metode HOPACH *clustering* dalam menggolongkan virus ebola (Muradi, Bustamam, & Lestari, 2016). Berdasarkan penelitian yang dilakukan oleh Aruna Bhat terlihat bahwa PAM atau K-Medoid lebih baik dibandingkan algoritma *clustering k-means*. Algoritma PAM lebih efektif dalam hal akurasi dan waktu dengan set data berukuran sedang dan berkinerja lebih baik daripada teknik pengelompokan *k-means* dalam hal *noise* dan *outlier* (Bhat, 2014). Oleh karena itu, dalam penelitian ini digunakan metode *spectral clustering* dengan menggunakan algoritma PAM.

METODE PENELITIAN

Pada penelitian ini data yang digunakan adalah 7.087 *microarray* adenoma yang diambil berdasarkan nilai kesamaan. *Microarray* adenoma diperoleh melalui laman <http://genomics-pubs.princeton.edu/oncology/> dengan data berupa format .xls. Matriks baris pada data terdiri dari objek data dan matriks kolom pada data terdiri dari kondisi eksperimen yaitu positif adenoma ataupun normal. Gambar 1. merupakan contoh 16 data *microarray* adenoma dan normal pada manusia dengan baris merepresentasikan gen atau DNA yang akan di *clustering*, sedangkan pada kolom merepresentasikan kondisi eksperimental.

Accession Number	Description	Adenoma 1	Adenoma 2	Adenoma 3	Adenoma 4	Normal (Ad) 1	Normal (Ad) 2	Normal (Ad) 3	Normal (Ad) 4
A23102	"Human GABA _A receptor alpha-3 subunit. - Also Similar To: AF002255, L41347, AF002256, L79661"	29.84	46.64	53.59	31.87	44.88	45.30	54.31	21.80
A2000114	"Human mRNA for "osteomodulin," complete cds"	0.93	2.17	11.03	-0.85	13.60	17.38	10.46	-2.42
A2000115	"Human mRNA, complete cds"	3.26	-1.83	1.58	5.52	7.77	8.95	6.48	1.94
A2000220	"Human mRNA for semaphorin "E," complete cds"	10.72	20.61	14.71	-5.52	8.74	18.96	28.90	31.49
A2000031	"Human DNA for GPI-anchored molecule-like "protein," complete cds - Also Represents: D84290"	3.26	4.34	8.93	4.25	11.65	4.21	-1.49	-2.91
A2000409	"Human mRNA for "MNK1," complete cds"	-0.93	-9.76	-7.36	-4.67	-3.40	3.16	7.47	-6.30
A2000410	"Human HOGG1 mRNA, complete cds - Also Represents: U96710, Y11838, Y11731"	1.40	-15.73	-2.10	0.00	-2.43	-3.69	2.49	1.45
A2000449	"Human mRNA for "VRK1," complete cds"	9.79	9.22	2.63	-3.40	3.88	-6.85	-3.49	4.36
A2000450	"Human mRNA for "VRK2," complete cds"	89.52	17.35	15.76	33.15	20.40	14.75	34.88	20.35
A2000460	"Human mRNA, clone "RES4-22B," complete cds"	94.65	126.90	142.90	78.62	138.40	132.21	151.48	134.69
A2000462	"Human mRNA for SH3 binding "protein," clone "RES4-23A," complete cds"	-0.93	7.05	8.41	0.42	4.37	-2.11	-4.48	1.45
A2000464	"Human mRNA, clone "RES4-24A," exon "1," "2," "3," "4"	48.95	66.71	73.03	66.30	69.44	93.23	48.33	36.82
A2000466	"Human mRNA, clone "RES4-24C," exon "1," "2," "3"	-55.02	-45.55	-74.60	-62.05	-65.07	-60.57	-62.29	-52.81
A2000467	"Human mRNA, clone "RES4-25," partial cds"	-54.55	-71.59	-78.81	-69.27	-39.82	-64.26	-72.75	-46.99
A2000468	"Human mRNA for zinc finger "protein," clone "RES4-26," complete cds"	28.91	21.69	11.03	18.27	20.88	25.28	6.48	15.50
A2000469	"Human mRNA for TGF-beta superfamily "protein," complete cds"	187.43	179.51	145.53	184.02	41.76	17.38	13.45	31.01

