

## BAB 2 LANDASAN KEPUSTAKAAN

### 2.1 Kajian Pustaka

Kajian pustaka yang dilakukan dalam penelitian ini adalah dengan cara membandingkan dua penelitian yang telah dilakukan sebelumnya mengenai deteksi penyakit kanker berdasarkan kandungan protein dalam tubuh dalam penelitian yang pernah dilakukan Arintha (2016) yaitu Penentuan Jenis Kanker Berdasarkan Struktur Protein Menggunakan Algoritma Modified K-Nearest Neighbor (MKNN) dan implementasi metode K-Medoids yang berdasarkan pada jurnal yang membahas Clustering Tag Status Facebook Dengan Menggunakan Algoritma K-Medoids (Candra, Sefia, dkk., 2012). Perbandingan kedua penelitian tersebut ditampilkan dalam Tabel 2.1.

**Tabel 2.1 Tabel Perbandingan 2 Penelitian**

No.	Judul	Obyek	Metode	Output
		<i>Input</i> dan Parameter	Proses	Hasil Penelitian
1.	Penentuan Jenis Kanker Berdasarkan Struktur Protein Menggunakan Algoritma Modified K-Nearest Neighbor (MKNN)	Kanker	Algoritma Modified K-Nearest Neighbor (MKNN)	Kelompok pasien yang menderita kanker dan non-kanker
		Struktur Protein	Klasifikasi pasien berdasarkan struktur protein menggunakan algoritma MKNN	Tingkat akurasi variabel K dari data struktur protein
2.	Clustering Tag Status Facebook Dengan Menggunakan Algoritma K-Medoids	Tag Status Facebook	Algoritma K-Medoids	Cluster teman pada tag facebook
		Source, Target, Visited friends	Menghitung jarak setiap data menggunakan metode K-Medoids	Mampu menemukan kelompok-kelompok dalam daftar teman facebook.

## 2.2 Landasan Teori

Landasan teori yang digunakan dalam penelitian ini berasal dari jurnal-jurnal, buku teks referensi, dan dari penelitian yang pernah dilakukan sebelumnya. Teori yang diambil tidak secara keseluruhan tetapi hanya teori yang dijadikan sebagai dasar penelitian saja. Dalam bab ini akan dijelaskan beberapa teori yang telah dipilih yaitu teori tentang penyakit kanker, protein tubuh, P53, Point Accepted Mutation, data mining, clustering, dan K-Medoids.

### 2.2.1 Kanker

Kanker adalah istilah umum untuk pertumbuhan sel tidak normal. Kanker sendiri digunakan untuk menggambarkan ratusan penyakit berbeda dengan fitur tertentu yang sama. Kanker dimulai dengan perubahan dalam struktur dan fungsi sel yang menyebabkan sel tumbuh dengan sangat cepat, membelah, dan menggandakan diri tanpa terkontrol yang dapat menyusup ke jaringan tubuh normal dan menekan jaringan tubuh sehingga mempengaruhi fungsi tubuh (Kelvin, Joanne Frankel., dkk., 2011).

Istilah tumor tidak sama dengan kanker. Tumor adalah istilah umum untuk setiap benjolan abnormal. Sedangkan kanker adalah tumor yang bersifat ganas (Diananda, Rama, 2011). Kanker menjadi momok bagi semua orang, hal ini karena angka kematian akibat kanker yang sangat tinggi. Tidak hanya di Indonesia melainkan juga di berbagai negara.

Pada tahun 2011, diperkirakan ada 1.334.100 kasus dengan angka kematian sebanyak 556.500 orang. Sedangkan di Eropa, terdapat tiga juta kasus kanker baru tiap tahun dengan angka kematian sebesar dua juta. Angka harapan hidup penderita kanker hanya 60% dibandingkan dengan bukan penderita (Diananda, Rama, 2011).

Dengan melihat sampel sel tumor di bawah mikroskop, ahli patologi dapat menentukan apakah itu merupakan tumor *benigna* (jinak) yang bersifat bukan kanker atau tumor *malignan* (ganas) yang bersifat kanker. Ahli patologi juga dapat menentukan *grade* (tingkat) tumor, yang merupakan ukuran seberapa abnormal sel terlihat jika diamati di bawah mikroskop; dalam beberapa kasus, grade memprediksi seberapa agresif kanker (Kelvin, Joanne Frankel., dkk., 2011). Untuk menentukan seberapa jauh sel abnormal menyebar, dokter dapat melakukan berbagai tes seperti tes darah, tomografi komputer (CT), pencitraan resonansi magnetik (MRI), *scan* radionuklida, tomografi emisi positron (PET), dan dalam beberapa jenis kanker, sampel sumsum tulang.

