

Target Baru Pengobatan Meningitis Berdasarkan Centrality Measure Jaringan Protein dan Self Organizing Map

Siti Amiroch¹, Mohammad Hamim Zajuli Al Faroby², Muhammad Dzulfikar Fauzi³

¹Department of Mathematics, Universitas Islam Darul 'ulum Lamongan, Indonesia.

^{2,3}School of Computing, Telkom University, Bandung, Indonesia.

e-mail: alfarobymhz@telkomuniversity.ac.id

Diajukan: 30 September 2024, Diperbaiki: 5 Nopember 2024, Diterima: 28 Nopember 2024

Abstrak

Meningitis adalah ancaman serius bagi kesehatan dengan konsekuensi berpotensi fatal. Pemahaman interaksi protein terkait kondisi kronis sangat krusial untuk pengembangan pengobatan yang efektif. Analisis secara *In Silico* dinilai memiliki keefektifan lebih besar karena melakukan simulasi melalui komputasi dan mencoba berbagai kemungkinan dengan biaya lebih murah. Penelitian ini bertujuan untuk menganalisis interaksi protein-protein yang berkaitan dengan Meningitis dengan teknik analisis kluster pada undirected graph. Metode yang diusulkan yakni algoritma Self Organizing Map (SOM) sebagai pengkluster. Algoritma ini mampu mengelompokkan data interaksi protein berbasis undirected graph. Data Protein yang terlibat pada penyakit Meningitis berasal dari OMIM. Dari data tersebut, protein-protein yang tergolong dalam locus gene dieksplorasi interaksinya sehingga menghasilkan data interaksi berupa undirected graph. Kombinasi centrality measure digunakan untuk rekayasa fitur pada data undirected graph. Kandidat protein utama berpotensi terletak pada model kluster 1 dengan nilai Silhouette score terbesar (0.359) dan Davies-Bouldin Index (1.667). Kluster tersebut memiliki 18 protein signifikansi tertinggi terhadap Meningitis. Dari hasil perankingan overall centrality diperoleh tiga protein signifikansi tertinggi yaitu CISH (3.921222), TNFSF10 (3.403541), dan ICAM3 (2.623702) yang berpotensi menjadi protein target. Protein CISH memiliki nilai skor overall centrality paling tinggi dibandingkan yang lain, sehingga protein CISH dimungkinkan bisa menjadi alternatif baru dalam treatment Meningitis.

Kata Kunci: *Self-organizing Map, Centrality Measure, Interaksi-interaksi antar Protein, Penanganan Meningitis.*

Abstract

Meningitis is a serious threat to health with potentially fatal consequences. Understanding protein interactions related to chronic conditions is crucial for the development of effective treatments. In silico analysis is considered to have greater effectiveness because it simulates through computation and tries various possibilities at a lower cost. This study aims to analyze protein-protein interactions related to Meningitis with cluster analysis techniques on undirected graphs. The proposed method is the Self Organizing Map (SOM) algorithm as a cluster. This algorithm can cluster undirected graph-based protein interaction data. Protein data involved in Meningitis disease comes from OMIM. From this data, proteins belonging to the gene locus are explored for their interactions, resulting in interaction data in the form of an undirected graph. The combination of centrality measure is used for feature engineering on undirected graph data. The main protein candidates are potentially located in the Cluster 1 model with the largest silhouette score (0.359) and Davies-Bouldin Index (1.667). The cluster has 18 proteins with the highest significance to Meningitis. From the overall centrality ranking results, the three highest significance proteins are CISH (3.921222), TNFSF10 (3.403541), and ICAM3 (2.623702) which have the potential to become Meningitis target proteins. CISH protein has the highest overall centrality score value compared to the others, so CISH protein may be a new alternative in the treatment of Meningitis.

Keywords: *Self-organizing Map, Centrality Measurement, Protein-Protein Interaction, Meningitis Treatment.*

1 Pendahuluan

Pusat syaraf tubuh manusia terletak pada otak. Kelainan, penyakit maupun luka yang menyasar pada bagian otak menjadi hal yang sangat fatal. Salah satu penyakit yang berakibat fatal dan berkaitan langsung dengan otak adalah Meningitis. Penyakit ini menyerang selaput otak yang disebabkan oleh patogen yang menginfeksi seperti bakteri, jamur, maupun virus. Meningitis merupakan keadaan gawat darurat medis yang harus mendapat prioritas pengobatan. Hal ini dikarenakan kasus ringan pada Meningitis dapat mengakibatkan kecacatan permanen seperti hilangnya pendengaran [1]. Sementara pada kasus berat, dapat menyebabkan kematian. Treatment yang dilakukan untuk pengobatan Meningitis sangat tergantung dari jenis Meningitis. Selain itu, dalam proses pengobatan dibutuhkan pengetahuan terkait protein target apa yang akan menjadi tempat senyawa obat melekat [2]. Pengobatan secara kimiawi membutuhkan prediksi yang tepat dalam menentukan protein target Meningitis yang mampu mengikat senyawa [3]. Protein dalam tubuh akan mengeluarkan sinyal darurat terkait kondisi tubuh ketika mengalami infeksi. Patogen dari luar penyebab Meningitis bisa jadi menyerang protein target tersebut untuk dijadikan sebagai inang [4]. Pengobatan yang ada untuk Meningitis tergantung jenis patogen yang menginfeksi. Rata-rata obat-obatan untuk mengatasi Meningitis berupa antibiotik maupun antiviral. Bentuk pengobatan Meningitis bergantung dengan protein target yang menjadi inang patogen. Sedangkan, untuk pencegahan yang ada hanya untuk vaksinasi Meningitis bakteri seperti vaksin Hib, meningokokus, dan pneumokokus[5]. Selanjutnya, analisis terkait protein-protein sangat penting dilakukan, agar dapat menemukan protein utama yang menjadi inang patogen Meningitis.

Protein-protein yang berkaitan dengan penyakit Meningitis dapat membentuk suatu koneksi antara protein satu dengan protein lainnya, yang biasa dikenal dengan interaksi protein-protein (PPI). Konektivitas antar protein dapat dianalisis untuk mendapatkan protein vital yang menjadi target utama dari patogen Meningitis. Konektivitas antar protein direpresentasikan sebagai suatu graf tidak berarah, dimana protein sebagai node (n) dan konektivitasnya sebagai verteks (v). Analisis terhadap konektivitas protein ini dapat mengadopsi metode centrality measure pada graf. Metode ini dinilai dapat menghitung bobot sentral dari suatu graf dimana dapat dianalisis kecenderungan dari setiap node. Kecenderungan node tersebut mengartikan bahwa seberapa besar node tertentu mempengaruhi node yang lain. Perhitungan bobot sentralitas ini tergantung dengan metode pembobotan yang digunakan. Metode pembobotan bisa menggunakan jarak terdekat, derajat graf, konektivitas antar node, dan pembobotan yang lainnya yang berbasis analisis graf. Centrality measure dapat memberikan informasi yang sangat berharga tentang peran dan kepentingan relatif dari setiap elemen dalam jaringan hubungan antara individu-individu atau gejala-gejala penyakit [6]. Dalam kasus Meningitis, sentralitas dapat membantu dalam

mengidentifikasi elemen-elemen yang memiliki pengaruh besar dalam penyebaran atau perkembangan penyakit ini. Sentralitas dapat dilihat dari berbagai aspek, seperti sentralitas degree (jumlah koneksi langsung), sentralitas closeness (kedekatan dengan elemen-elemen lain), dan sentralitas betweenness (keterlibatan dalam jalur-jalur komunikasi atau penyebaran) [4]. Dengan menganalisis sentralitas, kita dapat mengidentifikasi individu yang berperan sebagai titik fokus utama penyebaran Meningitis, sehingga upaya pencegahan atau penanganan dapat diarahkan dengan lebih efektif. Dengan demikian, centrality measure memberikan pandangan yang lebih holistik dan kontekstual terhadap dinamika penyakit Meningitis dalam suatu populasi atau jaringan kesehatan.

