



Penerapan K-Means dan Analisis Jaringan untuk Identifikasi Protein-Meningitis Signifikan Berdasarkan Interaksi Antar Protein

Mohammad Hamim Zajuli Al Faroby¹, Amalia Nur Alifah², Bernadus Anggo Seno Aji³, Irma Octavia Chaniago⁴

^{1,2,4}Program Studi Sains Data, Institut Teknologi Telkom Surabaya

³Program Studi Teknologi Informasi, Institut Teknologi Telkom Surabaya

¹alfaroby@ittelkom-sby.ac.id

²amaliaifah@ittelkom-sby.ac.id

³bernadus.seno@ittelkom-sby.ac.id

⁴octaviairma99@gmail.com

Corresponding author email: alfaroby@ittelkom-sby.ac.id

Abstract: Inflammation of brain membranes and spinal cord becomes serious. The symptoms caused by this disease cause the patient to experience severe headaches, fever, and a stiff neck. So far, the best prevention is the meningitis vaccine. It will also be necessary to obtain a specific drug compound to treat the condition of meningitis that has already occurred. This study aimed to receive significant proteins from meningitis based on protein-protein interactions. K-means and network clustering analysis are very helpful in determining protein dominance in the interaction network. The analysis results obtained from the open database (OMIM) show that meningitis protein interactions form a graph with 118 nodes and 789 edges. The k-means algorithm groups the data into 5 clusters. One of the five clusters formed was chosen for further analysis with graph properties. After getting the graph characteristics of the nodes, the following process is to calculate the overall value of the data for ranking. The significant protein that had the greatest overall value was CD4. This protein is a strong candidate for the most significant protein when the meningitis pathogen invades the human body.

Abstrak: Peradangan pada selaput otak dan sumsum tulang belakang menjadi hal serius. Gejala yang disebabkan oleh penyakit ini mengakibatkan pasien mengalami sakit kepala yang parah, demam dan disertai leher kaku. Selama ini pencegahan terbaik yakni dengan vaksin meningitis. Akan perlu juga untuk mendapatkan senyawa obat yang spesifik untuk menanggulangi kondisi meningitis yang sudah terjadi. Tujuan dari penelitian ini adalah mendapatkan protein signifikan dari meningitis berdasarkan interaksi protein-protein. Analisis klusterisasi k-means dan jaringan sangat membantu dalam menentukan dominasi protein dalam jaringan interaksi. Dari hasil analisis yang didapatkan dari basis data terbuka (OMIM), interaksi protein meningitis membentuk graf dengan *node* sebanyak 118 dan *edge* sebanyak 789. Algoritma k-means mengelompokkan data-data tersebut menjadi 5 kluster. Lima kluster yang terbentuk dipilih salah satu untuk dianalisis lebih lanjut dengan sifat-sifat graf. Setelah mendapatkan karakteristik graf dari *node*, proses selanjutnya adalah menghitung nilai keseluruhan data untuk di rangking. Protein signifikan yang memiliki nilai keseluruhan terbesar yakni CD4. Protein ini menjadi kandidat kuat protein yang paling signifikan Ketika pathogen meningitis menyerang kedalam tubuh manusia.

Kata kunci: Interaksi Protein-Protein, Analisis Jaringan, K-means, Meningitis.

I. PENDAHULUAN

Infeksi yang disebabkan oleh pathogen yang menyerang selaput otak dan sumsum tulang belakang mengakibatkan peradangan serius. Pathogen peradangan ini biasanya adalah virus, akantetapi dapat juga disebabkan oleh bakteri ataupun jamur. Penyakit peradangan selaput otak dan sumsum tulang belakang tersebut dikenal dengan Meningitis. Penderita meningitis akan mengalami gejala sakit kepala yang sangat parah, demam dan disertai leher kaku [1]. Penelitian secara in vivo menghasilkan terdapat beberapa protein utama yang diserang oleh pathogen penyebab Meningitis [2]. Akan tetapi, belum diteliti sampai mendalam mengenai seberapa krusial protein tersebut. Vaksin meningitis sendiri sudah ada dan terbukti efektif sebagai pencegahan awal terhadap pathogen Meningitis. Sebagai pengobatan dibutuhkan antibiotic maupun inhibitor khusus yang efektif menghambat pathogen meningitis [3]. Untuk menghasilkan obat tersebut, farmasis membutuhkan pengetahuan mengenai protein target dari suatu virus penyebab penyakit untuk dapat membuat suatu inhibitor virus yang menginjeksi protein [4].