Gambar 1. Contoh 16 data *microarray* adenoma dan normal

Metode yang digunakan dalam penelitian ini adalah metode *spectral clustering* dengan menggunakan algoritma PAM. Metode ini mampu mempartisi data dengan struktur yang lebih rumit

dibandingkan dengan metode *cluster* tradisional seperti *k-means*, *fuzzy c-means*, SOM, dan lain sebagainya. Pada penelitian ini alat bantu yang digunakan adalah software RStudio 1.1.456 dan Microsoft Excel 2016.

1. Normalisasi Data

Dataset yang diperoleh mungkin saja memiliki nilai yang sangat bervariasi mulai dari data yang terlalu besar ataupun data yang terlalu kecil. Oleh karena itu, perlu dilakukan normalisasi data untuk menghindari dimensi data yang terlalu besar ataupun terlalu kecil. Normalisasi data dilakukan dengan menggunakan *min-max normalization*. Berikut merupakan rumus *min-max normalization* (Wulandari, 2020):

$$v' = \frac{v - \min}{\max - \min} (new_{\max} - new_{\min}) + new_{\min} \quad (1)$$

Keterangan:

- v : nilai elemen hasil ekstraksi *n-mers frequency* sebelum dinormalisasi
- v' : nilai elemen hasil ekstraksi *n-mers frequency* setelah dinormalisasi
- \min : nilai minimum matriks
- \max : nilai maksimum matriks
- new_{\max} : nilai range maksimum terbaru
- new_{\min} : nilai range minimum terbaru

2. Spectral Clustering

Metode *spectral clustering* merupakan metode *clustering* yang berdasarkan pada kesamaan antara setiap ekspresi gen atau data. Metode ini menggunakan informasi yang diperoleh dari nilai eigen dan vektor eigen dari kedekatan matriks untuk mempartisi *graph* (Sharma et al., 2016). Dimana, vertex atau simpul dari *graph* tersebut merupakan setiap objek pada data. Sedangkan, sisi atau edge dari *graph* tersebut adalah hubungan antara data yang diperoleh dengan menghitung jarak dari dua data yang berhubungan (Yusuf, Ginardi, & Arieshanti, 2012). *Graph* pada *spectral clustering* merupakan *graph* tidak berarah $G = (V, E)$, dan pada penelitian ini *graph* yang digunakan adalah *graph* terhubung sepenuhnya yang menghubungkan semua titik data dengan similaritas positif. Berikut merupakan langkah-langkah dalam melakukan *spectral clustering* (Yusuf et al., 2012):

- a. Mengkonstruksi *graph* similaritas dari dataset dengan vertex atau simpul pada *graph* tersebut merupakan objek pada dataset. Sedangkan, bobot pada setiap edge atau sisi merupakan jarak antara satu vertex dengan vertex lainnya. Perhitungan jarak dihitung dengan fungsi kesamaan Gauss:

$$s(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (2)$$

Parameter σ merepresentasikan “lebar” dari neighborhood. Untuk menentukan nilai optimum dari σ diperoleh berdasarkan nilai *average silhouette* tertinggi dari seluruh kumpulan dataset yaitu $\sigma = 0,7233979114$.

- b. G merupakan grafik tak berarah dengan matriks bobot W untuk setiap edge $(i, j) \in E$, diberikan bobot w_{ij} dengan nilai $w_{ij} = w_{ji} \geq 0$. Matriks kesamaan

$$W = (w_{ij})_{i,j=1}^n \quad (3)$$

Dengan bobot dari edge (i, j) merepresentasikan kesamaan antara data x_i dan data x_j . Hitung d_i yang merupakan *degree* dari vertex $v_i \in V$ yang didefinisikan sebagai

$$d_i = \sum_{j=1}^n w_{ij} \quad (4)$$

Dari *degree* vertex tersebut dapat dibentuk matriks degree yang merupakan matriks diagonal yang berisi bobot setiap vertex.