Bila sebuah sel mengalami perubahan dalam unit penyusunannya, yang disebut DNA, sel ini dapat keluar dari siklus hidup yang diatur secara ketat ini. Perubahan DNA ini juga disebut sebagai mutasi, yang dapat membuat sel-sel terus tumbuh dan mengalami pembelahan diri (Dizon, Don S., dkk., 2011). Sel-sel tersebut tidak lagi merespon terhadap sinyal yang dikirimkan badan untuk menghentikan pembelahan diri dan proses dari pembelahan sel yang tidak terkendali ini menghasilkan massa sel dari satu jenis. Bila sebuah sel tumor

terlepas dari tempat asalnya, sel itu dapat terbawa oleh aliran darah dan sampai di tempat lain dari organ tubuh seseorang yang jauh letaknya dan mulai tumbuh di tempat tersebut, keadaan ini disebut metastatik. Dua ciri ini (pertumbuhan sel tidak terkendali dan kemampuan untuk mengalami metastatik) yang dapat menetapkan kanker.

## 2.2.2 Protein Tubuh

Protein berasal dari bahasa Yunani yaitu *proteus* yang berarti “yang pertama” atau “yang terpenting”. Seorang ahli kimia dari Belanda bernama Mulder, mengisolasi susunan tubuh yang mengandung nitrogen dan menamakannya protein, terdiri dari satuan dasarnya asam amino. Molekul protein tersusun dari satuan-satuan dasar kimia yaitu asam amino. Satu molekul protein dapat terdiri dari 12 sampai 20 macam asam amino dan dapat mencapai jumlah ratusan asam amino (Wibisono, Yudi, 2011). Hubungan dari kode genetik terhadap protein sebagai pembawa sinyal, mengangkut molekul seperti oksigen, mengatur proses sel dan sebagai mekanisme pertahanan (Kurnianti, Ria, 2013)..

### 2.2.2.1 Kode Genetik Protein

Semenjak tahun 1960 semakin nyata bahwa ada paling sedikit tiga residu nukleotida DNA diperlukan untuk mengkode untuk masing-masing asam amino. Empat huruf kode DNA yaitu A, T, G dan C tersusun membentuk tiga huruf yang disebut dengan kodon. Kode genetik triplet tersebut dinamakan “bahasa mRNA” (messenger RNA) karena mRNA yang membawa *copy* dari kode genetik yang terdapat pada DNA. Dalam mRNA, timin (T) diganti dengan urasil (U), jadi kode singlet pada mRNA menjadi Adenin (A), Guanin (G), Cytosin (C), dan Urasil (U). Susunan tersebut terlihat pada Tabel 2.2.

**Tabel 2.2 Kode Genetik**

	U	C	A	G		
U	UUU UUC	UCU UCC UCA UCG	UAU UAC	UGU UGC		
	UUA UUG		UAA UAG	UGA UGG		
	C		CCU CCC CCA CCG	CAU CAC	CGU CGC CGA CGG	
				CAA CAG		
A		ACU ACC ACA ACG		AAU AAC		AGU AGC
				AAA AAG		AGA AGG
	G		GCU GCC GCA GCG	GAU GAC	GGU GGC GGA GGG	
				GAA GAG		

Pada prosesnya di dalam sel, terjadi proses transkripsi yaitu sintesis RNA dengan DNA sebagai cetakannya. RNA yang membawa sandi alias pesan yang sama dengan resep pada DNA ini kemudian bertindak sebagai cetakan untuk sintesis protein. Setiap kodon mengkodekan 1 asam amino. Sementara itu jumlah asam amino penyusun protein diketahui hanya 20 saja (dengan beberapa tambahan asam amino yang jarang). Dengan demikian berarti ada asam amino yang dikodekan oleh lebih dari satu kodon (Rahman, Dila, 2012).