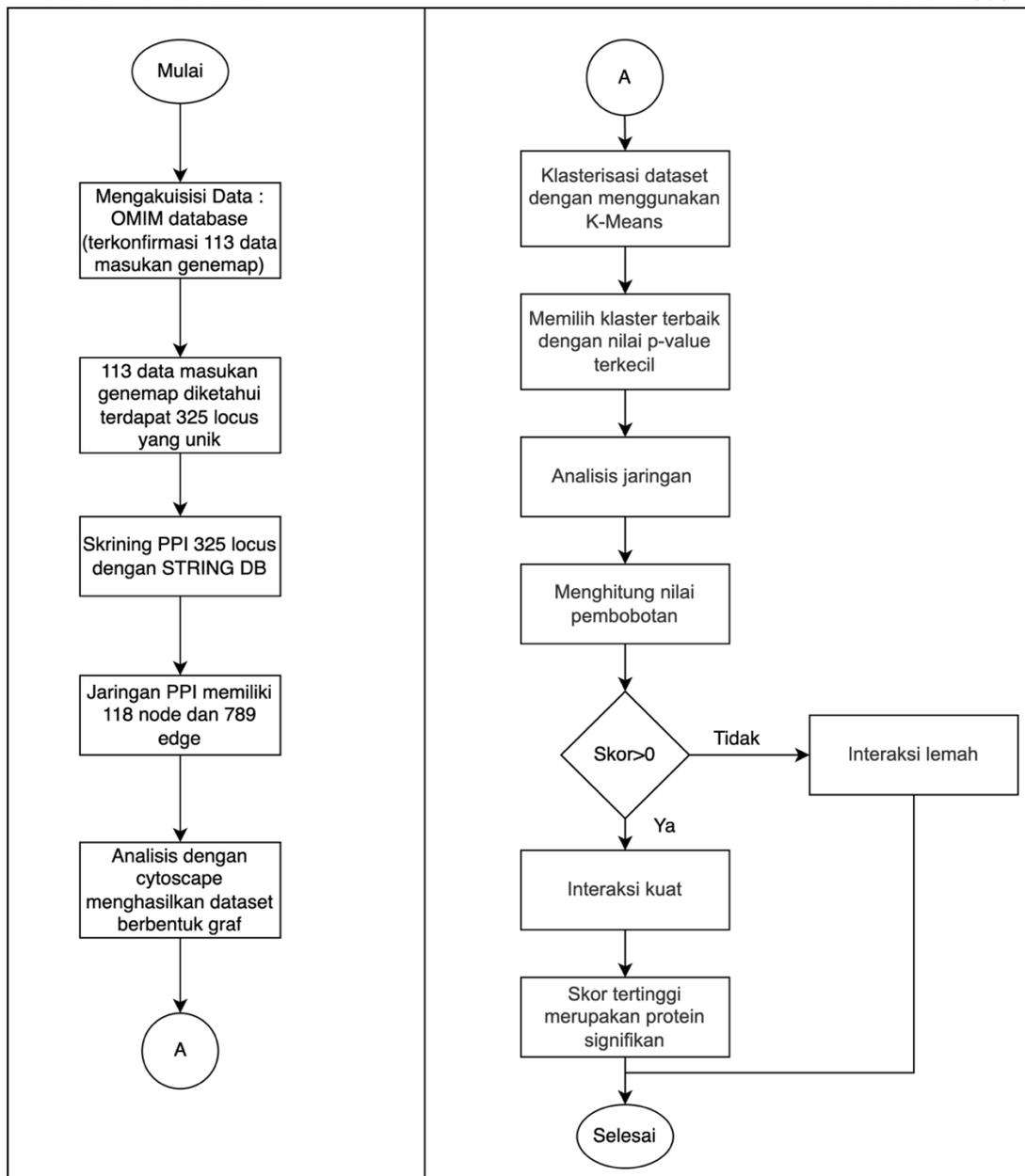


Protein sendiri merupakan pembentuk jaringan utama dalam metabolisme tubuh. Hampir semua fungsional organ, enzyme dan hormone manusia dibangun oleh protein [5]. Fungsional protein tidak bekerja secara independent, melainkan saling mendukung dan berrelasi terhadap protein-protein lainnya. Relasi antar protein secara fungsionalitas ini dinamakan Interaksi Protein-Protein (PPI). Beberapa penelitian terkait menganalisis protein signifikan menggunakan algoritma ClusterOne [6]. Teknik pembelajaran mesin sangat membantu mempercepat analisis dalam memprediksi kemungkinan data-data yang ada.

Penelitian ini bertujuan untuk mendapatkan protein signifikan dari meningitis berdasarkan interaksi antar protein. Analisis yang digunakan yakni berbasis data yang didapatkan dari basis data terbuka. Basis data tersebut dapat diakses secara gratis untuk tujuan penelitian. Untuk menganalisis data-data tersebut algoritma pengelompokan berbasis pembelajaran mesin diusulkan. Algoritma k-means menjadi algoritma sederhana dalam pengelompokan data. Hasil dari pengelompokan itu dianalisis dengan menggunakan teori graf untuk mendapatkan fitur-fitur yang terdapat pada jaringan interaksi protein. Protein signifikan dapat ditentukan dengan merangkin data-data yang didapat pada analisis sebelumnya dengan menggunakan perhitungan nilai eigen dan vector eigen.

II. METODOLOGI PENELITIAN

Penelitian ini bertujuan untuk mendapatkan protein signifikan dari penyakit Meningitis. Pencarian protein signifikan tersebut berdasarkan data interaksi antar protein yang terlibat dalam peradangan pathogen yang menyebabkan meningitis. Protein yang signifikan diduga menjadi sentral dari infeksi yang terjadi.



Gambar 1. Diagram alir penelitian untuk mendapatkan protein signifikan dari Meningitis

Apabila dapat ditemukan penawar untuk mereduksi infeksi pada protein signifikan tersebut, memungkinkan untuk mereduksi infeksi protein lainnya yang saling terkoneksi berdasarkan fungsional protein. Tahap pelaksanaan penelitian pada Gambar 1 menunjukkan proses pencarian protein Meningitis yang signifikan dari akuisisi data, eksplorasi data, klusterisasi data, analisis jaringan hingga perangkaian untuk menentukan protein yang paling signifikan.