Metode sentralitas digunakan sebagai alat untuk rekayasa fitur pada graf yang terbentuk dari jaringan interaksi protein. Fitur-fitur tersebut bermanfaat untuk analisis lebih lanjut menggunakan pendekatan komputasi. Pendekatan komputasi berupa pembelajaran mesin yang pernah dilakukan dengan memanfaatkan ClusterOne pada Cytoscape [7]. ClusterOne mengelompokkan node-node pada graf berdasarkan kedekatan dengan yang lain. Pendekatan dengan pembelajaran mesin memiliki keuntungan dalam efisiensi dan pencirian yang mendetail sebagai upaya menganalisis node dominan pada graf. Akan tetapi, data graf yang direkayasa dengan metode sentralitas tidak memiliki kebenaran dasar. Sehingga teknik pembelajaran mesin yang memungkinkan adalah metode klaster (unsupervised learning). Pengelompokan dengan pendekatan jarak memungkinkan untuk dilakukan, karena pada graf ada biaya (jarak) yang dapat dihitung [8]. Metode sentralitas memberikan solusi lain, dimana graf dapat direkayasa fiturnya agar ciri-ciri dari setiap node dapat diketahui. Pendekatan neural network sangat cocok dalam hal penyeleksian fitur data. Sehingga, metode pengelompokan berbasis neural network, seperti SOM memungkinkan hasil lebih baik.

Self-Organizing Maps (SOM) adalah pendekatan yang penting dan inovatif dalam analisis Meningitis. Penyakit Meningitis melibatkan dinamika kompleks yang melibatkan interaksi antar individu dan manifestasi penyakit dalam populasi. Metode tradisional seringkali tidak mampu mengatasi kompleksitas ini. Oleh karena itu, penggunaan pembelajaran mesin dapat membuka peluang baru untuk mengeksplorasi pola dan hubungan yang sulit ditemukan menggunakan metode analisis tradisional. Self-Organizing Maps (SOM) dipilih sebagai alat analisis karena kemampuannya mengungkap struktur non linier dan pola tersembunyi dalam data. SOM dapat membentuk representasi spasial dari hubungan antar individu dalam jaringan dan gejala penyakit, memungkinkan peneliti untuk memvisualisasikan dan menafsirkan struktur kompleks dengan lebih efektif [9]. SOM dapat digunakan untuk mengidentifikasi fokus utama penyebaran Meningitis dan mengklasifikasikan pola interaksi yang dapat mempengaruhi penularan penyakit. Penerapan pembelajaran mesin, khususnya pada SOM, juga menawarkan manfaat dalam hal

kemampuan beradaptasi terhadap data baru dan kompleksitas yang berubah seiring waktu. Hal ini meningkatkan akurasi analisis dan memberikan wawasan yang lebih dinamis mengenai perkembangan Meningitis pada populasi. Oleh karena itu, penggunaan machine learning, khususnya pendekatan SOM, diharapkan dapat memberikan kontribusi yang signifikan terhadap pemahaman dan upaya kita dalam menangani Meningitis secara lebih holistik dan spesifik.

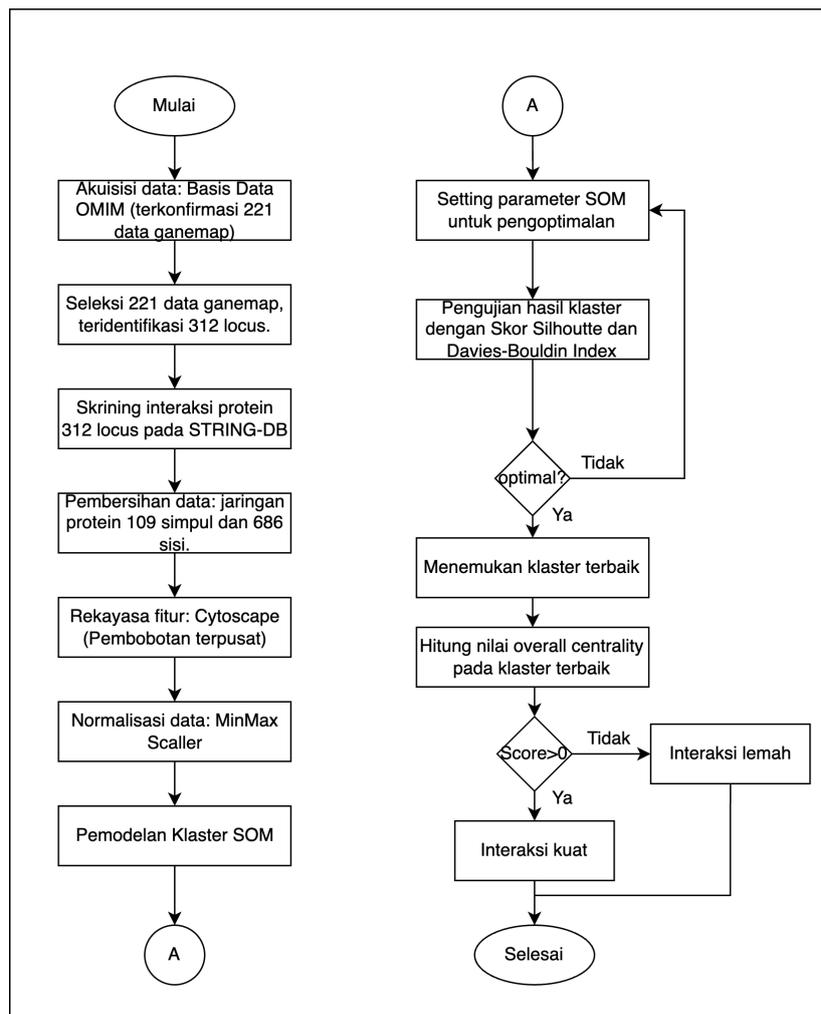
Analisis jaringan protein menuntut ketelitian dalam mencirikan protein-protein utama yang terlibat dalam penyakit Meningitis. Karenanya, penelitian ini bertujuan untuk menganalisis protein-protein Meningitis yang ditransformasikan kedalam bentuk undirected graf. Rekayasa fitur data graf memanfaatkan sentralitas pada node-node graf dan dianalisis dengan mengelompokkan menggunakan SOM. Penelitian ini berkontribusi dalam penemuan kandidat protein utama yang terdampak Meningitis dengan pendekatan komputasi. Hasil dari penelitian ini dalam bentuk kandidasi yang memungkinkan untuk diteliti lebih lanjut di bidang keilmuan kedokteran. Analisis konektivitas pada interaksi protein juga bermanfaat dalam mengetahui dampak protein sekunder maupun tersier yang mungkin saja bisa terdampak apabila protein utama mengalami gangguan akibat Meningitis. Akan tetapi, penelitian ini terbatas pada protein-protein yang dinyatakan berkaitan dengan Meningitis saja. Meskipun memungkinkan protein tersebut berkaitan dengan kondisi penyakit lain, tetapi hal tersebut di luar cakupan penelitian yang dikerjakan. Hasil analisis penelitian ini memungkinkan tenaga medis dapat memprediksi kondisi komplikasi apabila Meningitis yang dialami pasien pada level kronis.

2 Metode Penelitian

Penelitian ini dimulai dengan pengumpulan dataset yang berisi informasi tentang analisis interaksi protein-protein yang relevan dengan jaringan biologi, khususnya yang berkaitan dengan Meningitis. Proses penelitian secara singkat dan menyeluruh seperti pada Gambar 1. Dataset dikumpulkan dari basis data *Online Mendelian Inheritance in Man* (OMIM). Pencarian data pada basis data menggunakan kata kunci “Meningitis” untuk mendapatkan data *genemap* yang didalamnya mengandung informasi *locus* protein yang berkaitan dengan Meningitis[10]. Hasil pencarian pada OMIM mendapatkan 221 *genemap*, dimana terdapat 312 *locus* yang unik. Untuk mendapatkan bentuk interaksi protein pada setiap locus, pencarian data interaksi dilakukan pada Mbasis data STRING-DB [11]. Untuk mewakili koneksi protein yang terlibat, maka ditetapkan ambang batas agar data yang diperoleh harus memberikan landasan yang kuat untuk menganalisis pola dan karakteristik khusus dari jaringan biologi yang terlibat dalam kondisi ini. Hasil pencarian dari STRING-DB didapatkan 109 simpul yang mencerminkan protein dan 686 tepian yang

mencerminkan interaksi antara protein-protein. Hasil pencarian jaringan interaksi tersebut dianalisis menggunakan graf untuk mendapatkan nilai centrality measure pada setiap protein. Cytoscape digunakan untuk memahami struktur jaringan protein yang berupa graf, sehingga mendapatkan nilai centrality measure.

