

IMPLEMENTASI KLASIFIKASI DEMAM BERDARAH DENGUE MENGGUNAKAN ALGORITMA NAÏVE BAYES

Ngurah Agus Sanjaya ER

Program Studi Teknik Informatika, Jurusan Ilmu Komputer

Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Udayana

Email : agus.sanjaya@cs.unud.ac.id

Abstrak: Demam berdarah dengue (DBD) adalah suatu penyakit yang sering menyerang penduduk Indonesia hampir sepanjang tahun terutama ketika musim hujan tiba. Walaupun telah dilakukan berbagai pencegahan termasuk pengasapan dan pembersihan tempat yang tergenang air, penyakit ini tetap saja menjadi salah satu penyebab kematian terbesar di Indonesia. Di lain pihak, masyarakat tidak tanggap terhadap gejala-gejala penyakit DBD sehingga tidak menyadari dirinya mengidap penyakit tersebut.

Penelitian yang dilakukan bertujuan untuk membangun suatu aplikasi klasifikasi menggunakan algoritma Naïve Bayes yang dapat memberikan informasi kelas penyakit DBD yang diidapnya. Algoritma Naïve Bayes menggunakan suatu model probabilitas dalam menentukan kelas penyakit DBD seseorang. Klasifikasi didasarkan pada beberapa gejala penyakit DBD seperti suhu tubuh, demam hari, frekuensi nadi, mual, muntah, nyeri pada sendi, pusing, lemah, nafsu makan, BAB, bintik kemerahan pada kulit, mimisan dan pendarahan gusi.

Hasil ujicoba memperlihatkan bahwa aplikasi yang dibangun mendapatkan persentase yang baik dalam menentukan kelas penyakit DBD serta dapat digunakan sebagai deteksi dini untuk mengetahui apakah seseorang mengidap penyakit DBD atau tidak.

Kata Kunci : Demam Berdarah Dengue (DBD), Klasifikasi, Naïve Bayes

The Implementation of the Dengue fever using the dengue naive bayes algorithm

Abstract: Dengue hemorrhagic fever (DHF) is a disease that often attacks the almost year-round population of Indonesia, especially when the rainy season comes. The although it has performed a variety of prevention, including fumigation and cleaning of the flooded places, the disease remains to be one of the big cause of death in Indonesia. On the other hand, the people do not respond to the symptoms of dengue disease.

Research carried out aiming to build a classification application using the Naive Bayes algorithm that can provide DHF disease grade information that they suffered. Naive Bayes algorithm uses a probability model in determining a person's class of dengue disease. Classification is based on some symptoms of dengue fever such as body temperature, fever days, pulse rate, nausea, vomiting, pain in joints, dizziness, weakness, appetite, bowel movement, reddish spots on the skin, nosebleeds and bleeding gums.

Test results showed that the built application get a good percentage in determining the class of dengue fever and can be used as early detection to determine whether a person suffering from dengue fever or not.

Keywords: Dengue Hemorrhagic Fever (DHF), Classification, Naive Bayes.

I. PENDAHULUAN

Data mining merupakan suatu bidang dalam ilmu komputer yang bertujuan untuk menggali informasi dari kumpulan data yang besar. Teknik-teknik data mining diaplikasikan pada kumpulan data tersebut untuk mencari pola-pola baru dan berguna yang sebelumnya tidak diketahui (Tan et al, 2005). Penemuan informasi yang berguna tersebut dapat digunakan sebagai dasar untuk mengambil suatu keputusan.

Salah satu teknik data mining yang sangat berguna adalah klasifikasi. Klasifikasi bertujuan untuk menentukan kategori atau kelas dari suatu obyek berdasarkan atribut-atributnya. Kategori dan atribut pada permasalahan yang diangkat telah ditentukan sebelumnya. Sebagai contoh: klasifikasi dapat digunakan untuk mengkategorikan apakah suatu email termasuk spam atau tidak. Dalam permasalahan ini email adalah obyek, spam atau bukan spam adalah kelas dari obyek email tersebut. Atribut-atribut yang digunakan untuk proses klasifikasi mencakup subyek, pengirim, isi bahkan catatan kaki dari email.

Teknik klasifikasi ini sangat cocok untuk digunakan pada bidang kedokteran seperti mendiagnosis kelas penyakit Demam Berdarah Dengue (DBD). Penyakit ini senantiasa ada sepanjang tahun di Indonesia dan banyak menelan korban. Masyarakat Indonesia masih beranggapan bahwa penyakit DBD hanya menyerang kalangan anak-anak. Padahal, setiap orang dapat mengidap DBD, baik kalangan tua maupun muda. Disamping itu, gejala-gejala penyakit DBD yang mirip dengan flu biasa atau tifus menambah ketidakawasan penduduk terhadap penyakit berbahaya ini. Dengan dibangunnya suatu model klasifikasi dari data-data pasien pengidap penyakit DBD maka deteksi awal terhadap ada atau tidaknya penyakit ini pada tubuh seseorang dapat dilakukan.