Dengan dipecahkannya misteri kode genetik, kini kita bisa mengetahui protein apa yang dihasilkan suatu gen tanpa harus menganalisis proteinnya secara langsung, urutan basa DNA bisa diterjemahkan menjadi urutan asam amino protein (Rahman, Dila, 2012).

### **2.2.2.2 Gen P53**

Kode genetik yang mengatur pembentukan sel-sel kanker berhasil dipecahkan. Meski bentuk pengobatan dengan target gen masih terus dikembangkan, temuan ini membawa harapan baru untuk melawan penyakit mematikan tersebut.

Peta genetik yang dibuat para ilmuwan Amerika Serikat menunjukkan setidaknya terdapat sekitar 200 gen termutasi yang berperan mengatur pembentukan, pertumbuhan, dan penyebaran tumor (Diananda, Rama, 2011). Sebagian besar gen yang termutasi ini belum diketahui sebelumnya.

Penemuan ini bisa menjadi dasar pada ahli kedokteran untuk mengembangkan metode diagnosis kanker sejak stadium awal sehingga dapat segera ditangani. Selain itu, informasi genetik juga menjadi dasar untuk menyusun pengobatan yang tepat untuk setiap penderita kanker.

Menurut laporan dari beberapa percobaan, tes antibodi yang digunakan untuk melihat respon kekebalan tubuh terhadap kanker mungkin berguna untuk mengetahui resiko seseorang terhadap penyakit tersebut.

Menurut Dr. Thierry Soussi, dari Institut Curie di Paris, pengujian terhadap antibodi juga berfungsi untuk memonitor perkembangan seseorang yang tengah menjalani pengobatan kanker (Diananda, Rama, 2011). Ditemukan sejenis protein yang dikenal dengan nama "P53" berhubungan kuat dengan kanker.

Proses mutasi dalam gen P53 yang mampu mengontrol produksi protein, dipercaya sebagai penyebab dari 50% kasus keganasan, diantaranya kanker paru-paru, usus besar, esophagus, perut, payudara, dan beberapa jenis kanker otak. Gen P53 merupakan gen penekan tumor yang mengkode atau mengekspresikan protein 53, nama ini diambil dari berat molekulnya yang sebesar 53 kilodalton. Protein 53 merupakan faktor transkripsi terhadap gen-gen yang terlibat dalam regulasi siklus sel, induksi apoptosis, repair DNA, dan stabilitas genome.

Penemuan antibodi P53 dalam darah seseorang yang memiliki resiko tinggi kanker, seperti perokok berat, mengindikasikan bahwa para ahli memberikan harapan dalam pendeteksian kanker secara dini (Diananda, Rama, 2011). Hal itu

ditulis Soussi dalam jurnal Riset Kanker edisi Juni yang dipublikasikan *American Association for Cancer Research*.

Zat antibodi dihasilkan dari sistem kekebalan untuk melindungi tubuh dari serangan penyakit atau kerusakan akibat mutasi sel kanker. Menurut Soussi, para ahli belum dapat memahami mengapa tubuh dapat mengenali P53. Sedangkan di dalam kanker, P53 tidak dapat dikenali.

Proses mutasi ini dapat diwariskan seseorang secara genetik ataupun terjadi melalui kontak dengan bahan kimia, radiasi, dan kerusakan lainnya. Soussi mengatakan dia hanya menemukan satu penelitian yang mengungkapkan seseorang dipertimbangkan memiliki resiko tinggi terkena kanker. Kemudian dua perokok berat yang terbukti positif memiliki antibodi P53. Kedua perokok itu mengidap penyakit kanker (Diananda, Rama, 2011).

### **2.2.2.3 Point Accepted Mutation**

PAM (*Point Accepted Mutation*) merupakan sekumpulan PAM1 – PAM250 yang berasal dari penurunan *sequence* yang memiliki hubungan kekerabatan yang dekat (Kurnianti, Ria, 2013). Pada Tabel PAM titik yang termutasi pada protein adalah perubahan pada salah satu asam amino, yang terpilih secara alami.

Terdapat dua proses yang berbeda ketika terjadi perubahan pada asam amino, yaitu:

1. Yang pertama terjadinya mutasi pada bagian gen yang memproduksi asam amino dari protein.
2. Yang kedua terjadinya mutasi oleh jenis baru yang lebih dominan. Agar dapat diterima oleh asam amino yang lain biasanya asam amino yang baru ini membuat dirinya mirip dan mempunyai fungsi yang sama dengan asam amino yang lama.

