BAB 2
TINJAUAN PUSTAKA

2.1. Tinjauan Tempat Penelitian

Tinjauan tempat penelitian adalah untuk mengetahui keadaan di tempat penelitian di antaranya adalah sejarah berdirinya, visi dan misi, struktur organisasi, serta deskripsi kerja dari pihak yang nantinya terlibat di dalam sistem.

2.1.1. Profil Puskesmas

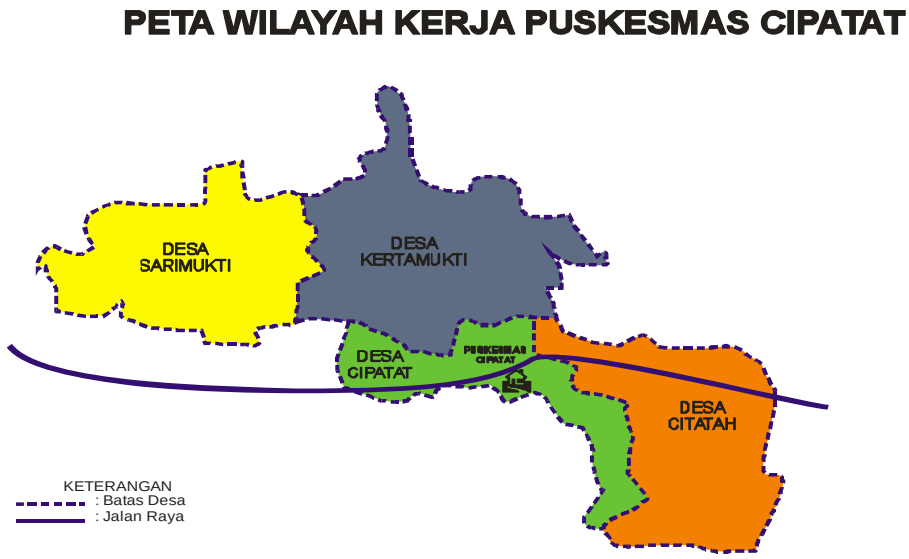
Pada bagian ini akan di jelaskan gambaran umum mengenai Puskesmas Cipatat Kabupaten Bandung Barat.

1. Geografis

Puskesmas Cipatat terletak di Jalan Raya Cipatat No. 28 Desa Cipatat Kecamatan Cipatat Kabupaten Bandung Barat yang berjarak ± 15 km dari Pusat Pemerintahan Kabupaten Bandung Barat, dengan jarak tempuh ± 30 menit menggunakan kendaraan roda 4. Secara administratif pemerintahan Puskesmas Cipatat memiliki cakupan wilayah kerja terdiri dari Desa Cipatat, Desa Citatah, Desa Kertamukti, dan Desa Sarimukti.

Luas wilayah kerja Puskesmas Cipatat yaitu 4.355,94 Ha dengan batas wilayah administratif sebagai berikut :

a. Sebelah Utara berbatasan dengan Kecamatan Cipendeuy, Wilayah Kerja Puskesmas Cipendeuy.
b. Sebelah Selatan berbatasan dengan Desa Cikande, Wilayah Kerja Puskesmas Batujajar.
c. Sebelah Barat berbatasan dengan Desa Ciptaharja, Wilayah Kerja Puskesmas DTP Rajamandala.
d. Sebelah Timur berbatasan dengan Desa Gunung Masigit, Wilayah Kerja Puskesmas Sumur Bandung.



Gambar 2.1 Peta Wilayah Kerja Puskesmas Cipatat

2. Demografis

Jumlah penduduk di wilayah Puskesmas Cipatat adalah 41.704 jiwa, adapun penyebarannya sebagai berikut :

a. Desa Cipatat : 3399 KK, jumlah penduduk 12.150 jiwa
b. Desa Citatah : 5164 KK, jumlah penduduk 16.562 jiwa
c. Desa Kertamukti : 2299 KK, jumlah penduduk 7.638 jiwa
d. Desa Sarimukti : 1654 KK, jumlah penduduk 5.354 jiwa

3. Sumber daya

Puskesmas Cipatat dikepalai oleh seorang Sarjana Kesehatan Masyarakat yang memimpin, mengawasi, dan mengkoordinasikan kegiatan puskesmas. Dalam melaksanakan tugas tersebut kepala puskesmas dibantu bagian tata usaha yang bertugas di bidang kepegawaian, keuangan, dan surat menyurat serta pencatatan dan pelaporan. Keadaan pegawai di Puskesmas Cipatat berjumlah 30 orang dengan jenis tenaga Dokter, Dokter Gigi, Sarjana Kesehatan Masyarakat, Perawat, Perawat Gigi, Bidan, Asisten Apoteker, Sanitarian, Gizi, Laboratorium ditambah dengan pelaksana Tata Usaha.

3.1.1. Lambang Puskesmas

Gambar 2.2 merupakan lambang Puskesmas terbaru melalui Peraturan Menteri Kesehatan Nomor 75 Tahun 2014 tentang Pusat Kesehatan Masyarakat.



Gambar 2.2 Logo Puskesmas

Pengertian dari lambang tersebut tentunya berdasarkan PMK Nomor 75 tahun 2014, sebagai berikut:

Bentuk segi enam (*hexagonal*), melambangkan:

a. Keterpaduan dan kesinambungan yang terintegrasi dari 6 prinsip yang melandasi penyelenggaraan Puskesmas.
b. Makna pemerataan pelayanan kesehatan yang mudah di akses masyarakat.
c. Pergerakan dan pertanggung jawaban Puskesmas di wilayah kerjanya.

Irisan dua buah bentuk lingkaran melambangkan dua unsur upaya kesehatan yaitu:

a. Upaya kesehatan Masyarakat (UKM) untuk memelihara dan meningkatkan kesehatan serta mencegah menanggulangi timbulnya masalah kesehatan masyarakat.
b. Upaya Kesehatan Perseorangan (UKP) untuk memelihara dan meningkatkan kesehatan serta mencegah dan menanggulangi timbulnya masalah kesehatan perorangan.

Stilasi bentuk sebuah bangunan, melambangkan Puskesmas sebagai tempat/wadah diberlakukannya semua prinsip dan upaya dalam proses penyelenggaraan kesehatan.

Bidang segitiga mewakili tiga faktor yang memengaruhi status derajat kesehatan masyarakat yaitu genetik, lingkungan, dan perilaku.

Bentuk palang hijau berbentuk segi enam melambangkan pelayanan kesehatan yang mengutamakan promotif preventif.

Warna hijau melambangkan tujuan pembangunan kesehatan yang diselenggarakan Puskesmas, dalam rangka mencapai derajat kesehatan masyarakat yang setinggi-tingginya.

Warna putih melambangkan pengabdian luhur Puskesmas [6].