Setelah pengumpulan data, langkah berikutnya adalah melakukan Analisis Data Eksploratif (EDA) untuk memahami lebih dalam tentang dataset. Proses ini melibatkan pemeriksaan korelasi antar variabel dan identifikasi outlier. Pemahaman korelasi antar kolom menjadi kunci untuk pemilihan fitur yang akurat dalam pengelompokan nantinya. Meskipun beberapa kolom menunjukkan korelasi yang moderat, keputusan diambil untuk mempertahankan semua variabel pada tahap ini agar informasi yang relevan tetap dapat digunakan dalam proses pengelompokan. Sebelum dilakukan kluster SOM, normalisasi data dilakukan untuk menyesuaikan skala fitur dalam rentang yang seragam yaitu skala fitur menjadi rentang 0 hingga 1, hal ini untuk memastikan semua variabel memiliki pengaruh yang setara dalam proses klusterisasi.



Gambar 1. Flowchart alur pelaksanaan metodologi penelitian

Penerapan Self-Organizing Map (SOM) melibatkan penggunaan library `sklearn-som` dalam bahasa pemrograman Python. Proses pengklasteran awal menggunakan parameter $m = 3$ (jumlah baris), $n = 3$ (jumlah kolom), dan $dim = 9$ (dimensi fitur untuk setiap neuron dalam peta SOM). Proses pelatihan dilakukan menggunakan data yang telah dinormalisasi, memungkinkan model untuk menangkap pola-pola yang ada dalam jaringan protein. Pertama, melakukan Inisialisasi Bobot Neuron. Setiap neuron pada peta SOM memiliki vektor bobot yang merepresentasikan lokasi di ruang input. Bobot ini diinisialisasi secara acak atau menggunakan teknik seperti Principal Component Analysis (PCA) [12]. Setelah itu, menghitung nilai pada setiap neuron dengan sistem Kompetisi. Pada setiap iterasi, data input dikompetisikan dengan neuron-neuron untuk menentukan neuron mana yang paling mendekati data input. Bobot neuron pemenang dan tetangganya diperbarui untuk mendekati nilai data input. Langkah-langkah tersebut diulang hingga konvergensi atau sampai iterasi mencapai jumlah maksimal yang ditentukan.

Setelah pelatihan, model SOM digunakan untuk melakukan prediksi dan mengklasifikasikan setiap protein ke dalam neuron yang sesuai pada peta SOM. Evaluasi dilakukan untuk memeriksa kualitas kluster yang dihasilkan oleh SOM. Metrik evaluasi termasuk Silhouette Score dan Davies-Bouldin Index (DBI) yang memberikan gambaran tentang seberapa baik data dikelompokkan [13]. Setelah proses pengelompokan menggunakan SOM, visualisasi dilakukan melalui scatter plot dengan menggunakan PCA untuk mereduksi dimensi. Reduksi dimensi ini mempermudah interpretasi dan visualisasi kelompok hasil pengelompokan. Interpretasi kelompok hasil dari SOM melibatkan identifikasi karakteristik unik atau peran fungsional dari protein-protein dalam setiap kelompok. Misalnya, kluster 0 mungkin terdiri dari protein-protein dengan fungsi struktural tertentu, sementara kluster 1 dapat menunjukkan keterkaitan dalam jalur biologis tertentu. Interpretasi lebih lanjut tentang makna biologis dari setiap kelompok diperlukan untuk memahami implikasinya dalam konteks Meningitis.

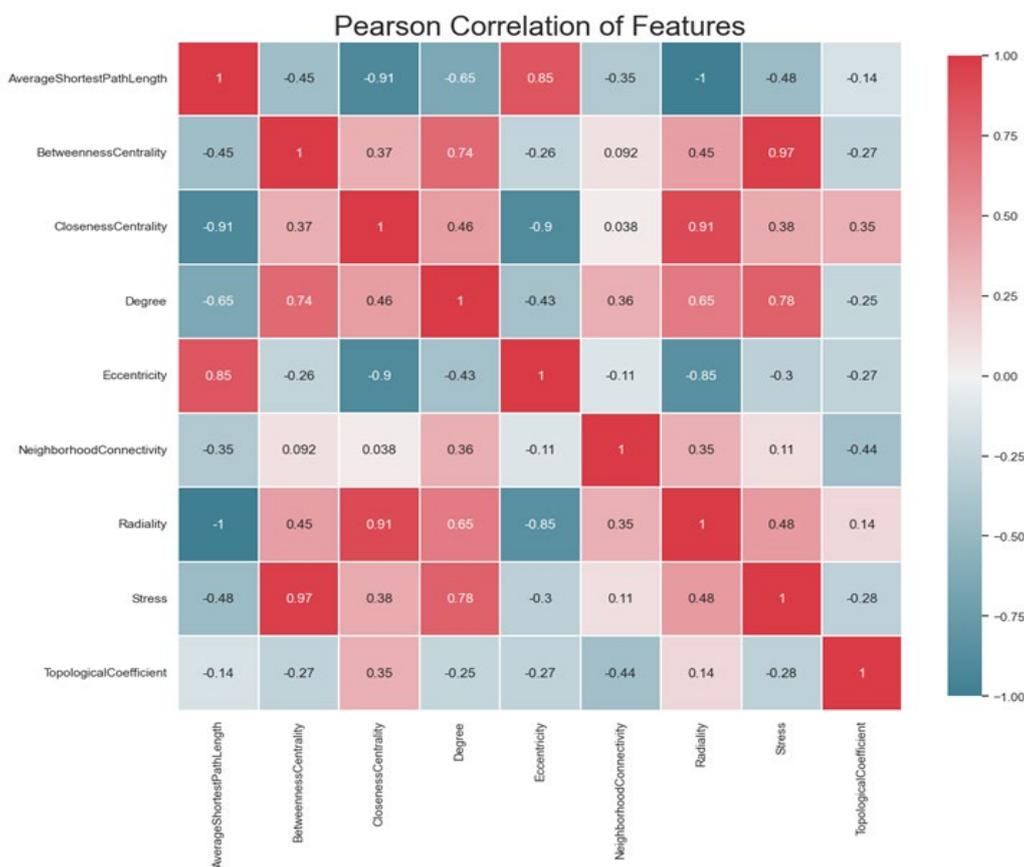
3 Hasil dan Pembahasan

3.1 Analisis dan Eksplorasi Data

Pemahaman dataset adalah langkah krusial untuk menentukan langkah-langkah selanjutnya dalam penelitian ini. Eksplorasi data memberikan gambaran yang cukup realistis terkait kondisi data dan hubungan antara fitur data. Dataset terdiri dari Sembilan fitur, dimana setiap fitur tersebut merupakan karakteristik jaringan dalam bentuk graf. Kesembilan fitur pada jaringan protein berbentuk graf tersebut adalah *average shortest path length* (rata-rata seluruh panjang jalur), *betweenness centrality* (centrality node berdasarkan lintasan terpendek), *closeness centrality* (jarak

rata-rata dari node awal ke semua node lain), *degree* (derajat node), *eccentricity* (jumlah subgraf yang mempunyai himpunan titik yang sama), *neighborhood connectivity* (himpunan semua tetangga dari node graf), *radiality* (radialitas node), *stress*, dan *topological coefficient* (nilai kontribusi node). Fitur-fitur ini dinilai mampu menjadi perekayasa sintesis untuk mencirikan kondisi keterkaitan node satu dengan yang lainnya pada suatu jaringan [6].

Salah satu eksplorasi data untuk mengetahui pengaruh antar fitur dengan pendekatan statistik dengan pemeriksaan korelasi. Pada Gambar 2 di bawah ini merupakan hasil uji korelasi di antara setiap fitur data yang terdiri dari sembilan fitur. Memahami korelasi antar kolom sangat penting untuk melakukan pemilihan fitur yang tepat dalam proses klusterisasi. Korelasi yang tinggi atau rendah antar kolom dapat memiliki dampak signifikan pada hasil klaster, dan membimbing pemilihan fitur yang lebih baik [14].