Jumlah dari matriks PAM (PAM1, PAM250) menunjukkan sebuah evolusi jarak. Semakin besar jumlahnya maka semakin besar pula jaraknya (Dor, Shifra Ben, 2007). Untuk memperoleh nilai pada PAM maka dilakukan perkalian, contoh PAM2 diperoleh dari perkalian antara PAM1 dan PAM1, begitu pula dengan PAM3 diperoleh dari perkalian antara PAM1 dan PAM2, begitu seterusnya.

Matrik PAM1 merupakan dasar untuk menghitung matrik yang lain dengan anggapan mutasi yang berulang akan mengikuti aturan yang sama dengan matrik PAM1, dengan logika tersebut dapat diperoleh matrik PAM250 (Kurnianti, Ria, 2013).

Dari Tabel 2.3 diketahui terdapat 20 macam asam amino yang digunakan untuk mensintesis protein di ribosom. Sebab kedua puluh asam amino ini saja yang memiliki sandi genetik (Uyha, 2010).

Berikut ini adalah Gambar Tabel PAM :

Tabel 2.3 Matrik PAM250

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	2	-2	0	0	-2	0	0	1	-1	0	-2	-1	-1	-3	1	1	1	-6	-4	0
R	-2	6	0	-1	-4	1	-1	-3	2	-2	-3	3	0	-4	0	0	-1	2	-5	-2
N	0	0	2	2	-3	1	2	1	2	-2	-3	1	-2	-3	0	1	1	-4	-2	-2
D	0	-1	2	4	-5	2	3	1	1	-2	-4	0	-3	-5	-1	0	0	-7	-4	-2
C	-2	-3	-4	-5	12	-5	-5	-4	-3	-3	-6	-5	-5	-4	-2	0	-2	-8	0	-2
Q	0	1	1	2	-5	4	2	-1	3	-2	-2	1	-1	-4	0	-1	-1	-5	-4	-2
E	0	-1	1	3	-5	2	4	0	1	-2	-3	0	-2	-5	0	0	0	-7	-4	-2
G	1	-3	0	1	-3		0	5	-2	-2	-4	-2	-3	-5	0	1	0	-7	-5	-1
H	-1	1	1	1	-3	3	0	-3	6	-3	-3	0	-3	-2	0	-1	-1	-3	0	-3
I	-1	-2	-2	-2	-2	-2	-2	-3	-3	4	2	-2	2	1	-2	-1	0	-5	-1	4
L	-2	-3	-3	-4	-6	-2	-3	-3	-2	2	6	-2	4	2	-2	-3	-2	-2	-1	2
K	-1	3	1	0	-5	1	0	-2	0	-2	-3	5	0	-5	-1	0	0	-4	-5	-2
M	-1	-1	-2	-3	-5	-1	-2	-3	-2	2	4	1	6	0	-2	-2	0	-4	-3	2
F	-3	-4	-3	-5	-4	-4	-5	-4	-2	1	2	-5	0	9	-5	-3	-3	0	7	-1
P	1	0	0	-1	-3	0	0	0	0	-2	-2	-1	-2	-4	6	1	0	-6	-5	-1
S	2	1	2	1	1	0	1	2	0	-1	-2	1	-1	-2	2	2	2	-2	-2	0
T	0	-2	0	-1	-3	-2	-1	-1	-2	-1	-2	-1	-1	-4	0	1	2	-6	-4	0
W	-6	2	-5	-7	-7	-6	-7	-7	-5	-6	-7	-4	-6	1	-6	-2	-5	17	1	-8
Y	-3	-5	-2	-4	1	-4	-4	-5	0	-1	-1	-5	-2	7	-5	-3	-3	0	10	-2
V	0	-2	-2	-2	-2	-2	-2	-2	-2	4	2	-2	2	-1	-1	-1	0	-6	-3	4

Macam-macam residu asam amino yang diurutkan dari atas ke bawah yaitu Sistein Alanin (Ala) = A, Arginin (Arg) = R, Asparagin (Asn) = N, Asam Aspartat (Asp) = D, (Cys) = C, Glutamin (Gin) = Q, Asam Glutamat (Glu) = E, Glisin (Gly) = G, Histidin (His) = H, Isoleusin (Ile) = I, Leusin (Leu) = L, Lisin (Lys) = K, Metionin (Met) = M, Fenilalanin (Phe) = F, Prolin (Pro) = P, Serin (Ser) = S, Treonin (Tgr) = T, Triptofan (Trp) = W, Tirosin (Tyr) = Y, Valin (Val) = V (Kurnianti, Ria, 2013).