2.1.2. Visi dan Misi

Visi adalah pandangan jauh tentang suatu perusahaan ataupun lembaga dan lain-lain, visi juga dapat diartikan sebagai tujuan perusahaan atau lembaga dan apa yang harus dilakukan untuk mencapai tujuan pada masa yang akan datang. Misi adalah suatu pernyataan tentang apa yang harus dikerjakan oleh perusahaan atau lembaga dalam usaha mewujudkan visi tersebut. Adapun Visi dan Misi dari Puskesmas Cipatat Kabupaten Bandung Barat sebagai berikut:

a. Visi Puskesmas

Menjadi Puskesmas Dambaan Masyarakat, ditunjang dengan Pelayanan Prima, Menuju Masyarakat Kecamatan Cipatat Sehat Mandiri.

b. Misi Puskesmas

1) Menggerakkan pembangunan berbagai sektor di Wilayah Cipatat agar senantiasa bertumpu pada Wawasan Kesehatan
2) Mendorong Kemandirian Keluarga dan Masyarakat Cipatat untuk hidup sehat, terutama dengan berperan aktif dalam upaya promotif dan preventif.
3) Memelihara dan meningkatkan mutu pelayanan kesehatan serta memperluas jangkauan pelayanan kesehatan kepada masyarakat.

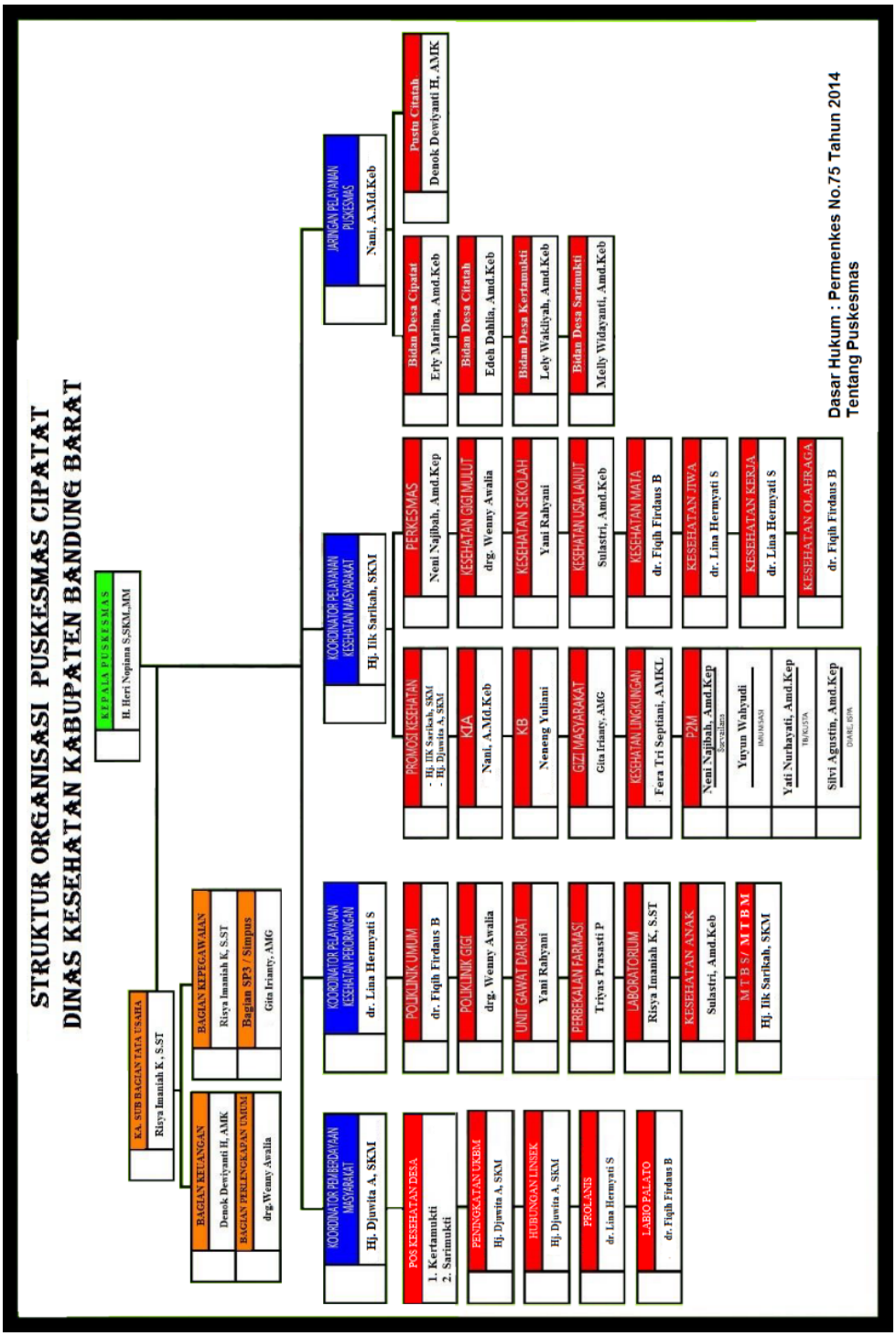
c. Moto

”Tiada hari yang lebih baik untuk memulai perubahan, kecuali hari ini.”

2.1.3. Struktur Organisasi

Struktur organisasi merupakan salah satu kerangka yang menggambarkan hubungan diantar bidang-bidang kerja maupun orang-orang yang mempunyai kedudukan, wewenang, dan tanggung jawab dari bidang kerja yang dipimpinnya alam suatu sistem kerja sama. Struktur organisasi Puskesmas Cipatat Kabupaten Bandung Barat dapat dilihat pada .

Gambar 2.3 Struktur Organisasi Puskesmas Cipatat

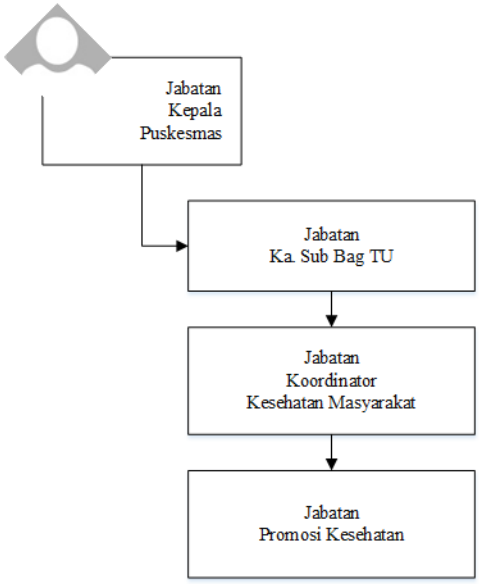


2.1.4. Deskripsi Kerja

Deskripsi kerja berisi informasi jabatan analis pengembangan sistem informasi karena prioritas hasil akhir dari penelitian ini bertujuan untuk membantu bagian atau orang yang bekerja pada bagian tersebut.

1. Promosi Kesehatan

a. Kedudukan dalam Struktur Organisasi



Gambar 2.4 Kedudukan Struktur Organisasi Promosi Kesehatan

b. Ikhtisar Jabatan

Melaksanakan kegiatan *avokasi*, pembinaan dan gerakan pemberdayaan masyarakat, melakukan penyebarluasan informasi, membuat rancangan media, melakukan pengkajian/ penelitian perilaku masyarakat yang berhubungan dengan kesehatan, serta merencanakan *intervestasi* dalam rangka pengembangan periku masyarakat yang mendukung kesehatan.

c. Uraian Tugas

1) Menyusun rencana kegiatan pelayan promosi kesehatan berdasarkan data program puskesmas dan ketentuan perundang-undangan yang berlaku sebagai pedoman kerja.

