

**HASIL PENELITIAN**

**Pendekatan Teknik Data Mining  
Pada Pusat Data Kesehatan Nasional  
Menggunakan *Map Visualisation***



**OLEH  
M. ADNAN NUR  
P2700211461**

**PROGRAM STUDI TEKNIK ELEKTRO  
KONSENTRASI TEKNIK INFORMATIKA**

**PROGRAM PASCASARJANA  
UNIVERSITAS HASANUDDIN  
MAKASSAR  
2013**

## KATA PENGANTAR

Puji dan syukur kehadirat Allah SWT, karena berkat rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan tesis yang berjudul : “Pendekatan Teknik Data Mining Pada Pusat Data Kesehatan Nasional Menggunakan *Map Visualisation*” ini. Penyusunan tesis ini merupakan salah satu syarat untuk memperoleh gelar Magister dalam program studi Teknik Informatika pada Program Pascasarjana Universitas Hasanuddin Makassar.

Dalam penyusunan tesis ini, berbagai pihak telah banyak memberikan dorongan, bantuan serta masukan sehingga dalam kesempatan ini penulis menyampaikan terima kasih yang sebesar-besarnya kepada :

1. Seluruh Dosen dan Staf Program Studi Teknik Informatika Program Pascasarjana Universitas Hasanuddin.
2. Bapak Dr. Adnan, ST.,MT. dan Dr.Eng. Armin Lawi ,M.Eng. pembimbing yang telah memberikan pengetahuan dan bimbingannya yang sangat bermanfaat bagi penyusunan tesis ini.
3. Ayah dan Ibu saya tercinta, Drs. H. A. Almin Nur dan A. Sitti Aminah.
4. Istri dan Anakku tersayang Reski Ismaya Nur dan Ayra Izzatunnisa yang selalu memberikan dukungan moril sehingga tugas akhir ini dapat terlaksana dengan baik.

5. Saudara-saudaraku A. Linda, Ardi, Alam, dan Nilma atas dorongan dan perhatiannya selama ini serta yang terpenting kebersamaan yang terbaik dalam situasi yang paling buruk yang pernah dihadapi.
6. Sahabat dan teman-temanku PascaMelek 2011 yang tidak mungkin saya sebutkan satu persatu, atas segala dukungan, bantuan dan sarannya sehingga tesis ini dapat terselesaikan dengan baik.
7. Kepada semua pihak yang membantu terlaksananya tesis ini, terima kasih atas dukungan dan doanya selama ini.

Penulis menyadari bahwa penulisan tesis ini masih banyak kekurangan, oleh karenanya kritik dan saran sangat penulis harapkan guna menyempurnakan penulisan ini.

Akhir kata penulis mengucapkan banyak terima kasih dan semoga tesis ini dapat berguna bagi kita semua.

Makassar, 20 Juli 2013

Penulis

## **Pendekatan Teknik Data Mining Pada Pusat Data Kesehatan Nasional Menggunakan *Map Visualisation***

### **ABSTRAK**

Terbitnya UU Nomor 36 Tahun 2009 Pasal 169 mendorong pemanfaatan Teknologi Informasi dan Komputer dalam menyebarluaskan informasi kesehatan. Keberadaan sistem Bank Data Kesehatan Nasional Kementerian Kesehatan RI telah memberikan kemudahan bagi masyarakat dalam memperoleh informasi kesehatan yang mencakup hingga tingkat Kabupaten/Kota. Namun, informasi yang disajikan masih berupa informasi kuantitatif dimana seluruh data disajikan dalam tabulasi angka berdasarkan tahun dengan atribut indikatornya. Sistem ini belum mampu menyajikan informasi lebih dalam yang bersifat kualitatif, misalnya bagaimana mengetahui kecenderungan dan keterkaitan antar indikator kesehatan pada suatu wilayah tertentu. Penelitian ini menerapkan dan mengkombinasikan beberapa teknik *data mining* dengan mengkategorikan sistem yang sedang berjalan sebagai sebuah *data warehouse*. Kombinasi penerapan teknik *data mining* meliputi klusterisasi (*clustering*) menggunakan algoritma *k-means*, klasifikasi (*classification*) menggunakan algoritma *naive bayes* dan pola asosiasi (*association rules*) menggunakan algoritma *FP-Growth* yang bekerja secara dinamis untuk menjawab kebutuhan pengetahuan informasi kesehatan yang lebih cepat dan mendalam secara kualitatif. Hasil dari analisis *data mining* selanjutnya divisualisasi dalam sebuah Sistem Informasi Geografis (SIG) yang dalam penelitian ini disebut sebagai *Map Visualisation* untuk memudahkan dalam memetakan informasi kesehatan berdasarkan provinsi dan kabupaten/kota.

**Kata Kunci:** Bank data kesehatan, *data warehouse*, *data mining*, *map visualisation*.

## DAFTAR ISI

HALAMAN JUDUL .....	i
KATA PENGANTAR .....	ii
ABSTRAK .....	iv
DAFTAR ISI .....	v
DAFTAR TABEL .....	x
DAFTAR GAMBAR.....	xi
BAB I. PENDAHULUAN	
A. LATAR BELAKANG .....	1
B. RUMUSAN MASALAH .....	3
C. TUJUAN PENELITIAN .....	3
D. MANFAAT PENELITIAN.....	4
E. BATASAN MASALAH.....	4
BAB II. TINJAUAN PUSTAKA	
A. DATA WAREHOUSE DAN DATA MINING	
1. <i>Data Warehouse</i> .....	6
2. Implementasi dan Strategi Arsitektur Data Warehouse	7
3. <i>Data Mining</i> .....	9

4. Teknik <i>Mining</i> .....	10
5. Klasterisasi ( <i>Clustering</i> ) Menggunakan Algoritma <i>K-Means</i> .....	12
6. Klasifikasi ( <i>Classification</i> ) Menggunakan Algoritma Naïve Bayes .....	13
7. Pola Asosiasi ( <i>Association Rules</i> ) Menggunakan Algoritma FP-Growth .....	15
 <b>B. MAP VISUALISATION</b>	
1. Sistem Informasi Geografis .....	18
2. <i>Map Server</i> .....	18
3. <i>Open Layers</i> .....	19
 <b>C. APLIKASI PENUNJANG</b>	
1. <i>Database Manegement System</i> menggunakan <i>MySQL</i> .....	20
2. <i>Web Server</i> Menggunakan <i>Apache</i> .....	21
3. Pengolahan Data Spasial Menggunakan <i>MAPServer</i> .....	22
 <b>D. BAHASA PEMPROGRAMAN</b>	
1. Server Side Menggunakan PHP .....	24
2. HTML (Hypertext Markup Language) .....	25

3. CSS (Cascading Style Sheet) .....	26
4. Javascript .....	26

#### E. PENELITIAN TERKAIT

1. Tantangan Informatika Kesehatan Berbasis Data Mining.....	27
2. Aplikasi Data Mining Dalam Bidang Kesehatan Studi Kasus Demam Berdarah .....	28
3. Aplikasi Sistem Informasi Geografis Pelayanan Kesehatan Kota Depok Bernasis Web Menggunakan Quantum GIS .....	29
4. Sistem Informasi Gografis Dalam Bidang kesehatan Masyarakat.....	29
5. Perumusan Pola Penyebaran Demam Berdarah Melalui Data Mining Pada Database Dinas kesehatan DKI Jakarta .....	31
6. Sistem Informasi Geografis Untuk Pemetaan Daerah Rawan Demam Berdarah Dengue .....	32