Banyak algoritma yang dapat digunakan untuk membangun model klasifikasi penyakit. Salah satu algoritma tersebut adalah Naïve Bayes yang menggunakan model probabilitas dalam data uji coba (Carabin et al, 2003). Metode ini biasa digunakan dalam menyelesaikan permasalahan yang berkaitan dengan pengambilan keputusan serta cocok untuk menyelesaikan permasalahan dimana diperlukan penanganan data yang tidak konsisten sedangkan permasalahan tersebut memerlukan data yang konsisten.

Jesmin dan Yi-Ping menggunakan algoritma Naïve Bayes untuk melakukan diagnosis terhadap penyakit kanker payudara. Pada penelitiannya metode ini mencapai nilai akurasi tertinggi yaitu 96,85% saat melakukan klasifikasi dibandingkan dengan algoritma klasifikasi C4.5 yang memiliki nilai akurasi sebesar 93,84% (Nahar J dan Yi-Ping, 2007).

II. TINJAUAN PUSTAKA

2.1 Teorema Bayes

Jika X dan Y merupakan pasangan variabel random, maka probabilitas gabungannya (*joint probability*), $P(X=x, Y=y)$ adalah probabilitas variabel X memiliki nilai x dan variabel Y memiliki nilai Y . Probabilitas bersyarat (*conditional probability*) adalah probabilitas suatu variabel random akan mendapatkan suatu nilai tertentu jika hasil dari nilai variabel random lain diketahui. Sebagai contoh, probabilitas bersyarat $P(Y=y|X=x)$ adalah probabilitas dari variabel Y

mendapatkan nilai y jika diketahui variabel X memiliki nilai x . Hubungan dari probabilitas gabungan dan bersyarat adalah sebagai berikut:

$$P(X,Y) = P(Y|X) \times P(X) = P(X|Y) \times P(Y) \dots\dots\dots (1)$$

Dengan menulis ulang dua ekspresi terakhir pada persamaan (1) di atas akan didapatkan teorema Bayes:

$$P(Y|X) = \frac{P(Y|X)P(Y)}{P(X)} \dots\dots\dots (2)$$

2.2 Teorema Bayes Untuk Klasifikasi

Diasumsikan X merupakan kumpulan atribut dan Y adalah kelas dari variabel. Jika kelas dari variabel tidak memiliki hubungan deterministik dengan atribut-atributnya maka X dan Y dapat dianggap sebagai variabel random. Hubungan X dan Y dapat diukur secara probabilitas menggunakan $P(Y|X)$. Probabilitas bersyarat Y , $P(Y|X)$ disebut juga dengan *posterior probability* dari Y sedangkan $P(Y)$ dinamakan *prior probability*.

Pada tahap pelatihan, model klasifikasi yang dibangun harus dapat mempelajari *posterior probability* $P(Y|X)$ untuk semua kombinasi X dan Y berdasarkan informasi yang didapat dari data latih. Dengan mengetahui probabilitas ini maka suatu data uji X' dapat diklasifikasikan dengan mencari kelas Y' yang memaksimalkan *posterior probability* $P(Y'|X')$.

2.3 Naïve Bayes Classifier

Naïve Bayes merupakan salah satu metode *machine learning* yang menggunakan konsep probabilitas. Naïve Bayes adalah model klasifikasi yang diperoleh dengan menerapkan metode yang relatif sederhana untuk sebuah set data latih (Mitchell, 1997). Naïve Bayes *Classifier* mengestimasi probabilitas bersyarat dari suatu kelas dengan mengasumsikan bahwa atribut-atributnya bebas secara bersyarat jika diberikan suatu nilai kelas y . Asumsi bebas secara bersyarat ini dapat ditulis sebagai berikut:

$$P(X|Y = y) = \prod_{i=1}^d P(X_i|Y = y) \dots\dots\dots (3)$$

Dimana masing-masing set atribut $X = \{X_1, X_2, \dots, X_d\}$ terdiri atas d atribut.

