

BAB II

TINJAUAN PUSTAKA

2.1 Tinjauan Umum Perusahaan

Tinjauan ini bertujuan untuk memahami situasi di Klinik Keluarga, yang bermula dari sejarahnya, logo perusahaan dan makna di balik logo tersebut, kemudian struktur organisasi perusahaan.

2.1.1 Sejarah Perusahaan

Klinik Keluarga didirikan pada tahun 2004 yang berlokasi di Kampung Maleber, Ciherang, Pacet, Kabupaten Cianjur. Kemudian untuk memberikan kemudahan akses pelayanan kesehatan dan mengurasi biaya transportasi saat berobat, di wilayah Perkebunan The Ciseureuh, Loji – Cipanas pada tahun 2008 didirikan Klinik Tazkiyah.[5]

Untuk memenuhi kebutuhan akan pelayanan kesehatan yang lebih baik, Klinik ini harus memperluas layanannya dengan menyediakan pasien rawat inap sementara (*observasi*) dan tindakan operasi minor. Namun, fasilitas yang ada tidak cukup memadai untuk memenuhi kebutuhan tersebut. Sebagai solusinya, didirikanlah Klinik Keluarga di Kampung Cigombong, Ciherang, Pacet, Kabupaten Cianjur. Klinik Keluarga memiliki lahan seluas 1125 m² dan diresmikan pada tanggal 17 April 2013. Layanan ini bertujuan untuk membantu pasien yang kesulitan dalam mendapatkan akses pelayanan kesehatan yang memuaskan[5].

Klinik menyatakan bahwa tujuan mereka bukanlah dimaksudkan sebagai tempat pelayanan kesehatan yang setara dengan rumah sakit daerah atau klinik medis yang ada saat ini, melainkan sebagai tempat penyehatan atau rumah sehat. Klinik ini akan menjadi basis bagi pasien yang ingin meningkatkan kesehatannya, dan juga menjadi tempat berkumpul keluarga, di mana terjalin hubungan kekeluargaan dan kasih sayang yang mendorong semangat untuk menjaga kesehatan[5].

2.1.2 Logo Perusahaan

Logo merupakan identitas diri suatu perusahaan. Berikut merupakan logo Klinik Keluarga yang dapat dilihat pada Gambar 2.1.