Proses penelitian ini menggunakan beberapa aplikasi untuk mendukung analisis. Pada bagian visualisasi graf, aplikasi yang digunakan yakni STRING-DB dan Cytoscape. Proses klusterisasi data kami menggunakan bantuan Bahasa pemrograman Python. Sedangkan pada proses analisis jaringan hingga akhir, aplikasi yang kami gunakan yakni Matlab dan Python.

2.1. Akuisisi Data

Data bioinformatika banyak tersedia didatabase terbuka. Salah satu database terbuka yang berkaitan dengan protein-protein manusia yakni Online Mendelian Inheritance in Man (OMIM)[7], halaman ini menyediakan berbagai protein yang berkaitan dengan berbagai fungsionalitas kerja protein di sistem jaringan manusia. Fungsionalitas ini bisa berkaitan dengan metabolisme, genetika,



kelainan maupun penyakit yang disebabkan dari pathogen. Data pada OMIM discraping berdasarkan signifikan entry data yang diinginkan, misal menggunakan kata kunci ‘Meningitis’. Dari hasil akuisisi data pada basis data OMIM, kami mendapatkan data sebanyak 113 data masukan yang berkaitan dengan meningitis. Data-data tersebut perlu di bersihkan untuk mendapatkan kode-kode unik dari protein yang berkaitan. Pembersihan data meliputi pemilihan data kolom locus/gen, penghapusan data duplikat, dan penyesuaian format data. Pembersihan data yang kami lakukan menunjukkan sebanyak 325 locus/gane unik yang juga merupakan kode protein. Data-data ini yang kami gunakan untuk mencari koneksi-koneksi interaksi protein.

Akuisisi pada database OMIM hanya untuk mendapatkan kode unik dari protein-protein yang berrelasi dengan meningitis. Data interaksi antar protein didapatkan pada basis data terbuka STRING-DB [8], untuk mencari koneksi-koneksi protein yang telah didapatkan sebelumnya pada basis data OMIM. Data-data interaksi protein inilah yang dianalisis dengan menggunakan algoritma k-means dan analisis jaringan. Struktur interaksi antar protein berbentuk graf, sehingga analisis jaringan dibutuhkan untuk mendapatkan nilai fitur data dengan menerapkan beberapa sifat-sifat pada graf. Kode protein unik yang didapatkan pada proses seblumnya dijadikan entri pada STRING-DB. Hasil pencarian interaksi antar protein, kami mendapatkan hanya 118 data protein yang berinteraksi dan terdapat 789 interaksi yang dihasilkan. Data interaksi yang didaptkan pada STRING-DB ini yang akan dianalisis dengan algoritma k-means dan analisis jaringan.

2.2. K-means

Algoritma k-means tergolong dalam metode *unsupervised learning*. Algoritma ini mengelompokkan data yang dianggap sejenis dengan proses iterative kedalam sejumlah K kluster yang telah ditentukan diawal. Algoritma ini tergolong sederhana dan tidak membutuhkan waktu yang lama dalam proses pelatihan. Basis dari k-means adalah menghitung jarak antara data dan titik pusat yang dipilih. Pemilihan titik pusat sendiri dapat dilakukan dengan random, pemilihan secara geometris ataupun menggunakan algoritma optimasi untuk membangkitkan nilai awal [9].

Misalkan diberikan sampel data sebanyak N yang akan diklasterkan ke dalam K kluster. Algoritma k-means dimulai dari [10], 1) Tentukan jumlah kluster K yang diinginkan; 2) Inisialisasi titik pusat kluster. Menentukan pusat kluster cara yang paling adalah dengan random atau acak. 3) Hitung jarak setiap data ke titik pusat yang diberikan. Pengelompokan data ditentukan dengan kedekatan data dengan titik pusat. Jarak paling antara satu data dengan satu cluster tertentu akan menentukan suatu data masuk dalam kluster mana. Perhitungan jarak sendiri ada bermacam-macam definisi yang dapat digunakan. Definisi jarak yang paling terkenal yakni Euclidian dan Manhattan. Proses tersebut diulang-ulang sebanyak iterasi yang diinginkan. Kondisi berhenti iterasi juga dapat dibatasi dengan menentukan nilai error terkecil dari suatu kluster.

