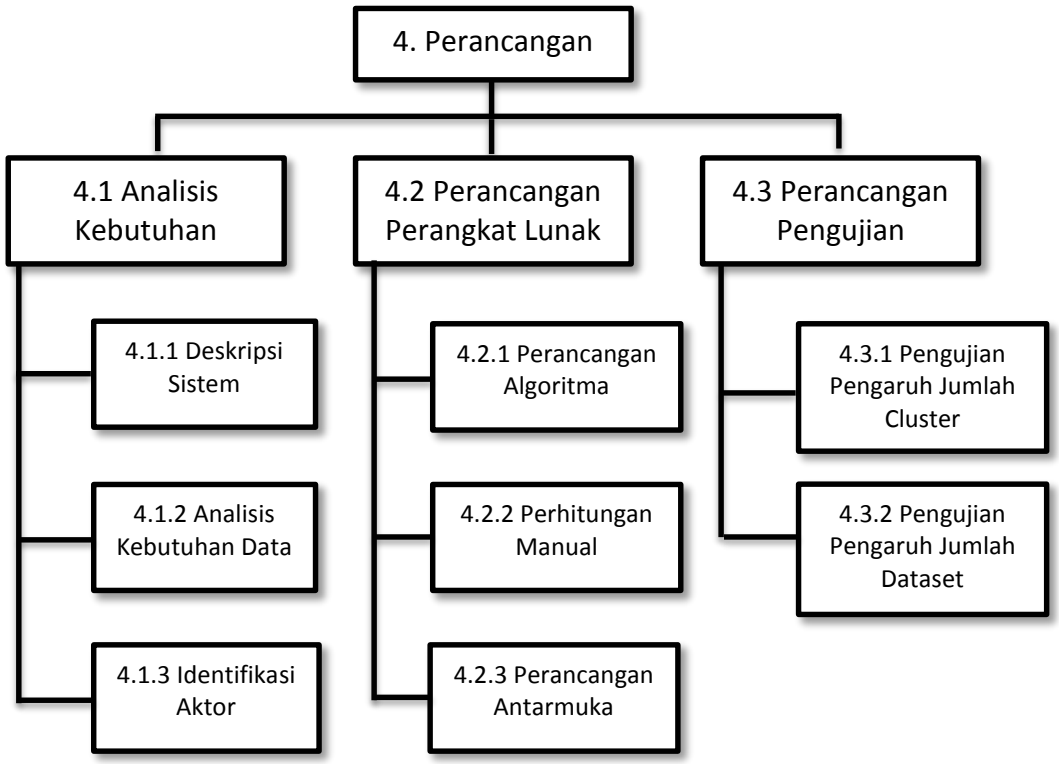
BAB 4 PERANCANGAN

Pada BAB ini membahas tentang perancangan pada “Clustering Pasien Kanker Berdasarkan Struktur Protein Dalam Tubuh Menggunakan Metode K-Medoids”. Pohon perancangan pada sistem ini meliputi tiga tahap yaitu analisis kebutuhan yang terdiri atas deskripsi sistem, analisis kebutuhan data, dan identifikasi aktor. Tahap selanjutnya adalah tahap perancangan perangkat lunak yang terdiri dari perancangan algoritma, perhitungan manual, dan perancangan antarmuka. Tahap yang terakhir adalah perancangan pengujian terdiri dari pengujian pengaruh jumlah cluster dan pengujian pengaruh jumlah dataset. Pohon perancangan pada penelitian ini ditunjukkan pada Gambar 4.1 berikut.



Gambar 4.1 Pohon Perancangan

4.1 Analisis Kebutuhan

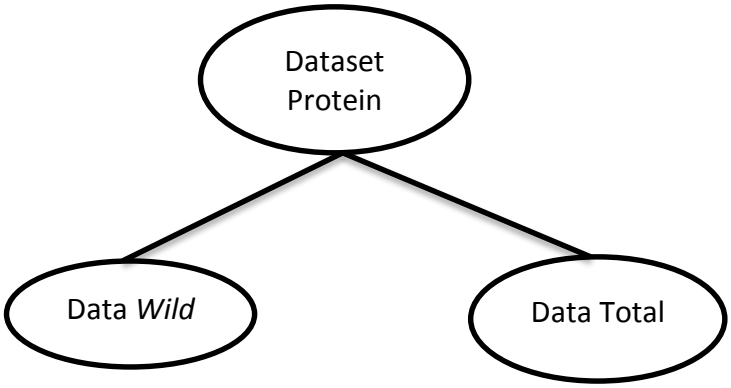
Tahap awal pada perancangan sistem clustering pasien kanker berdasarkan struktur protein dalam tubuh menggunakan metode k-medoids adalah analisis kebutuhan perangkat lunak. Analisis ini terbagi menjadi tiga yaitu deskripsi sistem yang dibangun, analisis kebutuhan data yang digunakan dalam sistem clustering, dan identifikasi aktor yang berperan dalam penggunaan sistem clustering pasien kanker.

4.1.1 Deskripsi Sistem

Penelitian ini memiliki tujuan untuk membuat sebuah aplikasi desktop yang berfungsi melakukan clustering pada pasien kanker, data yang digunakan adalah struktur protein dalam tubuh. Data yang digunakan dalam sistem ini dibagi menjadi dua, data *wild* dan data total. Data *wild* adalah data acuan yang digunakan sebagai proses konversi dataset protein dari data yang bertipe string menjadi data bertipe integer agar dapat dilakukan proses perhitungan clustering menggunakan metode K-Medoids. Data total adalah dataset protein yang akan dilakukan proses clustering.

Dalam proses pengclusteran metode yang digunakan adalah metode k-medoids. Metode k-medoids adalah metode partisi *clustering* untuk mengelompokkan sekumpulan *n* objek menjadi sejumlah *k cluster.* Sehingga dari data total yang dimasukkan akan dilakukan proses clustering sesuai dengan jumlah cluster yang diinginkan oleh pengguna aplikasi.

4.1.2 Analisis Kebutuhan Data



Gambar 4.2 Kebutuhan Data

Dalam penelitian ini terdapat dua jenis dataset yaitu data *wild* dan data total. Data *wild* digunakan sebagai data acuan atau data dasar yang akan digunakan sebagai bahan perbandingan dalam proses konversi data total menggunakan matriks PAM250. Sedangkan data total adalah dataset protein yang telah diketahui pengidentifikasian kelasnya dan merupakan data yang akan diolah menggunakan metode K-medoids.

Pada tahap pengumpulan data ini dataset tentang struktur protein dalam tubuh ini didapat dari http://www.uniprot.org/ yaitu merupakan gabungan dari

20 residu asam amino yang terdiri dari data String sepanjang 393 karakter. Sedangkan data protein didapatkan dari http://p53.free.fr/ , data ini digunakan untuk menentukan data yang bersifat kanker maupun non-kanker.

Pada sistem clustering pasien kanker berdasarkan struktur protein dalam tubuh ini terdapat 848 dataset yang digunakan dan memiliki panjang data yang sama dengan data *wild* yaitu sebanyak 393 karakter. Jika panjang data antara data *wild* dan dataset protein tidak sama maka program tidak bisa berjalan. Dari 848 dataset terdapat empat jenis kelas yang telah diidentifikasi yaitu 147 data termasuk dalam *Lung Cancer* atau kanker paru-paru, 147 data termasuk dalam *Colorectal Cancer* atau kanker usus, 147 data termasuk dalam *Breast Cancer* atau kanker payudara, dan 407 lainnya termasuk dalam *Non-Cancer*. Dari 848 data tersebut akan dilakukan proses clustering dengan metode k-medoids dan akan dilakukan uji kualitas cluster yang telah dilakukan dengan metode *silhouette coefficient*.

Dataset protein yang digunakan masih berupa data fisik berbentuk String yang harus dikonversikan terlebih dahulu ke dalam data numerik bertipe integer agar data tersebut dapat diolah dan dilakukan perhitungan untuk melakukan proses clusterisasi. Contoh data protein yang diperoleh dapat dilihat pada Tabel 4.1.