2) Melaksanakan kegiatan promosi kesehatan meliputi penyuluhan kesehatan, pembinaan PSM/UKBM, pembinaan PHBS dan fasilitator desa siaga serta koordinasi lintas program terkait sesuai dengan prosedur dan ketentuan perundang-undangan yang berlaku.
3) Mengevaluasi hasil kegiatan pelayanan promosi kesehatan secara keseluruhan.
4) Membuat catatan dan laporan kegiatan di bidang tugasnya sebagai bahan informasi dan pertanggung jawaban kepada atasan.
5) Melaksanakan tugas lain yang diberikan oleh atasan.

d. Tanggung Jawab

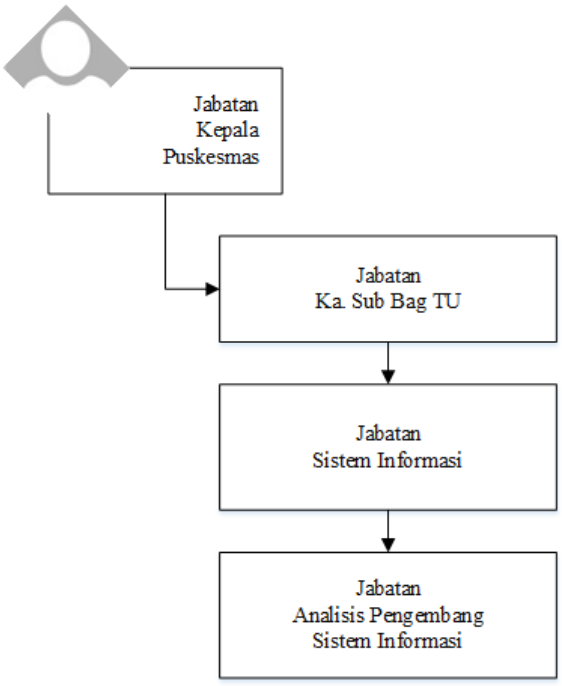
Penyuluhan kesehatan masyarakat/promosi kesehatan.

e. Wewenang

Melakukan kegiatan penyuluhan kesehatan masyarakat/promosi kesehatan secara *professional*.

2. Analis Pengembangan Sistem Informasi

a. Kedudukan dalam Struktur Organisasi



Gambar 2.5 Kedudukan Struktur Organisasi Analisis Pengambang SI

b. Ikhtisar Jabatan

Melakukan pengumpulan, mengidentifikasi, menganalisis data teknologi sistem informasi, melakukan koordinasi, membuat laporan dan melaksanakan tugas ke dinas sesuai ketentuan yang berlaku agar sasaran pengembangan sistem informasi dapat tercapai.

c. Uraian Tugas

1) Mengumpulkan data sistem informasi berdasarkan prosedur yang berlaku sebagai bahan untuk penyelesaian pelaksanaan tugas,
2) Mengidentifikasi data sistem informasi sesuai dengan permasalahan berdasarkan prosedur yang berlaku agar sasaran pengembangan sistem informasi dapat tercapai,
3) Menganalisis pengembangan sistem informasi sesuai kebutuhan berdasarkan ketentuan yang berlaku agar kekurangan sistem informasi dapat disempurnakan,
4) Melakukan koordinasi dengan unit kerja instansi sesuai degan prosedur yang berlaku untuk kelancaran pelaksanaan tugas,

5) Membuat laporan hasil pelaksanaan tugas sesuai dengan prosedur yang berlaku sebagai bahan evaluasi dan pertanggung jawaban, dan
6) Melaksanakan tugas kedinasan lain yang diperintahkan oleh pimpinan baik secara tertulis maupun lisan.

d. Bahan kerja

No	Bahan Kerja	Penggunaan Dalam Tugas
1	Data sistem informasi	Untuk dikumpulkan, diidentifikasi dan dianalisis.
2	Notulen	Hasil koordinasi
3	Surat Tugas	Petunjuk kerja
4	SOP	Petunjuk kerja

e. Perangkat/Alat Kerja

No	Perangkat Kerja	Digunakan Untuk Tugas
1	Komputer	Untuk melaksanakan tugas kedinasan
2	Jaringan Internet	Sarana informasi
3	Telepon	Sarana komunikasi
4	*Printer*	Untuk mencetak laporan
5	ATK	Untuk menunjang pelaksanaan tugas

f. Wewenang

1) Memberikan data sistem informasi
2) Menggunakan komputer dengan baik
3) Meminta data-data sistem informasi
4) Menilai data-data sistem informasi

2.2. Landasan Teori

Pada bagian landasan teori ini akan menjelaskan teori-teori yang akan digunakan sebagai acuan dalam penelitian yang akan dilakukan.

2.2.1. 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. Rekam medis harus dibuat secara tertulis, lengkap dan

jelas atau secara elektronik. Rekam medis berisi tentang semua tindakan yang dilakukan terhadap pasien baik pasien rawat jalan, rawat inap, dan gawat darurat.

Isi rekam medis untuk pasien rawat inap dan perawatan satu hari sekurang-kurangnya memuat:

a) Identitas Pasien,
b) Tanggal dan waktu,
c) Hasil *anamnesis*, mencakup sekurang-kurangnya keluhan dan riwayat penyakit,
d) Hasil pemeriksaan fisik dan penunjang medik,
e) Diagnosis,
f) Rencana penatalaksanaan,
g) Pengobatan atau tindakan,
h) Pelayanan lain yang telah diberikan kepada pasien,
i) Untuk pasien kasus gigi dilengkapi dengan *odontogram* klinik, dan
j) Persetujuan tindakan bila diperlukan.

Pemanfaatan rekam medis dapat dipakai sebagai :

a) Pemeliharaan kesehatan dan pengobatan pasien,
b) Alat bukti dalam proses penegakan hukum, disiplin kedokteran dan kedokteran gigi dan penegakan etika kedokteran dan etika kedokteran gigi,
c) Keperluan pendidikan dan penelitian,
d) Dasar pembayaran biaya pelayanan kesehatan, dan
e) Data statistik kesehatan.

Pemanfaatan rekam medis untuk keperluan pendidikan dan penelitian tidak diperlukan persetujuan pasien, bila dilakukan untuk kepentingan negara [7].