2.3. Analisis Jaringan

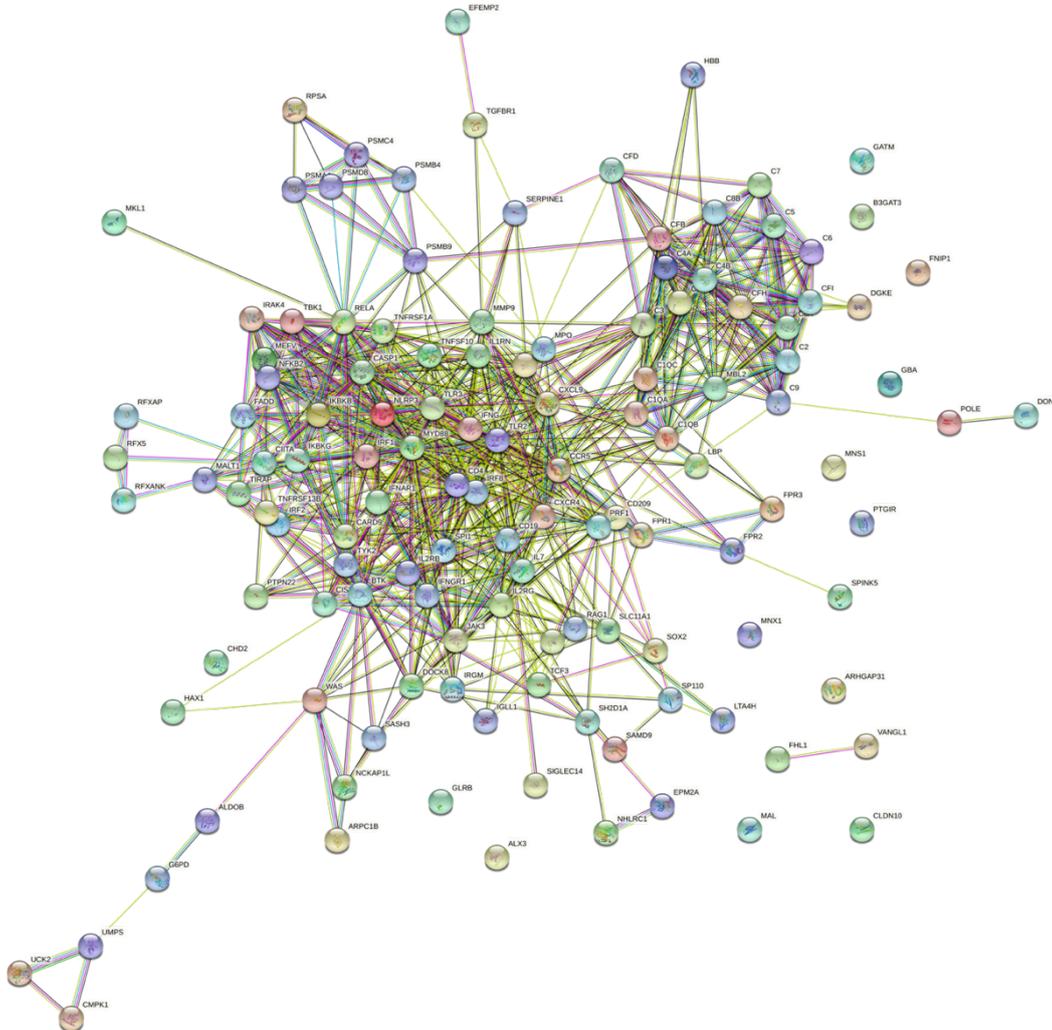
Bentuk visual dari interaksi antar protein adalah graf. *Node* dari graf merupakan protein itu sendiri, sedangkan *edge* dari graf merupakan representasi bahwa antar protein terdapat interaksi [6]. Graf yang dihasilkan dari interaksi antar protein tergolong kompleks menyerupai jaringan. Sehingga analisis jaringan yang digunakan pada penelitian ini adalah analisis graf itu sendiri. Graf sendiri memiliki karakteristik-karakteristik yang dapat diekstrak. Ekstraksi karakteristik sendiri berdasarkan *node* pada jaringan protein yang terbentuk [11].

Analisis jaringan yang kami gunakan menggunakan bantuan aplikasi cytoscape. Ekstraksi karakteristik dari graf yang terdapat pada aplikasi cytoscape terdiri dari, centrality, derajat, radiality dan stress pada jaringan berbentuk graf tersebut [12].

III. HASIL DAN PEMBAHASAN

3.1. Graf Interaksi Antar Protein Meningitis

Hasil eksplorasi data menunjukkan bahwa terdapat 118 protein yang memiliki interaksi dan 789 data interaksi yang terjadi antara protein-protein tersebut. Bentuk graf yang terjadi dari interaksi protein-protein tersebut di visualkan pada Gambar 2.



Gambar 2. Visualisasi interaksi antar protein-meningitis dengan mengambil threshold interaksi sebesar 0,4 pada database STRING-DB.

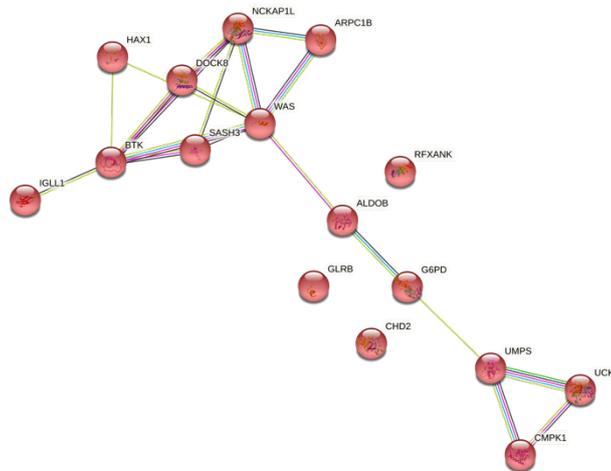
Node mewakili semua protein yang dihasilkan oleh satu lokus gen penyandi protein yang berkaitan dengan meningitis, sedangkan untuk edge merupakan interaksi yang terjadi antara dua protein. Pada Gambar 2, terlihat edge yang terbentuk memiliki warna yang berbeda. Warna dari edge tersebut representasi dari dasar menentukan interaksi antar protein. Secara fungsionalitas warna-warna tersebut mewakili interaksi yang terjadi berdasarkan gene neighborhood, gene fusion, gene homology dan gene co-occurrence.