Tabel 4.1 Sampel Data Protein

Data Protein	Kelas
MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPDDIE QWFTEDPGPDEAPRMPEAAPPVAPAPAAPTPAAPAPAPSWPLSSSVPSQK TYQGSYGFRLGFLHSGTAKSVTCTYSPALNKMFCQLAKTCPVQLWVDSTPP PGTRVRAMAIYKQSQHMTEVVRRCPHHERCSDSDGLAPPQHLIRVEGNLR VEYLDDRNTFRHSVVVPYEPPEVGSDCTTIHYNYMCNSSCMGGMNRRPILT IITLEDSSGNLLGRNSFEVRVCACPGRDRRTEEENLRKKGEPHHELPPGSTKR ALPNNTSSSPQPKKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEP GGSRAHSSHLKSKKGQSTSRHKKLMFKTEGPDSD	*Non-cancer*
MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPDDIE QWFTEDPGPDEAPRMPEAAPPVAPAPAAPTPAAPAPAPSWPLSSSVPSQK TYQGSYGFRLGFLHSGTAKSVTCTYSPALNKMFCQLAKTCPVQLWVDSTPP PGTRVRAMAIYKQSQHMTKVVRRCPHHERCSDSDGLAPPQHLIRVEGNLR VEYLDDRNTFRHSVVVPYEPPEVGSDCTTIHYNYMCNSSCMGGMNRRPILT IITLEDSSGNLLGRNSFEVRVCACPGRDRRTEEENLRKKGEPHHELPPGSTKR ALPNNTSSSPQPKKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEP GGSRAHSSHLKSKKGQSTSRHKKLMFKTEGPDSD	*Breast Cancer*
MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPDDIE QWFTEDPGPDEAPRMPEAAPPVAPAPAAPTPAAPAPAPSWPLSSSVPSQK TYQGSYGFRLGFLHSGTAKSVTCTYSPALNKMFCQLAKTCPVQLWVDSTPP PGTRVRAMAIFKQSQHMTEVVRRCPHHERCSDSDGLAPPQHLIRVEGNLR VEYLDDRNTFRHSVVVPYEPPEVGSDCTTIHYNYMCNSSCMGGMNRRPILT IITLEDSSGNLLGRNSFEVRVCACPGRDRRTEEENLRKKGEPHHELPPGSTKR	*Colorectal Cancer*

Data Protein	Kelas
ALPNNTSSSPQPKKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEP GGSRAHSSHLKSKKGQSTSRHKKLMFKTEGPDSD	
MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPDDIE QWFTEDPGPDEAPRMPEAAPPVAPAPAAPTPAAPAPAPSWPLSSSVPSQK TYQGSYGFRLGFLHSGTAKSVTCTYSPALNKMFCQLAKTCPVQLWVDSTPP PGTRVRAMAIYKQSQHMTEVVRRCPHHERCSDSDGLAPPQHLIRVEENLRV EYLDDRNTFRHSVVVPYEPPEVGSDCTTIHYNYMCNSSCMGGMNRRPILTII TLEDSSGNLLGRNSFEVRVCACPGRDRRTEEENLRKKGEPHHELPPGSTKRA LPNNTSSSPQPKKKPLDGEYFTLQIRGRERFEMFRELNEALELKDAQAGKEP GGSRAHSSHLKSKKGQSTSRHKKLMFKTEGPDSD	*Lung Cancer*

Terdapat empat kelas dalam dataset protein, kelas *non-cancer* (NC) yang disimbolkan dengan 0, kelas *Breast Cancer* (BC) atau Kanker Payudara yang disimbolkan dengan 1, kelas *Colorectal Cancer* (CC) atau Kanker Usus yang disimbolkan dengan 2, dan kelas *Lung Cancer* (LC) atau Kanker Paru-paru yang disimbolkan dengan 3.

4.1.3 Identifikasi Aktor

Aktor adalah seseorang yang berperan dan memiliki hak untuk mengoperasikan sebuah sistem yang dibuat. Dalam sistem clustering pasien kanker berdasarkan struktur protein dalam tubuh menggunakan metode k-medoids ini hanya ada satu aktor yang berperan yaitu pengguna. Pengguna dapat melakukan tiga aksi di dalam sistem. Aksi yang bisa dilakukan adalah melakukan clusterisasi data, menentukan jumlah cluster yang diinginkan, melihat data latih, melihat hasil perhitungan k-medoids, dan melihat kualitas cluster yang dihasilkan.

Tabel 4.2 Identifikasi Aktor

Aktor	Deskripsi
Pengguna	Pengguna merupakan aktor yang memiliki hak akses sepenuhnya terhadap aplikasi clustering. Aksi yang bisa dilakukan adalah mengunggah data *wild*, dataset protein , dan jumlah nilai k (jumlah cluster) ke dalam sistem serta melakukan proses perhitungan k-medoids dan perhitungan silhouette coefficient.

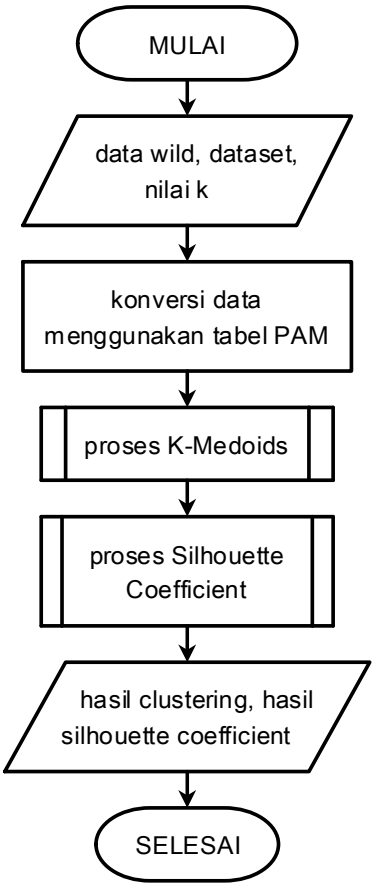
4.2 Perancangan Perangkat Lunak

Dalam membangun sebuah sistem perlu dilakukan proses perancangan perangkat lunak terlebih dahulu. Hal ini bertujuan untuk dapat dijadikan sebagai acuan atau panduan dalam proses pembangunan sistem. Perancangan perangkat

lunak yang harus dilakukan adalah perancangan algoritma, perhitungan manual, dan perancangan antarmuka.

4.2.1 Perancangan Algoritma

Setelah data yang dibutuhkan tersedia maka proses pengclusteran pasien kanker berdasarkan struktur protein dalam tubuh bisa dilakukan. Perancangan algoritma ditampilkan dalam Gambar 4.3.