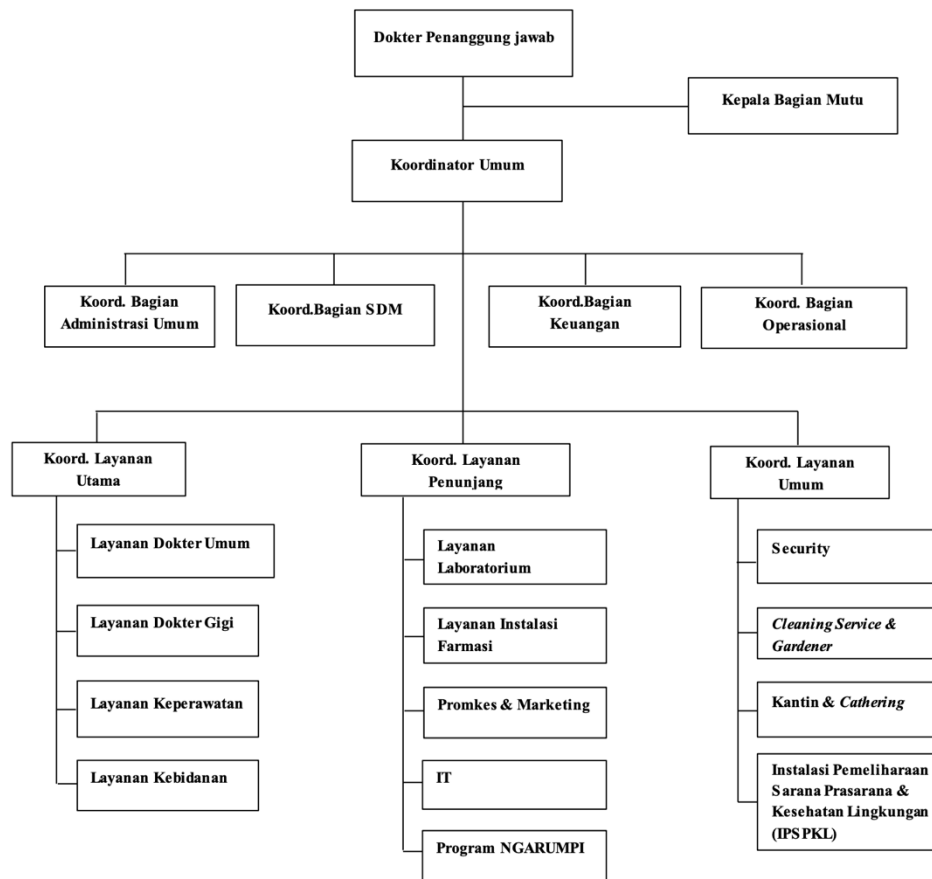
Gambar 2.1 Logo Perusahaan

Logo tersebut memiliki makna bahwa tiga bunga membentuk satu keluarga yang terdiri dari ayah, ibu, dan anak. Bunga tulip yang dipilih sebagai bentuk bunga dalam logo melambangkan "Cinta yang Sempurna" dan melambangkan cinta dalam keluarga. Klinik keluarga menekankan pentingnya hubungan keluarga dalam memberikan pelayanan terbaik kepada masyarakat. Logo ini mencerminkan bahwa Klinik Keluarga tidak hanya tempat pengobatan tetapi juga tempat komunikasi dan pertukaran pengetahuan tentang cara hidup sehat keluarga. Klinik keluarga bertujuan untuk mempromosikan semangat hidup sehat dan bukan hanya fokus pada pengobatan pasien. Lambang ini secara khusus menampilkan bagian dari tubuh manusia yang terikat dalam satu kesatuan kecil (team)[5].

Warna biru dalam logo melambangkan sifat tenang dan profesional, yang juga diasosiasikan dengan kepercayaan, keandalan, dan sifat maskulin. Di sisi lain, warna hijau melambangkan pertumbuhan dan juga melambangkan nilai-nilai agama yang dipromosikan oleh Klinik Keluarga dengan mengedepankan akhlakul karimah. Warna hijau juga menggambarkan kenyamanan dan ketenangan yang dapat ditemukan di Klinik Keluarga, dengan lingkungan yang bersih dan asri. Warna merah dalam logo melambangkan femininitas dan digunakan untuk memacu semangat serta membangkitkan gairah dalam memberikan pelayanan terbaik bagi pasien[5].

2.1.3 Struktur Organisasi Klinik Keluarga

Klinik Keluarga merupakan bagian dari PT. Abdi Yakin Anugerah. Berikut merupakan struktur organisasi di Klinik Keluarga yang dapat dilihat pada Gambar 2.2



Gambar 2.2 Struktur Organisasi Perusahaan

Struktur organisasi yang diambil adalah bagian Promkes & Marketing yang berada dibawah Koord. Layanan Penunjang sebagai bagian yang mendukung dalam pembangunan sistem data mining ini. Adapun daftar pekerjaan dari bidang Promkes adalah sebagai berikut:

1. Melaksanakan penyusunan IKU Divisi.
2. Melaksanakan pengisian penilaian IKI Divisi.
3. Melaksanakan pengisian penilaian Service.
4. Melakukan penyusunan PICA.
5. Melaksanakan rapat bulanan.

6. Melaksanakan rapat divisi dan coffee break mingguan.
7. Melaksanakan kunjungan offline penyuluhan kesehatan ke daerah, sekolah dan hotel.
8. Melaksanakan pengelolaan share edukasi group komunitas telegram.
9. Melaksanakan share informasi layanan (group komunitas).

2.2 Landasan Teori

Landasan teori merupakan penjelasan seluruh konsep dasar dan teori-teori yang terstruktur secara sistematis mengenai topik yang berkaitan dalam sebuah penelitian. Landasan teori memiliki nilai yang penting penting dalam sebuah penelitian terutama dalam penulisan skripsi atau tugas akhir. Konsep dan teori yang tercakup dalam landasan teori meliputi data *mining*, metode penelitian, algoritma data *mining*. Berikut penjelasannya:

2.2.1 Data

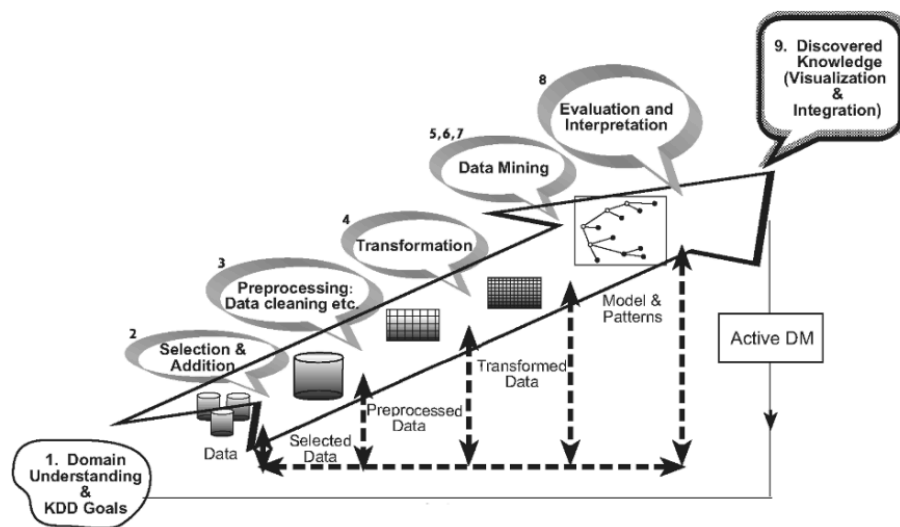
Data adalah sekumpulan keterangan ataupun fakta yang dibuat dengan kata-kata, kalimat, simbol, angka, dan lainnya. Data disini didapatkan melalui sebuah proses pencarian dan juga pengamatan yang tepat berdasarkan sumber-sumber tertentu. Adapun pengertian lain dari data yaitu sebagai suatu kumpulan keterangan atau deskripsi dasar yang berasal dari obyek ataupun kejadian [6]. Data sendiri tidak memiliki makna, sehingga kita masih perlu memproses dan memanipulasi data tersebut. Data memiliki kumpulan item, yang terdiri dari objek dan atribut. Atribut-atribut adalah karakteristik atau sifat dari objek yang disebut variabel atau properti.

2.2.2 Data Mining

Data *mining* merupakan proses penggalian informasi dan pola yang bermanfaat dari data yang sangat besar. Data *mining* mencakup pengumpulan data, ekstraksi data, analisis data, dan statistik data. Data *mining* juga dikenal sebagai *knowledge discovery*, *knowledge extraction*, *data/pattern analysis*, *information harvesting*, dan lain-lain. Data *mining* juga merupakan proses untuk menemukan informasi yang berguna. Setelah ditemukan informasi dan pola dapat digunakan untuk alat pendukung dalam pengambilan keputusan [7].

2.2.3 Metodologi Knowledge Discovery in Databases (KDD)

Pada penelitian ini metode yang digunakan dalam penyelesaian data mining adalah kerangka kerja *Knowledge Discovery in Databases (KDD)*. KDD adalah proses untuk menemukan pola atau informasi yang menarik atau berguna dari berbagai sumber data seperti database, teks, gambar, web, dll. Hasil pola atau informasi yang didapat harus valid, memiliki potensi manfaat, dan mudah dipahami [4].



Gambar 2.3 Alur Proses Metodologi KDD

Berikut adalah tahapan dari KDD :

1. *Domain Understanding & KDD Goals*, yaitu membangun pemahaman tentang domain aplikasi. Hal ini dilakukan sebagai persiapan awal untuk memahami apa yang harus dilakukan dengan banyak pertimbangan seperti transformasi, algoritma, representasi, dan yang lainnya yang diperlukan. Orang yang bertanggung jawab atas proyek KDD perlu memahami dan menentukan tujuan pengguna akhir serta lingkungan di mana proses penemuan pengetahuan akan dilakukan, termasuk pengetahuan sebelumnya yang relevan. Selama proses KDD berlangsung, kemungkinan akan ada revisi dan penyesuaian dalam langkah ini.