Gambar 2. Correlation HeatMap setiap fitur data

Hasil uji korelasi memperlihatkan bahwa beberapa kolom memiliki hubungan yang cukup tinggi dengan kolom lainnya, seperti Betweenness Centrality yang memiliki korelasi sebesar 97% dengan Stress, Closeness Centrality dengan Radiality sebesar 91%, dan masih banyak korelasi lainnya. Meskipun beberapa kolom terlihat memiliki korelasi yang tidak terlalu tinggi, berkisar antara 74%, 78%, 85%, dan sebagainya. Hampir semua fitur memiliki suatu nilai uji korelasi tinggi terhadap suatu fitur lainnya. Sehingga, berdasarkan uji korelasi ini, semua fitur data memiliki

keterikatan dengan yang lain, meskipun tidak semua, akan tetapi selalu ada pada salah satu fitur. Hasil uji tersebut menjadi landasan bahwa semua fitur digunakan pada penelitian ini. Fitur yang memiliki korelasi kuat satu sama lain seperti Betweenness Centrality dengan Stress atau Average Shortest Path Length dengan Closeness Centrality dapat dipertimbangkan untuk analisis jaringan yang lebih dalam, karena fitur-fitur ini saling terkait dalam menggambarkan sifat struktur jaringan. Fitur dengan korelasi rendah atau negatif mungkin menangkap aspek jaringan yang berbeda, dan dapat digunakan untuk memberikan perspektif berbeda dalam memahami struktur jaringan. Keterkaitan node dalam jaringan dipengaruhi oleh beberapa metrik ini, yang dapat membantu dalam pemodelan, klasifikasi, atau prediksi sifat jaringan lebih lanjut. Secara keseluruhan, Gambar 2 menunjukkan bahwa terdapat hubungan linier yang signifikan antar beberapa metrik jaringan yang penting untuk dipahami dalam analisis atau pengambilan keputusan berbasis jaringan. Penting untuk dicatat bahwa penghapusan kolom yang memiliki korelasi tinggi dapat berdampak pada hasil kluster dan interpretasi yang benar dari struktur data [15]. Oleh karena itu, pada tahap ini, peneliti memilih untuk mempertahankan semua kolom dalam analisis untuk memastikan bahwa seluruh informasi yang relevan dapat digunakan dalam proses klusterisasi.

Tabel 1. Dataset hasil dari rekayasa fitur menggunakan centrality measure pada graf.

Nama Protein	Average Shortest Path Length	Betweenness Centrality	Closeness Centrality	Degree	Eccentricity	Neighborhood Connectivity	Radiality	Stress	Topological Coefficient
ARPC1B	3.0638	0.000000	0.326389	3	5	10.3333	0.963146	0	0.607843
NCKAPIL	2.5851	0.005364	0.386831	7	4	11.8571	0.971695	336	0.296429
...
PTGIR	2.5851	0.000000	0.386831	1	4	46.0	0.971695	0	0.000000
SAMD9	3.4787	0.000000	0.287462	1	5	7.0	0.955737	0	0.000000

Penerapan Self-Organizing Map (SOM) dalam proses kluster membutuhkan normalisasi data karena sensitivitas SOM terhadap skala variabel. Proses normalisasi data ini menggunakan MinMaxScaler. Pada Tabel 1 terdapat data asli sebelum dilakukan normalisasi, sedangkan Tabel 2 merupakan hasil normalisasi yang dapat membantu mencegah dominasi variabel dengan skala besar terhadap data dengan skala kecil, menjaga keseimbangan dan memastikan setiap variabel memiliki rentang nilai yang seragam. Nilai interval data setelah normalisasi berada pada selang $[0,1]$, artinya nilai minimum setiap kolom adalah 0 dan nilai maksimumnya adalah 1. Dengan normalisasi, SOM dapat mempertahankan struktur intrinsik dalam data dengan tepat, mencapai konvergensi yang lebih baik selama pelatihan, dan tidak terlalu dipengaruhi oleh unit pengukuran spesifik.

Tabel 2. Bentuk dataset setelah proses normalisasi data

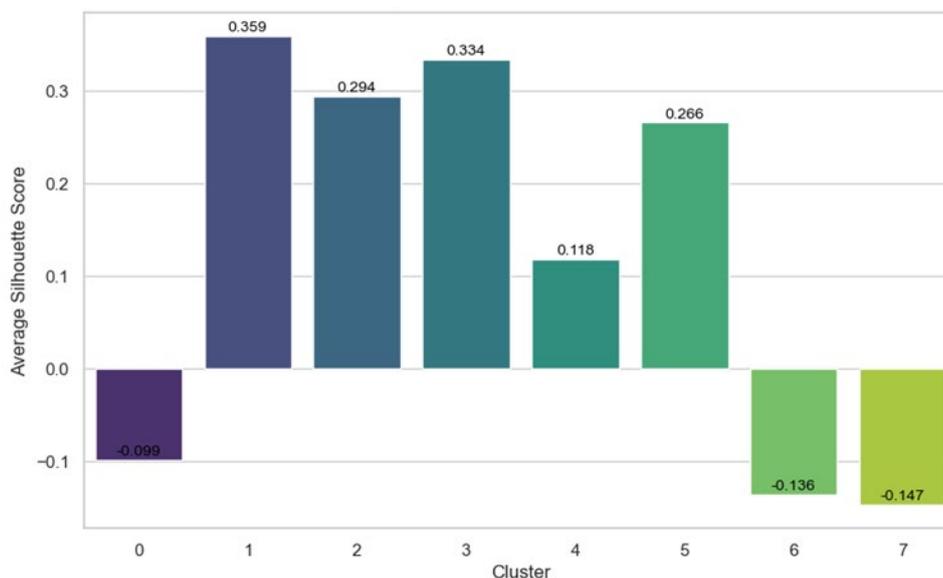
Nama Protein	Average Shortest Path Length	Betweenness Centrality	Closeness Centrality	Degree	Eccentricity	Neighborhood Connectivity	Radiality	Stress	Topological Coefficient
ARPC1B	0.788618	0.000000	0.068993	0.036364	1.00	0.189394	0.211382	0.000000	0.607843
NCKAP1L	0.605691	0.044252	0.152531	0.109091	0.75	0.224026	0.394309	0.036818	0.296429
...
PTGIR	0.605691	0.000000	0.152531	0.000000	0.75	1.000000	0.394309	0.000000	0.000000
SAMD9	0.947154	0.000000	0.015191	0.000000	1.00	0.113636	0.052846	0.000000	0.000000

Dalam penelitian ini, penerapan Self-Organizing Map (SOM) untuk kluster pada data yang telah diskalakan dengan memanfaatkan library Python, yaitu `sklearn-som`. Proses pemodelan data melibatkan percobaan beberapa kali. Hasil terbaik dari beberapa kali percobaan yang digunakan untuk membangun model yakni $m = 2$, $n = 4$, dan $dim = 9$. Hasil prediksi ini menghasilkan prediksi 8 kluster pada keseluruhan data. Hasil klusterisasi ini menjadi dasar untuk analisis lebih lanjut terkait struktur dan pola dalam data yang diamati. Hasil kluster berdasarkan SOM terdapat pada Gambar 3. Grafik ini menunjukkan distribusi elemen dalam setiap kluster sebagai hasil dari algoritma SOM. Kluster 1 merupakan kluster terbesar, sementara Kluster 2 adalah kluster terkecil, dengan perbedaan jumlah elemen di antara kluster lainnya. Hasil klusterisasi menunjukkan distribusi elemen yang tidak seimbang. Sebagai contoh, kluster 1 memiliki elemen terbanyak (18 elemen), sedangkan kluster 2 memiliki elemen paling sedikit (8 elemen). Ketidakseimbangan dalam distribusi ini bisa menjadi indikasi bahwa SOM berhasil mengidentifikasi beberapa kluster yang jauh lebih dominan atau lebih terpisah dari yang lain. Hal ini bisa berarti bahwa kluster 1 mungkin mengandung data yang lebih mirip satu sama lain, atau bahwa sebagian besar data berada di area yang dekat pada ruang vektor fitur yang digunakan oleh SOM [16].