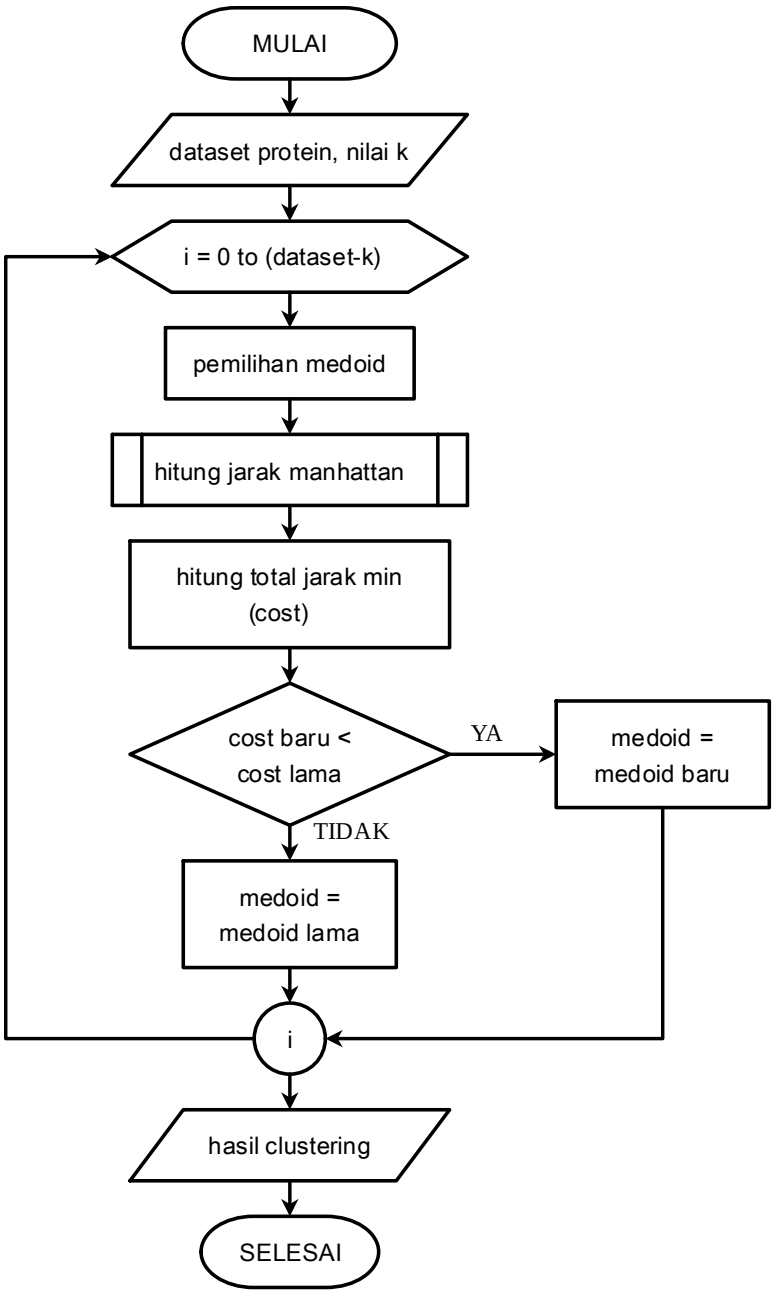


Gambar 4.3 Diagram Alir Sistem

Pengguna harus melakukan proses input data berupa dataset protein yang telah disiapkan, data *wild* protein yang telah disiapkan, dengan syarat panjang data *wild* harus sama dengan panjang data pada dataset protein. Pengguna juga memasukkan jumlah k yang diinginkan. Jumlah k ini akan menentukan banyak cluster yang harus diproses oleh sistem. Proses yang pertama dilakukan adalah mengkonversi data fisik berupa karakter menjadi data numerik bertipe integer menggunakan Tabel PAM250 agar dapat dilakukan perhitungan. Kemudian menekan tombol proses maka proses clustering akan berjalan. Setelah clusterisasi berhasil dilakukan maka program akan menampilkan hasilnya berupa data yang telah diproses, jumlah data yang diproses, dan jumlah data pada masing-masing cluster. Langkah selanjutnya adalah melakukan proses evaluasi cluster untuk mengetahui kualitas pada masing-masing cluster yang telah dilakukan.

4.2.1.1 Proses *K-Medoids Clustering*

Pada proses ini, algoritma K-Medoids akan mulai dijalankan untuk mengelompokkan dataset protein yang telah disiapkan. Dalam proses ini data yang diterima sudah bertipe integer karena telah dikonversikan pada proses sebelumnya. Data numerik yang telah dikonversikan ini hanya bersifat virtual atau hanya ada dalam sistem saja sehingga tidak disimpan dalam database. Pada akhir proses akan ditampilkan hasil data yang telah dilakukan proses klaterisasi sesuai dengan jumlah cluster yang telah diinputkan oleh pengguna. Diagram metode K-Medoids ditampilkan pada Gambar 4.4.



Gambar 4.4 Diagram Alir Metode K-Medoids

Langkah-langkah dari metode K-Medoids adalah :

a. Menentukan jumlah cluster yang diinginkan.

b. Langkah 1

Menentukan secara acak medoid awal yang berbeda sebanyak jumlah cluster yang telah diinputkan.

c. Langkah 2

Menghitung jarak setiap data dengan medoid awal. Perhitungan jarak dilakukan dengan menggunakan jarak manhattan (*manhattan distance)*. Rumus untuk perhitungan jarak manhattan dapat dilihat pada persamaan 2.1.

d. Langkah 3

Menghitung jarak masing-masing baris data dengan persamaan 2.2. Setelah itu menghitung total jarak seluruh data sehingga diketahui *total cost* dari medoid awal. Perhitungan *total cost* dapat dihitung dengan menggunakan persamaan 2.3.

e. Langkah 4

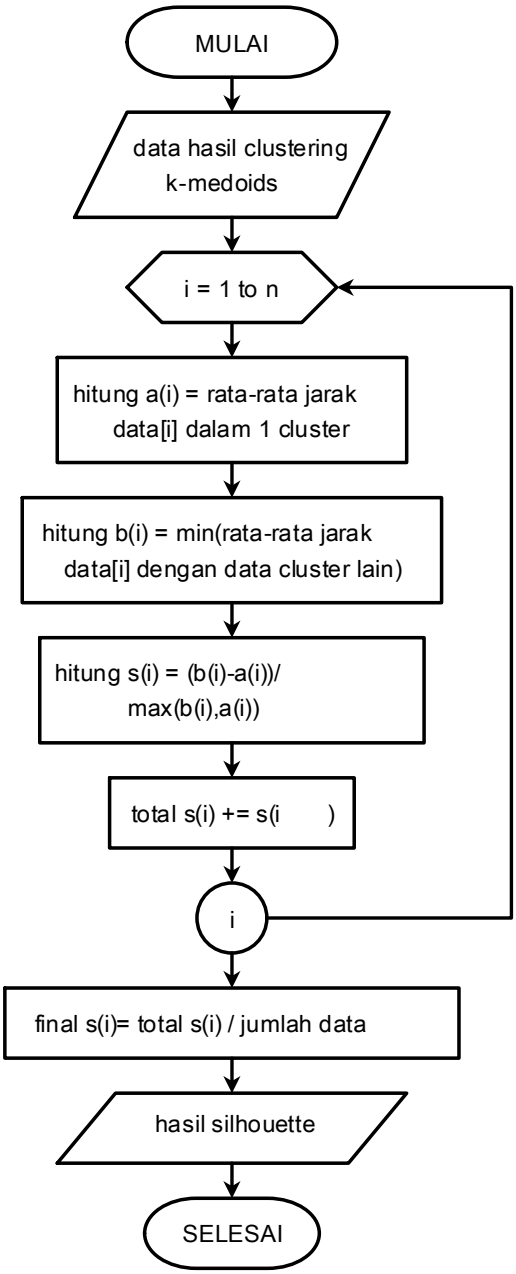
Jika total cost baru lebih kecil dari total cost sebelumnya maka ganti medoid awal dengan medoid yang baru, tetapi jika total cost baru lebih besar dari total cost sebelumnya maka medoid tetap. Lalu lanjutkan perhitungan dengan menghitung jarak manhattan lagi seperti proses sebelumnya.

f. Langkah 5

Ulangi langkah-langkah 2-4 sampai semua data pernah menjadi medoid dan diperoleh hasil *clustering* dengan total cost yang minimal.

4.2.1.2 Proses *Silhouette Coefficient*

Silhouette Coefficient merupakan algoritma yang berfungsi untuk melakukan pengujian kualitas cluster yang dihasilkan oleh metode K-Medoids. Diagram alir proses *Silhouette Coefficient* ini ditunjukkan pada Gambar 4.5.



Gambar 4.5 Diagram Alir *Silhouette Coefficient*

Langkah-langkah dari perhitungan *Silhouette Coefficient* adalah :

a. Menghitung nilai rata-rata jarak data ke-*i* terhadap semua data yang berada pada satu cluster yang sama. Rata-rata jarak pertama ini disebut dengan $a(i)$.
b. Menghitung nilai rata-rata jarak data ke-i terhadap semua data pada setiap cluster lain. Cari jarak rata-rata yang memiliki nilai terkecil diantara semua cluster. Rata-rata jarak kedua ini disebut dengan $b(i)$.
c. Menghitung nilai silhouette coefficient menggunakan persamaan 2.4 atau dengan persamaan 2.5.

4.2.2 Perhitungan Manual

Pada subbab perhitungan manual ini diambil beberapa sampel data dari dataset protein. Satu data *wild* yang digunakan sebagai data perbandingan dalam proses konversi data dan 12 dataset protein yang diambil secara acak dari data total protein tubuh yang akan dilakukan proses klasterisasi menggunakan metode K-Medoids. Data yang diambil memiliki panjang data 10 karakter. Dari dataset tersebut telah diketahui kelas masing-masing yaitu *non-cancer* (NC), *breast cancer* (BC), *colorectal cancer* (CC) dan *lung cancer* (LC).

4.2.2.1 Konversi Data

Dari dataset yang ada yaitu berupa data fisik yang masih bertipe String harus dikonversikan terlebih dahulu menjadi data numerik bertipe integer.

Pengkonversian data dengan cara mencocokkan dataset dengan data *wild* yang telah ada. Pengkonversian data berdasarkan pada matriks PAM250 yang telah ditampilkan pada Gambar 2.2.

Berikut ini ditampilkan data *wild* pada Tabel 4.3 dan dataset protein pada Tabel 4.4.