Keterangan gambar pada *node* yang merepresentasikan protein juga memiliki makna tersendiri. Misalkan *node* dengan warna dan *node* yang tidak berwarna (putih). *Node* berwarna mengartikan

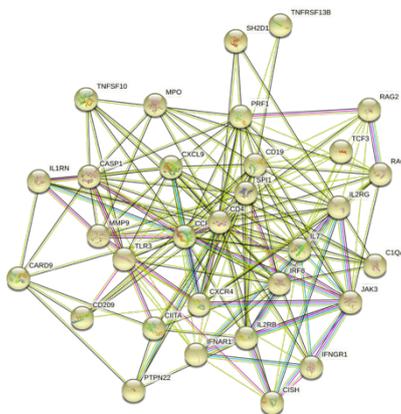
bahwa protein-protein tersebut merupakan protein utama dalam interaksi sedangkan warna putih menyatakan protein pelapis. *Node* yang memiliki isi mengartikan bahwa struktur 3D protein diketahui atau dapat diprediksi, sedangkan *node* yang tidak memiliki isi, struktur 3D belum diketahui.

3.2. Hasil dan Analisis Klusterisasi dengan K-Means

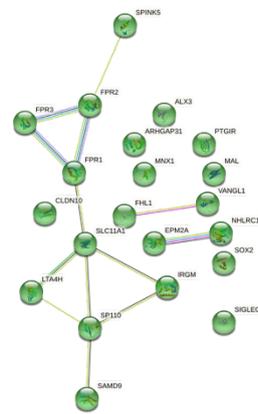
Data interaksi protein diklusterkan untuk mengelompokkan protein-protein yang memiliki interaksi sejenis. Algoritma yang kami gunakan untuk mengkluster adalah k-means. Mula-mula kluster yang kami bentuk sebanyak 5 kelompok. Hasil kluster data interaksi protein dengan menggunakan k-means digambarkan pada Gambar 3. Kluster 1 ditunjukkan pada Gambar 3a, yang divisualkan dengan warna merah. Pada kluster 1 memiliki anggota *node* sebanyak 16 buah, akantetapi tidak semua anggota *node* saling terkoneksi. Terdapat tiga *node* yang tidak memiliki *edge*, sehingga protein-protein tersebut tidak saling berinteraksi dengan anggota lain pada klasternya. Hal serupa juga terjadi pada kluster 3 (Gambar 3c), kluster 4 (Gambar 3d) dan kluster 5 (Gambar 3e). Beberapa *node* pada kluster-kluster tersebut tidak saling terkoneksi dengan anggota *node* lainnya pada kluster tersebut.



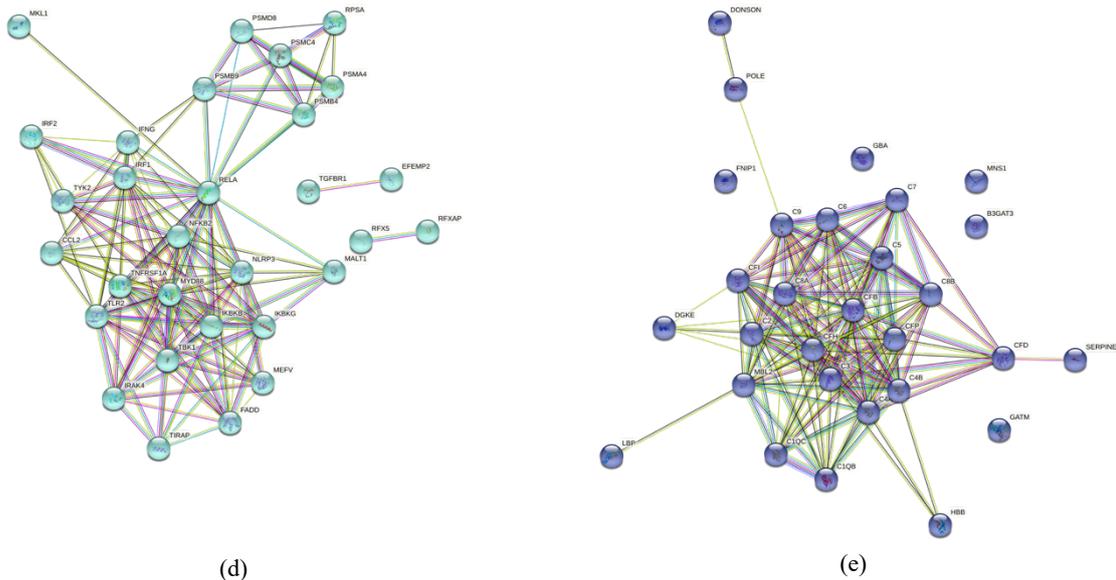
(a)



(b)



(c)



Gambar 3. Hasil klusterisasi data interaksi protein-meningitis menggunakan algoritma k-menas dengan menetapkan kluster sebanyak lima golongan besar. (a) visualisai dari kluster 1, (b) visualisasi dari kalster 2, (c) visualisasi dari kluster 3, (d) visualisasi dari kluster 4, dan (e) visualisasi dari kluster 5.

Kluster 2 memiliki bentuk graf lebih komplit dibanding kluster lainnya. Pada kluster 2, memiliki anggota *node* sebanyak 31 dengan jumlah *edge* sebanyak 170. Hal ini menunjukkan kondisi kluster 2 lebih baik dibandingkan kluster lainnya. Analisis lain terkait kluster terbaik juga didukung dengan perbandingan nilai *p – value* dari setiap kluster. Pada Table 1, nilai *p – value* $< 1,0e - 16$ menunjukkan kluster 2 tersebut sangat bagus dan antara anggota kluster yang terbentuk kemiripannya sangat dekat [12]. Kondisi ini memungkinkan bahwa kluster 2 dapat dijadikan sebagai kluster utama dalam mengidentifikasi protein signifikan dari meningitis.