- c. Partisi data menjadi k cluster yang saling lepas, dengan k merupakan jumlah cluster dengan cara menormalisasi dengan matriks Laplacian menggunakan matriks bobot W dan matriks *degree* D yang dihitung dengan persamaan:

$$L_{\text{sym}} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} = I - D^{-\frac{1}{2}} W D^{-\frac{1}{2}} \quad (5)$$

Dengan $L = D - W$, dimana D adalah matriks diagonal dengan anggotanya berasal dari persamaan (4) merupakan bobot dari semua yang bertetangga dengan vertex i dan W dari persamaan (3).

- d. Hitung k -eigen vektor pertama dari matriks Laplacian. Maka terbentuklah matriks k -eigen yang merupakan k -eigen vektor pertama dari matriks Laplacian. Matriks k -eigen berukuran $n \times k$, dengan n merupakan jumlah objek pada dataset.
- e. Normalisasi data dengan matriks k -eigen sehingga terbentuk k kolom yang merepresentasikan setiap nilai normalisasi eigen pada setiap kolomnya
- f. Hasil dari data normalisasi kemudian dilakukan *clustering* dengan menggunakan algoritma partisi *clustering*. Dalam penelitian ini dilakukan dengan menggunakan algoritma PAM. Data normalisasi mewakili masukan data latih. Data latih ke- i akan dimasukkan pada suatu *cluster* jika dan hanya jika data hasil normalisasi ke- i masuk pada *cluster* yang sama.

3. PAM

Misalkan D_{ij} merupakan ketidaksamaan antara gen i dan gen j dimana setiap gen diwakili oleh vector dimensi k . Pencarian untuk k medoid dari dataset merupakan dasar dari algoritma PAM (Cahyaningrum et al., 2017). Langkah-langkah algoritma *clustering* PAM yaitu:

- a. Inisialisasi pusat *cluster* sebanyak k (jumlah *cluster*)
- b. Alokasikan setiap data normalisasi ke *cluster* terdekat menggunakan persamaan ukuran jarak *Euclidian Distance* dengan persamaan (Pramesti, Lahan, Tanzil Furqon, & Dewi, 2017):

$$d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} ; i = 1, 2, 3, \dots, n \quad (6)$$
- c. Pilih secara acak objek pada masing-masing *cluster* sebagai kandidat medoid baru.
- d. Hitung jarak setiap objek yang berada pada masing-masing *cluster* dengan kandidat medoid baru.
- e. Hitung total simpangan (S) dengan menghitung nilai total *distance* baru dikurang total *distance* lama. Jika $S < 0$, maka tukar objek dengan data *cluster* untuk membentuk sekumpulan k objek baru sebagai medoid.
- f. Ulangi langkah c sampai e hingga tidak terjadi perubahan medoid, sehingga didapatkan *cluster* beserta anggota *cluster* masing-masing.

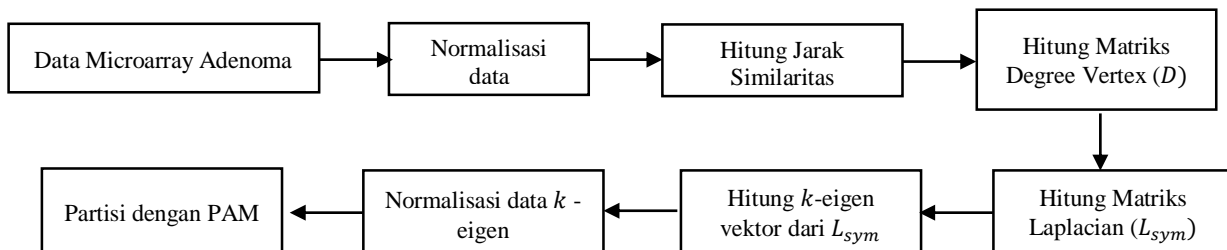
4. Average Silhouette

Metode average silhouette merupakan metode yang pertama kali diperkenalkan oleh Peter J Rousseuw. Metode ini digunakan untuk mengukur akurasi dari hasil *clustering*. Nilai silhouette didasarkan pada pengukuran seberapa mirip suatu objek dengan kelompoknya dibandingkan dengan kelompok lain. Persamaan Silhouette (Siswantining, Wulandari, & Bustamam, 2018):