Tabel 4.3 Data *Wild* bentuk Fisik

Variabel	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
Data *Wild*	Y	K	Q	S	T	E	V	V	R	R

Tabel 4.4 Dataset bentuk Fisik

Dataset	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	Kelas
D1	Y	K	Q	S	T	E	V	V	R	R	NC
D2	Y	K	Q	S	S	E	V	V	R	R	BC
D3	Y	K	Q	S	T	E	V	V	R	C	CC
D4	Y	K	Q	L	T	E	V	V	R	R	LC
D5	Y	K	Q	S	T	E	V	V	R	C	CC
D6	Y	M	Q	S	T	E	V	V	R	R	CC
D7	Y	K	Q	S	L	E	V	V	R	R	BC
D8	Y	K	Q	S	G	E	V	V	R	R	BC
D9	Y	K	Q	M	T	E	V	L	R	R	LC
D10	Y	K	Q	L	T	E	V	L	R	R	LC
D11	Y	K	Q	S	T	E	V	V	R	R	NC

Dataset	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	Kelas
D12	Y	K	Q	S	T	E	V	V	R	R	NC

Hasil konversi data fisik menjadi data virtual numerik ditampilkan pada Tabel 4.5. Untuk kelas dikonversikan menjadi 0 = NC (*Non Cancer*), 1 = BC (*Breast Cancer*), 2 = CC (*Colorectal Cancer*), dan 3 = LC (*Lung Cancer*).

Tabel 4.5 Dataset Hasil Konversi

Dataset	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	Kelas
D1	10	5	4	2	2	4	4	4	6	6	0
D2	10	5	4	2	2	4	4	4	6	6	1
D3	10	5	4	2	2	4	4	4	6	-3	2
D4	10	5	4	-3	2	4	4	4	6	6	3
D5	10	5	4	2	2	4	4	4	6	-3	2
D6	10	1	4	2	2	4	4	4	6	6	2
D7	10	5	4	2	-2	4	4	4	6	6	1
D8	10	5	4	2	0	4	4	4	6	6	1
D9	10	5	4	-2	2	4	4	2	6	6	3
D10	10	5	4	-3	2	4	4	2	6	6	3
D11	10	5	4	2	2	4	4	4	6	6	0
D12	10	5	4	2	2	4	4	4	6	6	0

4.2.2.2 Perhitungan K-Medoids

Dalam sistem clustering pasien kanker dengan metode metode K-Medoids perhitungan dimulai dari menginputkan nilai *k* atau jumlah cluster yang diinginkan, langkah selanjutnya pemilihan secara acak medoid awal sebanyak nilai *k* yang diinginkan dari dataset sampai menghitung total cost. Perhitungan dilakukan berulang sampai semua dataset pernah menjadi medoid. Total cost yang memiliki nilai paling kecil berikutnya akan dilakukan proses perhitungan evaluasi kualitas cluster.

a. Pada perhitungan manual nilai k yang digunakan adalah 3. Sehingga data yang diperoses adalah mengelompokkan dataset sebanyak 3 cluster.

b. Langkah 1

Menentukan secara acak medoid awal yang berbeda sebanyak 3 medoid dari dataset protein yang telah dikonversikan. Dalam perhitungan ini medoid awal yang digunakan adalah D1, D4, dan D8.

Tabel 4.6 Dataset Medoid Awal

Cluster	Dataset	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	Kelas
C1	D1	10	5	4	2	2	4	4	4	6	6	0
C2	D4	10	5	4	-3	2	4	4	4	6	6	3
C3	D8	10	5	4	2	0	4	4	4	6	6	1

c. Langkah 2

Menghitung jarak setiap data dengan medoid awal C1, C2 dan C3 menggunakan rumus jarak manhattan (*Manhattan Distance*). Rumus jarak manhattan telah ditampilkan pada persamaan 4.1.

Sebagai contoh ditampilkan perhitungan secara detail pada data D2 yang diambil dari dataset protein. Data D2 akan dilakukan perhitungan dengan data C1, C2 dan C3 pada data cluster medoid awal.

Perhitungan pertama untuk D2:C1.

D2:C1 = |10-10|+ |5-5|+|4-4|+|2-2|+|2-2|+|4-4|+|4-4|+|4-4|+|6-6|+|6-6| = 0

Perhitungan pertama untuk D2:C2.

D2:C2 = |10-10|+ |5-5|+|4-4|+|2-(-3)|+|2-2|+|4-4|+|4-4|+|4-4|+|6-6|+|6-6| = 5

Perhitungan pertama untuk D2:C3.

D2:C3 = |10-10|+ |5-5|+|4-4|+|2-2|+|2-0|+|4-4|+|4-4|+|4-4|+|6-6|+|6-6| = 2

Perhitungan yang serupa dilakukan pada semua dataset protein yang digunakan. Sehingga akan mendapatkan hasil yang terlihat pada Tabel 4.7.

Tabel 4.7 Hasil Perhitungan Jarak Manhattan pada Medoid Awal

Dataset	Jarak C1	Jarak C2	Jarak C3
D2	0	5	2
D3	9	14	11
D5	9	14	11
D6	4	9	6
D7	4	9	2
D9	6	3	8
D10	7	2	9
D11	0	5	2
D12	0	5	2

d. Langkah 3

Langkah selanjutnya adalah menghitung nilai cost dari tiga cluster pada setiap data yang telah dihitung jaraknya dengan medoid awal. Rumus untuk menghitung nilai cost dapat dilihat pada persamaan 4.2. Contoh perhitungan pada D2 adalah sebagai berikut.

Cost (D2) = min {0,5,2} = 0

Hasil dari perhitungan cost minimal pada data D2 adalah 0. Dan perhitungan dilanjutkan pada semua dataset protein yang digunakan. Hasil perhitungan untuk semua dataset protein dapat dilihat pada Tabel 4.8.

Tabel 4.8 Nilai Cost Cluster pada Medoid Awal

Dataset	Jarak C1	Jarak C2	Jarak C3	Cost
D2	0	5	2	0
D3	9	14	11	9
D5	9	14	11	9
D6	4	9	6	4
D7	4	9	2	2
D9	6	3	8	3
D10	7	2	9	2
D11	0	5	2	0
D12	0	5	2	0
Total Cost				29

Setelah diketahui nilai cost minimal pada setiap data maka nilai cost tersebut dijumlahkan dan dapat diketahui anggota data pada setiap cluster. Maka didapatkan anggota Cluster 1 : D1, D2, D3, D5, D6, D11 dan D12, anggota Cluster 2 : D4, D9, dan D10, anggota Cluster 3 : D7 dan D8. Lalu total cost yang didapatkan adalah 29. Tujuan dari penjumlahan nilai cost ini adalah untuk mengetahui total cost pada iterasi ke berapakah yang memiliki nilai paling kecil. Dari total cost paling kecil itu maka clusterisasi berhasil dilakukan dan dapat dilanjutkan pada proses evaluasi kualitas cluster.

e. Langkah 4

Pada langkah ini iterasi pertama mulai dilakukan. Perhitungan dilakukan dengan mengganti salah satu medoid. Medoid baru yang digunakan adalah salah satu data non medoid yang berada pada cluster yang sama pada medoid awal. Dalam perhitungan ini medoid yang diganti adalah cluster 3, yaitu D8 diganti dengan D7.

Tabel 4.9 Medoid Iterasi 1

Cluster	Dataset	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	Kelas
C1	D1	10	5	4	2	2	4	4	4	6	6	0
C2	D4	10	5	4	-3	2	4	4	4	6	6	3
C3	D7 (medoid baru)	10	5	4	2	-2	4	4	4	6	6	1

Jika medoid iterasi 1 telah ditentukan maka proses perhitungan dilakukan sama persis pada proses sebelumnya. Yang pertama adalah menghitung jarak manhattan antara dataset dan medoid iterasi 2. Setelah selesai maka dilanjutkan dengan perhitungan cost pada setiap data. Langkah akhir adalah menghitung total cost dan menentukan anggota cluster.

f. Langkah 5

Langkah selanjutnya adalah mengulangi langkah 2-4 sampai semua data pernah menjadi medoid.

Setelah medoid iterasi 1 ditentukan maka perhitungan jarak manhattan dan proses selanjutnya dapat dilakukan, sehingga mendapatkan hasil yang terlihat pada Tabel 4.10.