### 2.2.3 Data Mining

Data mining adalah serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual. Data mining merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar. Data mining sering juga disebut *knowledge discovery in database (KDD)*, adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari *data mining* ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan. *Data mining* adalah kegiatan menemukan pola yang menarik dari data dalam jumlah besar, data dapat disimpan dalam *database*, *data warehouse*, atau penyimpanan informasi lainnya. *Data mining* berkaitan dengan bidang ilmu-ilmu

lain, seperti *database sistem*, *data warehousing*, statistik, *machine learning*, *information retrieval*, dan komputasi tingkat tinggi. Selain itu, *data mining* didukung oleh ilmu lain seperti *neural network*, pengenalan pola, *spatial data analysis*, *image database*, *signal processing* (Han, J., M. Kamber, 2006). *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar.

Masalah-masalah yang sesuai untuk diselesaikan dengan teknik *data mining* dapat dicirikan dengan:

- Memerlukan keputusan yang bersifat *knowledge-based*.
- Mempunyai lingkungan yang berubah.
- Metode yang ada sekarang bersifat sub-optimal.
- Tersedia data yang bisa diakses, cukup dan relevan.
- Memberikan keuntungan yang tinggi jika keputusan yang ambil tepat.

Berdasarkan beberapa pengertian tersebut dapat ditarik kesimpulan bahwa *data mining* adalah suatu teknik menggali informasi berharga yang terpendam atau tersembunyi pada suatu koleksi data (*database*) yang sangat besar sehingga ditemukan suatu pola yang menarik yang sebelumnya tidak diketahui. Kata *mining* sendiri berarti usaha untuk mendapatkan sedikit barang berharga dari sejumlah besar material dasar. Karena itu *data mining* sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (*artificial intelligent*), *machine learning*, statistik dan *database*. Beberapa metode yang sering disebut-sebut dalam literatur *data mining* antara lain *clustering*, *classification*, *association rules mining*, *neural network*, *genetic algorithm* dan lain-lain.

*Data mining* sering digunakan untuk membangun model prediksi/inferensi yang bertujuan untuk memprediksi tren masa depan atau perilaku berdasarkan analisis data terstruktur. Dalam konteks ini, prediksi adalah pembangunan dan penggunaan model untuk menilai kelas dari contoh tanpa label, atau untuk menilai jangkauan nilai atau contoh yang cenderung memiliki nilai atribut. Klasifikasi dan regresi adalah dua bagian utama dari masalah prediksi, dimana klasifikasi digunakan untuk memprediksi nilai diskrit atau nominal sedangkan regresi digunakan untuk memprediksi nilai terus-menerus atau nilai yang ditentukan.

### **2.2.3.1 Clustering**

*Clustering* adalah metode analisis data, yang sering dimasukkan sebagai salah satu metode *data mining*. Tujuannya yaitu untuk mengelompokkan data dengan karakteristik yang sama ke suatu 'wilayah' yang sama dan data dengan karakteristik yang berbeda ke 'wilayah lain. Langkah awal dari metode ini adalah dengan mendapatkan subset-subset yang terhubung untuk kemudian mencari defisini dari masing-masing subset. *Clustering* dan segmentasi sebenarnya mempartisi *database*, karena itu setiap partisi atau group menurut kriteria atau matrix tertentu. Jika pengukuran kesamaan tersedia, maka terdapat sejumlah teknik untuk membentuk *cluster* (Shibab, A., 2000).