### BAB III. METODE PENELITIAN

#### A. JENIS DAN SUMBER DATA

1. Data <i>Primer</i> .....	33
2. Data <i>Sekunder</i> .....	33

B. ANALISIS DATA	
1. <i>Clustering</i> /Klasterisasi.....	34
2. <i>Classification</i> /Klasifikasi .....	34
3. <i>Association</i> /Asosiasi.....	36
C. PERANCANGAN SISTEM	
1. <i>Use Case Diagram</i> .....	37
2. <i>Class Diagram</i> .....	38
3. <i>Sequence Diagram</i> .....	40
4. <i>Activity Diagram</i> .....	53
D. INSTRUMEN PENELITIAN	
1. Perangkat Lunak .....	54
2. Perangkat Keras.....	55
BAB IV. METODE PENELITIAN	
A. HASIL PENELITIAN	
1. Klasterisasi ( <i>Clustering</i> ) Data.....	56
2. Klasifikasi ( <i>Classification</i> ) Data .....	61
3. Pola Asosiasi ( <i>Association Rule</i> ) Data .....	67

B. PEMBAHASAN DAN PENGUJIAN	
1. Pemilihan Jenis Penyajian Informasi.....	75
2. Pemilihan Proses Mining.....	75
3. Form Input Data .....	77
4. Pengujian Proses Penyajian Informasi dan Waktu Eksekusi.....	85
BAB V. KESIMPULAN DAN SARAN	
A. KESIMPULAN .....	98
B. SARAN .....	99
DAFTAR PUSTAKA .....	100
LAMPIRAN .....	105

## DAFTAR TABEL

TABEL 4.1. Data Awal Untuk Proses Klasterisasi.....	57
TABEL 4.2. Jumlah dan label Klaster .....	57
TABEL 4.3. Posisi Klaster Awal Untuk Setiap Wilayah .....	58
TABEL 4.4. Centroid (hasil perhitungan rata-rata) dari setiap Klaster ...	58
TABEL 4.5. Posisi Klaster Awal Untuk Setiap Wilayah .....	60
TABEL 4.6. Hasil Akhir Perhitungan Centroid Klaster.....	60
TABEL 4.7. Hasil Akhir Posisi Klaster Setiap Wilayah .....	61
TABEL 4.8. Data awal untuk proses Klasifikasi .....	62
TABEL 4.9. Centroid Kategori.....	62
TABEL 4.10. Jumlah dan label Kategori .....	63
TABEL 4.11. Hasil Klaster untuk Indikator Label Kategori .....	63
TABEL 4.12. Perhitungan Mean dan Standar Deviasi Untuk Indikator Jumlah Penduduk .....	64
TABEL 4.13. Perhitungan Mean dan Standar Deviasi Untuk Indikator Jumlah Kasus Campak .....	64
TABEL 4.14. Data Baru Indikator Jumlah Penduduk dan Jumlah Kasus Campak .....	64

TABEL 4.15. Hasil Prediksi Persentasi Imunisasi Campak Bayi.....	67
TABEL 4.16. Indikator yang akan diasosiakan.....	68
TABEL 4.17. Indikator Hasil Preprocessing .....	69
TABEL 4.18. Record Item Set.....	70
TABEL 4.19. Frekuensi Kemunculan Tiap Indikator .....	70
TABEL 4.20. Hasil Pengurutan Item .....	71
TABEL 4.21. Hasil Conditional Pattern Base dan frequent itemset.....	72
TABEL 4.22. Support Count dan Confidence dari Conditional Pattern Base.....	73
TABEL 4.23. Hasil Frequent Item Set .....	73

## DAFTAR GAMBAR

GAMBAR 2.1. Arsitektur Data <i>Warehouse</i> .....	7
GAMBAR 2.2. <i>K-Means</i> .....	13
GAMBAR 2.3. Propabilitas Bersyarat .....	14
GAMBAR 2.4. Contoh Algoritma <i>FP-Growth</i> .....	17
GAMBAR 3.1. Klasterisasi Data.....	34
GAMBAR 3.2. Klasifikasi Data .....	35

GAMBAR 3.3. Asosiasi Data.....	36
GAMBAR 3.4. <i>Use Case</i> Publik .....	38
GAMBAR 3.5. <i>Use Case</i> Administrator.....	38
GAMBAR 3.6. <i>Class Diagram</i> Publik .....	39
GAMBAR 3.7. <i>Class Diagram</i> Administrator .....	40
GAMBAR 3.8. <i>Sequence Diagram Input User</i> .....	41
GAMBAR 3.9. <i>Sequence Diagram Input Pulau</i> .....	41
GAMBAR 3.10. <i>Sequence Diagram Input Provinsi</i> .....	41
GAMBAR 3.11. <i>Sequence Diagram Input Kabupaten</i> .....	42
GAMBAR 3.12. <i>Sequence Diagram Input Kategori Indikator</i> .....	42
GAMBAR 3.13. <i>Sequence Diagram Input Indikator</i> .....	42
GAMBAR 3.14. <i>Sequence Diagram Input Data Tahunan</i> .....	43
GAMBAR 3.15. <i>Sequence Diagram Peta Data View</i> .....	43
GAMBAR 3.16. <i>Sequence Diagram Peta Data Cluster</i> .....	44
GAMBAR 3.17. <i>Sequence Diagram Peta Data Classification</i> .....	45
GAMBAR 3.18. <i>Sequence Diagram Peta Data Association</i> .....	46
GAMBAR 3.19. <i>Sequence Diagram Grafik Data View</i> .....	47

GAMBAR 3.20. <i>Sequence Diagram Grafik Data Cluster</i> .....	47
GAMBAR 3.21. <i>Sequence Diagram Grafik Data Classification</i> .....	48
GAMBAR 3.22. <i>Sequence Diagram Grafik Data Association</i> .....	49
GAMBAR 3.23. <i>Sequence Diagram Tabel Data View</i> .....	49
GAMBAR 3.24. <i>Sequence Diagram Tabel Data Cluster</i> .....	50
GAMBAR 3.25. <i>Sequence Diagram Tabel Data Classification</i> .....	51
GAMBAR 3.26. <i>Sequence Diagram Tabel Data Association</i> .....	52
GAMBAR 3.27. <i>Activity Diagram Peta</i> .....	53
GAMBAR 3.28. <i>Activity Diagram Grafik</i> .....	53
GAMBAR 3.29. <i>Activity Diagram Tabel</i> .....	54
GAMBAR 4.1. <i>Fp-Tree Data Transaksi</i> .....	71
GAMBAR 4.2. <i>Arsitektur Diagram Sistem</i> .....	74
GAMBAR 4.3. <i>Menu Utama</i> .....	75
GAMBAR 4.4. <i>Halaman Pemilihan Proses Mining Untuk Penyajian</i> <i>Informasi Peta</i> .....	76
GAMBAR 4.5. <i>Halaman Pemilihan Proses Mining Untuk Penyajian</i> <i>Informasi Grafik</i> .....	76

GAMBAR 4.6. Halaman Pemilihan Proses Mining Untuk Penyajian	
Informasi Tabel.....	77
GAMBAR 4.7. Halaman Input Data Jenis Informasi Peta Dengan Proses	
Mining Data <i>View</i> .....	78
GAMBAR 4.8. Halaman Input Data Jenis Informasi Peta Dengan Proses	
Mining Data <i>Cluster</i> .....	78
GAMBAR 4.9. Halaman Input Data Jenis Informasi Peta Dengan Proses	
Mining Data <i>Classification</i> .....	79
GAMBAR 4.10. Halaman Input Data Jenis Informasi Peta Dengan Proses	
Mining Data <i>Association</i> .....	79
GAMBAR 4.11. Halaman Input Data Jenis Informasi Grafik Dengan	
Proses Mining Data <i>View</i> .....	80
GAMBAR 4.12. Halaman Input Data Jenis Informasi Grafik Dengan	
Proses Mining Data <i>Cluster</i> .....	81
GAMBAR 4.13. Halaman Input Data Jenis Informasi Grafik Dengan	
Proses Mining Data <i>Classification</i> .....	82
GAMBAR 4.14. Halaman Input Data Jenis Informasi Grafik Dengan	
Proses Mining Data <i>Association</i> .....	82