Tabel 4.10 Hasil Jarak Manhattan dan Nilai Cost Iterasi 1

Dataset	Jarak C1	Jarak C2	Jarak C3	Cost
D2	0	5	4	0
D3	9	14	13	9
D5	9	14	13	9
D6	4	9	8	4
D8	2	7	2	2
D9	6	3	10	3
D10	7	2	11	2
D11	0	5	4	0
D12	0	5	4	0
Total Cost				29

Hasil perhitungan jarak manhattan dan nilai cost pada setiap data di iterasi 1 dapat dilihat pada Tabel 4.10. Maka didapatkan anggota Cluster 1 : D1, D2, D3, D5, D6, D11 dan D12, anggota Cluster 2 : D4, D9, dan D10, anggota Cluster 3 : D7 dan D8. Lalu total cost yang didapatkan adalah 29. Karena total cost tidak lebih kecil

dari total sebelumnya maka medoid dan anggota cluster tidak berubah yaitu Cluster 1 : D1, D2, D3, D5, D6, D11 dan D12, anggota Cluster 2 : D4, D9, dan D10, anggota Cluster 3 : D7 dan D8.

Perhitungan dilanjutkan pada iterasi 2, dengan medoid baru pada cluster 2.

Tabel 4.11 Medoid Iterasi 2

Cluster	Dataset	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	Kelas
C1	D1	10	5	4	2	2	4	4	4	6	6	0
C2	D9 (medoid baru)	10	5	4	-2	2	4	4	2	6	6	3
C3	D8	10	5	4	2	0	4	4	4	6	6	1

Dilanjutkan dengan perhitungan jarak manhattan dan nilai cost setiap data.

Tabel 4.12 Hasil Jarak Manhattan dan Nilai Cost Iterasi 2

Dataset	Jarak C1	Jarak C2	Jarak C3	Cost
D2	0	6	2	0
D3	9	15	11	9
D5	9	15	11	9
D6	4	10	6	4
D7	4	10	2	2
D4	5	3	7	3
D10	7	1	9	1
D11	0	6	2	0
D12	0	6	2	0
Total Cost				28

Hasil perhitungan jarak manhattan dan nilai cost pada setiap data di iterasi 2 dapat dilihat pada Tabel 4.12. Maka didapatkan anggota Cluster 1 : D1, D2, D3, D5, D6, D11 dan D12, anggota Cluster 2 : D4, D9 dan D10, anggota Cluster 3 : D7 dan D8. Lalu total cost yang didapatkan adalah 28. Karena total cost lebih kecil dari total cost medoid sebelumnya maka medoid dan anggota cluster berubah.

Perhitungan dilanjutkan pada iterasi 3, dengan medoid baru pada cluster 2.

Tabel 4.13 Medoid Iterasi 3

Cluster	Dataset	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	Kelas
C1	D1	10	5	4	2	2	4	4	4	6	6	0

Cluster	Dataset	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	Kelas
C2	D10 (medoid baru)	10	5	4	-3	2	4	4	2	6	6	3
C3	D8	10	5	4	2	0	4	4	4	6	6	1

Dilanjutkan dengan perhitungan jarak manhattan dan nilai cost setiap data.

Tabel 4.14 Hasil Jarak Manhattan dan Nilai Cost Iterasi 3

Dataset	Jarak C1	Jarak C2	Jarak C3	Cost
D2	0	7	2	0
D3	9	16	11	9
D5	9	16	11	9
D6	4	11	6	4
D7	4	11	2	2
D4	5	2	7	2
D9	6	1	8	1
D11	0	7	2	0
D12	0	7	2	0
Total Cost				27

Hasil perhitungan jarak manhattan dan nilai cost pada setiap data di iterasi 3 dapat dilihat pada Tabel 4.14. Maka didapatkan anggota Cluster 1 : D1, D2, D3, D5, D6, D11 dan D12, anggota Cluster 2 : D4, D9 dan D10, anggota Cluster 3 : D7 dan D8. Lalu total cost yang didapatkan adalah 27. Karena total cost lebih kecil dari total cost medoid sebelumnya maka medoid dan anggota cluster berubah.

Perhitungan dilanjutkan pada iterasi 4, dengan medoid baru pada cluster 1.

Tabel 4.15 Medoid Iterasi 4

Cluster	Dataset	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	Kelas
C1	D2 (medoid baru)	10	5	4	2	2	4	4	4	6	6	1
C2	D10	10	5	4	-3	2	4	4	2	6	6	3
C3	D8	10	5	4	2	0	4	4	4	6	6	1

Dilanjutkan dengan perhitungan jarak manhattan dan nilai cost setiap data.

Tabel 4.16 Hasil Jarak Manhattan dan Nilai Cost Iterasi 4

Dataset	Jarak C1	Jarak C2	Jarak C3	Cost
D1	0	7	2	0
D3	9	16	11	9
D5	9	16	11	9
D6	4	11	6	4
D7	4	11	2	2
D4	5	2	7	2
D9	6	1	8	1
D11	0	7	2	0
D12	0	7	2	0
Total Cost				27

Hasil perhitungan jarak manhattan dan nilai cost pada setiap data di iterasi 4 dapat dilihat pada Tabel 4.16. Maka didapatkan anggota Cluster 1 : D1, D2, D3, D5, D6, D11 dan D12, anggota Cluster 2 : D4, D9 dan D10, anggota Cluster 3 : D7 dan D8. Lalu total cost yang didapatkan adalah 27. Karena total cost tidak lebih kecil dari total cost medoid sebelumnya maka medoid dan anggota cluster tidak berubah yaitu Cluster 1 : D1, D2, D3, D5, D6, D11 dan D12, anggota Cluster 2 : D4, D9 dan D10, anggota Cluster 3 : D7 dan D8.

Perhitungan dilanjutkan pada iterasi 5, dengan medoid baru pada cluster 1.

Tabel 4.17 Medoid Iterasi 5

Cluster	Dataset	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	Kelas
C1	D3 (medoid baru)	10	5	4	2	2	4	4	4	6	-3	2
C2	D10	10	5	4	-3	2	4	4	2	6	6	3
C3	D8	10	5	4	2	0	4	4	4	6	6	1

Dilanjutkan dengan perhitungan jarak manhattan dan nilai cost setiap data.

Tabel 4.18 Hasil Jarak Manhattan dan Nilai Cost Iterasi 5

Dataset	Jarak C1	Jarak C2	Jarak C3	Cost
D1	9	7	2	2
D2	9	7	2	2
D5	0	16	11	0

Dataset	Jarak C1	Jarak C2	Jarak C3	Cost
D7	13	11	2	2
D4	14	2	7	2
D9	15	1	8	1
D11	9	7	2	2
D12	9	7	2	2
Total Cost				19

Hasil perhitungan jarak manhattan dan nilai cost pada setiap data di iterasi 5 dapat dilihat pada Tabel 4.18. Maka didapatkan anggota Cluster 1 : D3 dan D5, anggota Cluster 2 : D4, D9 dan D10, anggota Cluster 3 : D1, D2, D6, D7, D8, D11 dan D12. Lalu total cost yang didapatkan adalah 19. Karena total cost lebih kecil dari total cost medoid sebelumnya maka medoid dan anggota cluster berubah.

Perhitungan dilanjutkan pada iterasi 6, dengan medoid baru pada cluster 1.

Tabel 4.19 Medoid Iterasi 6

Cluster	Dataset	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	Kelas
C1	D5 (medoid baru)	10	5	4	2	2	4	4	4	6	-3	2
C2	D10	10	5	4	-3	2	4	4	2	6	6	3
C3	D8	10	5	4	2	0	4	4	4	6	6	1

Dilanjutkan dengan perhitungan jarak manhattan dan nilai cost setiap data.

Tabel 4.20 Hasil Jarak Manhattan dan Nilai Cost Iterasi 6

Dataset	Jarak C1	Jarak C2	Jarak C3	Cost
D1	9	7	2	2
D2	9	7	2	2
D3	0	16	11	0
D6	13	11	6	6
D7	13	11	2	2
D4	14	2	7	2
D9	15	1	8	1
D11	9	7	2	2
D12	9	7	2	2
Total Cost				19

Hasil perhitungan jarak manhattan dan nilai cost pada setiap data di iterasi 6 dapat dilihat pada Tabel 4.20. Maka didapatkan anggota Cluster 1 : D3 dan D5, anggota Cluster 2 : D4, D9 dan D10, anggota Cluster 3 : D1, D2, D6, D7, D8, D11 dan D12. Lalu total cost yang didapatkan adalah 19. Karena total cost tidak lebih kecil dari total cost medoid sebelumnya maka medoid dan anggota cluster tidak berubah yaitu Cluster 1 : D3 dan D5, anggota Cluster 2 : D4, D9 dan D10, anggota Cluster 3 : D1, D2, D6, D7, D8, D11 dan D12.