2. *Selection & Addition*, setelah menentukan tujuan, data yang akan digunakan untuk penemuan pengetahuan harus ditentukan. Ini termasuk mencari tahu data apa yang tersedia, memperoleh data tambahan yang diperlukan, dan kemudian mengintegrasikan semua data untuk penemuan pengetahuan ke dalam satu kumpulan data, termasuk atribut yang akan dipertimbangkan untuk proses. Proses ini sangat penting karena data *mining* belajar dan menemukan dari data yang tersedia. Ini adalah dasar bukti untuk membangun model. Jika beberapa atribut penting hilang, maka seluruh studi mungkin gagal. Dari keberhasilan proses, ada baiknya mempertimbangkan sebanyak mungkin atribut pada tahap ini. Di sisi lain, untuk mengumpulkan, mengatur, dan mengoperasikan repositori data yang kompleks itu mahal, dan ada pertukaran dengan peluang untuk memahami fenomena dengan baik. Tradeoff ini merupakan aspek di mana aspek interaktif dan iteratif dari KDD berlangsung. Ini dimulai dengan kumpulan data terbaik yang tersedia dan kemudian berkembang dan mengamati efeknya dalam hal penemuan dan pemodelan pengetahuan.
3. *Preprocessing, Data Cleaning* pada langkah ini, keandalan data ditingkatkan. Hal ini bagian dari pembersihan data, seperti menangani nilai yang hilang dan menghilangkan *noise* atau *outlier*.
4. *Transformation*, pada tahap transformasi dalam data *mining*, data yang akan digunakan dalam penambangan data dipersiapkan dan dikembangkan untuk menjadi lebih baik. Metode ini termasuk pengurangan dimensi data seperti pemilihan dan ekstraksi fitur, serta pengambilan sampel rekaman, dan transformasi atribut seperti diskritisasi atribut numerik dan transformasi fungsional juga dilakukan. Langkah ini sering kali penting untuk keberhasilan proyek KDD, tetapi biasanya sangat spesifik untuk proyek tertentu. Misalnya, dalam pemeriksaan medis, hasil bagi atribut mungkin menjadi faktor yang paling penting. Dalam pemasaran, efek di luar kendali serta upaya dan masalah temporal perlu dipertimbangkan. Meskipun transformasi yang tepat mungkin tidak digunakan pada awalnya, proses KDD dapat

mengarah pada pemahaman tentang transformasi yang dibutuhkan melalui iterasi berikutnya.

5. *Data Mining* memilih tugas data mining yang sesuai. Memutuskan jenis data mining yang akan digunakan, misalnya, klasifikasi, regresi, atau pengelompokan. Ini sebagian besar tergantung pada tujuan KDD, dan juga pada langkah-langkah sebelumnya. Ada dua tujuan utama dalam data *mining*: prediksi dan deskripsi. *Prediction* sering disebut sebagai supervised data *mining*, sedangkan deskriptif data *mining* mencakup aspek *unsupervised* dan visualisasi dari data *mining*. Sebagian besar teknik data *mining* didasarkan pada pembelajaran induktif, di mana model dibangun secara eksplisit atau implisit dengan menggeneralisasi dari sejumlah contoh pelatihan yang memadai. Asumsi yang mendasari pendekatan induktif adalah bahwa model yang dilatih dapat diterapkan untuk kasus-kasus masa depan. Strategi ini juga memperhitungkan tingkat *meta-learning* untuk kumpulan data tertentu yang tersedia.
6. *Data Mining* kemudian tahap data *mining* yang kedua adalah memilih metode spesifik yang akan digunakan untuk pola pencarian (termasuk beberapa penginduksi). Untuk setiap strategi *meta-learning* ada beberapa kemungkinan bagaimana hal itu dapat dicapai. *Meta-learning* berfokus untuk menjelaskan apa yang menyebabkan suatu algoritma data *mining* berhasil atau tidak dalam suatu masalah tertentu. Dengan demikian, pendekatan ini mencoba untuk memahami kondisi di mana algoritma data *mining* paling tepat. Setiap algoritma memiliki parameter dan taktik pembelajaran.
7. *Data Mining* tahap terakhir dalam langkah ini kita mungkin perlu menggunakan algoritma beberapa kali sampai hasil yang memuaskan diperoleh, misalnya dengan menyetel parameter kontrol algoritma, seperti jumlah minimum instance dalam satu daun pohon keputusan.
8. *Evaluation*. Pada tahap ini, kami mengevaluasi dan menafsirkan pola-pola yang ditemukan (aturan, kehandalan, dll.), dengan mengacu pada tujuan yang didefinisikan pada langkah pertama. Di sini kami mempertimbangkan tahap *preprocessing* dengan mengacu pada efeknya

pada hasil algoritma penambahan data (misalnya, menambah fitur pada langkah 4, dan memulai dari sana). Tahap ini memfokuskan pada kejelasan dan kegunaan model yang dihasilkan. Pada tahap ini, pengetahuan yang ditemukan juga didokumentasikan untuk digunakan kembali.