3.2 Evaluasi Model SOM

Evaluasi model unsupervised learning, seperti pada Self-Organizing Map (SOM), umumnya melibatkan penggunaan metrik seperti Silhouette Score dan Davies-Bouldin Index (DBI). Tujuan dari evaluasi ini adalah untuk mendapatkan pemahaman tentang kemampuan model dalam memetakan objek ke dalam kluster, sejauh mana kluster terpisah, dan seberapa baik struktur kluster tersebut. Metrik evaluasi ini memberikan pandangan tentang kualitas pembentukan kluster dan kemampuan model untuk menggambarkan struktur intrinsik dalam data tanpa adanya label ground truth [17]. Dalam konteks hasil evaluasi kluster yang diberikan (DBI: 1.6662), nilai tersebut mengindikasikan bahwa kluster-kelompok cenderung memiliki jarak yang memadai satu sama lain dan memiliki ukuran yang seimbang. Nilai DBI mendekati 0 mencerminkan pembentukan kluster yang sangat baik dan terpisah, sementara nilai yang tinggi dapat menunjukkan tumpang tindih antar kluster yang dapat merugikan kualitas kluster. Dengan nilai DBI sekitar 1.6662, dapat

disimpulkan bahwa hasil klaster ini menunjukkan pembentukan klaster yang baik dan ukuran yang seimbang, menggambarkan kualitas klaster yang memuaskan secara umum.



Gambar 3. Rata-rata pengujian nilai sillhouete pada setiap klaster

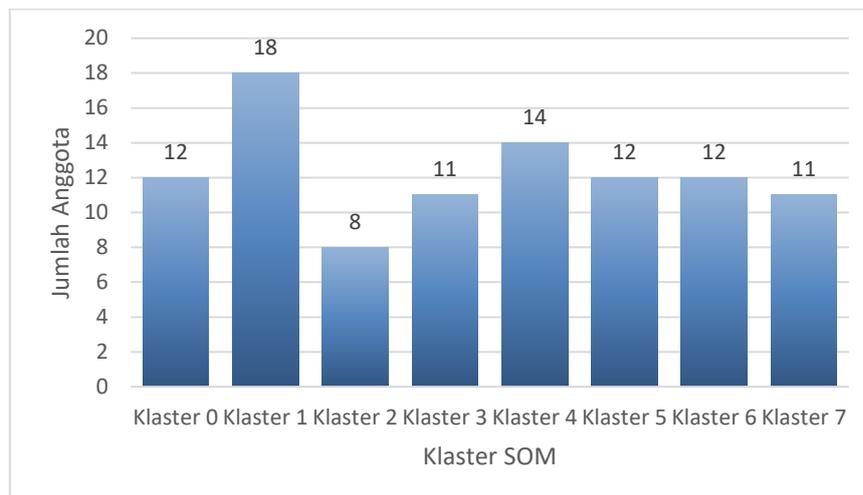
Selain DBI, Silhouette score juga memiliki peran penting dalam menilai kinerja model. Silhouette score adalah metrik yang digunakan untuk mengevaluasi kualitas pengelompokan dengan mengukur seberapa mirip suatu titik data dengan klasternya sendiri dibandingkan dengan klaster lain. Ini membantu mengevaluasi sejauh mana setiap data sesuai dengan klaster tempat mereka berada, memberikan wawasan tentang kecocokan setiap sampel dengan klaster yang telah dibentuk. Semakin tinggi Silhouette score rata-rata suatu klaster, semakin baik kohesi dan pemisahannya dari klaster lain. Skor yang mendekati 1 menunjukkan bahwa titik-titik data dalam klaster tersebut dikelompokkan dengan baik dan berbeda dari klaster lainnya. Sebaliknya, skor yang mendekati -1 menunjukkan bahwa titik data mungkin salah klasifikasi atau klasternya tidak terdefinisi dengan baik. Klaster dengan skor siluet rata-rata negatif menunjukkan adanya outlier atau titik data yang tidak cocok dengan klaster mana pun [18].

Berdasarkan hasil evaluasi Silhouette score pada Gambar 3, klaster 0 menunjukkan Silhouette score sebesar -0.098812, mengindikasikan adanya tumpang tindih dan ketidakjelasan dalam pemisahan objek di dalamnya. Mungkin diperlukan peninjauan ulang terhadap parameter klaster atau preprocessing data untuk meningkatkan pemisahan. Sebaliknya, Klaster 1 memiliki Silhouette score paling besar yaitu 0.358929, menunjukkan pemisahan yang paling baik dan homogen antara objek dalam klaster tersebut. Klaster 2 hingga 5 menunjukkan pemisahan yang baik, homogen, dan mewakili kelompok yang terpisah. Meskipun Klaster 4 dan 5 memiliki pemisahan yang baik, namun menunjukkan sedikit tumpang tindih antar klaster. Sementara Klaster 6 dan 7 menunjukkan adanya tumpang tindih dan tidak jelas [19]. Hasil ini memberikan dasar

untuk perbaikan parameter kluster guna meningkatkan kualitas pemisahan kluster, sehingga pemahaman struktur dan perbedaan dalam jaringan protein Meningitis dapat lebih akurat.

3.3 Interpretasi Model SOM

Hasil pengelompokan dari model SOM, setiap klasternya memiliki jumlah elemen antara 11 hingga 14, yaitu Klaster 0, 3, 4, 5, 6, dan 7. Ini menunjukkan bahwa metode SOM dapat membagi data ke dalam kelompok yang relatif seimbang, meskipun terdapat beberapa kluster dominan dan minor. Klaster minor merupakan jumlah elemen paling sedikit. Ini bisa berarti bahwa Klaster minor mewakili kelompok data yang lebih spesifik atau outlier. Klaster ini mungkin mengandung data yang memiliki karakteristik unik yang jauh berbeda dibandingkan dengan kluster lainnya. Klaster 1 yang memiliki 18 elemen merupakan kluster dominan dalam hasil klusterisasi ini. Hal ini mungkin menunjukkan bahwa terdapat grup data yang memiliki karakteristik yang sangat mirip, yang dikelompokkan dengan jelas oleh SOM [20].

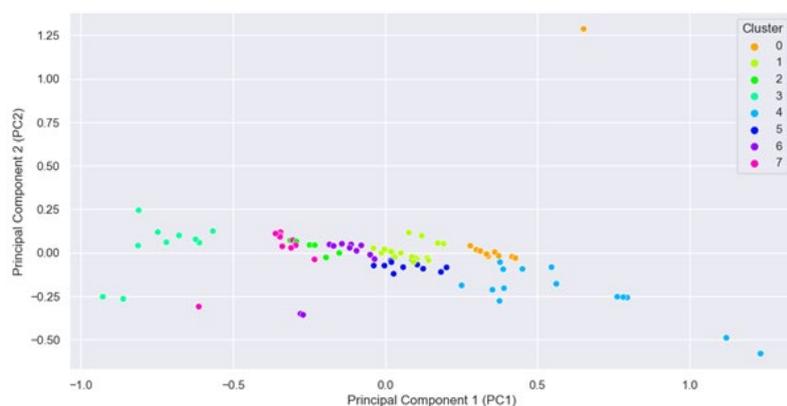


Gambar 4. Visualisasi jumlah anggota dan kluster yang terbentuk

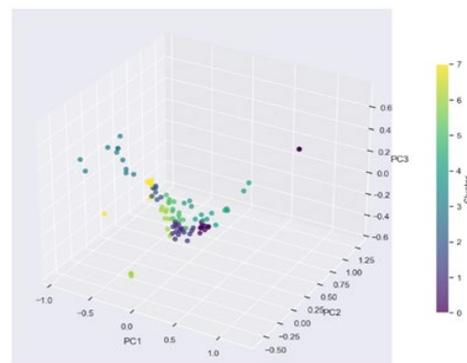
Berdasarkan kluster SOM yang dihasilkan, interpretasi kluster lebih mudah diidentifikasi dengan divisualisasikan menggunakan grafik scatter plot 2 dan 3 dimensi. Kondisi ini membutuhkan proses reduction kolom hingga tersisa 2 dan 3 kolom dari total 9 kolom. Metode Principal Component Analysis (PCA) dapat memaksimalkan reduksi dimensi tanpa harus kehilangan terlalu banyak informasi. Hal tersebut mempermudah kita untuk mengetahui apakah 2 kolom tersebut mewakili berapa persen informasi terhadap data keseluruhan berdasarkan parameter explained variance ratio. Hasil reduksi menunjukkan nilai kumulatif dari variansi yang dijelaskan oleh komponen-komponen ini yaitu sebesar 91.19%. Artinya, komponen reduksi secara kolektif dapat menjelaskan 91.19% variabilitas data. Hal ini menunjukkan bahwa komponen reduksi cukup untuk merepresentasikan hampir seluruh informasi dalam data dengan pengurangan dimensi yang cukup baik. Dalam PCA, semakin besar jumlah variansi yang dijelaskan oleh