2.2.2. ICD 10

International Statistical Classification of Diseases and Related Health Problems, yang biasanya disebut dengan nama pendek *International Classification of Diseases* (ICD), adalah alat diagnostik standar untuk epidemiologi, manajemen kesehatan dan tujuan klinis. ICD dikelola oleh Organisasi Kesehatan Dunia WHO (World Health Organization), yang bertugas mengarahkan dan mengkoordinasikan

kewenangan untuk kesehatan dalam sistem organisasi Persatuan Bangsa-bangsa (PBB). ICD ini dirancang sebagai sistem klasifikasi perawatan kesehatan, menyediakan sistem kode diagnostik untuk mengklasifikasikan penyakit, termasuk klasifikasi berbagai tanda gejala, temuan abnormal, keluhan, situasi sosial, dan penyebab eksternal cedera atau penyakit. System ini dirancang untuk memetakan kondisi kesehatan yang sesuai kategori umum bersama-sama dengan variasi tertentu, dengan menggunakan kode sampai enam karakter. Dengan kategori utama dirancang untuk menyertakan kesatuan penyakit yang serupa

The *International Classification of Diseases* ini diterbitkan oleh Organisasi Kesehatan Dunia WHO (World Health Organization) dan digunakan di seluruh dunia untuk morbiditas dan mortalitas statistik, sistem pembayaran klaim asuransi, dan mendukung pengambilan keputusan yang otomatis dalam perawatan kesehatan. Sistem ini dirancang untuk mempromosikan sistem internasional dalam pengumpulan, pengolahan, klasifikasi, dan penyajian statistik morbiditas dan mortalitas. ICD adalah inti berbasis sistem statistik diagnostik klasifikasi untuk masalah-masalah yang berhubungan dengan perawatan kesehatan. ICD direvisi secara berkala dan saat ini sedang dalam revisi yang kesepuluh. ICD-11 direncanakan akan diterbitkan pada tahun 2017.

Penggunaan ICD-10 dimulai pada tahu 1983, dan revisi baru disahkan oleh *the Forth third World Health Assembly* pada bulan Mei 1990. Versi terbaru mulai digunakan di negara-negara anggota WHO mulai tahun 1994. System klasifikasi ICD-10 ini memungkinkan lebih dari 155.000 kode yang berbeda dan berbagai diagnosa dan prosedur baru, penambahan yang signifikan pada 17.000 kode yang tersedia di ICD-9. Pengaplikasian ICD-10 ini relatif cepat di sebagian besar dunia. Beberapa bahan materi tentang ICD tersedia secara online oleh WHO untuk memudahkan Penggunaannya, termasuk manual, pedoman pelatihan, *browser*, dan *file* untuk *download*. Penggunaan ICD 10 di Indonesia tertuang dalam Keputusan Direktur Jenderal Pelayanan Medik (Dirjen Yanmed) Nomor : HK 00.05.1.4.0074 mengenai Penggunaan Klasifikasi Internasional penyakit revisi yang ke sepuluh (ICD-10) di Rumah Sakit tertanggal 19 Februari 1996 [8]. Berikut merupakan pembagian Bab pada ICD-10 :

Tabel 2.1 Penggolongan Penyakit dengan Kode ICD-10

BAB	PENGGOLONGAN PENYAKIT	Kode ICD-10
I	Penyakit infeksi dan parasitic tertentu	A00-B99
II	Neoplasms (kanker)	C00-D48
III	Penyakit darah dan organ pembentuk darah serta gangguan terkait mekanisme imunitas	D50-D89
IV	Penyakit endokrin , nutrisional dan metabolic	E00-E90
V	Gangguan mental dan perilaku	F00-F99
VI	Penyakit sistem syaraf	G00-G99
VII	Penyakit mata dan Adneksa mata	H00-H59
VIII	Penyakit telinga dan proses mastoid	H60-H95
IX	Penyakit sistem sirkulasi	I00-I99
X	Penyakit sistem Respirasi	J00-J99
XI	Penyakit sistem digestif	K00-K93
XII	Penyakit kulit dan jaringan subkutan	L00-L99
XIII	Penyakit sistem otot-kerangka tulang dan jaringan ikat	M00-M99
XIV	Penyakit sistem kemih-kelamin	N00-N99
XV	Kehamilan, kelahiran, dan Nifas	O00-O99
XVI	Kondisi tertentu yang berasal dalam periode perinatal	P00-P96
XVII	Malformasi, deformasi dan abnormal kromosomal kongenital	Q00-Q99
XVIII	Simtoma, tanda-tanda dan temuan klinis dan laboratories yang abnormal NEC	R00-R99
XIX	Cedera, keracunan, dan konsekuensi lain akibat sebab luar tertentu	S00-T98
XX	Kausa external morbiditas dan mortalitas	V01-Y98
XXI	Faktor-faktor yang mempengaruhi status kesehatan	Z00-Z99
XXII	Penyakit khusus tertentu	U00-U99

2.2.3. Data

Data adalah segala fakta, angka, atau teks yang dapat diproses oleh komputer. Saat ini, akumulasi pertumbuhan jumlah data berjalan dengan cepat dalam format dan basis data yang berbeda. Data tersebut, antara lain, adalah

a. Data operasional atau transaksional, seperti penjualan, inventaris, penggajian, akuntansi dan sebagainya;
b. Data operasional, seperti industri penjualan (supermarket), peramalan, dan data ekonomi makro;
c. *Metadata*, adalah data mengenai data itu sendiri, seperti desain logika basis data atau definisi kamus data.

2.2.4. Basis Data

Database (Basis Data) merupakan kumpulan data yang saling berhubungan. Terdapat struktur baris dan kolom dalam *database* yang akan menampung *record-*

record data *website*, contoh *database* Management System yang sering digunakan adalah MySQL, PostGres,Microsoft SQL, ADO DB, SybBase SQL, Oracle.

Basis data atau *database* adalah kumpulan dari data atau dapat dideskripsikan sebagai kumpulan dari objek, kejadian, maupun peristiwa yang saling berkaitan untuk memenuhi kebutuhan tertentu yang disimpan secara elektronik agar dapat dimanfaatkan kembali dengan cepat dan mudah [9].

a. Operasi Dasar Basis Data

Pada media penyimpanan, basis data dapat dibuat maupun dapat dihapus. Dalam media penyimpanan juga dapat disimpan lebih dari satu basis data yang digunakan untuk banyak keperluan. Karena itulah, operasi-operasi yang dapat dilakukan berkenaan dengan basis data antara lain adalah:

1) Pembuatan basis data baru (*create database*), yaitu untuk membuat basis data baru.
2) Penghapusan basis data (*drop database*), yaitu penghapusan basis data sekaligus dengan isi data yang ada di dalamnya.
3) Pembuatan tabel baru ke suatu basis data (*create table*), yaitu untuk penambahan tabel baru ke basis data yang telah ada.
4) Penghapusan tabel dari suatu basis data (*drop table*), yaitu penghapusan terhadap tabel lama yang ada di basis data.
5) Penambahan atau pengisian data baru ke sebuah tabel di sebuah basis data (*insert*), yaitu penambahan data ke dalam tabel.
6) Pengambilan data dari sebuah tabel (*query*), yaitu pencarian data dari tabel-tabel data.
7) Pengubahan data dari sebuah tabel (*update*), yaitu perbaikan data yang ada pada tabel.
8) Penghapusan data dari sebuah tabel (*delete*), yaitu penghapusan data yang ada pada tabel yang terdapat dalam basis data.

b. Objektif Basis Data

Selain untuk mempermudah dan mempercepat pemrosesan data, manfaat adanya basis data antara lain:

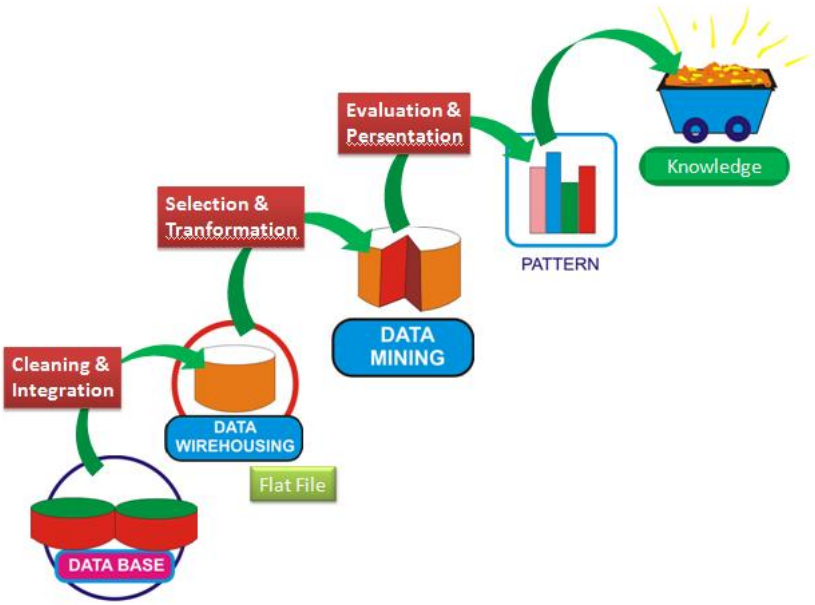
1) Kecepatan dan Kemudahan (*Speed*)
2) Efisiensi Ruang Penyimpanan (*Space*)
3) Keakuratan (*Accuracy*)
4) Ketersediaan (*Availability*)
5) Kelengkapan (*Completeness*)
6) Keamanan (*Security*)
7) Kebersamaan Pemakaian (*Sharability*)

2.2.5. Data mining

Tan (2006) mendefinisikan *data mining* sebagai proses untuk mendapatkan informasi yang berguna dari gedung basis data yang besar. *Data mining* juga dapat diartikan sebagai pengekstrak informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Istilah *data mining* kadang disebut juga *knowledge discovery.*

Salah satu teknik yang dibuat dalam *data mining* adalah bagaimana menelusuri data yang ada untuk membangun sebuah model, kemudian menggunakan model tersebut agar dapat mengenali pola data yang lain yang tidak berada dalam basis data yang tersimpan. Kebutuhan untuk prediksi juga dapat memanfaatkan teknik ini. Dalam *data mining*, pengelompokan data juga bisa dilakukan. Tujuannya adalah agar kita dapat mengetahui tindak lanjut berikutnya yang dapat diambil. Semua hal tersebut bertujuan mendukung kegiatan operasional perusahaan sehingga tujuan akhir perusahaan diharapkan dapat tercapai [3].

Data mining memiliki banyak sebutan atau istilah lain, diantranya *knowledge mining from data, knowledge extraction, data/pattern analysis*, data *archeology* dan data *dredging* [10]. Banyak orang yang menggunakan istilah populer terhadap data mining, *Knowledge Discovery from data*, atau KKD. Data mining hanya dilihat sebagai langkah penting dalam sebuah proses penemuan pengetahuan. Tahapan proses data mining digambarkan pada Gambar 2.6.



Gambar 2.6 Langkah Data Mining dalam proses KDD

Proses dalam *data mining* pada Gambar 2.6 dapat dijelaskan sebagai berikut:

a. Data Selection

Pemilihan (seleksi) data baru sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD (*Knowledge Discovery from Data*) dimulai. Data hasil seleksi yang akan digunakan untuk proses *data mining*, disimpan dalam suatu berkas, terpisah dari basis data operasional.

b. Preprocessing/Cleaning

Sebelum proses *data mining* dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus KDD. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak.

c. Transformasi

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining*. Proses *coding* dalam KDD merupakan proses kreatif dan sangat bergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

d. *Data mining*

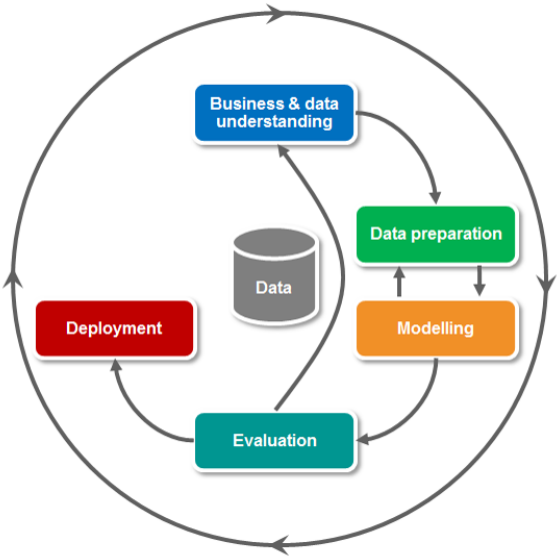
Data mining adalah proses mencari pola atau informasi menarik dalam data yang terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode atau algoritma dalam *data mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

e. *Interprestasi* / Evaluasi

Pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut dengan *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya.

2.2.6. Metode CRISP-DM

CRISP-DM (*Cross-Industry Standard of Data Mining).* CRISP-DM merupakan metodologi *data mining* yang digambarkan dengan model proses dan di dalamnya terdapat tugas-tugas yang harus dilakukan untuk melakukan proyek *data mining* [11]. Model proses CRISP-DM terdiri dari enam tahap yang fleksibel. Model ini telah menjadi proses standar proyek *data mining* di banyak perusahaan untuk mendukung keputusan bisnis. Berikut adalah tahap-tahap pada model proses CRISP-DM.



Gambar 2.7 Model proses CRISP-DM

a. Tahap Pemahaman Bisnis

Tahap pemahaman bisnis fokus pada mengidentifikasi tujuan proyek dan kebutuhan dari sudut pandang bisnis. Kemudian mendefinisikan masalah *data mining* dan menyusun rencana untuk mencapai tujuan proyek. Beberapa kegiatan pada tahap pemahaman bisnis sebagi berikut.

1. Melakukan identifikasi tujuan bisnis,
2. Melakukan penilaian situasi,
3. Menentukan tujuan *data mining*, dan
4. Menyusun rencana proyek yang akan dilakukan.

b. Tahap Pemahaman Data

Tahap pemahaman data dimulai dengan pengumpulan data awal, mempelajari data-data tersebut, dan menilai kualitas data. Kemudian dilakukan pendeteksian bagian dari data yang mungkin mengandung informasi tersembunyi. Beberapa kegiatan pada tahap pemahaman data adalah sebagai berikut.

1. Mengumpulkan data awal,
2. Mendeskripsikan data yang diperoleh,
3. Mengeksplorasi data, dan

4. Melakukan verifikasi terhadap kualitas data.

c. Tahap Persiapan Data

Tahap persiapan data digunakan untuk membangun data final dari sekumpulan data mentah. Pada tahap ini dilakukan *data preprocessing* agar data dapat digunakan untuk proses pemodelan. Kegiatan pada tahap ini mungkin dilakukan beberapa kali dan tidak dalam urutan yang ditentukan. Beberapa kegiatan tersebut adalah sebagai berikut.