GAMBAR 4.15. Halaman Input Data Jenis Informasi Tabel Dengan Proses Mining Data <i>View</i> .....	83
GAMBAR 4.16. Halaman Input Data Jenis Informasi Tabel Dengan Proses Mining Data <i>Cluster</i> .....	84
GAMBAR 4.17. Halaman Input Data Jenis Informasi Tabel Dengan Proses Mining Data <i>Classification</i> .....	85
GAMBAR 4.18. Halaman Input Data Jenis Informasi Tabel Dengan Proses Mining Data <i>Association</i> .....	85
GAMBAR 4.19. Halaman Informasi Peta Hasil <i>Cluster</i> .....	86
GAMBAR 4.20. Halaman Informasi Peta Hasil <i>Classification</i> .....	89
GAMBAR 4.21. Halaman Informasi Peta Hasil <i>Association</i> .....	90
GAMBAR 4.22. Halaman Informasi Grafik Hasil <i>Cluster</i> .....	91
GAMBAR 4.23. Halaman Informasi Grafik Hasil <i>Classification</i> .....	92
GAMBAR 4.24. Halaman Informasi Grafik Hasil <i>Association</i> .....	93
GAMBAR 4.25. Halaman Informasi Tabel Hasil <i>Cluster</i> .....	94
GAMBAR 4.26. Halaman Informasi Tabel Hasil <i>Classification</i> .....	95
GAMBAR 4.27. Halaman Informasi Tabel Hasil <i>Association</i> .....	96

# **BAB I**

## **PENDAHULUAN**

### **A. LATAR BELAKANG**

Keberadaan teknologi sebagai media dalam pendistribusian informasi dalam bidang kesehatan belakangan ini semakin dibutuhkan. Jumlah data kesehatan yang terus meningkat dan kebutuhan penyajian informasi yang cepat dan akurat mendorong penerapan teknologi diberbagai aspek bidang kesehatan. Di Indonesia, terbitnya UU Nomor 36 Tahun 2009 Pasal 169 yang menyatakan bahwa “Pemerintah memberikan kemudahan kepada masyarakat untuk memperoleh akses terhadap informasi kesehatan dalam upaya meningkatkan derajat kesehatan masyarakat” menjadi dasar dalam penerapan teknologi untuk penyajian informasi kesehatan tersebut. Penerapan teknologi ini tentunya membutuhkan metode-metode baru dalam pengolahan dan penyajian informasinya agar dapat dimanfaatkan oleh berbagai kalangan seperti akademisi, pemerintahan dan masyarakat umum.

Saat ini, Departemen Kesehatan RI telah memiliki sebuah sistem yang disebut Bank Data Kesehatan Nasional yang dapat diakses melalui websitenya <http://www.bankdata.depkes.go.id>. Sistem tersebut menyediakan informasi kuantitatif berdasarkan indikator kependudukan, pendidikan, penyakit, upaya kesehatan, kesehatan lingkungan, tenaga kesehatan dan sarana kesehatan. Dimana setiap informasi tersebut dapat

disajikan menurut wilayah provinsi dan kabupaten/kota untuk setiap tahunnya. Dari segi ketersediaan informasi yang ada, sistem tersebut masih dikategorikan sebagai sebuah data warehouse dan belum menyajikan fasilitas penggalian informasi tertentu untuk kebutuhan pengambilan kebijakan. Sehingga membutuhkan pengolahan data lebih lanjut yang tentunya membutuhkan waktu. Teknik data mining yang meliputi klasterisasi (*clustering*), klasifikasi (*classification*) dan pola asosiasi antar data (*association rules*) dianggap perlu diterapkan secara fleksibel pada Bank Data Kesehatan tersebut. Kedinamisan penggunaan teknik data mining memungkinkan penyajian berbagai jenis kebutuhan informasi yang belum diprediksi sebelumnya dapat dilakukan, misalnya:

- a. Pengelompokan tingkat penderita penyakit tertentu berdasarkan indikator jumlah penderita penyakit untuk setiap wilayah provinsi maupun kota/kabupaten.
- b. Mencari keterkaitan antara indikator kesehatan lingkungan dan indikator sarana kesehatan terhadap tingkat penderita penyakit tertentu pada suatu wilayah.
- c. Mencari keterkaitan antara indikator tenaga kesehatan dan indikator sarana kesehatan terhadap data upaya kesehatan.

Dari segi visualisasi, pendekatan geografis lebih memudahkan dalam penyajian informasi menurut penyebaran lokasi atau wilayah. Pemetaan informasi kesehatan berdasarkan provinsi maupun kabupaten/kota

secara mendalam dapat dilakukan dengan cepat melalui penyajian peta/*map* yang interaktif.

Implementasi beberapa teknik data mining secara fleksibel berdasarkan Bank Data Kesehatan yang tersedia dan selanjutnya divisualisasikan melalui peta (*map*) interaktif , grafik dan tabel diharapkan mampu memenuhi kebutuhan informasi kesehatan bagi semua kalangan dan sebagai pendukung keputusan yang akurat bagi penentu kebijakan.

## **B. RUMUSAN MASALAH**

- a. Bagaimana mengkombinasikan dan mengimplementasikan algoritma *K-mean*, *Naive Bayes* dan *Fp-Growth* untuk klasterisasi (*clustering*), klasifikasi (*classification*) dan pola asosiasi (*association rule*) sebagai pendekatan teknik data *mining* secara dinamis sesuai ketersediaan data ?
- b. Bagaimana menyajikan informasi hasil olahan teknik data mining berdasarkan ketersediaan data melalui peta/*map* interaktif, tabel serta grafik berbasis *web* ?

## **C. TUJUAN PENELITIAN**

- a. Mengkombinasikan dan mengimplementasikan beberapa teknik data mining diantaranya klasterisasi (*clustering*) menggunakan algoritma *K-mean*, klasifikasi (*classification*) menggunakan kombinasi algoritma

*Naive Bayes* dan *K-mean* serta pola asosiasi (*association rules*) menggunakan kombinasi algoritma *FP-Growth* dan *K-mean*.

- b. Penyajian data maupun informasi hasil pengolahan teknik data mining melalui Peta (*Map*) interaktif, tabel serta grafik berbasis web.

#### **D. MANFAAT PENELITIAN**

- a. Dari segi metodologi: menguji keefektifan dan performa algoritma untuk masing-masing teknik data mining yang disediakan berdasarkan ketersediaan data dari Bank Data Kesehatan serta menerapkan teknologi mapping/pemetaan yang sesuai.
- b. Dari segi substansi: Sebagai pendukung keputusan bagi penentu kebijakan dalam perencanaan, mendiagnosis, dan intervensi dari masalah kesehatan untuk mencapai tujuan pembangunan kesehatan.