$$S_k = \frac{b_k - a_k}{\max(a_k, b_k)} \quad (7)$$

- a. Untuk setiap gen k hitung a_k , yang merupakan rata-rata jarak dari gen k dengan anggota lainnya pada kelompok tersebut.
- b. Untuk setiap gen k dan masing-masing kelompok k yang bukan anggota kelompok j , hitung rata-rata jarak dari gen j dengan anggota kelompok k yaitu (b_{kj}) . Misalkan $b_k = \min_k (b_{kj})$

Rata-rata pemecahan dari setiap *cluster* induk dinamakan *split silhouette* (SS_j) $j = 1, 2, \dots, c$. *Split silhouette* merupakan ukuran dari *cluster* yang heterogen.



Gambar 2. Tahap-Tahap *Spectral Clustering* menggunakan PAM

HASIL DAN PEMBAHASAN

Microarray Adenoma yang diperoleh melalui laman <http://genomics-pubs.princeton.edu/oncology/> diolah menggunakan software R 1.1.456. Data yang diperoleh kemudian dilakukan normalisasi data menggunakan min-max normalization untuk menghindari data yang terlalu besar ataupun terlalu kecil. Pada Gambar 3. Terlihat bahwa dataset yang awalnya terlalu besar dan terlalu kecil, dengan dilakukan normalisasi data maka dataset menjadi normal.

Adenoma.1	Adenoma.2	Adenoma.3	Adenoma.4
-0.0730044229	-0.1314971288	-0.082860970	-0.059878453
-0.0695381378	-0.1419442021	-0.102954528	-0.052377744
-0.0584400749	-0.0808013310	-0.075036177	-0.065377403
-0.0695381378	-0.1255313001	-0.087326205	-0.053873176
-0.0757715003	-0.1642954405	-0.121963672	-0.064376524
-0.0723052152	-0.1807083425	-0.110779321	-0.058877574
-0.0598236136	-0.1121150586	-0.100721911	-0.062881092
0.0587887937	-0.0897638202	-0.072803560	-0.019843273
0.0664205716	0.2114143060	0.197534531	0.033697893
-0.0757715003	-0.1180808873	-0.088431882	-0.058383022
-0.0015662215	0.0459381634	0.048969780	0.019191028
-0.1562398949	-0.2626903756	-0.264936247	-0.131941778
-0.1555406872	-0.3342803199	-0.273887980	-0.140443366
-0.0313792485	-0.0778321628	-0.082860970	-0.037364552

Gambar 3. Hasil Normalisasi 14 *Microarray* Adenoma

Kemudian langkah selanjutnya adalah melakukan *clustering* menggunakan *spectral clustering* dan dilakukan partisi *clustering* menggunakan PAM. Pada proses *spectral clustering* parameter σ merepresentasikan “lebar” dari neighborhood. Saat dilakukan *clustering* menggunakan *spectral clustering* dilakukan *cluster* dengan nilai $k = 2, 3, \dots, 10$ untuk mencari nilai k terbaik yang diperoleh dengan mencari nilai *average silhouette* tertinggi. Untuk menentukan nilai optimum dari σ diperoleh berdasarkan nilai *average silhouette* tertinggi dari seluruh kumpulan dataset. Dalam penelitian ini, diperoleh nilai dari neighborhood tertinggi yaitu $\sigma = 0,7233979114$ yang diperoleh dari dataset ke- 3052 (“Human APC gene “mRNA,” complete cds - Also Represents: M73548”) yang merupakan *cluster* pertama dengan neighborhood cluster kedua. Sedangkan, normalisasi data dengan matriks k -eigen sehingga terbentuk $k = 3$ kolom yang merepresentasikan setiap nilai normalisasi eigen yang merupakan dataset baru yang dihasilkan pada setiap kolomnya dapat dilihat pada Gambar 4.