1. Melakukan pemilihan data,
2. Melakukan pembersihan data,
3. Melakukan pembangunan data,
4. Mengintegrasikan data dari berbagai sumber, dan
5. Mentransformasikan data sehingga siap untuk diproses.

d. Tahap Pemodelan

Tahap pemodelan merupakan tahap utama pada proses *data mining.* Pada tahap ini diterapkan teknik pemodelan pada data yang telah disiapkan. Beberapa kegiatan pada tahap pemodelan adalah sebagai berikut.

1. Memilih teknik pemodelan yang sesuai dengan data,
2. Menjelaskan prosedur teknik pemodelan yang digunakan,
3. Melakukan penerapan teknik pemodelan, dan
4. Menilai model yang dihasilkan.

e. Tahap Evaluasi

Setelah model terbentuk, perlu dilakukan evaluasi terhadap langkah-langkah yang dilakukan sebelumnya. Hal tersebut dilakukan untuk memastikan model sesuai dengan tujuan bisnis yang ditetapkan. Beberapa kegiatan pada tahap evaluasi adalah sebagai berikut.

1. Mengevaluasi model yang dihasilkan,
2. Mengkaji ulang proses-proses yang dilakukan, dan
3. Menentukan keputusan Penggunaan hasil *data mining.*

f. Tahap Penerapan

Tahap ini merupakan tahap implementasi hasil proses *data mining.* Model yang telah dihasilkan perlu diorganisir dan disajikan kepada Aktor. Hal tersebut biasanya berguna untuk proses pengambilan keputusan perusahaan. Beberapa kegiatan pada tahap penerapan adalah sebagai berikut,

1. Menentukan rencana penerapan hasil *data mining,*
2. Menentukan rencana pengawasan dan pemeliharaan,
3. Membuat laporan akhir, dan
4. Melakukan ulasan terhadap proyek yang telah dilakukan.

2.2.7. Data *Preprocessing*

Dalam *data mining*, kualitas data yang akan digunakan perlu diperhatikan. Ada beberapa faktor yang mempengaruhi kualitas data, antara lain keakuratan, keutuhan, konsisten, aktualitas, dan penafsiran. Data *preprocessing* dapat memperbaiki kualitas data, sehingga dapat meningkatkan keakuratan dan efisiensi hasil dari *data mining* [10]. Beberapa kegiatan ada *preprocessing* yang akan dilakukan dalam penelitian ini adalah sebagai berikut.

a. Menangani data yang tidak konsisten dan *missing value*

Pada data rekam medis bulan Agustus 2016 atribut yang mengandung data teks bebas (*Free text entries*) yang akan dikodekan atau dikonversikan ke bentuk numerik. Kemudian *record* yang memiliki *missing value* yang ditandai dengan *null*, 0 atau "-" pada atribut akan dihapus, karena atribut yang nilainya lengkap akan menghasilkan nilai akhir yang lebih akurat dan pada batasan penelitian hanya data pasien yang memiliki nomor rekam medis dan datanya lengkap yang akan digunakan dalam proses *clustering*.

b. Menangani *redudansi* data

Pada data rekam medis bulan Agustus 2016 ditemukan data pasien yang sama atau redundansi data maka dari itu dilakukannya proses *remove duplicate* data pada atribut NO RM.

c. Menangani *outliers*

Outliers pada setiap atribut data rekam medis akan tetap diproses karena data *outlier* tersebut merupakan data yang valid. Algoritma yang digunakan akan tahan dengan adanya *outlier*. Analisa deskriptif merupakan cara yang digunakan untuk mengetahui nilai minimum, maksimum, rata-rata, standar deviasi, batas atas dan batas bawah untuk menentukan batasan *outlier*.

2.2.8. Clustering

Proses pengelompokan dari kumpulan objek fisik atau abstrak ke dalam kelas-kelas yang memiliki kemiripan disebut *clustering.* Menurut Jiawei Han dan Micheline Kamber, sebuah *cluster* adalah kumpulan objek data yang memiliki kemiripan dengan objek data lainnya yang berada dalam *cluster* yang sama dan tidak sama dengan objek yang berada di *cluster* berbeda. Sebuah *cluster* dari objek-objek dapat dilihat sebagai satu grup dan banyak yang menganggapnya sebagai bentuk dari kompresi data [10].

Jiawei Han dan Micheline Kamber mengatakan bahwa secara umum, metode utama dalam *clustering* dikelompokkan ke dalam beberapa kategori sebagai berikut [10].

a. Metode Partisi / *Partition Methods*

n objek atau *tupel* data dari sebuah database, metode partisi membuat *k* partisi dari data, di mana setiap partisi merepresentasikan sebuah *cluster* dan $k \leq n$, dengan kata lain, ha tersebut mengklasifikasi data ke dalam *k* grup, yang bersama-sama mengikuti persyaratan sebagai berikut: 1) setiap grup harus mengandung setidaknya satu objek, dan 2) setiap objek harus milik tepat satu grup. *k* merupakan jumlah partisi yang harus dibuat. Kemudian menggunakan teknik relokasi berulang (*iterastive relocation technique*) yang bertujuan untuk menyesuaikan partisi dengan memindahkan objek dari satu grup ke grup lain.

b. Metode *Hirarkis* / *Hierarchical Methods*

Metode *hirarkis* menghasilkan sebuah dekomposisi *hirarkis* dari himpunan objek data. Metode hirarkis dapat diklasifikasikan ke dalam *agglomerative* (*aglomeratif*) atau *desisive* (pemecah belah), berdasarkan bagaimana dekomposisi

hirarkisnya terbentuk. Pendekatan *aglomeratif*, atau disebut juga pendekatan *bottom-up*, dimulai dengan setiap grup membentuk kelompok yang terpisah. Pendekatan ini berturut-turut menyatukan objekobjek atau grup-grup yang berdekatan, sampai semua grup menyatu (level teratas dari *hirarki*), atau sampai kondisi *terminasi* terpenuhi. Pendekatan *devisive* atau pemecah belah, atau disebut juga pendekatan *top-down*, dimulai dengan semua objek berada di dalam *cluster* yang sama. Dalam setiap iterasi, sebuah *cluster* terbagi menjadi *cluster* yang lebih kecil, sampai satu objek berada dalam satu *cluster*, atau sampai kondisi terminasi terpenuhi.

c. Metode Berbasis *Dentitas* / *Desnsity-based Methods*

Metode *clustering* telah dikembangkan berdasarkan gagasan dari *density /*kepadatan. ide utama dari metode ini adalah untuk mengembangkan *cluster* sesuai dengan densitas di sekitarnya melebihi ambang batas tertentu. Algoritma yang bisa digunakan antara lain: DBSCAN, OPTICS, dan DENCLUE. DBSCAN dan OPTICS, merupakan metode berbasis densitas yang khas yang menumbuhkan *cluster* sesuai dengan analisis konektivitas berbasis densitas. DENCLUE adalah sebuah metode yang objek *cluster*nya berdasarkan pada analisis nilai distribusi dari fungsi densitas.

d. Metode Berbasis *Grid* / *Grid-based Methods*

Metode berbasis *grid* mengkuantisasi ruang objek menjadi jumlah sel yang terbatas yang membentuk sebuah struktur kisi / *grid*. Semua operasi dilakukan di dalam struktur kisi. Keuntungan utama dari pendekatan ini adalah waktu pemrosesan yang cepat, yang biasanya tidak bergantung pada jumlah data objek dan bergantung pada jumlah sel dari setiap dimensi dalam ruang terkuantisasi. Contoh algoritma yang bisa digunakan untuk metode berbasis *grid* adalah algoritma STING.