Table 1. Parameter-parameter hasil analisis dari kluster-kluster yang dihasilkan oleh alhoritma k-means.

Kluster	Jumlah <i>node</i>	Jumlah <i>edge</i>	Rata-rata derajat <i>node</i>	Rata-rata koefisien kluster	<i>p-value</i>
1	16	20	2,5	0,542	$< 1,0e - 16$
2	31	170	11	0,675	$< 1,0e - 16$
3	21	13	1,24	0,492	$1,59e - 11$
4	30	136	9,07	0,828	$< 1,0e - 16$
5	29	141	9,72	0,677	$< 1,0e - 16$

Berdasarkan hasil analisis dari kluster yang dihasilkan oleh algoritma k-means diatas, maka analisis selanjutnya lebih ditekankan pada kluster 2. Pada kluster 2 ini, diekstrak nilai fitur-fitur dari nodenya. Fitur-fitur tersebut akan menjadi parameter penentu signifikansi protein utama dari meningitis.

3.3. Analisis Jaringan: Karakteristik Graf

Ekstraksi fitur yang diambil terdiri dari tujuh fitur yakni *betweenness centrality*, *degree centrality*, *closeness centrality*, *clustering coefficient*, *average shortest path*, *radiality* dan *stress* [13]. Analisis graf ini menggunakan bantuan aplikasi cytoscape dan matlab. Hasil dari ekstraksi karakteristik graf kluster 2 dapat dilihat pada Tabel 2.

Table 2. Hasil ekstraksi fitur dengan menganalisis karakteristik dasar graf yang terdapat pada kluster 2

Nama	Betweenness Centrality	Closeness Centralit y	Degree Centralit y	Clustering Coefficient	Average Shortest Path Length	Radiality	Stress
C1QA	0,001043605	0,5555556	6	0,8	1,8	0,973333333	6
CD4	0,275054735	1	30	0,32183908	1	1	590



SPI1	0,048151797	0,6976744	17	0,455882353	1,433333333	0,985555556	148
IRF8	0,016863711	0,625	12	0,560606061	1,6	0,98	58
CCR5	0,038247583	0,7142857	18	0,516339869	1,4	0,986666667	148
CXCL9	0,022181171	0,6666666	15	0,561904762	1,5	0,983333333	92
IL2RG	0,02109013	0,6521739	14	0,615384615	1,533333333	0,982222222	70
CARD9	0,006130268	0,5555556	6	0,533333333	1,8	0,973333333	14
IL1RN	0,003872469	0,5660377	7	0,714285714	1,766666667	0,974444444	12
CD209	0,003639847	0,5555556	6	0,733333333	1,8	0,973333333	8
PTPN22	0,001340996	0,5454545	5	0,8	1,833333333	0,972222222	4
CIITA	0,004367816	0,5660377	7	0,666666667	1,766666667	0,974444444	14
CASP1	0,029979931	0,6521739	14	0,483516484	1,533333333	0,982222222	94
MPO	0,00173326	0,5555556	6	0,733333333	1,8	0,973333333	8
TNFSF10	0	0,5555556	6	1	1,8	0,973333333	0
IFNAR1	0,009206349	0,6122449	11	0,636363636	1,633333333	0,978888889	40
TLR3	0,018940887	0,6521739	14	0,593406593	1,533333333	0,982222222	74
MMP9	0,005765371	0,5882352	9	0,638888889	1,7	0,976666667	26
PRF1	0,02648513	0,6666667	15	0,580952381	1,5	0,983333333	88
IL2RB	0,01430031	0,6122449	11	0,636363636	1,633333333	0,978888889	40
IL7	0,036183178	0,7142857	18	0,54248366	1,4	0,986666667	140
CD19	0,055245393	0,7317073	19	0,485380117	1,366666667	0,987777778	176
JAK3	0,011845466	0,6382978	13	0,653846154	1,566666667	0,981111111	54
CXCR4	0,011335523	0,6382979	13	0,666666667	1,566666667	0,981111111	52
TNFRSF13B	0	0,5172414	2	1	1,933333333	0,968888889	0
TCF3	0,001404853	0,5660377	7	0,857142857	1,766666667	0,974444444	6
RAG1	0,006634738	0,6122449	11	0,745454545	1,633333333	0,978888889	28
RAG2	0,001787995	0,5882353	9	0,861111111	1,7	0,976666667	10
SH2D1A	0	0,5263158	3	1	1,9	0,97	0
CISH	0,003035942	0,5769231	8	0,75	1,733333333	0,975555556	14
IFNGR1	0,002292465	0,5769231	8	0,785714286	1,733333333	0,975555556	12

Dari Tabel 2 diatas, kolom nama merupakan kode protein yang sudah baku. Sehingga jika ingin mengetahui protein yang dimaksud dapat memasukkan kode protein tersebut ke dalam database protein seperti Uniprot dan Protein Databank. Kolom lainnya merupakan hasil ekstraksi data graf. Dari tabel tersebut kita memiliki 31 data dan 7 fitur data untuk kemudian dianalisis mendapatkan protein signifikan berdasarkan perangkingan karakteristik *node* pada graf.