sejumlah komponen yang sedikit, semakin efektif proses pengurangan dimensi. Dalam hal ini, hanya dengan tiga komponen, kita sudah dapat menjelaskan 91.19% variansi data, sehingga komponen-komponen ini sudah cukup representatif. Interpretasi dari klaster SOM dalam bentuk scatter plot 2D dan 3D disajikan pada Gambar 5 dan Gambar 6. Titik-titik data tampaknya membentuk kelompok yang berbeda. Hal ini menunjukkan bahwa algoritma klaster mampu mengidentifikasi kelompok titik data dengan karakteristik serupa. Beberapa klaster lebih terpisah dibandingkan yang lain. Hal ini mungkin menunjukkan bahwa titik data di klaster ini lebih berbeda satu sama lain dibandingkan dengan titik data di klaster lain. Kepadatan titik data dalam setiap klaster dapat bervariasi. Klaster yang padat mungkin menunjukkan konsentrasi titik data yang lebih tinggi di wilayah tersebut [16]. Hasil klaster juga menunjukkan terdapat beberapa titik data yang tampak jauh dari klaster utama. Hal ini dapat dianggap outlier atau anomali.

Berdasarkan hasil keluaran dari Self-Organizing Map (SOM), interpretasi hasil klaster untuk data menunjukkan pengelompokan berdasarkan karakteristik dan peran potensial dari protein-protein dalam masing-masing klaster. Kelompok klaster 0 seperti IRF8[21], CD19[22], dan NLRP3[23] memiliki peran dan karakteristik yang sejenis. Mereka memiliki potensi dalam regulasi respons imun terhadap Meningitis, terlibat dalam diferensiasi sel darah putih, dan modulasi inflamasi. Klaster 1, yang melibatkan CARD9[24], TIRAP[25], dan IRAK4[25], menunjukkan keterlibatan yang mungkin dalam jalur sinyal dan respon inflamasi, dengan peran potensial dalam aktivasi makrofag dan respons antibakteri. Klaster 2 dengan protein SASH3[26], C2[27], dan CFB[27], mungkin terlibat dalam modulasi aktivitas biologis, termasuk sistem komplemen dan respons imun. Klaster 3, yang melibatkan ARPC1B[28], DGKE[29], dan EFEMP2[30], menunjukkan potensi regulasi pada proses biologis terkait Meningitis, seperti regulasi respon inflamasi dan proliferasi sel. Sementara itu, klaster-kluster lain seperti klaster 4 (BTK, CD4, NFkB2), klaster 5 (WAS, DOCK8, IL2RG), klaster 6 (TNFRSF13B, PTPN22, MALT1), dan klaster 7 (NCKAP1L, IGLL1, C8A) juga memiliki peran khusus protein dalam respons imun dan pertahanan tubuh terhadap Meningitis[4]. Analisis lebih lanjut pada setiap klaster dan protein individu diharapkan akan memberikan pemahaman yang lebih mendalam dan kontekstual terhadap mekanisme kompleks dalam jaringan biologis yang terlibat dalam kondisi Meningitis. Informasi ini dapat memberikan landasan untuk pengembangan strategi diagnostik atau terapeutik yang lebih efektif.



Gambar 5. Plot 2D dari data hasil klaster SOM

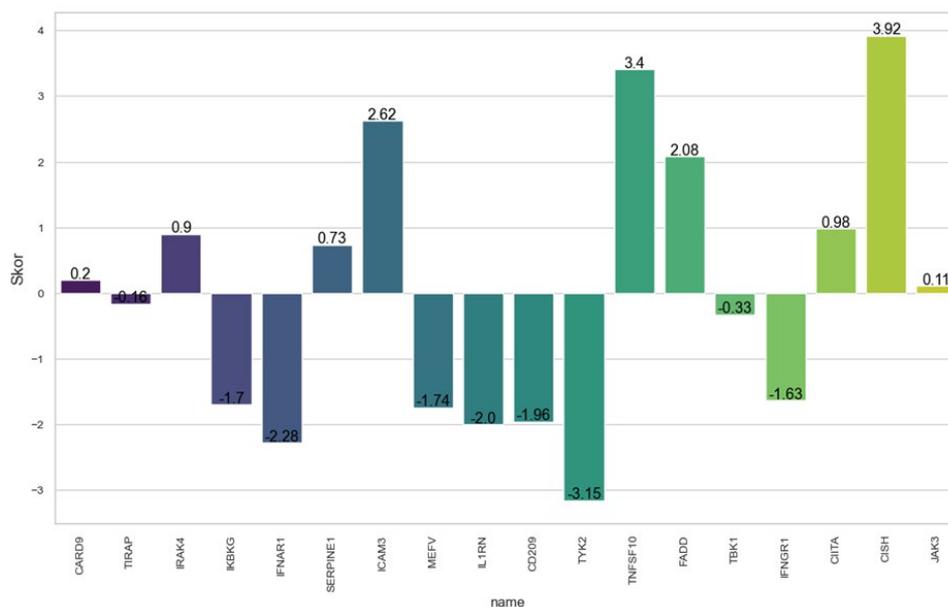


Gambar 6. Plot 3D dari data hasil klaster SOM

Setiap klaster dapat memiliki makna biologis yang unik atau mungkin memainkan peran tertentu dalam suatu proses biologis. Interpretasi lebih lanjut dapat dilakukan dengan merinci informasi tentang setiap protein dan analisis lebih lanjut terkait dengan konteks biologisnya.

3.4 Perhitungan *Overall Centrality*

Overall centrality adalah metrik dalam analisis jaringan yang mengukur pentingnya suatu node (simpul) dalam jaringan tersebut. Semakin tinggi skor overall centrality, semakin penting node tersebut dalam mempengaruhi atau terpengaruh oleh node lain dalam jaringan. Perhitungan skor overall centrality, tujuan utamanya adalah untuk mendapatkan pemahaman yang lebih baik tentang struktur dan variabilitas data awal, serta untuk mengidentifikasi apakah ada pola atau informasi yang signifikan yang dapat dijelaskan oleh komponen utama pertama. Perhitungan skor overall centrality melibatkan standarisasi fitur, dilanjutkan dengan PCA pada fitur-fitur yang telah distandarkan. Eigenvalue dan rasio varians dari komponen utama pertama diperoleh sebagai hasil dari analisis PCA tersebut sebagai koefisien untuk setiap fitur dalam komponen utama pertama. Skor overall centrality kemudian dihitung untuk setiap baris dalam Dataframe yang telah distandarkan sebagai kombinasi linear dari fitur-fitur yang telah distandarkan menggunakan koefisien komponen utama pertama. Node dengan skor positif memiliki pengaruh yang lebih besar dalam jaringan. Mereka cenderung menjadi pusat koneksi dan memiliki banyak hubungan dengan node lain. Node dengan skor negatif memiliki pengaruh yang lebih kecil atau bahkan cenderung menghambat aliran informasi dalam jaringan. Mereka mungkin berada di pinggiran jaringan atau memiliki sedikit koneksi [31].



Gambar 7. Nilai *overall centrality* pada protein-protein yang tergolong dalam kluster 1.

Dari Gambar 7 diketahui bahwa protein TNFSF10 (3.403541), ICAM3 (2.623702), dan CISH (3.921222) menunjukkan kontribusi positif yang besar, menandakan peran sentralnya dalam merepresentasikan variasi dalam data. Di sisi lain, protein-protein dengan nilai *overall centrality* negatif, seperti TYK2 (-3.152865) dan IFNAR1 (-2.279826), menunjukkan perbedaan yang signifikan dengan protein-protein lainnya dan memiliki skor negatif yang cukup besar. Hal ini menunjukkan bahwa kedua node ini mungkin memiliki peran yang lebih terbatas atau bahkan menghambat aliran informasi dalam jaringan. Gen/protein dengan skor tinggi memiliki kemungkinan memiliki peran penting dalam regulasi banyak proses biologis atau terlibat dalam banyak interaksi protein-protein[32]. Mereka bisa menjadi target yang baik untuk studi lebih lanjut atau pengembangan obat. Gen/protein dengan skor rendah, memungkinkan memiliki peran yang lebih spesifik atau terbatas dalam sel. Mutasi pada gen-gen ini mungkin memiliki dampak yang lebih kecil dibandingkan dengan mutasi pada gen dengan skor tinggi.