e. Metode Berbasis Model / *Model-bases Methods*

Metode berbasis model dapat menemukan *cluster* dengan membangun sebuah fungsi densitas yang mencerminkan distribusi tata ruang dari poin data. Algoritma yang biasa digunakan untuk metode berbasis model antara lain: EM, COBWEB, dan SOM. EM adalah algoritma yang melakukan analisis *expectation-*

maximization berdasarkan pemodelan statistik. COBWEB adalah algoritma pembelajaran konseptual yang melakukan analisis probabilitas dan mengambil konsep sebagai model untuk *cluster*. SOM (*Self-Organizing feature Map*) merupakan algoritma berbasis *neural network* di mana *cluster* dipetakan ke dalam data dimensional tinggi yaitu peta 2-D atau 3-D , yang juga berguna untuk visualisasi data.

f. *Clustering high-dimensionaal* data

Clustering high-dimensional data merupakan tugas yang penting dalam analisis *cluster* karena banyak aplikasi yang membutuhkan analisis dari objek-objek yang mengandung jumlah dimensi yang besar. Banyak dimensi yang mungkin tidak relevan. Seiring bertambahnya jumlah dimensi, data menjadi semakin jarang dan menyebabkan pengukuran jarak antara pasangan dari titik menjadi tidak berarti dan kepadatan rata-rata dari titik mana saja yang di dalam data cenderung rendah.

g. Pengelompokan berbasis pembatas / *Constraint-based clustering*

Merupakan pendekatan yang melakukan pengelompokan dengan penggabungan *user* yang ditetapkan atau pembatas orientasi aplikasi. *Constraint* (pembatas) mengekspresikan harapan *user* atau mendeskripsikan "sifat" dari hasil pengelompokan yang diinginkan, dan memberikan cara yang efektif untuk berkomunikasi dengan proses *clustering*.

Hal yang penting dalam *clustering* adalah mengukur kemiripan atau ke tidak miripan data yang akan dikelompokkan. Selanjutnya akan terbentuk matriks jarak yang menjelaskan kemiripan antara satu data dengan data lainnya atau jarak antara satu data dengan data lainnya. Persamaan yang biasa digunakan untuk menghitung jarak antara dua data adalah *Euclidean Distance*. Berikut adalah persamaan *Euclidean Distance*.

$$d(i,j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \cdots + (x_{in} - x_{jn})^2} \qquad (2.1)$$

d(i,j) = jarak antara data ke i dan data ke j

xi1 = nilai atribut ke satu dari data ke i

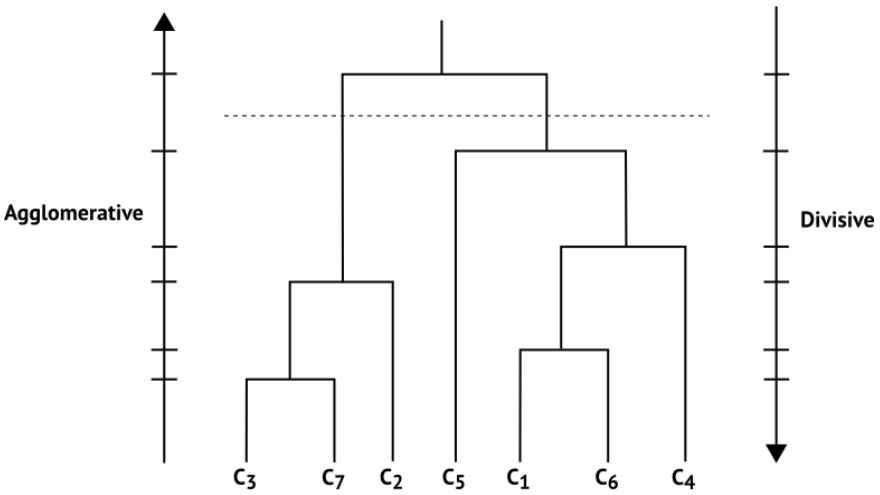
xj1 = nilai atribut ke satu dari data ke j

n = jumlah atribut yang digunakan

2.2.9. Hierarchical Clustering

Hierarchical clustering merupakan metode *clustering* yang dapat melakukan pengelompokan objek pada data ke dalam sebuah hierarki. Terdapat dua teknik pengelompokan objek pada *hierarchical clustering* yaitu secara *agglomerative (bottom-up)* dan *divisive (top-down). Hierarchical agglomerative clustering* menggabungkan setiap objek hingga menjadi satu kelompok, sedangkan *hierarchical divisive clustering* memisahkan semua objek pada sebuah kelompok besar menjadi kelompok yang hanya memiliki satu objek.

Hasil pengelompokan *hierarchical clustering* dapat direpresentasikan pada sebuah *dendrogram*. *Dendrogram* tersebut merupakan visualisasi struktur pengelompokan data dan dapat memberikan deskripsi yang informatif. Jumlah kelompok yang diinginkan dapat diperoleh dengan memotong *dendrogram* pada suatu jarak tertentu. Terdapat banyak metode dalam menentukan jumlah kelompok yang diinginkan dari sebuah *dendrogram*. Gambar 2.8 berikut merupakan contoh *dendrogram* hasil pengelompokan menggunakan metode hierarchical *cluster*ing.



Gambar 2.8 Dendrogram hasil pengelompokan hierarchical clustering

2.2.10. Agglomerative Hierarchical Clustering (AHC)

Pengelompokan hierarki (*hierarchical clustering*) adalah metode analisis kelompok yang berusaha untuk membangun sebuah hierarki kelompok. Strategi untuk pengelompokan hierarki umumnya jatuh ke dalam dua jenis: *agglomerative* dan *divisif*.

Pengelompokan hierarki *agglomeratif* merupakan metode pengelompokan hierarki dengan pendekatan bawah-atas (*bottom up*). Proses pengelompokan dimulai dari masing-masing data sebagai satu buah kelompok, kemudian secara *rekursif* mencari kelompok terdekat sebagai pasangan untuk bergabung sebagai satu kelompok yang lebih besar. Proses tersebut diulang terus sehingga tampak bergerak ke atas membentuk jenjang (hierarki). Cara ini membutuhkan parameter kedekatan kelompok (*cluster proximity*) [3].