#### **E. BATASAN MASALAH**

Peneliti membatasi ruang lingkup masalah dalam beberapa hal, yaitu:

1. Data yang akan digunakan pada sistem yang dikembangkan diperoleh dari website Bank Data Kesehatan Nasional yaitu <http://www.bankdata.depkes.go.id> yang selanjutnya dikategorikan sebagai data *warehouse*.
2. Data yang akan diolah dalam hal ini klusterisasi, klasifikasi dan asosiasi adalah nilai indikator kesehatan menurut wilayah provinsi dan

kabupaten/kota setiap tahunnya yang disesuaikan dengan ketersediaan data pada website Bank Data Kesehatan Nasional tersebut.

3. Teknik data mining yang disediakan antara lain pola asosiasi (*association rule*) menggunakan algoritma *FP-Growth*, klasifikasi (*Classification*) menggunakan algoritma *Naive Bayes* dan klasterisasi (*Clustering*) menggunakan algoritma *K-mean*.
4. Penyajian data dan informasi melalui Peta/*Map*, tabel serta grafik mencakup wilayah provinsi dan kabupaten/kota.

## BAB II TINJAUAN PUSTAKA

### A. DATA WAREHOUSE DAN DATA MINING

#### 1. Data Warehouse

Gagasan dari data *warehouse* adalah untuk menciptakan ruang penyimpanan permanen untuk data yang diperlukan dalam mendukung pelaporan, analisis dan fungsi *Business Intelligent* lainnya. Data *warehouse* biasanya:

- a. Berada pada komputer yang didedikasikan untuk fungsi tersebut.
- b. Jalan pada Database Management System (DBMS) seperti *Oracle*, *Microsoft* dan *IBM*.
- c. Mempertahankan data untuk jangka waktu yang lama.
- d. Konsolidasi data diperoleh dari berbagai sumber.
- e. Dibangun dengan rancangan sebuah model yang mengubah data produksi atau transaksi ke sebuah data yang mampu dijadikan acuan untuk mengambil keputusan dengan cepat. (Chieft Architect Dkk, 2011)

Pengembangan data *warehouse* melibatkan tahapan pembersihan data dan integrasi antar data. Tahapan tersebut disebut sebagai *preprocessing*. Namun, data *warehouse* bukanlah merupakan persyaratan utama untuk data mining. (Joyce Jackson, 2002).



**Gambar 2.1.** *Arsitektur Data Warehouse*

Umumnya data *warehouse* mempunyai tiga komponen yaitu

- a. *Data acquisition software*. Data ini diperoleh dari sistem lain atau sumber eksternal. Data dikonsolidasikan, dirangkum serta diekstrak kedalam data warehouse.
- b. *Data warehouse* itu sendiri terdiri atas data dan perangkat lunak *database management system* terkait yang biasa disebut “database target”.
- c. Disisi aplikasi klien, memungkinkan pengguna dan aplikasi lain (misalnya DSS dan EIS) untuk mengakses dan menganalisis data *warehouse*.

Ketiga komponen tersebut dapat berada pada *platform* yang berbeda. Terlepas dari itu, ketiga komponen tersebut saling keterkaitan. (Joyce Jackson, 2002)

## **2. Implementasi dan Strategi Arsitektur Data Warehouse**

Implementasi data *warehouse* merupakan aktivitas kompleks yang terbagi dua tahap utama. Pada tahap pertama, terdapat konfigurasi sistem dimana model data warehouse konseptual didirikan sesuai dengan kebutuhan pengguna (desain data warehouse). Kemudian sumber data ditetapkan, serta cara penggalan dan perolehan data.

Selanjutnya, teknologi penyimpanan dipilih dan diputuskan terhadap metode data yang akan diakses. (Manole Velicanu, 2007)

Data warehouse dapat dibangun baik dengan pendekatan *top-down* atau *pendekatan bottom-up*. Namun, dari segi implementasi pendekatan bottom-up lebih realistis tetapi integrasi setiap data *mart*s dapat dilakukan lebih mudah jika perencanaan dilakukan dengan teliti. (Prabhu C.S.R, 2006)

Pendekatan *top-down* digunakan ketika masalah teknologi dan ekonomi diketahui dengan baik. Pendekatan ini dapat mencapai sinergi antara subyek bisnis dan menyediakan hasil tunggal. Ini merupakan metode sistemik yang meminimalkan masalah integrasi, tetapi mahal, membutuhkan waktu yang cukup lama dan memiliki fleksibilitas yang rendah.

Pendekatan top-down sesuai dengan visi *Bill Inmon*, yang menganggap bahwa data warehouse harus menanggapi kebutuhan semua pengguna dalam organisasi, dan bukan hanya kelompok tertentu. Sedangkan Untuk pendekatan bottom-up, metode yang digunakan lebih cepat dan didasarkan pada percobaan dan prototipe. Ini adalah metode yang fleksibel yang memungkinkan organisasi untuk melangkah lebih jauh dengan biaya yang lebih rendah, untuk membangun data *mart* independen. Namun, masalah timbul ketika mencoba untuk mengintegrasikan data *mart* dengan data *warehouse* yang konsisten. Pendekatan ini dimulai dari persyaratan bisnis, sedangkan pendekatan

*top-down* memiliki pandangan dalam integrasi data dan konsistensi pada tingkat seluruh perusahaan. (Manole Velicanu, 2007)

### 3. Data Mining

Data mining merupakan perluasan dari analisis data tradisional dan pendekatan statistik dalam menggabungkan beberapa teknik analisis dari berbagai disiplin ilmu. Ada dua jenis pendekatan pada data mining yaitu menciptakan sebuah model dan menemukan sebuah pola. Data mining membantu sebuah organisasi agar fokus terhadap informasi penting yang terdapat pada database. Namun, data mining hanyalah sebuah *tool* dan hal tersebut tidak menghilangkan kebutuhan atas analisa bisnis, pemahaman data serta untuk memahami metode analisis yang terlibat secara konvensional. (Joyce Jackson, 2002).

Untuk meningkatkan keefektifan data mining, salah satu kebutuhan yang perlu ditelusuri adalah jenis fitur apa yang akan diterapkan untuk memperoleh informasi baru dan tantangan apa yang dihadapi ketika menerapkan teknik data mining.

- a. Penanganan berbagai jenis data
- b. Efisiensi dan Skalabilitas penggunaan algoritma data *mining*.
- c. *Usefulness*, *certainly* dan *expressiveness* dari hasil data mining.
- d. *Expression* dari berbagai jenis hasil data mining.
- e. Penggalan Informasi yang dinamis terhadap berbagai level abstraksi.

- f. Penggalian Informasi dari berbagai sumber data.
- g. Proteksi terhadap privasi dan keamanan data. (Ming-Syan Chen Dkk, 1996)

#### 4. Teknik Mining

Data Mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, antara lain (Kusrini Dkk 2009):

##### a. Deskripsi

Terkadang peneliti dan analis secara sederhana ingin mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat pada data. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

##### b. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model dibangun menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi.

##### c. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan datang dimasa mendatang.

Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan untuk prediksi

d. Klasifikasi

Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori yaitu pendapatan tinggi, pendapatan sedang dan pendapatan rendah.

e. Pengklusteran

Pengklusteran merupakan pengelompokan *record*, pengamatan atau memperlihatkan dan membentuk kelas objek-objek yang memiliki kemiripan. Kluster adalah kumpulan *record* yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan *record-record* dalam *cluster* lain.

Pengklusteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam pengklusteran. Pengklusteran tidak mencoba untuk melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target. Akan tetapi, algoritma pengklusteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok yang memiliki kemiripan.

f. Asosiasi

Tugas asosiasi dalam data mining adalah menemukan atribut yang muncul dalam suatu waktu.