9. *Integration*. Dalam langkah ini siap untuk memasukkan pengetahuan ke dalam sistem lain untuk tindakan lebih lanjut. Pengetahuan menjadi aktif dalam arti bahwa kita dapat membuat perubahan pada sistem dan mengukur efeknya. Sebenarnya, keberhasilan langkah ini menentukan efektivitas dari seluruh proses KDD. Ada banyak tantangan dalam langkah ini, seperti kehilangan kondisi "laboratorium" di bawah mana kita telah beroperasi. Misalnya, pengetahuan ditemukan dari gambar diam tertentu (biasanya sampel) dari data, tetapi sekarang data menjadi dinamis. Struktur data dapat berubah (beberapa atribut menjadi tidak tersedia), dan domain data dapat dimodifikasi (seperti, atribut mungkin memiliki nilai yang tidak diasumsikan sebelumnya).

2.2.4 Rekam Medis

Rekam Medis adalah berkas yang berisikan catatan dan dokumen tentang identitas pasien, pemeriksaan, pengobatan, tindakan dan pelayanan lain yang telah diberikan kepada pasien[8]

2.2.5 Clustering

Clustering (Cluster Analysis) adalah proses pembagian sejumlah objek data (atau observasi) menjadi *subset*. Setiap *subset* adalah sebuah *cluster*, sehingga objek dalam satu *cluster* mirip satu sama lain, tetapi tidak mirip dengan objek dalam *cluster* lain. Kelompok *cluster* yang dihasilkan dari analisis *cluster* disebut *clustering*. Dalam konteks ini, metode *clustering* yang berbeda dapat menghasilkan *clustering* yang berbeda pada data yang sama. Pembagian ini tidak dilakukan oleh manusia, tetapi oleh algoritma *clustering*. Oleh karena itu, *clustering* berguna karena dapat menemukan kelompok yang sebelumnya tidak dikenal dalam data [3]. Pada *clustering* terdapat beberapa metode salah satu diantaranya yaitu metode *K-*

Means. Metode tersebut memiliki tujuan untuk meminimalkan jarak antara titik yang ditentukan sebagai pusat *cluster* dan titik data *cluster* [9].

2.2.6 K-Means

K-Means adalah salah satu metode *clustering* yang termasuk kedalam pendekatan *partitioning*. Partisi data set dilakukan untuk menentukan karakteristik masing-masing cluster, sehingga cluster yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster dan yang memiliki karakteristik yang berbeda dikelompokkan ke dalam cluster lainnya [10]. Algoritma K-Means merupakan suatu model *centroid*. Model *centroid* merupakan model yang menggunakan *centroid* untuk membuat suatu *cluster*. Centroid merupakan titik tengah suatu *cluster* yang berupa *centroid* nilai untuk digunakan dalam menghitung jarak suatu obyek data terhadap *centroid*. Suatu obyek data akan dikelompokkan ke dalam *cluster* apabila memiliki jarak terpendek terhadap *centroid*. Algoritma K-Means merupakan algoritma pembelajaran sederhana yang bertujuan untuk memecahkan suatu permasalahan pengelompokkan yang bertujuan untuk meminimalkan kesalahan ganda. Nilai *centroid* pada interaksi pertama dilakukan secara acak [11]. Secara umum penerapan algoritma kmeans dapat dilakukan dalam beberapa tahapan yaitu sebagai berikut:

1. Menentukan banyaknya jumlah k yaitu *cluster* yang telah ditentukan menggunakan *Elbow Method* dengan menghitung nilai dari *Sum of Square Error* (SSE), berikut adalah rumus SSE yang digunakan:

$$SSE = \sum_{k=1}^K \sum_{X_i \in S_k} ||X_i - C_k||^2 \quad (2.1)[12]$$

Keterangan:

K = jumlah *cluster*

X_i = data ke - i

C_k = *centroid cluster*

Sum of Square Error (SSE) merupakan rumus yang digunakan untuk mengukur perbedaan antara data yang diperoleh dengan model perkiraan yang telah dilakukan sebelumnya[13]. SSE sering digunakan sebagai acuan penelitian terkait dalam menentukan optimal *cluster*[14].

2. Memilih nilai awal *centroid* secara acak pada dataset[15].
3. Menghitung masing-masing jarak setiap data terhadap *centroid* yang telah ditentukan menggunakan persamaan *Euclidean Distance*. Persamaanya adalah sebagai berikut:

$$d(x, y) = \sqrt{\sum_{t=1}^n (x_t - y_t)^2} \quad (2.2)$$

Keterangan:

$d(x, y)$ = Jarak data ke x pusat *cluster*

x_i = Data ke-i pada atribut ke-n

y_i = Data ke-j pada atribut ke-n

n = Banyaknya data

Pada persamaan diatas data akan dikelompokkan berdasarkan karakteristik yang sama dan akan dimasukkan pada kelompok yang sama sehingga data terkelompok [16].