3.4. Perangkingan Protein

Analisis untuk mendapatkan rangking dari kluster 2 melalui beberapa tahap. Pertama data distandarkan, sehingga semua kolom terdistribusi normal. Setelah data terdistribusi normal data ditransformasikan membentuk suatu matriks dengan mempertahankan indeks datanya. Setelah itu dicari nilai eigen dan vector eigen dari matriks data tersebut [14]. Perhitungan mengenai nilai eigen dan vector eigen dengan menggunakan bantuan pemrograman python. Setelah mendapatkan nilai eigen dan vector eigen, kemudian menghitung nilai keseluruhan skor dengan menjumlahkan hasil perkalian vector eigen dengan semua kolom data. Hasil dari perangkingan protein meningitis yang signifikan digambarkan pada Gambar 4. Hasil analisis tersebut menunjukkan protein CD4 (*cluster of differentiation 4*) memiliki skor paling tinggi dibandingkan protein lainnya. Protein CD4 merupakan



glikoprotein yang berfungsi sebagai koreseptor untuk reseptor sel T. CD4 terdapat pada permukaan sel imun seperti sel T helper, monosit, makrofag, dan sel dendritik. Hal ini sejalan dengan perannya sebagai pertahanan terdepan tubuh kita apabila ada pathogen yang masuk kedalam tubuh. Berdasarkan hasil analisis ini protein signifikan yang menjadi faktor utama infeksi penyebab pathogen meninges adalah protein CD4.

Hasil ini dapat dijadikan studi lanjutan apabila ingin mencari ligan tepat untuk inhibitor virus ataupun pathogen lain yang menyebabkan terjadinya penyakit meninges. Karena untuk mendapatkan senyawa obat yang optimal, dibutuhkan analisis protein target yang tepat agar efektifitas pengobatan lebih baik.

IV. KESIMPULAN

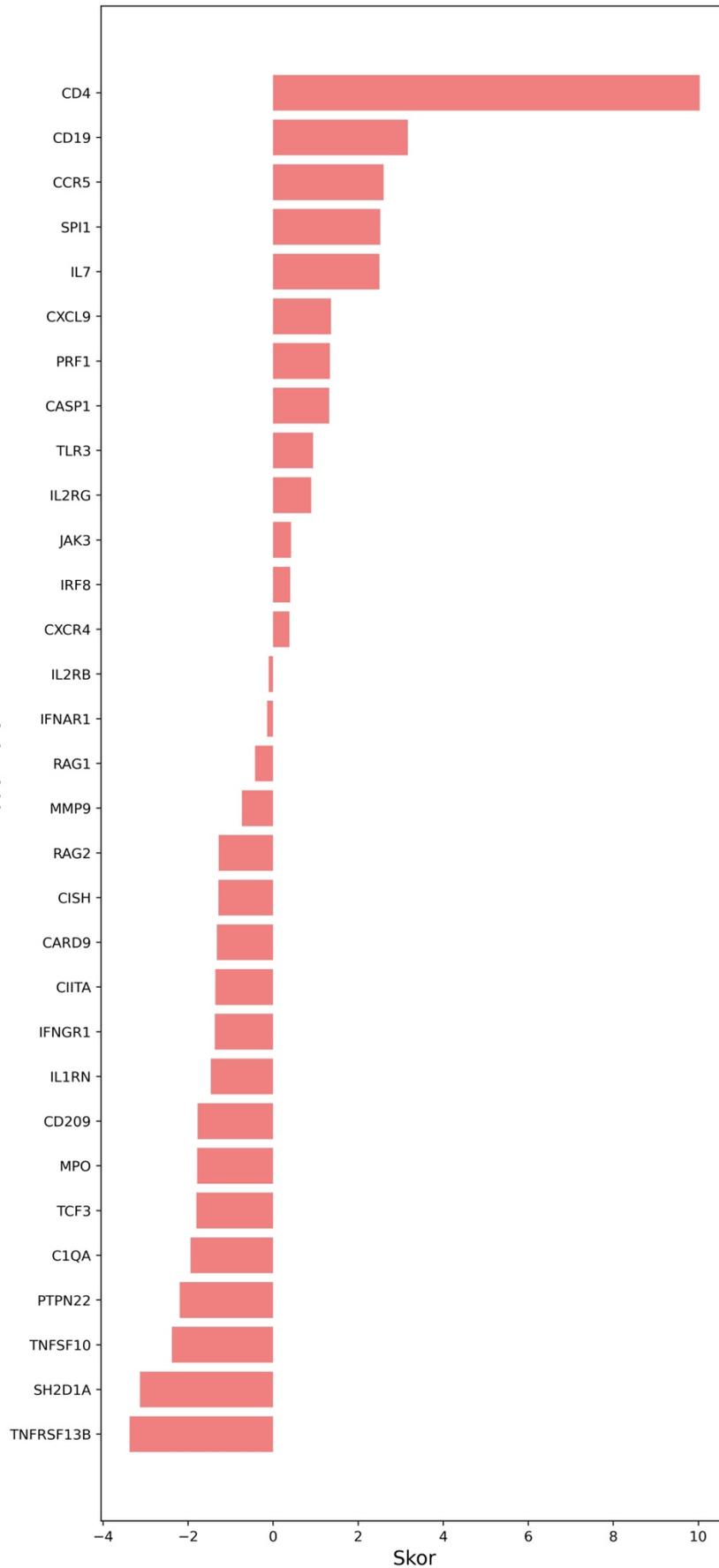
Analisis interaksi protein berguna untuk mengidentifikasi protein signifikan dari meninges. Dengan memilah data berdasarkan kemiripan fungsional protein kita dapat mendapatkan kelompok data yang tergolong paling dominan. Analisis pengelompokan data menggunakan k-means dapat membantu dengan cepat menganalisis data interaksi protein. Kluster terbaik dari data berbentuk graf yakni kluster yang memiliki nilai p – value paling kecil dan semua *node* terhubung satu sama lain. Kluster 2 dari hasil analisis tersebut menjadi kandidat kelompok protein signifikan. Untuk mendapatkan protein tunggal yang signifikan perhitungan dengan mencari nilai eigen dan vector eigen dari matriks data diperlukan. Hasil akhir menunjukkan protein CD4 memiliki total skor paling tinggi dibandingkan 30 data protein lainnya pada kluster 2. Jika dilihat dari fungsionalitas protein CD4, skor yang dihasilkan dari perhitungan berbanding lurus dengan bukti fungsionalitas protein dalam tubuh yang merupakan pertahanan terdepan dalam mendeteksi adanya pathogen yang masuk dan menyerang bagian selaput otak atau sumsum tulang belakang yang menyebabkan meninges.