## 5. Klasterisasi (*Clustering*) Menggunakan Algoritma *K-Means*

*Clustering* bertujuan untuk melakukan analisis *cluster* secara otomatis dalam menemukan kelompok/*cluster* konseptual berdasarkan karakteristik objek. *Clustering* menyajikan abstraksi prototype melalui representasi objek yang mewakili setiap objek dari masing-masing *cluster*. Selanjutnya, prototype/perwakilan objek tersebut menjadi dasar dari sejumlah teknik pengolahan data lainnya seperti *summarization*, *compression and nearest neighbor finding*. (Juinjie Wu, 2012).

*K-Means* merupakan *clustering* berbasis prototipe. Sebuah algoritma partisi *cluster* sederhana yang membagi  $K$  *non-overlapping cluster*. Setiap *cluster* direpresentasikan oleh sebuah *centroid*. Alur Kerja *K-Means* antara lain:

- a. Memilih sejumlah  $K$  sebagai centroid dimana  $K$  ditentukan oleh pengguna dan menunjukkan jumlah *cluster* yang ingin dibentuk.
- b. Selanjutnya setiap nilai dari objek dibandingkan dengan masing-masing *centroid*. *Centroid* yang memiliki jarak terdekat dengan nilai objek akan dimasukkan ke *cluster* dari *centroid* tersebut tersebut.
- c. *Centroid* diperbaharui dengan mencari rata-rata nilai objek yang menjadi anggota *cluster centroid*.
- d. Proses ini terus berlangsung hingga nilai *centroid* tidak berubah.

*K-Means* dinyatakan sebagai suatu fungsi yang bergantung pada kedekatan data terhadap *cluster* sebagai berikut :

$$\min_{\{m_k\}, 1 \leq k \leq K} \sum_{k=1}^K \sum_{x \in C_k} \pi_x \text{dist}(x, m_k)$$

**Gambar 2.2.** *K-Means*

Dimana  $\pi_x$  adalah bobot  $x$ ,  $n_k$  jumlah objek data dari *cluster*  $C_k$ ,  $m_k = \sum_{x \in C_k} \frac{\pi_x x}{n_k}$  adalah *centroid* dari *cluster*  $C_k$ ,  $K$  adalah jumlah *cluster* yang ditentukan dan fungsi “dist” menghitung jarak antara objek  $x$  dan *centroid*  $m_k$ ,  $1 \leq k \leq K$ . (Jinjie Wu, 2012)

Algoritma K-means dinilai cukup efisien yang ditunjukkan dengan kompleksitasnya  $O(\mathbf{tkn})$ , dengan catatan  $n$  adalah banyaknya objek data,  $k$  adalah jumlah *cluster* yang dibentuk dan  $t$  adalah banyaknya iterasi. Biasanya, nilai  $k$  dan  $t$  jauh lebih kecil daripada nilai  $n$ . Selain itu, algoritma ini akan terhenti dalam kondisi optimum lokal dan bekerja pada attribute numerik. (Sri Andayani, 2007)

## 6. Klasifikasi (*Classification*) Menggunakan Algoritma Naive Bayes

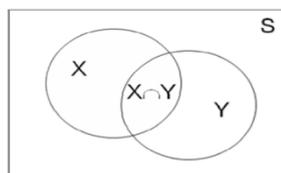
Naive Bayes merupakan salah satu algoritma klasifikasi pembelajaran induktif yang paling efektif dan efisien untuk *machine learning* dan data mining. Performas *naïve bayes* yang kompetitif dalam proses klasifikasi walaupun menggunakan asumsi keindependen *attribute* (tidak ada kaitan antar attribute). Asumsi Keindependenan attribute ini pada data jarang terjadi, namun walaupun asumsi keindependenan attribute tersebut dilanggar performa pengklasifikasian

*naïve bayes* cukup tinggi. Hal ini dibukti berbagai penelitian empiris. (M. Ammar Shadiq, 2009)

*The Naïve Bayes Classifiers* juga biasa dikenal dengan algoritma klasifikasi simple Bayesian. Naive Bayes Classifier merupakan metode terbaru yang digunakan untuk mengklasifikasikan sekumpulan dokumen. Algoritma ini memanfaatkan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya. Tetapi yang membuat algoritma ini populer adalah pendekatan yang dilakukan oleh Paul Graham. Naïve Bayes banyak digunakan karena terbukti efektif untuk kategorisasi teks, sederhana, cepat dan akurasi tinggi. (Hariadi Mochammad Dkk, 2011)

Metode *Bayes* menggunakan propabilitas bersyarat sebagai dasarnya. Dalam ilmu probabilitas bersyarat dinyatakan sebagai:

$$P(X|Y) = \frac{P(X \cap Y)}{P(Y)}$$



**Gambar 2.3.** *Probabilitas Bersyarat*

Probabilitas X di dalam Y adalah probabilitas inteseksi X dan Y dari probabilitas Y, atau dengan bahasa lain  $P(X|Y)$  adalah prosentase

banyaknya  $X$  di dalam  $Y$ . Untuk bentuk data kontinu, digunakan *Dentitas Gaussian* dengan persamaan:

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi}\sigma_{ij}} \exp \left( -\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2} \right)$$

Dimana  $\exp = 2,7183$   
 $\pi = 3,14$   
 $\sigma = \text{standar deviasi}$   
 $\mu = \text{mean}$   
 $x = \text{data}$

$$\text{Untuk } \mu = \frac{1}{n} \sum_{i=1}^n x_i \text{ dan } \sigma = \sqrt{\frac{\sum (x_1 - x)^2}{n}}$$

## 7. Association Rules menggunakan algoritma FP-Growth

Algoritma *FP-Growth* merupakan pengembangan dari algoritma *Apriori*. Sehingga kekurangan dari algoritma *Apriori* diperbaiki oleh algoritma *FP-Growth*. *Frequent Pattern Growth (FP-Growth)* adalah salah satu alternatif algoritma yang dapat digunakan untuk menentukan himpunan data yang paling sering muncul (*frequent itemset*) dalam sebuah kumpulan data. Pada algoritma *Apriori* diperlukan *generate candidate* untuk mendapatkan *frequent itemsets*. Akan tetapi, di algoritma *FP-Growth* *generate candidate* tidak dilakukan karena *FP-Growth* menggunakan konsep pembangunan *tree* dalam pencarian *frequent itemsets*. Hal tersebutlah yang menyebabkan algoritma *FP-Growth* lebih cepat dari algoritma *Apriori*. (Erwin, 2009)

Karakteristik algoritma *FP-Growth* adalah struktur data yang digunakan adalah *tree* yang disebut dengan *FP-Tree*. Dengan menggunakan *FP-Tree*, algoritma *FP-growth* dapat langsung mengekstrak *frequent Itemset* dari *FP-Tree*. Penggalan itemset yang

*frequent* dengan menggunakan algoritma *FP-Growth* akan dilakukan dengan cara membangkitkan struktur data *tree* atau disebut dengan *FPTree*.