4. Menentukan kelompok setiap data berdasarkan jarak terdekat terhadap nilai *centroid*-nya.
5. Memperbaharui nilai *centroid* baru dengan persamaan berikut:

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj} \quad (2.3)$$

Keterangan:

V_{ij} = Nilai *centroid* rata-rata *cluster* ke-i untuk variabel ke-i

N_i = Nilai jumlah *cluster* ke-i

i, k = Nilai indeks dari *cluster*

j = Nilai indeks dari variabel

X_{kj} = Nilai data ke-k variabel ke-j dalam *cluster*

6. Mengulang proses menghitung jarak data terhadap *centroid* baru yang terbentuk hingga tidak ada lagi perubahan posisi pada nilai *centroid*.

2.2.7 Davies Bouldin Index (DBI)

Davies Bouldin Index merupakan salah satu metode yang dipakai untuk mengevaluasi *cluster* yang diperkenalkan oleh David L.Davies dan Donald W.Bouldin[17]. DBI adalah salah satu cara untuk mengukur kualitas *cluster* pada *clustering* [18]. DBI merupakan suatu fungsi rasio dari sejumlah distribusi kedalam

cluster sebagai pemisah antar *cluster*. Bentuk pendekatan pada pengujian nilai DBI meliputi nilai separasi dan kohesi. Separasi ialah jarak antara pusat *cluster* dari *cluster*. Kohesi berupa jumlah dari kemiripan data terhadap pusat *cluster* dari *cluster* tersebut. *Cluster* dikatakan optimal jika *cluster* tersebut memiliki nilai kohesi yang rendah sedangkan nilai separasi yang tinggi. Ada empat tahapan dalam menghitung DBI diantaranya yaitu [19]:

1. Menghitung *Sum Of Square Within Cluster (SSW)* adalah keterikatan anggota satu *cluster* atau seberapa mirip antara anggota satu dan dua dan semakin kecil semakin bagus dikarenakan semakin mirip. SSW dihitung untuk mengetahui matrik/kohesi/homogenitas. Kohesi merupakan keterikatan anggota *cluster* dalam satu *cluster*.

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i) \quad (2.4)$$

Keterangan :

m_i = jumlah data dalam *cluster* ke-i

x = data dalam *cluster*

$d(x, c)$ = jarak data ke *centroid*

x_j = data pada *cluster* tersebut

c_i = *centroid cluster* ke-i

2. Menghitung *Sum Of Square between cluster (SSB)* merupakan jarak antar *cluster* cukup besar sehingga terpisah ke dalam kelompok lain. SSB bertujuan untuk mengetahui separasi/heterogenitas. Separasi merupakan perbedaan antara satu *cluster* dengan *cluster* lainnya.

$$SSB_{i,j} = d(c_i, c_j) \quad (2.5)$$

Keterangan :

C_i = *Cluster* 1

C_j = *Cluster* lainnya

$d(c_i, c_j)$ = Jarak antara *centroid* satu dengan lainnya

3. Menghitung Rasio berfungsi untuk dapat mengetahui seberapa bagus nilai perbandingan *cluster* satu dengan *cluster* lainnya. Jumlah kohesi harus kecil sedangkan jumlah separasi harus lebih besar.

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{ij}} \quad (2.6)$$

Keterangan :

$R_{i,j}$ = Rasio antar *cluster*

SSW_i = *cluster* 1

SSW_j = *cluster* 2

SSB_{ij} = separasi dari *cluster* 1 dan 2

4. Menghitung Davies Bouldin Indeks

Faktanya jika hasil dari perhitungan DBI yang diperoleh semakin kecil mendekati nol akan tetapi tidak negatif (non-negatif ≥ 0), maka nilai hasil *clustering* semakin baik.

$$DBI_i = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} R_{i,j} \quad (2.7)$$

Keterangan :

k = *cluster* yang ada

$R_{i,j}$ = rasio antara *cluster* i dan j

Max = dicari rasio antar *cluster* yang terbesar

2.2.8 Black Box Testing

Black Box Testing merupakan pengujian yang dilakukan hanya dengan mengamati hasil eksekusi melalui data uji dan memeriksa fungsional dari perangkat lunak. Pengujian ini tujuannya untuk mengetahui bahwa bagian-bagian dalam sistem aplikasi telah benar menampilkan sesuai dengan perancangan program [20].