2.4 Bebas Bersyarat

Diasumsikan bahwa X , Y dan Z merupakan tiga set variabel random. Variabel-variabel di X dikatakan bebas bersyarat terhadap Y jika Z diketahui dan kondisi berikut dipenuhi:

$$P(X|Y,Z) = P(X|Z) \dots\dots\dots (4)$$

Kondisi bebas bersyarat antara X dan Y dapat ditulis ke dalam bentuk yang hampir sama dengan persamaan (3) di atas:

$$P(X,Y|Z) = P(X|Z) \times P(Y|Z) \dots\dots\dots (5)$$

2.5 Penggunaan Naive Bayes Classifier

Dengan digunakannya asumsi bebas secara bersyarat maka probabilitas bersyarat suatu kelas tidak perlu dihitung untuk setiap X melainkan cukup dihitung estimasi probabilitas bersyarat untuk tiap X_i jika diberikan kelas Y . Untuk mengklasifikasikan suatu data uji, maka Naive Bayes Classifier hanya perlu menghitung *posterior probability* untuk tiap kelas Y :

$$P(Y|X) = \frac{P(Y) \prod_{i=1}^d P(X_i|Y)}{P(X)} \dots\dots\dots(6)$$

Karena nilai $P(X)$ adalah tetap untuk tiap nilai Y , kelas dari data uji tersebut didapatkan cukup dengan mencari kelas yang memaksimalkan nilai pembilang pada persamaan (6) di atas.

2.6 Menentukan Maximum Conditional Expectation Berdasarkan Bayesian Posterior Distribution

Adapun langkah-langkah dalam menentukan nilai maksimum *Posterior Distribution* atau dalam menentukan kelas suatu data sampel yang belum diketahui kelasnya berdasarkan *Bayesian Classification* adalah sebagai berikut (Tan *et al*, 2006):

Setiap sampel data diwakili dengan sebuah *n-dimensional feature vector*, $X=(x_1, x_2, \dots, x_n)$.

Jika ada m kelas, maka $Y_1, Y_2, \dots, Y_m$.

Kemungkinan prior kelas dapat diestimasi dengan persamaan berikut :

$$P(Y_i) = \frac{S_i}{S} \dots\dots\dots(7)$$

dimana S_i adalah jumlah dari sampel-sampel percobaan kelas Y_i , dan S adalah jumlah dari sampel-sampel percobaan.

1. Gunakan fungsi distribusi Gaussian dalam mempresentasikan probabilitas kelas bersyarat untuk atribut yang kontinyu. Distribusi ini memiliki dua parameter diantaranya mean (μ) dan variance (σ^2). Untuk masing-masing kelas Y_i , peluang kelas bersyarat untuk atribut X_i adalah sesuai dengan persamaan berikut:

$$P(X_k|Y_i) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} \exp \left(-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2} \right) \dots\dots\dots(8)$$

Dimana σY_i dan μY_i adalah rata-rata dan *covariance* untuk sampel-sampel percobaan dari kelas Y_i . Berikut merupakan persamaan dalam menentukan *mean* dan *covariance* untuk sampel-sampel percobaan dari kelas Y_i :

$$\mu = \frac{\sum_{i=1}^n X_i}{N} \dots\dots\dots (9)$$

$$\sigma = \frac{\sum (X - \mu)^2}{N} \dots\dots\dots (10)$$

Untuk mengklasifikasi suatu sampel X yang tidak diketahui, maka dicari nilai ekspektasi dari sampel X_i di dalam kelas Y , dengan menggunakan persamaan berikut :

$$E(Y|X) = P(Y) \prod_{i=1}^d P(X_i|Y) \dots\dots\dots (11)$$

2. Setelah mendapatkan nilai ekspektasi diatas, maka selanjutnya dicari nilai maksimumnya diantara kelas-kelas yang dibandingkan dengan menggunakan persamaan perbandingan berikut:

$$E(Y_i|X) > E(Y_j|X); 1 \leq j \leq m, j \neq i \dots\dots\dots (12)$$

Dengan demikian dapat diketahui data sampel X termasuk golongan kelas yang memiliki nilai ekspektasi maksimum.

III. HASIL DAN PEMBAHASAN

Perancangan dan implementasi dari penelitian yang dilakukan menggunakan metode SDLC (*software development life cycle*). Tahapan-tahapan yang dilalui adalah tahap analisis, desain, implementasi dan pengujian.