3.5 Diskusi

Protein-protein yang telah dihasilkan melibatkan sejumlah proses biologis yang berperan penting dalam regulasi respons seluler dan imun. CISH (Cytokine-inducible SH2-containing protein) termasuk dalam keluarga SOCS yang mengatur transduksi sinyal sitokin melalui jalur JAK-STAT5 [33]. Jalur ini merupakan salah satu jalur transduksi sinyal yang paling umum digunakan oleh sitokin untuk menginduksi ekspresi gen target. Ketika suatu sitokin berikatan dengan reseptornya, akan mengaktifkan kinase JAK (Janus kinase) yang kemudian akan memfosforilasi protein STAT (Signal Transducer and Activator of Transcription). STAT yang terfosforilasi akan dimerisasi dan masuk ke dalam nukleus untuk mengaktifkan transkripsi gen

target. CISH memiliki peran penting dalam mencegah respons imun berlebihan, mempertahankan homeostasis dan membantu meredakan respons imun dan mengembalikan tubuh ke kondisi normal. CISH juga terlibat dalam regulasi pertumbuhan sel. Mutasi atau disregulasi CISH dapat berkontribusi pada perkembangan penyakit seperti kanker [34]. TNFSF10 (Tumor necrosis factor ligand superfamily member 10) berfungsi sebagai sitokin yang dapat memicu apoptosis dan terlibat dalam keluarga tumor necrosis factor. TNFSF10 merupakan molekul yang menarik dengan potensi besar dalam terapi kanker dan penyakit lainnya. TNFSF10 berikatan dengan reseptor kematian spesifik (death receptors) pada permukaan sel, seperti DR4 (Death Receptor 4) dan DR5 (Death Receptor 5). Ikatan ini akan memicu pembentukan kompleks kematian yang mengaktifkan kaspase dan pada akhirnya menyebabkan apoptosis sel. TNFSF10 juga berperan dalam regulasi sistem imun dengan mempengaruhi aktivitas sel-sel imun seperti sel natural killer (NK) dan sel T sitotoksik [35]. Pemahaman yang lebih mendalam tentang mekanisme kerja TNFSF10 dan pengembangan strategi untuk mengatasi resistensi akan membuka jalan bagi pengembangan terapi yang lebih efektif di masa depan. Meskipun ICAM3 tidak memiliki informasi yang disediakan, pemahaman mengenai fungsi protein ini mendukung penelitian lebih lanjut dalam pemahaman mekanisme regulasi dan respons biologis sel.

Hasil skrining dengan model kluster memberikan kesimpulan bahwa protein CISH memiliki pengaruh paling dominan diikuti dengan TNFSF10 dan ICAM3 dalam jaringan protein yang terjadi pada kasus Meningitis. Hal ini dibuktikan dengan skor overall centrality yang lebih tinggi dibandingkan skor protein-protein lainnya pada Kluster 1. Melihat peran penting CISH dalam mengatur respons imun pada Meningitis, protein ini menjadi target yang menarik untuk pengembangan terapi baru. Beberapa potensi terapi yang melibatkan CISH antara lain, meningkatkan ekspresi CISH atau menghambat Protein yang Menghambat CISH. Meningkatkan ekspresi CISH, diharapkan dapat mengurangi peradangan yang terjadi pada Meningitis dan melindungi jaringan otak dari kerusakan. Menghambat Protein yang Menghambat CISH diharapkan dapat meningkatkan aktivitas CISH dan mengurangi peradangan. Meskipun potensi CISH sebagai target terapi sangat menjanjikan, masih diperlukan penelitian lebih lanjut untuk memahami secara mendalam mekanisme kerja CISH dalam Meningitis. Hasil penelitian ini dapat diuji coba pada model hewan. Melalui studi pada model hewan, dapat dipelajari efek dari manipulasi ekspresi CISH terhadap perkembangan Meningitis dan tingkat keparahan penyakit. Studi klinis pada manusia diperlukan untuk menguji keamanan dan efektivitas terapi yang menargetkan CISH pada pasien Meningitis.

4 Simpulan

Model kluster menggunakan SOM sangat efektif digunakan untuk melakukan screening pada data graf tidak berlabel. Hal ini dibuktikan dengan kemampuannya yang berhasil membangun delapan kluster dengan mempertahankan nilai DBI (DBI: 1.6662) dan Silhouette score cukup tinggi. Eksplorasi data sebelum pemodelan membantu menganalisis kondisi data, dimana menghasilkan korelasi yang linier antar fitur data. Kondisi ini membuktikan bahwa setiap fitur memiliki peran terhadap pembentukan model SOM yang dibangun. Kluster-kluster tersebut memberikan wawasan tentang keterlibatan protein-protein dalam respons imun, regulasi inflamasi, dan proses biologis terkait Meningitis. Kluster terbaik dipilih berdasarkan nilai Silhouette score tertinggi yakni pada kluster 1 (0.3589). Skrining potensial protein target difokuskan pada kluster 1, dimana kluster ini dianggap sebagai kluster inti yang kohesi dan pemisahannya dari kluster lain sangat baik berdasarkan nilai uji model. Protein potensial pada kluster 1 dapat dianalisis dengan menghitung nilai overall centrality. Protein-protein seperti CISH (3.92), TNFSF10 (3.40), dan ICAM3 (2.62) teridentifikasi memiliki peran penting dalam merepresentasikan variasi dalam data. Tiga protein ini memiliki nilai overall centrality tinggi dibandingkan calon protein-protein target lainnya. Kondisi ini membuat mereka cenderung menjadi pusat koneksi dan memiliki banyak hubungan dengan node lain. Skor tinggi memiliki kemungkinan memiliki peran penting dalam regulasi banyak proses biologis atau terlibat dalam banyak interaksi protein-protein. Mereka bisa menjadi target yang baik untuk studi lebih lanjut atau pengembangan obat. Penelitian ini berhasil memberikan pemahaman yang lebih baik tentang struktur dan karakteristik jaringan protein terkait Meningitis serta memberikan dasar untuk pengembangan strategi diagnostik atau terapeutik yang lebih efektif. Hasil penelitian ini berupa penemuan kandidat protein target untuk treatment Meningitis (CISH, TNFSF10, ICAM3) bisa menjadi alternatif dalam potensi treatment penyakit maupun obat baru untuk mencegah maupun mengatasi Meningitis.

5 Ucapan Terima Kasih

Terimakasih kepada Kementerian Pendidikan dan Kebudayaan Republik Indonesia berdasarkan No surat 0557/E5.5/AL.04/2023 yang telah memberikan pendanaan kepada penelitian ini. Semoga hasil penelitian ini memberikan manfaat untuk Masyarakat dan perkembangan ilmu pengetahuan.

6 Daftar Pustaka

- [1] A. Kohil, S. Jemmieh, M. K. Smatti, and H. M. Yassine, “Viral meningitis: an overview,” *Archives of Virology 2021 166:2*, vol. 166, no. 2, pp. 335–345, Jan. 2021, doi: 10.1007/S00705-020-04891-1.
- [2] N. Principi and S. Esposito, “Bacterial meningitis: 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] A. Pizzorno, B. Padey, O. Terrier, and M. Rosa-Calatrava, “Drug Repurposing Approaches for the Treatment of Influenza Viral Infection: Reviving Old Drugs to Fight Against a Long-Lived Enemy,” *Front Immunol*, vol. 10, no. MAR, p. 531, 2019, doi: 10.3389/FIMMU.2019.00531.
- [4] 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.
- [5] O. Hoffman and J. R. Weber, “Pathophysiology and treatment of bacterial meningitis,” *Ther Adv Neurol Disord*, vol. 2, no. 6, pp. 401–412, Nov. 2009, doi: 10.1177/1756285609337975.
- [6] 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*, AIP Publishing LLC/AIP Publishing, Aug. 2022, p. 020021. doi: 10.1063/5.0083336.
- [7] 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.
- [8] Z. Kang *et al.*, “Structured graph learning for clustering and semi-supervised classification,” *Pattern Recognit*, vol. 110, p. 107627, Feb. 2021, doi: 10.1016/j.patcog.2020.107627.
- [9] P. B. Lamichhane and W. Eberle, “Self-Organizing Map-Based Graph Clustering and Visualization on Streaming Graphs,” in *2022 IEEE International Conference on Data Mining Workshops (ICDMW)*, IEEE, Nov. 2022, pp. 706–713. doi: 10.1109/ICDMW58026.2022.00097.