Metode *FP-Growth* dapat dibagi menjadi 3 tahapan utama yaitu sebagai:

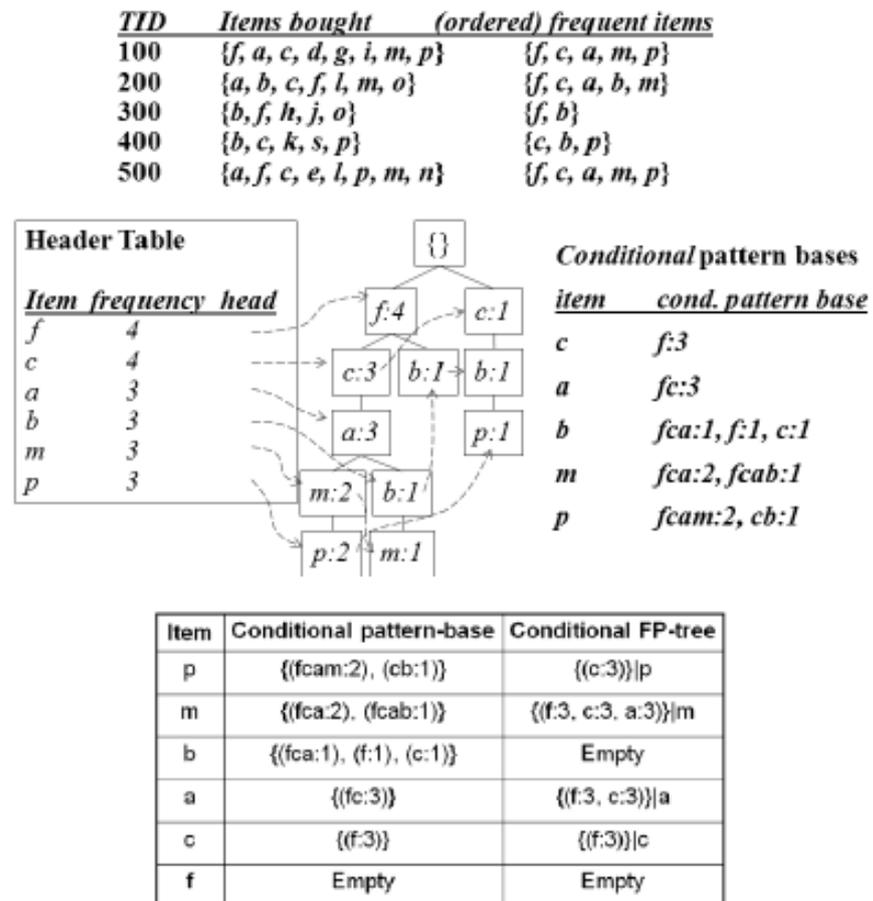
- a. Tahap pembangkitan *conditional pattern base*,
- b. Tahap pembangkitan *conditional FP-Tree*, dan
- c. Tahap pencarian *frequent itemset*.

Ketiga tahap tersebut merupakan langkah yang akan dilakukan untuk mendapat *frequent itemset*. *FP-Growth* memampatkan data transaksi yang memiliki item yang sama sehingga penggunaan memori lebih sedikit serta proses pencarian *frequent itemset* menjadi lebih cepat.

Penting tidaknya suatu pola asosiasi dapat diketahui dengan dua parameter yaitu:

- a. *Support Count* (Nilai Penunjang) yaitu persentase kombinasi item dalam database
- b. *Confidence* (Nilai Kepastian) yaitu kuatnya hubungan antar item dalam pola asosiasi.

Berikut adalah contoh penggunaan algoritma *fp-growth* pada data penjualan.



Gambar 2.4. Contoh Algoritma FP-Growth

## **B. MAP VISUALISATION**

*Map Visualisation* yang dimaksud oleh penulis dalam penelitian ini adalah penggunaan sistem informasi geografis dalam penyajian hasil olahan teknik data mining. Sistem informasi geografis tentunya akan memudahkan penyebaran informasi menurut wilayah.

### **1. Sistem Informasi Geografis**

Sistem Informasi Geografis merupakan perpaduan dari perangkat lunak komputer dan data yang digunakan untuk melihat dan mengelola informasi yang berhubungan dengan lokasi tertentu, menganalisis hubungan spasial dan model proses spasial. Teknologi Sistem Informasi Geografis mengintegrasikan operasi umum dari database seperti *query* dan analisis statistik dengan memvisualisasikan hasilnya melalui peta. (Esri, 2011)

### **2. Map Server**

*Map Server* merupakan *Open Source geographic data rendering engine* yang dibuat menggunakan bahasa C. *Map Server* menyediakan *platform* untuk penyajian data spasial dan aplikasi pemetaan interaktif berbasis web. Pengoperasiannya dijalankan sebagai sebuah program *CGI* atau melalui *Mapscript* yang

mendukung beberapa bahasa pemrograman. Secara umum, *Map Server* dijalankan pada sebuah server dan keluarannya berupa peta raster yang disajikan melalui web browser pada computer pengguna. *Map Server* dirilis dibawah lisensi MIT dan dapat berjalan disemua platform(*Windows, Linux, Max OS*). (*MapServer, 2013*)

### 3. ***Open Layers***

*Openlayers* merupakan sebuah *library* aplikasi berbasis *client side* menggunakan bahasa *javascript* untuk menampilkan peta pada *web browser*. *Openlayer* tidak bergantung pada penggunaan salah satu *Web Server* sehingga dapat digunakan secara dinamis dari sumber manapun. Selain itu, pemanggilan beberapa *Web Server* secara bersamaan dapat dilakukan. Lisensi *openlayers* yang *FreeBSD* membuat *openlayers* fleksibel untuk diterapkan di website-website yang membutuhkan visualisasi peta. (*OpenLayers, 2013*)

## C. APLIKASI PENUNJANG

### 1. *Database Manegement System* menggunakan *MySQL*

Mysql dibuat oleh Tcx dan telah dipercaya mengelola sistem dengan 40 buah database berisi 10.000 tabel dan 500 diantaranya memiliki 7 juta baris. Kemampuan paling menonjol *MySql Server* adalah dalam hal kecepatan yang sangat tinggi dalam melakukan proses data, multi-threaded, multi-user, dan sangat mudah dalam melakukan query dibandingkan *SQL server* yang lain. (Muhammad Syaukani, 2009)

Sebagai server database dengan konsep database modern, MySQL memiliki banyak sekali kelebihan. Beberapa kelebihan yang dimiliki MySQL antara lain:

#### a. *Portability*

Berfungsi dengan stabil tanpa kendala dan berlaku pada berbagai sistem operasi.

#### b. *Open Source*

*MySQL* merupakan database *opensource*, dibawah lisensi GPL.

#### c. *Multiuser*

MySQL merupakan database yang dapat digunakan untuk menangani beberapa *user*

dalam waktu bersamaan tanpa mengalami masalah atau konflik.

d. *Performance Tuning*

MySQL mempunyai kecepatan yang cukup baik dalam menangani *query-query* sederhana serta mampu memproses lebih banyak SQL persatuan waktu.

e. *Column Type*

Database MySQL didukung dengan *type data* yang sangat kompleks.

f. *Security*

Sistem *security* pada MySQL mempunyai beberapa lapisan sekuritas seperti level *subnetmask*, nama *host*, dan izin akses user dengan sistem perizinan yang mendetail serta *password* terenkripsi.

g. *Scalability dan Limit*

MySQL mempunyai kemampuan menangani database dalam skala cukup besar, dengan jumlah *record* lebih dari 50 juta dan 60 ribu tabel serta 5 miliar baris. Selain itu mampu menampung indeks sampai 32 indeks pada setiap tabelnya. . (Muhammad Syaukani, 2009)

## **2. Web Server Menggunakan Apache**

Apache merupakan suatu aplikasi yang bertindak sebagai *web server*. Apache berfungsi untuk melakukan parsing terhadap *file request* yang dilakukan melalui *browser* dan menyajikan hasilnya

melalui file tersebut. Beberapa fitur dan kemampuan yang terdapat pada apache diantaranya:

- a. Proteksi *password* terhadap halaman yang banyak digunakan oleh *user*.
- b. *Customized error pages*.
- c. Dapat menyajikan kode diberbagai versi HTML dan kemampuan untuk menentukan tingkatan *browser* dalam menampilkan *content*.
- d. Penggunaan dan *error log* dalam berbagai jenis dan dapat disesuaikan.
- e. *Virtual Hosting* untuk IP yang berbeda dapat dipetakan dalam server yang sama.
- f. Terdapat index terhadap direktori menyangkut file dari direktori tersebut.
- g. Tidak terdapat batasan terhadap *URL aliasing or rewriting* (Jason Gerber dkk, 2005).