REFERENSI

1. A. Kohil, S. Jemmieh, M. K. Smatti, and H. M. Yassine, “Viral meninges: an overview,” *Arch Virol*, vol. 166, no. 2, pp. 335–345, Jan. 2021, doi: 10.1007/S00705-020-04891-1.
2. N. Principi and S. Esposito, “Bacterial meninges: new treatment options to reduce the risk of brain damage,” *Expert Opin Pharmacother*, vol. 21, no. 1, pp. 97–105, Jan. 2019, doi: 10.1080/14656566.2019.1685497.
3. S. Zipursky, M. H. Djingarey, J.-C. Lodjo, L. Olodo, S. Tiendrebeogo, and O. Ronveaux, “Benefits of using vaccines out of the cold chain: Delivering Meningitis A vaccine in a controlled temperature chain during the mass immunization campaign in Benin,” *Vaccine*, vol. 32, no. 13, pp. 1431–1435, Mar. 2014, doi: 10.1016/j.vaccine.2014.01.038.
4. S. Amiroch, M. I. Irawan, I. Mukhlash, M. H. Z. al Faroby, and C. A. Nidom, “Machine Learning for the Prediction of Antiviral Compounds Targeting Avian Influenza A/H9N2 Viral Proteins,” *Symmetry (Basel)*, vol. 14, no. 6, p. 1114, May 2022, doi: 10.3390/sym14061114.
5. H. V. Cook, N. T. Doncheva, D. Szklarczyk, C. von Mering, and L. J. Jensen, “Viruses.STRING: A virus-host protein-protein interaction database,” *Viruses*, vol. 10, no. 10, p. 519, Oct. 2018, doi: 10.3390/v10100519.
6. M. H. Z. al Faroby, H. N. Fadhilah, and F. H. Sembiring, “Identifikasi Interaksi Protein-Protein Meningitis Menggunakan ClusterONE dan Analisis Jaringan,” *Journal of Advances in Information and Industrial Technology*, vol. 4, no. 1, pp. 17–28, May 2022, doi: 10.52435/jaiit.v4i1.180.



7. J. S. Amberger, C. A. Bocchini, A. F. Scott, and A. Hamosh, “OMIM.org: Leveraging knowledge across phenotype-gene relationships,” *Nucleic Acids Res*, vol. 47, no. D1, pp. D1038–D1043, Jan. 2019, doi: 10.1093/nar/gky1151.
8. D. Szklarczyk *et al.*, “The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets,” *Nucleic Acids Res*, vol. 49, no. D1, pp. D605–D612, Jan. 2021, doi: 10.1093/nar/gkaa1074.
9. Y. Li and H. Wu, “A Clustering Method Based on K-Means Algorithm,” *Phys Procedia*, vol. 25, pp. 1104–1109, 2012, doi: 10.1016/J.PHPRO.2012.03.206.
10. K. P. Sinaga and M. S. Yang, “Unsupervised K-means clustering algorithm,” *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.
11. M. H. Z. al Faroby, M. I. Irawan, and N. N. T. Puspansih, “XGBoost and Network Analysis for Prediction of Proteins Affecting Insulin based on Protein Protein Interactions,” *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, vol. 4, no. 5, pp. 253–262, Nov. 2020, doi: 10.22219/kinetik.v5i4.1076.
12. S. Amiroch, M. H. Z. al Faroby, M. I. Irawan, I. Mukhlash, and A. C. Nidhom, “Analysis of protein-protein interaction to obtain significant protein in influenza virus type A/H9N2,” in *AIP Conference Proceedings*, Aug. 2022, vol. 2498, no. 1, p. 020021. doi: 10.1063/5.0083336.
13. E. Ragueneau, A. Shrivastava, J. H. Morris, N. del-Toro, H. Hermjakob, and P. Porras, “IntAct App: a Cytoscape application for molecular interaction network visualization and analysis,” *Bioinformatics*, vol. 37, no. 20, pp. 3684–3685, Oct. 2021, doi: 10.1093/BIOINFORMATICS/BTAB319.
14. C. F. A. Negre *et al.*, “Eigenvector centrality for characterization of protein allosteric pathways,” *Proc Natl Acad Sci U S A*, vol. 115, no. 52, pp. E12201–E12208, Dec. 2018, doi: 10.1073/PNAS.1810452115.





Gambar 4. Hasil perangkaian protein yang terdapat pada kluster 2.