- [10] 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.
- [11] 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.
- [12] B. Liu, M. Lai, J.-L. Wu, C. Fu, and A. Binaykia, “Patent analysis and classification prediction of biomedicine industry: SOM-KPCA-SVM model,” *Multimed Tools Appl*, vol. 79, no. 15–16, pp. 10177–10197, Apr. 2020, doi: 10.1007/s11042-019-7422-x.
- [13] F. Ros, R. Riad, and S. Guillaume, “PDBI: A partitioning Davies-Bouldin index for clustering evaluation,” *Neurocomputing*, vol. 528, pp. 178–199, Apr. 2023, doi: 10.1016/j.neucom.2023.01.043.
- [14] M. Wang, Z. Cang, and G.-W. Wei, “A topology-based network tree for the prediction of protein–protein binding affinity changes following mutation,” *Nat Mach Intell*, vol. 2, no. 2, pp. 116–123, Feb. 2020, doi: 10.1038/s42256-020-0149-6.
- [15] M. C. R. Melo, R. C. Bernardi, C. de la Fuente-Nunez, and Z. Luthey-Schulten, “Generalized correlation-based dynamical network analysis: a new high-performance approach for identifying allosteric communications in molecular dynamics trajectories,” *J Chem Phys*, vol. 153, no. 13, Oct. 2020, doi: 10.1063/5.0018980.
- [16] G. Sivagamasundari and P. Latha, “Clustering Protein Super Secondary Structure with Artificial Neural Network Based Self Organizing Map Machine Learning Approach Using In-Memory Computing Environment,” *J Comput Theor Nanosci*, vol. 17, no. 5, pp. 2039–2042, May 2020, doi: 10.1166/jctn.2020.8846.
- [17] I. F. Ashari, E. Dwi Nugroho, R. Baraku, I. Novri Yanda, and R. Liwardana, “Analysis of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index Evaluation on K-Means Algorithm for Classifying Flood-Affected Areas in Jakarta,” *Journal of Applied Informatics and Computing*, vol. 7, no. 1, pp. 89–97, Jul. 2023, doi: 10.30871/jaic.v7i1.4947.
- [18] A. M. Bagirov, R. M. Aliguliyev, and N. Sultanova, “Finding compact and well-separated clusters: Clustering using silhouette coefficients,” *Pattern Recognit*, vol. 135, p. 109144, Mar. 2023, doi: 10.1016/j.patcog.2022.109144.

-
- [19] M. Shutaywi and N. N. Kachouie, “Silhouette Analysis for Performance Evaluation in Machine Learning with Applications to Clustering,” *Entropy*, vol. 23, no. 6, p. 759, Jun. 2021, doi: 10.3390/e23060759.
- [20] L. M. P. Mariño and F. de A. T. de Carvalho, “Two weighted c-medoids batch SOM algorithms for dissimilarity data,” *Inf Sci (N Y)*, vol. 607, pp. 603–619, Aug. 2022, doi: 10.1016/j.ins.2022.06.019.
- [21] Z. Wu *et al.*, “The roles of IRF8 in nonspecific orbital inflammation: an integrated analysis by bioinformatics and machine learning,” *J Ophthalmic Inflamm Infect*, vol. 14, no. 1, pp. 1–19, Dec. 2024, doi: 10.1186/S12348-024-00410-4/FIGURES/15.
- [22] C. He *et al.*, “CD19 CAR antigen engagement mechanisms and affinity tuning,” *Sci Immunol*, vol. 8, no. 81, Mar. 2023, doi: 10.1126/sciimmunol.adf1426.
- [23] L. Y. El-Sharkawy, D. Brough, and S. Freeman, “Inhibiting the NLRP3 Inflammasome,” *Molecules*, vol. 25, no. 23, p. 5533, Nov. 2020, doi: 10.3390/molecules25235533.
- [24] C. Ji, Z. Yang, X. Zhong, and J. Xia, “The role and mechanism of CARD9 gene polymorphism in diseases,” *Biomed J*, vol. 44, no. 5, pp. 560–566, Oct. 2021, doi: 10.1016/j.bj.2020.12.006.
- [25] D. Lai *et al.*, “SARS-CoV-2 N Protein Triggers Acute Lung Injury via Modulating Macrophage Activation and Infiltration in in vitro and in vivo,” *J Inflamm Res*, vol. Volume 16, pp. 1867–1877, Apr. 2023, doi: 10.2147/JIR.S405722.
- [26] O. M. Delmonte *et al.*, “SASH3 variants cause a novel form of X-linked combined immunodeficiency with immune dysregulation,” *Blood*, vol. 138, no. 12, pp. 1019–1033, Sep. 2021, doi: 10.1182/blood.2020008629.
- [27] M. R. Zabihi, B. Farhadi, and M. Akhoondian, “Complement protein expression changes in various conditions of breast cancer: in-silico analyses—experimental research,” *Annals of Medicine & Surgery*, vol. 86, no. 9, pp. 5152–5161, Sep. 2024, doi: 10.1097/MS9.0000000000002216.
- [28] Y. Gamallat *et al.*, “ARPC1B Is Associated with Lethal Prostate Cancer and Its Inhibition Decreases Cell Invasion and Migration In Vitro,” *Int J Mol Sci*, vol. 23, no. 3, p. 1476, Jan. 2022, doi: 10.3390/ijms23031476.
- [29] J. Li *et al.*, “A novel compound heterozygous mutation in DGKE in a Chinese patient causes atypical hemolytic uremic syndrome,” *Hematology*, vol. 25, no. 1, pp. 101–107, Jan. 2020, doi: 10.1080/16078454.2020.1731969.

- [30] X. Shen, X. Jin, S. Fang, and J. Chen, “EFEMP2 upregulates PD-L1 expression via EGFR/ERK1/2/c-Jun signaling to promote the invasion of ovarian cancer cells,” *Cell Mol Biol Lett*, vol. 28, no. 1, p. 53, Jul. 2023, doi: 10.1186/s11658-023-00471-8.
- [31] F. Bloch, M. O. Jackson, and P. Tebaldi, “Centrality measures in networks,” *Soc Choice Welfare*, vol. 61, no. 2, pp. 413–453, Aug. 2023, doi: 10.1007/s00355-023-01456-4.
- [32] M. S. Usman, W. A. Kusuma, F. M. Afendi, and R. Heryanto, “Identification of Significant Proteins Associated with Diabetes Mellitus Using Network Analysis of Protein-Protein Interactions,” *Computer Engineering and Applications*, vol. 8, no. 1, pp. 41–52, Feb. 2019.
- [33] W. Naser, S. Maymand, D. Dlugolenski, F. Basheer, and A. C. Ward, “The Role of Cytokine-Inducible SH2 Domain-Containing Protein (CISH) in the Regulation of Basal and Cytokine-Mediated Myelopoiesis,” *Int J Mol Sci*, vol. 24, no. 16, p. 12757, Aug. 2023, doi: 10.3390/ijms241612757.
- [34] S. Maymand *et al.*, “Role of Cytokine-Inducible SH2 Domain-Containing (CISH) Protein in the Regulation of Erythropoiesis,” *Biomolecules*, vol. 13, no. 10, p. 1510, Oct. 2023, doi: 10.3390/biom13101510.
- [35] A. Yadav *et al.*, “Early transcriptomic host response signatures in the serum of dengue patients provides insights into clinical pathogenesis and disease severity,” *Sci Rep*, vol. 13, no. 1, p. 14170, Aug. 2023, doi: 10.1038/s41598-023-41205-2.