### **3. Pengolahan Data Spasial Menggunakan MapServer**

Mapserver menyajikan gambar peta dari data atau informasi spasial yang disimpan dengan format digital dimana data tersebut dapat berupa data raster maupun data vektor. *Mapserver* dapat mengolah data spasial lebih dari 20 jenis data termasuk shapefile, *PostGIS* dan geometri *ArcSDE*, *opendap*, *arc/info coverages*, dan *census tiger files*.( Bill Kropla, 2005)

Mapserver dapat digunakan dengan dua cara yaitu *CGI* dan *MapScript*. Pada *CGI*, *Web Server* yang digunakan harus mendukung *script CGI* dengan pengaturan mudah dan sangat cepat. Untuk *MapScript*, *API MapServer* dapat diakses menggunakan bahasa *Perl*, *Phyton* atau *PHP*. Antarmuka dari *MapScript* memungkinkan kedinamisan visualisasi peta dalam mengelola fitur-fitur yang tersedia. (Bill Kropla, 2005)

## D. BAHASA PEMROGRAMAN

### 1. *Server Side* Menggunakan PHP (*Hipertext Preprocessor*)

PHP dikembangkan sepenuhnya untuk bahasa *script side-server scripting*. PHP bersifat open source dan dapat digabungkan dengan berbagai *server* yang berbeda-beda. PHP mempunyai kemampuan untuk mengakses database dan diintegrasikan dengan HTML. Berdasarkan hasil survey yang dilakukan *Netcraft*, PHP lebih populer dalam jumlah pemakai dibandingkan dengan modul perl, CGI dan ASP.

PHP memiliki beberapa kelebihan diantaranya:

- a. Mudah dibuat dan dijalankan
- b. Aplikasi PHP cukup cepat dibandingkan dengan aplikasi CGI dengan Perl atau Python bahkan lebih cepat dibanding dengan ASP maupun Java dalam berbagai aplikasi web
- c. Mampu berjalan pada web server dengan sistem operasi yang berbeda-beda.
- d. PHP bisa didapatkan dengan gratis (*freeware*).
- e. Dapat berjalan pada web server yang berbeda-beda seperti Microsoft Personal Web Server, Apache, IIS, Xitami dll.
- f. Dapat di-*embedded*. PHP dapat diletakkan pada file html. (Andi Sunyoto, 2007)

PHP bisa melakukan apa saja yang dapat dilakukan oleh CGI sebagai pendahulunya, seperti mengumpulkan data dari form, menghasilkan isi halaman web dinamis, dan kemampuan mengirim serta menerima cookies, bahkan lebih daripada kemampuan CGI yaitu tersedianya beberapa library untuk koneksi terhadap beberapa jenis database. Selain itu PHP dapat bekerja sebagai CGI processor.

## 2. HTML (Hypertext Markup Language)

Dokumen HTML adalah file teks murni yang dapat dibuat dengan editor teks sembarang. Dokumen ini dikenal sebagai *web page*. Dokumen HTML merupakan dokumen yang disajikan dalam *browser web*. Dokumen ini umumnya berisi informasi atau interface aplikasi dalam internet. Dokumen HTML disusun oleh elemen-elemen. Elemen merupakan istilah bagi komponen dasar pembentuk dokumen HTML. Beberapa contoh elemen adalah *head*, *body*, *table*, paragraf dan *list*. Elemen dapat berupa teks murni atau bukan teks atau keduanya. Untuk menandai berbagai elemen dalam suatu dokumen HTML digunakan *tag*. *Tag* HTML terdiri atas sebuah kurung sudut kiri ( <, tanda lebih kecil ), sebuah nama tag, dan sebuah kurung sudut kanan ( >, tanda lebih besar). Tag umumnya berpasangan, tag yang menjadi pasangan selalu diawali dengan karakter garis miring ( / ). Tag yang pertama menunjukkan tag awal yang berarti awal elemen, dan yang kedua menunjukkan tag akhir, berarti akhir elemen. (Abdul Kadir, 2009)

### **3. Cascading Style Sheet (CSS)**

*Cascading Style Sheet (CSS)* merupakan aturan untuk mengendalikan beberapa komponen dalam sebuah web sehingga akan lebih terstruktur dan seragam. CSS bukan merupakan bahasa pemrograman. Sama halnya *styles* dalam aplikasi pengolahan kata yang dapat mengatur beberapa *style*, misalnya *heading*, *subbab*, *bodytext*, *footer*, *images*, dan *style* lainnya untuk dapat digunakan bersama-sama dalam beberapa berkas (*file*). Pada umumnya CSS dipakai untuk memformat tampilan halaman web yang dibuat dengan bahasa HTML dan XHTML.

### **4. Javascript**

*Javascript* adalah bahasa *script* (bahasa yang kodenya ditulis menggunakan teks biasa) yang ditempelkan pada dokumen html dan diproses pada sisi klien. Dengan adanya bahasa ini, kemampuan dokumen HTML menjadi semakin luas. Sebagai Contoh dengan menggunakan javascript dimungkinkan untuk memvalidasi masukan-masukan pada formulir sebelum formulir dikirim ke *server*. Selain itu, dengan menggunakan *javascript* juga dimungkinkan untuk mengimplementasi tugas yang bersifat interaktif tanpa berhubungan dengan *server*. (Abdul Kadir, 2009)

## E. PENELITIAN TERKAIT

### 1. Tantangan Informatika Kesehatan Berbasis Data Mining

Keberhasilan penerapan aplikasi data mining terlihat jelas dalam berbagai bidang seperti *e-business*, *marketing*, dan *retail* yang memberikan dampak popularitas penggunaannya dalam *knowledge discovery in databases* (KDD). Dan salah satu sektor yang memanfaatkan data mining adalah bidang kedokteran dan kesehatan masyarakat.

Hasil penelitian ini menyajikan sebuah survey tentang teknik dalam *knowledge discovery in database* (KDD). Menggunakan alat bantu data mining dalam dunia kesehatan masyarakat. Penelitian ini juga membahas isu-isu penting dan tantangan yang berkaitan dengan data mining dan bidang kesehatan secara umum.

Penelitian ini menemukan semakin banyak aplikasi data mining yang dapat diterapkan diantaranya analisis data pusat kesehatan untuk mendukung informasi bagi penentu kebijakan, memprediksi wabah penyakit, memprediksi tingkat kematian pada suatu rumah sakit serta memprediksi penipuan pada asuransi. (Ruben D Canlas jr, 2009).

Penelitian ini fokus pada tahap analisis dan deskripsi bagaimana penerapan data mining dalam dunia kesehatan. Tahap Perancangan

dan implemtasi belum dilakukan pada penelitian untuk suatu kasus kesehatan tertentu secara komperhensif.

## **2. Aplikasi Data Mining Dalam Bidang Kesehatan Studi Kasus**

### **Demam Berdarah**

Dalam penelitian ini, dikembangkan model data mining untuk mendeteksi atau pengklasifikasian penderita demam berdarah menjadi kategori: sembuh, mulai sembuh, belum sembh, meninggal dengan melihat beberapa variabel input. Variabel-variabel itu antara lain kandungan hemoglobin, lekosit, hematokrit, trombosit, monosit dan eritrosit, yang diukur pada saat pasien pertama kali masuk dan keluar dari rumah sakit.