1	-0.0044753714	-0.0063929140	-2.624476e-03
2	-0.0077282273	-0.0076145595	-4.324357e-03
3	-0.0007297800	-0.0062319708	-8.976741e-04
4	-0.0075744039	-0.0074070678	-4.223029e-03
5	-0.0085158512	-0.0053191281	-4.964329e-03
6	-0.0092392038	-0.0052492000	-5.546480e-03
7	-0.0076221771	-0.0066940496	-4.282764e-03
8	0.0039671908	-0.0065848216	1.409084e-03
9	0.0164772241	0.0037415497	1.824170e-02
10	-0.0081151667	-0.0074376311	-4.527038e-03
11	0.0138808156	-0.0020206362	8.232889e-03
12	-0.0058946322	0.0154976295	-1.631714e-02
13	-0.0040390357	0.0142753201	-1.404636e-02
14	0.0002107957	-0.0068310371	-6.539197e-04
15	0.0157204570	0.0021637100	1.491923e-02

Gambar 4. Dataset 15 *Microarray* Adenoma baru dengan $k = 3$

Tabel 1. Hasil *Clustering Microarray* Adenoma menggunakan *Spectral Clustering* dengan Algoritma PAM

Cluster ke-	Jumlah Anggota Cluster	Nilai Average Silhouette Cluster	Medoid Cluster	Jarak maksimum	Jarak rata-rata	Diameter	Pemisah
1	4349	0,5912449	497	0,12156157	0,007268904	0,12836073	$7,308168 \times 10^{-5}$
2	2115	0,4515836	1139	0,06360155	0,009919601	0,07052616	$7,308168 \times 10^{-5}$
3	623	0,3567585	4138	0,06529478	0,015870512	0,08323241	$3,364821 \times 10^{-4}$

Tabel 1 menunjukkan hasil dari *clustering microarray* adenoma dengan menggunakan *spectral clustering*. Pada proses partisi, *clustering* dilakukan dengan menggunakan algoritma PAM. Banyaknya cluster yang terbentuk diperoleh berdasarkan nilai *average silhouette* tertinggi. Berdasarkan hasil tersebut diperoleh jumlah optimal *cluster* adalah tiga *cluster* dengan nilai *average silhouette* 0,5289522. Hasil *clustering* diperoleh 4.349 *microarray* adenoma pada *cluster* ke-1, 2.115 *microarray* adenoma pada *cluster* ke-2, dan 623 *microarray* adenoma pada *cluster* ke-3. Dengan medoid dari *cluster* ke-1 adalah *microarray* 497 ("Human mRNA for protein disulfide isomerase-related protein "P5," complete cds"), medoid dari *cluster* ke-2 adalah *microarray* 1139 ("Gtp Cyclohydrolase "I," Alt. Splice 1 - Also Represents: U19523"), dan medoid dari *cluster* ke-3 adalah *microarray* 4138 ("Human uridine diphosphoglucose pyrophosphorylase "mRNA," complete cds"). Pada Tabel 1. terlihat bahwa diameter untuk *cluster* ke-1 lebih besar dibandingkan dengan diameter *cluster* lainnya. Sedangkan, *cluster* ke-2 memiliki diameter lebih kecil dibandingkan dengan *cluster* lainnya. Jarak maksimum terbesar diperoleh dari *cluster* ke-1 dan jarak maksimum terkecil diperoleh pada *cluster* ke-2. Terlihat jelas pada Tabel 1 bahwa ketiga *cluster* yang dihasilkan bukan merupakan L *cluster*, hal tersebut dikarenakan nilai diameter dari masing-masing *cluster* lebih besar dibandingkan dari nilai pemisah ketiga *cluster*.