Pengelompokan data menggunakan *agglomerative hierarchical clustering* (AHC) diawali dengan merepresentasikan setiap objek pada data sebagai satu kelompok, kemudian dilakukan perhitungan jarak (*distance measure*) antar kelompok tersebut. Setelah itu dua kelompok yang memiliki jarak terdekat digabungkan menjadi sebuah kelompok baru. Kemudian jarak antara kelompok yang baru dengan kelompok lain dihitung menggunakan salah satu metode perhitungan jarak antar kelompok (*single linkage, complete linkage, average*

linkage, dll.) . Proses perhitungan jarak dan penggabungan dua kelompok dilakukan secara iteratif. Berikut ini merupakan beberapa metode yang dapat digunakan untuk menghitung jarak antar kelompok :

a. *Single Linkage* (Jarak Terdekat)

   Single linkage disebut juga minimum *link*, di mana *similaritas* dari dua *cluster* didasarkan pada dua titik paling dekat dari dua *cluster* yang berbeda. Untuk mencari *single linkage* dapat menggunakan persamaan berikut ini :

$$d_{uv} = \min\{d_{uv}\} \, , \, d_{uv} \in D \qquad (2.2)$$

b. *Complete Linkage* (Jarak Terjauh)

   Complete linkage disebut juga *maximum link*, di mana jarak antara dua buah *cluster* dihitung dari jarak terjauh antara anggota *cluster* satu dengan *cluster* kedua. Untuk mencari *complete linkage* dapat menggunakan persamaan berikut ini :

$$d_{uv} = \max\{d_{uv}\} \, , \, d_{uv} \in D \qquad (2.3)$$

c. *Average Linkage* (Jarak Rata-rata)

   Average linkage disebut juga *group average*, di mana jarak antara dua buah *cluster* dihitung dari rata-rata jarak antara anggota *cluster* satu dengan *cluster* yang kedua. Untuk mencari *average linkage* dapat menggunakan persamaan berikut ini :

$$d_{uv} = \text{average}\{d_{uv}\} \, , \, d_{uv} \in D \qquad (2.4)$$

Berikut ini merupakan langkah-langkah dalam algoritma *Agglomerative Hierarchical Clustering* (AHC) :

1. Setiap objek data dianggap sebagai *cluster*. Jika *n* = jumlah data dan *k* = jumlah *cluster*, maka *k* = *n*.
2. Menghitung matriks jarak antar *cluster* (menggunakan *euclidean distance*)

3. Cari dua *cluster* yang mempunyai jarak paling dekat (*single linkage*) lalu gabungkan dua *cluster* tersebut, sehingga $k = k$-1.
4. Jika $k > 1$, ulangi langkah 2 dan 3
5. Jika $k = 1$, perulangan selesai.
6. *Cluster* dihasilkan dari pemotongan *dendrogram* pada level yang tepat. Pemotongan *dendrogram* dapat dilakukan dengan berdasarkan pada *agglomeration schedule*. Jumlah *cluster* bisa didapatkan dengan mencari perbedaan *koofisien* jarak dari setiap iterasi.

2.2.11. Unified Modeling Language (UML)

Unified Modeling Language (UML) adalah notasi berbasis grafik yang digunakan untuk merancang perangkat lunak yang dibangun menggunakan pendekatan berorientasi objek [12]. UML dapat melakukan pemodelan struktur aplikasi, perilaku, arsitektur, proses bisnis dan struktur data. UML merupakan standar terbuka yang dikontrol oleh sebuah perusahaan konsorsium terbuka yaitu *Object Management Group* (OMG). OMG dibentuk untuk membangun standar yang mendukung *interoperability* sistem berorientasi objek. Tabel 2.2 merupakan beberapa diagram UML yang akan digunakan pada penelitian.

Tabel 2.2 Daftar diagram UML 2.0 yang digunakan pada penelitian

No	Nama Diagram	Fungsi
1	*Use case diagram*	Menggambarkan kebutuhan fungsional sistem dan interaksi antara Aktor dan sistem.
2	*Activity* diagram	Menggambarkan logika *procedural*, proses bisnis, dan alur kerja. Selain itu, diagram ini dapat menggambarkan kegiatan yang dilakukan secara bersamaan (paralel).
3	*Sequence diagram*	Menjelaskan prilaku dari sebuah *scenario* tunggal.
4	*Class diagram*	Menggambarkan tipe dan prilaku objek dalam sistem serta hubungan antar setiap objek.

2.2.12. Ukuran Pusat Data

Deskripsi data yaitu ukuran pusat data, baik dari data mentah (data yang belum dikelompokkan dan termasuk data yang terurut) maupun data yang telah diringkas menjadi distribusi frekuensi (data yang telah dikelompokkan). Ukuran deskripsi data ini sangat bermanfaat dalam analisis dan interpretasi data. Ada tiga

ukuran pusat data yang banyak digunakan, yaitu: rata-rata hitung (selanjutnya disebut rata-rata), media, dan mode [13].

2.2.12.1.Rata-Rata Hitung

Rata-rata hitung, atau lebih dikenal dengan rata-rata, merupakan ukuran pusat data yang paling sering digunakan, karena mudah dimengerti oleh siapa saja dan penghitungnya pun mudah. Rata-rata yang dihitung dari data sampel atau sebagai statistik sampel disimbulkan dengan $\overline{X}$ (baca: X-bar) dan jika dihitung dari data populasi atau sebagai parameter populasi disimbolkan dengan huruf Yunani μ_x (baca: ayu x).

Rata-rata dihitung dengan menjumlahkan seluruh angka data yang selanjutnya dibagi dengan banyaknya (jumlah) data, untuk data sampel disebut sebagai ukuran sampel yang disimbolkan dengan **n** dan untuk data populasi disebut sebagai ukuran populasi yang disimbolkan dengan **n**.

Jika $X_1, X_2, X_3, ..., X_n$ adalah angka-angka data yang banyaknya (jumlahnya) n, maka rata-ratanya dihitung:

$$\overline{X} = \frac{X_1+X_2+X_3+...+X_n}{n}$$

Atau dirumuskan sebagai berikut:

$$\overline{X} = \frac{\sum X_i}{n} \qquad (2.5)$$

$\overline{X}$: Rata-rata sampel

$\sum$: Huruf latin (sigma) yang menunjukkan penjumlahan

X_i : Data ke-i dari variabel acak X

n : Ukuran sampel (banyaknya data sampel)

2.2.12.2.Mode

Mode, sebagai ukuran pusat data, berbeda dengan rata-rata hitung dalam penentuan Mode lebih mirip media dalam penentuannya yang tidak melalui proses aritmetika seperti halnya penentuan rata-rata. Mode adalah suatu nilai yang terdapat dalam serangkaian data yang memiliki frekuensi tertinggi [13]. Mode dirumuskan:

$$mo = B_m + i\left[\frac{d_1}{d_1 + d_2}\right] \qquad (2.6)$$

mo : Mode

B_m : tepi batas kelas bawah pada kelas mode.

i : Interval kelas

d_1 : Frekuensi kelas mode kurangi frekuensi kelas sebelum kelas mode.

d_2 : Frekuensi kelas mode kurangi frekuensi kelas sesudah kelas mode.

2.2.12.3.Simpangan baku

Dalam praktisnya, ukuran variabilitas yang sering digunakan adalah simpangan baku (atau deviasi standar) yang merupakan akar kuadrat dari variasi. Hal ini disebabkan bahwa variasi tidak dap dinyatakan dalam satuan ukur apapun seperti rupiah, kilogram, ton, dan lain sebagainya. Berikut ini merupakan rumus dari simpangan baku [13].

$$s = \sqrt{\frac{\sum(X_i - \overline{X})^2}{n-1}} \qquad (2.7)$$

s : Simpangan baku sampel

X_i : Data ke-i dari variabel acak X

$\overline{X}$: Rata-rata sampel

n : Ukuran sampel