Pengklasifikasian/pemprediksian dilakukan dengan menggunakan teknik data mining, kernel-based method yang dalam waktu belakangan mendapat perhatian besar di kalangan peneliti di bidang data mining dan machine learning. Metoda tersebut adalah support vector machines. Sebagai pembanding, untuk melihat performansi, diterapkan teknik tradisional seperti Linear Discriminant Analysis (LDA), K-Nearest Neighbour dan Support Vector Machines (SVM). Ketiga metode tersebut memiliki kelebihan dan kekurangan. Dalam penelitian didapatkan metoda yang paling bagus dari segi akurasi adalah Support Vector Machine dengan linear kernel. (Budi Santoso, 2009)

Penelitian yang dilakukan oleh Budi Santoso fokus pada pencarian metode klasifikasi yang sesuai diterapkan pada data kasus demam berdarah dengan membandingkan beberapa algoritma .

### **3. Aplikasi Sistem Informasi Geografis Pelayanan Kesehatan Kota Depok Berbasis Web Menggunakan Quantum GIS**

Teknologi SIG (Sistem Informasi Geografis) / *Geographic Information System* (GIS) merupakan suatu teknologi mengenai informasi geografis yang telah sangat berkembang. Penelitian ini membahas pembuatan aplikasi Sistem Informasi Geografis berbasis Web, khususnya dalam bidang sarana pelayanan kesehatan Kota Depok, dengan menggunakan data-data yang telah diperoleh dari Dinas Kesehatan pemerintah Kota Depok dan media internet. Namun, pembahasannya dibatasi pada bagaimana website ini dapat menampilkan data-data tersebut ke dalam bentuk peta/data spasial sehingga dapat lebih mudah didapatkan dan dipahami oleh pengguna.

Implementasi website ini dibuat dengan menggunakan aplikasi QuantumGIS, MapServer dan PostgreSQL/PostGIS. Sedangkan bahasa pemrograman yang digunakan adalah PHP, HTML, Javascript dan CSS. (Endah Dharmaputeri, 2009)

Penelitian yang dilakukan oleh Endah Dharmaputeri fokus pada penyajian informasi melalui pemetaan. Tidak terdapat pengolahan khususnya data mining pada data dan informasi yang disajikan.

#### **4. Sistem Informasi Geografis Dalam Bidang Kesehatan Masyarakat**

Desentralisasi di bidang kesehatan telah memungkinkan untuk mengidentifikasi masalah kesehatan, karakteristik populasi, dan insiden lokal yang lebih spesifik. Masalah kesehatan dapat dikategorikan berdasarkan wilayah geografis (nasional, regional, dan lokal). Penelitian ini menyajikan kasus ditahun 1997 hingga 2001, pada penyebaran kasus malaria di Kabupaten Ciamis terkonsentrasi di beberapa desa di pantai selatan, di mana hutan mangrove dan laguna biasanya ditemukan. Ini cenderung menyebar dari barat ke timur. Kasus malaria kebanyakan menyebar pada ketinggian antara 0-100 meter di atas permukaan laut, dan hanya beberapa kasus yang ditemukan di daerah di lebih dari 100 meter di atas permukaan laut. Dengan kemampuan GIS untuk mengelola data spasial, memungkinkan untuk melakukan berbagai skenario seperti melakukan analisis strategi. Untuk meminimalkan bias informasi, direkomendasikan untuk melakukan analisis GIS dari berbagai multidisiplin. Tujuan akhir dari aplikasi GIS di bidang kesehatan adalah untuk meningkatkan kemampuan dalam perencanaan, diagnosis, dan intervensi dari masalah kesehatan di tingkat administrasi pemerintahan untuk mendukung pencapaian tujuan pembangunan kesehatan. (Endang Indarsih, 2008).

Mirip dengan penelitian yang dilakukan oleh Endah Dharmaputeri (2009), penelitian ini juga fokus pada penyajian informasi melalui

pemetaan. Walaupun dalam penelitian ini dijelaskan bahwa pemetaan informasi yang dihasilkan memudahkan dalam menganalisis keterkaitan antar indikator yang tersedia. Namun, analisis tersebut membutuhkan pengolahan lebih lanjut untuk menghasilkan nilai persentase keterkaitan indikator.

## **5. Perumusan Pola Penyebaran Demam Berdarah Melalui Data Mining Pada Database Dinas Kesehatan DKI Jakarta**

Tujuan penelitian ini adalah merumuskan suatu sistem peringatan dini prediksi meledaknya DBD dan cara penanggulangannya. Dalam penelitian ini digunakan datamining untuk menggali pola keterkaitan antar variabel pada database DBD khususnya di wilayah Jakarta Selatan dan Jakarta Pusat dan data cuaca pada wilayah yang sama. Metode yang digunakan adalah metode *Classification based on Predictive Association Rules* (CPAR). Output yang didapat dari pencarian dengan menggunakan CPAR tersebut berupa aturan – aturan dengan kaidah aturan IF – THEN, yang selanjutnya digunakan dalam membangun suatu sistem pakar untuk mengidentifikasi kemungkinan terjadi DBD di suatu wilayah didasarkan pada keadaan cuaca di wilayah tersebut. Sistem pakar berbasis logika fuzzy ini diberi nama **Aplikasi DBD**. Aplikasi DBD dibangun menggunakan Matlab 7.0 R 14, dengan menggunakan fungsi toolbox fuzzy untuk membangun SIF (Sistem Inferensi Fuzzy) dan fasilitas GUI (graphical users interface) untuk antarmukanya. (Hendra Lukito, 2007)

Penelitian ini fokus pada penerapan pola asosiasi yang selanjutnya digunakan sebagai aturan dalam prediksi berdasarkan indikator yang telah ditetapkan sebelumnya. Penyajian informasi hasil olahan pola asosiasi tentunya masih sederhana karena diimplementasikan menggunakan *matlab*.

## **6. Sistem Informasi Geografis Untuk Pemetaan Daerah Rawan Demam Berdarah Dengue**

Tujuan Penelitian ini adalah untuk mengetahui kemampuan citra penginderaan jauh dan SIG dalam penentuan factor-faktor lingkungan fisik untuk pemetaan daerah rawan DBD, sebaran DBD berdasarkan peta sebaran kasus, tempat, dan waktu mulai tahun 2005 sampai dengan 2010 serta menentukan tingkat kerawanan DBD berdasarkan variabel lingkungan dan kejadian DBD. Integrasi citra penginderaan jauh dan SIG dapat menganalisis faktor-faktor resiko lingkungan fisik yang terkait dengan penyebaran DBD diantaranya: penggunaan lahan, ketinggian, curah hujan, area terbang nyamuk dan area kepadatan jentik. (Andri Ruliansyah, 2011)

Penelitian ini merupakan penelitian observational dengan analisa pendekatan *cross sectional*. Hasil penelitian menunjukkan bahwa hasil digitasi Citra *Quickbird* cukup baik digunakan untuk menentukan faktor – faktor lingkungan fisik untuk penentuan daerah rawan DBD.

Dari beberapa paparan penelitian sebelumnya, belum terdapat penelitian yang menerapkan metode klasterisasi, klasifikasi dan asosiasi terhadap indikator yang tersedia secara dinamis serta bagaimana menyajikan informasi hasil olahan data mining tersebut melalui pemetaan.