PENUTUP

Pada penelitian ini, data *microarray* adenoma dapat dilakukan *clustering* menggunakan metode *spectral clustering* menggunakan algoritma partisi PAM. Tahapan pada penelitian ini adalah melakukan normalisasi data menggunakan *min-max normalization*, menghitung matriks similaritas antara data *microarray* adenoma, menghitung matrix degree vertex, membuat matriks Laplacian, menemukan k -eigen vektor dari nilai eigen terkecil, normalisasi data k -eigen, dan kemudian melakukan partisi dataset *microarray* adenoma menggunakan algoritma PAM. Hasil pada penelitian ini menunjukkan bahwa *microarray* adenoma menghasilkan tiga *cluster* dengan nilai *average silhouette* 0,5289522. Hasil *clustering* diperoleh 4.349 *microarray* adenoma pada *cluster* ke-1, 2.115 *microarray* adenoma pada *cluster* ke-2, dan 623 *microarray* adenoma pada *cluster* ke-3. Berdasarkan hasil penelitian ini, dapat dilakukan analisis lebih lanjut baik dari sisi biologis, laboratorium dari *microarray* adenoma, maupun algoritma partisi lainnya untuk menguji keakuratan hasil *clustering microarray adenoma*.

DAFTAR PUSTAKA

- Bhat, A. (2014). K-Medoids Clustering Using Partitioning Around Medoids for Performing Face Recognition. *International Journal of Soft Computing, Mathematics and Control*, 3(3), 1–12. <https://doi.org/10.14810/ijscmc.2014.3301>
- Cahyaningrum, R. D., Bustamam, A., & Siswantining, T. (2017). Implementation of spectral clustering with partitioning around medoids (PAM) algorithm on microarray data of carcinoma. *AIP Conference Proceedings*, 1825(March). <https://doi.org/10.1063/1.4978976>
- Hidayat, M. (2015). Adenoma Hipofisis. *Majalah Kedokteran Andalas*, 38(2), 130. <https://doi.org/10.22338/mka.v38.i2.p130-138.2015>
- Muradi, H., Bustamam, A., & Lestari, D. (2016). Application of hierarchical clustering ordered partitioning and collapsing hybrid in Ebola Virus phylogenetic analysis. *ICACISIS 2015 - 2015 International Conference on Advanced Computer Science and Information Systems, Proceedings*, (October), 317–323. <https://doi.org/10.1109/ICACISIS.2015.7415183>
- Pramessti, D. F., Lahan, Tanzil Furqon, M., & Dewi, C. (2017). Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 1(9), 723–732. <https://doi.org/10.1109/EUMC.2008.4751704>
- Rizby, L. P., Marji, & Muflikhah, L. (2018). Clustering Pasien Kanker Berdasarkan Struktur Protein Dalam Tubuh. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIHK) Universitas Brawijaya*, 2(10), 3810–3816.
- Sharma, A., Kumar, R., & Mansotra, V. (2016). Proposed Stemming Algorithm for Hindi Information Retrieval. *International Journal of Innovative Research in Computer and Communication Engineering (An ISO Certified Organization)*, 3297(6), 11449–11455.

<https://doi.org/10.15680/IJIRCCE.2016>

- Siswantining, T., Wulandari, S., & Bustamam, A. (2018). Collaboration and implementation of self organizing maps (SOM) partitioning algorithm in HOPACH clustering method. *AIP Conference Proceedings, 2014*, 020134-1-020134–020139. <https://doi.org/10.1063/1.5054538>
- Taslim, T., & Fajrizal, F. (2016). Penerapan algoritma k-mean untuk clustering data obat pada puskesmas rumbai. *Digital Zone: Jurnal Teknologi Informasi Dan Komunikasi*, 7(2), 108–114. <https://doi.org/10.31849/digitalzone.v7i2.602>
- Wulandari, S. (2020). Clustering Kecamatan Di Kota Bandung Berdasarkan Algoritma K-Means. *Seminar Nasional Riset Dan Teknologi (SEMNAS RISTEK) 2020, P-ISSN:252(3)*, 128–132.
- Yusuf, A., Ginardi, H., & Arieshanti, I. (2012). Pengembangan Perangkat Lunak Prediktor Nilai Mahasiswa Menggunakan Metode Spectral Clustering dan Bagging Regresi Linier. *Jurnal Teknik ITS*, 1(2), A246–A250. <https://doi.org/10.12962/j23373539.v1i1.645>