Perhitungan dilanjutkan pada iterasi 7, dengan medoid baru pada cluster 1.

Tabel 4.21 Medoid Iterasi 7

Cluster	Dataset	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	Kelas
C1	D6 (medoid baru)	10	1	4	2	2	4	4	4	6	6	2
C2	D10	10	5	4	-3	2	4	4	2	6	6	3
C3	D8	10	5	4	2	0	4	4	4	6	6	1

Dilanjutkan dengan perhitungan jarak manhattan dan nilai cost setiap data.

Tabel 4.22 Hasil Jarak Manhattan dan Nilai Cost Iterasi 7

Dataset	Jarak C1	Jarak C2	Jarak C3	Cost
D1	4	7	2	2
D2	4	7	2	2
D3	13	16	11	11
D5	13	16	11	11
D7	8	11	2	2
D4	9	2	7	2
D9	10	1	8	1
D11	4	7	2	2
D12	4	7	2	2
Total Cost				35

Hasil perhitungan jarak manhattan dan nilai cost pada setiap data di iterasi 7 dapat dilihat pada Tabel 4.22. Maka didapatkan anggota Cluster 1 : D6 dan D9, anggota Cluster 2 : D4, D9 dan D10, anggota Cluster 3 : D1, D2, D3, D7, D8, D11 dan D12. Lalu total cost yang didapatkan adalah 35. Karena total cost tidak lebih kecil dari total cost medoid sebelumnya maka medoid dan anggota cluster tidak berubah yaitu Cluster 1 : D3 dan D5, anggota Cluster 2 : D4, D9 dan D10, anggota Cluster 3 : D1, D2, D6, D7, D8, D11 dan D12.

Perhitungan dilanjutkan pada iterasi 8, dengan medoid baru pada cluster 1.

Tabel 4.23 Medoid Iterasi 8

Cluster	Dataset	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	Kelas
C1	D11 (medoid baru)	10	5	4	2	2	4	4	4	6	6	0
C2	D10	10	5	4	-3	2	4	4	2	6	6	3
C3	D8	10	5	4	2	0	4	4	4	6	6	1

Dilanjutkan dengan perhitungan jarak manhattan dan nilai cost setiap data.

Tabel 4.24 Hasil Jarak Manhattan dan Nilai Cost Iterasi 8

Dataset	Jarak C1	Jarak C2	Jarak C3	Cost
D1	0	7	2	0
D2	0	7	2	0
D3	9	16	11	9
D5	9	16	11	9
D7	4	11	2	2
D4	5	2	7	2
D9	6	1	8	1
D6	4	11	6	4
D12	0	7	2	0
Total Cost				27

Hasil perhitungan jarak manhattan dan nilai cost pada setiap data di iterasi 8 dapat dilihat pada Tabel 4.24. Maka didapatkan anggota Cluster 1 : D1, D2, D3, D5, D6, D11 dan D12, anggota Cluster 2 : D4, D9 dan D10, anggota Cluster 3 : D7 dan D8. Lalu total cost yang didapatkan adalah 27. Karena total cost tidak lebih kecil dari total cost medoid sebelumnya maka medoid dan anggota cluster tidak berubah yaitu Cluster 1 : D3 dan D5, anggota Cluster 2 : D4, D9 dan D10, anggota Cluster 3 : D1, D2, D6, D7, D8, D11 dan D12.

Perhitungan dilanjutkan pada iterasi 9, dengan medoid baru pada cluster 1.

Tabel 4.25 Medoid Iterasi 9

Cluster	Dataset	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	Kelas
C1	D12 (medoid baru)	10	5	4	2	2	4	4	4	6	6	0
C2	D10	10	5	4	-3	2	4	4	2	6	6	3
C3	D8	10	5	4	2	0	4	4	4	6	6	1

Dilanjutkan dengan perhitungan jarak manhattan dan nilai cost setiap data.

Tabel 4.26 Hasil Jarak Manhattan dan Nilai Cost Iterasi 9

Dataset	Jarak C1	Jarak C2	Jarak C3	Cost
D1	0	7	2	0
D2	0	7	2	0
D3	9	16	11	9
D5	9	16	11	9
D7	4	11	2	2
D4	5	2	7	2
D9	6	1	8	1
D6	4	11	6	4
D11	0	7	2	0
Total Cost				27

Hasil perhitungan jarak manhattan dan nilai cost pada setiap data di iterasi 9 dapat dilihat pada Tabel 4.26. Maka didapatkan anggota Cluster 1 : D1, D2, D3, D5, D6, D11 dan D12, anggota Cluster 2 : D4, D9 dan D10, anggota Cluster 3 : D7 dan D8. Lalu total cost yang didapatkan adalah 27. Karena total cost tidak lebih kecil dari total cost medoid sebelumnya maka medoid dan anggota cluster tidak berubah yaitu Cluster 1 : D3 dan D5, anggota Cluster 2 : D4, D9 dan D10, anggota Cluster 3 : D1, D2, D6, D7, D8, D11 dan D12.

Setelah semua dataset pernah menjadi medoid maka dapat ditemukan kelompok yang memiliki hasil cluster paling baik, yaitu yang memiliki total cost minimal sebesar 19 pada iterasi 5. Data cluster yang dihasilkan adalah :

1. Kelompok cluster 1 dengan medoid D3 yang memiliki anggota data 3 dan data 5.
2. Kelompok cluster 2 dengan medoid D10 yang memiliki anggota data 4, data 9, dan data 10.
3. Kelompok cluster 3 dengan medoid D8 yang memiliki anggota data 1, data 2, data 6, data 7, data 8, data 11, dan data 12.

4.2.2.3 Perhitungan *Silhouette Coefficient*

Tujuan *silhouette coefficient* adalah untuk melakukan evaluasi hasil clustering yaitu untuk mengetahui kualitas cluster yang telah dilakukan perhitungan pada proses sebelumnya. Hasil dari silhouette coefficient ini berada antara nilai -1 sampai. Jika *s(i)* = 1 maka data telah berada pada cluster yang tepat, jika *s(i)* = 0 maka data berada pada posisi di tengah, maksudnya data terdapat kemungkinan

berada pada cluster yang tepat tetapi bisa juga seharusnya berada pada cluster yang lain, jika *s(i)* = -1 maka data berada pada cluster yang salah sehingga seharusnya data berada pada cluster yang lain.

Perhitungan manual *silhouette coefficient* adalah sebagai berikut :

a. Langkah 1

Menghitung jarak D1 terhadap semua data yang berada pada cluster yang sama, yaitu cluster 3.

Tabel 4.27 Jarak D1 dengan Anggota Cluster 3

Dataset	Dataset	Jarak D1 dengan semua anggota C3
D1	D2	0
	D6	4
	D7	4
	D8	2
	D11	0
	D12	0
$a(i)$ (Rata-rata)		1,7

b. Langkah 2

Menghitung jarak D1 terhadap semua data yang berada pada cluster 1 dan cluster 2, kemudian cari nilai rata-rata jarak yang paling kecil.

Tabel 4.28 Jarak D1 dengan Anggota Cluster 1 dan Cluster 2

Dataset	Cluster	Dataset	Jarak D1 dengan semua data	Rata-rata Jarak	$b(i)$ (Rata-rata Minimal)
D1	1	D3	9	9	6
		D5	9		
	2	D4	5	6	
		D9	6		
		D10	7		

c. Langkah 3

Setelah rata-rata jarak D1 dengan semua data selesai dihitung maka perhitungan *silhouette coefficient* bisa dilakukan menggunakan persamaan 4.5. Diketahui $a(1) = 1{,}7$ dan $b(1) = 6$, sehingga $a(1) < b(1)$ nilai *silhouette coefficient*-nya adalah :

$$S = 1 - \frac{a(1)}{b(1)}$$

$$= 1 - \frac{1{,}7}{6}$$

$$= 1 - 0{,}28$$

$$= 0{,}72$$

Perhitungan serupa dilakukan pada semua dataset hingga mendapatkan hasil rata-rata *silhouette coefficient* dari semua cluster yang ada.