### 2.2.3.2 K-Medoids

Algoritma K-medoids atau dikenal pula dengan *PAM (Partitioning Around Medoids)* menggunakan metode partisi *clustering* untuk mengelompokkan sekumpulan  $n$  objek menjadi sejumlah  $k$  *cluster*. Algoritma ini menggunakan objek pada kumpulan objek untuk mewakili sebuah *cluster*. Objek yang terpilih untuk mewakili sebuah *cluster* disebut *medoid*. *Cluster* dibangun dengan menghitung kedekatan yang dimiliki antara *medoid* dengan objek *non-medoid* (Han, J., M. Kamber, 2006).

Menurut Han dan Kamber (2006) algoritma K-medoids adalah sebagai berikut.

1. Secara acak pilih  $k$  objek pada sekumpulan  $n$  objek sebagai *medoid*.
2. Ulangi:
3. Tempatkan objek *non-medoid* ke dalam *cluster* yang paling dekat dengan *medoid*. Lakukan perhitungan jarak menggunakan jarak *Manhattan*, dengan rumus :

$$d(i, j) = \sum_{i=1}^n |x(i_i) - x(j_j)| + |y(i_i) - y(j_j)| \quad (2.1)$$

4. Secara acak pilih  $o_{acak}$ , sebuah objek *non-medoid*.
5. Hitung nilai cost pada setiap objek dengan rumus :

$$cost(i) = \min\{dC1(i), dC2(i), \dots, dCn(i)\} \quad (2.2)$$

6. Dilanjutkan dengan menghitung total biaya atau total cost,  $S$ , dari pertukaran medoid  $o_j$  dengan  $o_{random}$  dengan rumus :

$$S (total\ cost) = \sum_{i=1}^n cost(i) \quad (2.3)$$

7. Jika  $S < 0$  maka tukar  $o_j$  dengan  $o_{acak}$  untuk membentuk sekumpulan  $k$  objek baru sebagai *medoid*.
8. Hingga tidak ada perubahan.

### 2.2.3.3 Silhouette Coefficient

Metode evaluasi yang akan digunakan pada sistem ini adalah metode *silhouette coefficient*. Metode ini berfungsi untuk menguji kualitas dari *cluster* yang dihasilkan. Setelah itu tahapan untuk menghitung nilai *silhouette coefficient* adalah sebagai berikut :

1. Untuk setiap objek  $i$ , hitung rata-rata jarak dari objek  $i$  dengan seluruh objek yang berada dalam satu *cluster*. Akan didapatkan nilai rata-rata yang disebut  $a_i$ .
2. Untuk setiap objek  $i$ , hitung rata-rata jarak dari objek  $i$  dengan objek yang berada di *cluster* lainnya. Dari semua jarak rata-rata tersebut ambil nilai yang paling kecil. Nilai ini disebut  $b_i$ .
3. Setelah itu maka untuk objek  $i$  memiliki nilai *silhouette coefficient* :

$$S(i) = \frac{b(i) - a(i)}{\max(b(i), a(i))} \quad (2.4)$$

atau bisa ditulis dengan,

$$S(i) = \begin{cases} 1 - \frac{a(i)}{b(i)}, & \text{if } a(i) < b(i) \\ 0, & \text{if } a(i) = b(i) \\ \frac{b(i)}{a(i)} - 1, & \text{if } a(i) > b(i) \end{cases} \quad (2.5)$$

Hasil perhitungan nilai *silhouette coefficient* dapat bervariasi antara -1 hingga 1. Hasil *clustering* dikatakan baik jika nilai *silhouette coefficient* bernilai positif ( $a_i < b_i$ ) dan  $a_i$  mendekati 0, sehingga akan menghasilkan nilai *silhouette coefficient* yang maksimum yaitu 1 saat  $a_i = 0$ . Maka dapat dikatakan, jika  $s_i = 1$  berarti objek  $i$  sudah berada dalam *cluster* yang tepat. Jika nilai  $s_i = 0$  maka objek  $i$  berada di antara dua *cluster* sehingga objek tersebut tidak jelas harus dimasukkan ke dalam *cluster* A atau *cluster* B. Akan tetapi, jika  $s_i = -1$  artinya struktur *cluster* yang dihasilkan *overlapping*, sehingga objek  $i$  lebih tepat dimasukkan ke dalam *cluster* yang lain. Nilai rata-rata *silhouette coefficient* dari tiap objek dalam suatu *cluster* adalah suatu ukuran yang menunjukkan seberapa ketat data dikelompokkan dalam *cluster* tersebut.