Perhitungan *silhouette coefficient* data 2:

Tabel 4.29 Jarak D2 dengan Semua Data

Dataset	Dataset C3	Jarak D2 : anggota C3	Dataset C1	Jarak D2 : anggota C1	Dataset C2	Jarak D2 : anggota C2
D2	D1	0	D3	9	D4	5
	D6	4	D5	9	D9	6
	D7	4			D10	7
	D8	2				
	D11	0				
	D12	0				
$a(2)$ (Rata-rata)		1,7	Rata^2	9	Rata^2	6
			$b(2)$	6		

Diketahui $a(2) = 1{,}7$ dan $b(2) = 6$, sehingga $a(2) < b(2)$ nilai *silhouette coefficient*-nya adalah :

$$S = 1 - \frac{a(2)}{b(2)}$$

$$= 1 - \frac{1{,}7}{6}$$

$$= 1 - 0{,}28$$

$$= 0{,}72$$

Perhitungan *silhouette coefficient* data 6:

Tabel 4.30 Jarak D6 dengan Semua Data

Dataset	Dataset C3	Jarak D6 : anggota C3	Dataset C1	Jarak D6 : anggota C1	Dataset C2	Jarak D6 : anggota C2
D6	D1	4	D3	13	D4	9
	D2	4	D5	13	D9	10
	D7	8			D10	11
	D8	6				
	D11	4				
	D12	4				
$a(6)$ (Rata-rata)		5	Rata^2	13	Rata^2	10
			$b(6)$	10		

Diketahui $a(6) = 5$ dan $b(6) = 10,$ sehingga $a(6) < b(6)$ nilai *silhouette coefficient*-nya adalah :

$$S = 1 - \frac{a(6)}{b(6)}$$

$$= 1 - \frac{5}{10}$$

$$= 1 - 0{,}5$$

$$= 0{,}5$$

Perhitungan *silhouette coefficient* data 7:

Tabel 4.31 Jarak D7 dengan Semua Data

Dataset	Dataset C3	Jarak D7 : anggota C3	Dataset C1	Jarak D7 : anggota C1	Dataset C2	Jarak D7 : anggota C2
D7	D1	4	D3	13	D4	9
	D2	4	D5	13	D9	10
	D6	8			D10	11
	D8	2				
	D11	4				
	D12	4				
$a(7)$ (Rata-rata)		4,3	Rata^2	13	Rata^2	10
			$b(7)$	10		

Diketahui $a(7) = 4{,}3$ dan $b(7) = 10$, sehingga $a(7) < b(7)$ nilai *silhouette coefficient*-nya adalah :

$$S = 1 - \frac{a(7)}{b(7)}$$

$$= 1 - \frac{4{,}3}{10}$$

$$= 1 - 0{,}43$$

$$= 0{,}57$$

Perhitungan *silhouette coefficient* data 8:

Tabel 4.32 Jarak D8 dengan Semua Data

Dataset	Dataset C3	Jarak D8 : anggota C3	Dataset C1	Jarak D8 : anggota C1	Dataset C2	Jarak D8 : anggota C2
D8	D1	2	D3	11	D4	7
	D2	2	D5	11	D9	8
	D6	6			D10	9
	D7	2				
	D11	2				
	D12	2				
$a(8)$ (Rata-rata)		2,7	Rata^2	11	Rata^2	8
			$b(8)$	8		

Diketahui $a(8) = 2{,}7$ dan $b(8) = 8$, sehingga $a(8) < b(8)$ nilai *silhouette coefficient*-nya adalah :

$$S = 1 - \frac{a(8)}{b(8)}$$

$$= 1 - \frac{2{,}7}{8}$$

$$= 1 - 0{,}34$$

$$= 0{,}66$$

Perhitungan *silhouette coefficient* data 11:

Tabel 4.33 Jarak D11 dengan Semua Data

Dataset	Dataset C3	Jarak D11 : anggota C3	Dataset C1	Jarak D11 : anggota C1	Dataset C2	Jarak D11 : anggota C2
D11	D1	0	D3	9	D4	5
	D2	4	D5	9	D9	6
	D6	4			D10	7
	D7	2				
	D8	0				
	D12	0				
$a(11)$ (Rata-rata)		1,7	Rata^2	9	Rata^2	6
			$b(11)$	6		

Diketahui $a(11) = 1{,}7$ dan $b(11) = 6,$ sehingga $a(11) < b(11)$ nilai *silhouette coefficient*-nya adalah :

$$S = 1 - \frac{a(11)}{b(11)}$$

$$= 1 - \frac{1{,}7}{6}$$

$$= 1 - 0{,}28$$

$$= 0{,}72$$

Perhitungan *silhouette coefficient* data 12:

Tabel 4.34 Jarak D12 dengan Semua Data

Dataset	Dataset C3	Jarak D12 : anggota C3	Dataset C1	Jarak D12 : anggota C1	Dataset C2	Jarak D12 : anggota C2
D12	D1	0	D3	9	D4	5
	D2	4	D5	9	D9	6
	D6	4			D10	7
	D7	2				
	D8	0				
	D11	0				
$a(12)$ (Rata-rata)		1,7	Rata^2	9	Rata^2	6
			$b(12)$	6		

Diketahui $a(12) = 1{,}7$ dan $b(12) = 6,$ sehingga $a(12) < b(12)$ nilai *silhouette coefficient*-nya adalah :

$$S = 1 - \frac{a(12)}{b(12)}$$

$$= 1 - \frac{1{,}7}{6}$$

$$= 1 - 0{,}28$$

$$= 0{,}72$$

Setelah perhitungan silhouette coefficient pada setiap data pada cluster 3 selesai dihitung, langkah selanjutnya adalah menghitung rata-rata nilai silhouette coefficient dari semua data.

$$S(3) = \frac{0{,}72 + 0{,}72 + 0{,}5 + 0{,}57 + 0{,}66 + 0{,}72 + 0{,}72}{7} = \frac{4{,}61}{7} = 0{,}66$$

Dilanjutkan dengan menghitung data pada cluster 2 terhadap semua data. Perhitungan *silhouette coefficient* data 4:

Tabel 4.35 Jarak D4 dengan Semua Data

Dataset	Dataset C2	Jarak D4 : anggota C2	Dataset C1	Jarak D4 : anggota C1	Dataset C3	Jarak D4 : anggota C3
D4	D9	3	D3	14	D1	5
	D10	2	D5	14	D2	5
					D6	9
					D7	9
					D8	7
					D11	5
					D12	5
$a(4)$ (Rata-rata)		2,5	Rata^2	14	Rata^2	6,4
			$b(4)$	6,4		

Diketahui $a(4) = 2{,}5$ dan $b(4) = 6{,}4$, sehingga $a(4) < b(4)$ nilai *silhouette coefficient*-nya adalah :

$$S = 1 - \frac{a(4)}{b(4)}$$

$$= 1 - \frac{2{,}5}{6{,}4}$$

$$= 1 - 0{,}39$$

$$= 0{,}61$$

Perhitungan *silhouette coefficient* data 9:

Tabel 4.36 Jarak D9 dengan Semua Data

Dataset	Dataset C2	Jarak D9 : anggota C2	Dataset C1	Jarak D9 : anggota C1	Dataset C3	Jarak D9 : anggota C3
D9	D4	3	D3	15	D1	6
	D10	1	D5	15	D2	6
					D6	10
					D7	10
					D8	8
					D11	6
					D12	6
$a(9)$ (Rata-rata)		2	Rata^2	15	Rata^2	7,4
			$b(9)$	7,4		

Diketahui $a(9) = 2$ dan $b(9) = 7{,}4$, sehingga $a(9) < b(9)$ nilai *silhouette coefficient*-nya adalah :

$$S = 1 - \frac{a(9)}{b(9)}$$

$$= 1 - \frac{2}{7{,}4}$$

$$= 1 - 0{,}27$$

$$= 0{,}73$$

Perhitungan *silhouette coefficient* data 10:

Tabel 4.37 Jarak D10 dengan Semua Data

Dataset	Dataset C2	Jarak D10 : anggota C2	Dataset C1	Jarak D10 : anggota C1	Dataset C3	Jarak D10 : anggota C3
D10	D4	2	D3	16	D1	7
	D9	1	D5	16	D2	7
					D6	11
					D7	11
					D8	9
					D11	7
					D12	7
$a(10)$ (Rata-rata)		1,5	Rata^2	16	Rata^2	8,4
			$b(10)$	8,4		

Diketahui $a(10) = 1{,}5$ dan $b(10) = 8{,}4$, sehingga $a(10) < b(10)$ nilai *silhouette coefficient*-nya adalah :

$$S = 1 - \frac{a(10)}{b(10)}$$

$$= 1 - \frac{1{,}5}{8{,}4}$$

$$= 1 - 0{,}18$$

$$= 0{,}82$$

Setelah perhitungan silhouette coefficient pada setiap data pada cluster 2 selesai dihitung, langkah selanjutnya adalah menghitung rata-rata nilai silhouette coefficient dari semua data.

$$S(2) = \frac{0{,}61 + 0{,}73 + 0{,}82}{3} = \frac{2{,}16}{3} = 0{,}72$$

Dilanjutkan dengan menghitung data pada cluster 1 terhadap semua data. Perhitungan *silhouette coefficient* data 3:

Tabel 4.38 Jarak D3 dengan Semua Data

Dataset	Dataset C1	Jarak D3 : anggota C1	Dataset C2	Jarak D3 : anggota C2	Dataset C3	Jarak D3 : anggota C3
D3	D5	0	D4	14	D1	9
			D9	15	D2	9
			D10	16	D6	13
					D7	13
					D8	11
					D11	9
					D12	9
$a(3)$ (Rata-rata)		0	Rata^2	15	Rata^2	10,4
			$b(3)$	10,4		

Diketahui $a(3) = 0$ dan $b(3) = 10{,}4$, sehingga $a(3) < b(3)$ nilai *silhouette coefficient*-nya adalah :

$$S = 1 - \frac{a(3)}{b(3)}$$

$$= 1 - \frac{0}{10{,}4}$$

$$= 1 - 0$$

$$= 1$$

Perhitungan *silhouette coefficient* data 5:

Tabel 4.39 Jarak D5 dengan Semua Data

Dataset	Dataset C1	Jarak D5 : anggota C1	Dataset C2	Jarak D5 : anggota C2	Dataset C3	Jarak D5 : anggota C3
D5	D3	0	D4	14	D1	9
			D9	15	D2	9
			D10	16	D6	13
					D7	13
					D8	11
					D11	9
					D12	9
$a(5)$ (Rata-rata)		0	Rata^2	15	Rata^2	10,4
			$b(5)$	10,4		

Diketahui $a(5) = 0$ dan $b(5) = 10{,}4$, sehingga $a(5) < b(5)$ nilai *silhouette coefficient*-nya adalah :

$$S = 1 - \frac{a(5)}{b(5)}$$

$$= 1 - \frac{0}{10{,}4}$$

$$= 1 - 0$$

$$= 1$$

Setelah perhitungan *silhouette coefficient* pada setiap data pada cluster 1 selesai dihitung, langkah selanjutnya adalah menghitung rata-rata nilai *silhouette coefficient* dari semua data.

$$S(1) = \frac{1+1}{2} = \frac{2}{2} = 1$$

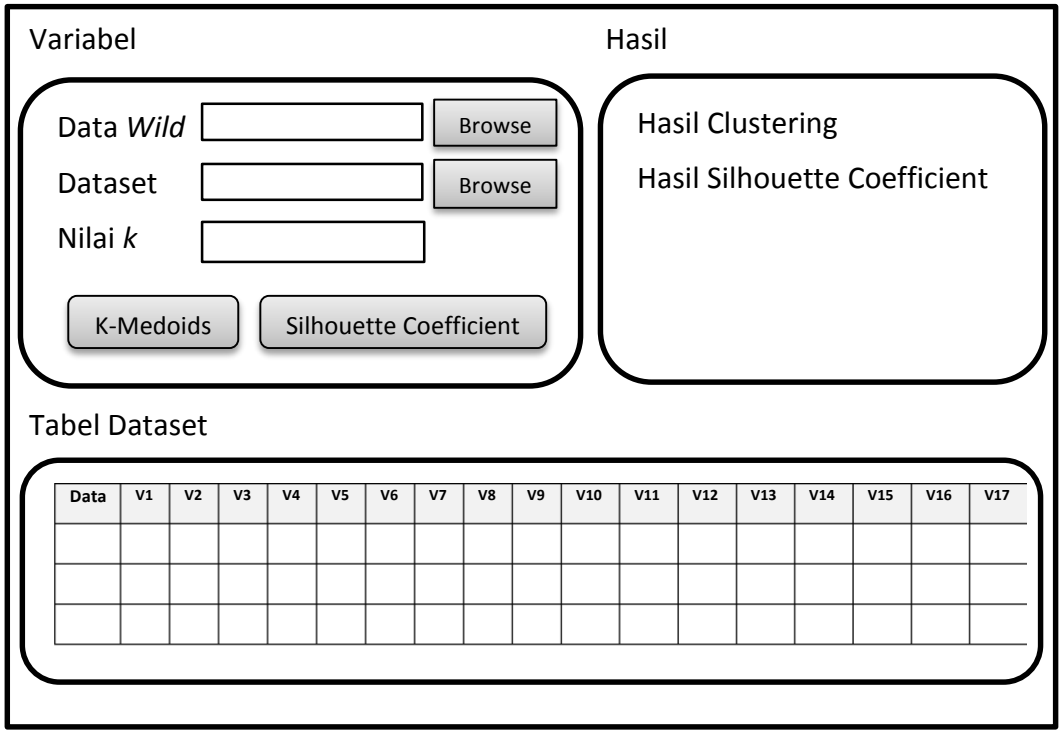
Semua data dari semua cluster telah selesai dilakukan perhitungan *silhouette coefficient*, sehingga nilai *silhouette coefficient* dari sistem K-Medoids bisa ditentukan rata-rata tiga cluster tersebut, sehingga :

$$S = \frac{0{,}66 + 0{,}72 + 1}{3} = \frac{2{,}38}{3} = 0{,}79$$

Nilai *silhouette coefficient* pada sistem clustering pasien kanker menggunakan metode K-Medoids adalah 0,79. Karena hasil yang didapatkan lebih mendekati 1 sehingga dapat disimpulkan bahwa semua data telah masuk pada cluster yang tepat.

4.2.3 Perancangan Antarmuka

Pada tahap ini dilakukan perancangan antarmuka sistem yang akan dibuat. Antarmuka sistem berfungsi untuk memudahkan pengguna dalam menggunakan sistem yang dibuat dan sebagai perantara antara sebuah sistem atau perangkat lunak komputer dengan pengguna.



Gambar 4.5 Rancangan Antarmuka Sistem Clustering

Dalam sistem ini hanya terdapat satu halaman kerja yang akan menampilkan dataset protein yang telah disimpan dalam file xml. Juga terdapat tombol "browse" yang digunakan untuk memasukkan data *wild* dan dataset protein yang telah disimpan terlebih dahulu, lalu juga memasukkan nilai *k* untuk menentukan jumlah cluster yang akan diproses. Setelah semua data dimasukkan maka tekan tombol "K-Medoids" untuk dilakukan proses perhitungan dengan metode K-Medoids dan tekan tombol "Silhouette Coefficient" untuk dilakukan proses evaluasi hasil clustering untuk mengetahui kualitas cluster. Setelah proses perhitungan selesai maka akan ditampilkan banyak cluster yang telah diproses dan anggota cluster. Dan juga akan dihitung nilai *Silhouette Coefficient* yaitu untuk mengetahui kualitas cluster yang dihasilkan metode K-Medoids.

4.3 Perancangan Pengujian

Pada tahap ini akan dijelaskan tentang proses pengujian yang dilakukan pada sistem clustering pasien kanker berdasarkan struktur protein dalam tubuh ini. Pengujian yang dilakukan ada dua macam yaitu pengujian atas pengaruh jumlah cluster dan pengujian pengaruh jumlah dataset protein.

4.3.1 Pengujian Pengaruh Jumlah Cluster

Pengujian pengaruh jumlah cluster ini bertujuan untuk mengetahui pada cluster ke berapa program bisa bekerja secara maksimal. Jumlah cluster yang dimaksud adalah nilai *k* yang diinputkan oleh user. Cluster yang baik dapat dilihat dari nilai *silhouette coefficient*. Semakin nilai *silhouette coefficient* mendekati angka 1 maka sistem yang berjalan semakin optimal.

4.3.2 Pengujian Pengaruh Jumlah Dataset

Pengujian pengaruh jumlah dataset ini bertujuan untuk mengetahui tingkat efisien sistem terhadap banyak data yang diproses. Pada tahap ini pengujian dilakukan sebanyak lima kali yaitu pengujian dengan dataset sebanyak 20%, 30%, 40%, 50% dan 60% dari data seluruhnya. Setelah dilakukan proses pengujian sebanyak lima kali maka akan didapatkan hasil yang paling baik yang diambil dari nilai *k* dan jumlah dataset yang memiliki nilai silhouette coefficient paling tinggi.