3.1 Analisis

Tahapan analisis dilakukan untuk mendapatkan gambaran atribut-atribut yang diperlukan untuk membedakan kelas-kelas yang ada pada penyakit DBD. Dari hasil wawancara yang dilakukan dengan beberapa dokter di rumah sakit yang terbiasa menangani pasien dengan penyakit demam berdarah maka didapatkan atribut-atribut sebagai berikut:

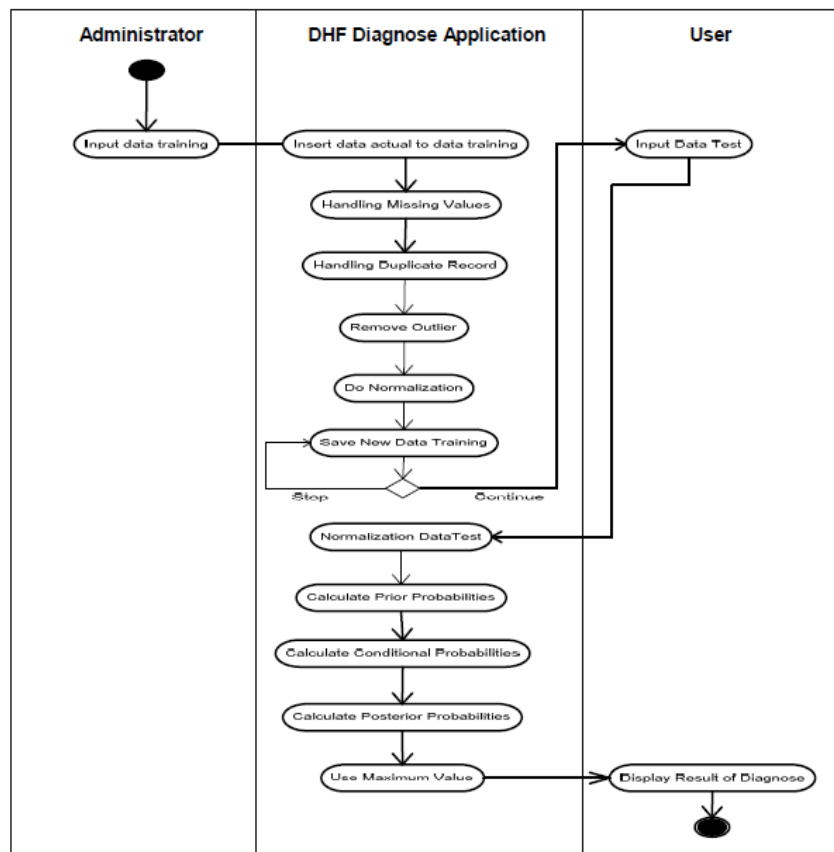
Tabel 1. Atribut-atribut yang Digunakan dalam Klasifikasi DBD

No	Atribut	Keterangan	Tipe Data	Nilai
1.	Suhu tubuh	Menyimpan data gejala besarnya suhu tubuh pada pasien	Kontinyu	34-41 ⁰ C
2.	Frekuensi nadi	Menyimpan data gejala besarnya frekuensi nadi pada pasien	Kontinyu	60-220 kali per menit
3.	Demam hari	Menyimpan data gejala lamanya gejala demam yang dialami pasien	Kontinyu	1-7 hari
4.	Nyeri pada sendi	Menyimpan data gejala nyeri pada sendi pasien	Kategorikal	Ya/Tidak
5.	Mual	Menyimpan data gejala mual pada pasien	Kategorikal	Ya/Tidak
6.	Muntah	Menyimpan data gejala muntah pada pasien	Kategorikal	Ya/Tidak
7.	Sakit kepala/pusing	Menyimpan data gejala sakit kepala/pusing pada pasien	Kategorikal	Ya/Tidak
8.	Lesu lemah	Menyimpan data gejala lemah/lesu pada pasien	Kategorikal	Ya/Tidak
9.	Nafsu makan	Menyimpan data gejala nafsu makan pada pasien	Kategorikal	Normal/Berkurang
10.	Buang air besar (BAB)	Menyimpan data gejala BAB pada pasien	Kategorikal	Normal/Gangguan
11.	Bintik kemerahan pada kulit	Menyimpan data gejala adanya bintik merah pada kulit pasien	Kategorikal	Ya/Tidak
12.	Mimisan	Menyimpan data gejala mimisan pada pasien	Kategorikal	Ya/Tidak
13.	Pendarahan gusi	Menyimpan data gejala adanya pendarahan pada gusi pasien	Kategorikal	Ya/Tidak

3.2 Desain Sistem

Aktivitas yang dilakukan pada sistem dapat dilihat pada *activity diagram* pada gambar 1. Berdasarkan *activity diagram* tersebut, aplikasi diagnosis demam berdarah *dengue* memiliki inputan yaitu data uji dari pengguna, data latih dari administrator dan output berupa nilai *posterior probability* masing-masing kelas sebagai penentu derajat/grade DBD dari yang diidap oleh pengguna.

Data latih yang digunakan untuk membangun sistem klasifikasi ini didasarkan pada data rekam medis pasien pada rumah sakit. Data yang dimasukkan pertama kali masih merupakan data mentah yang berisikan semua data rekam medis dari pasien-pasien rumah sakit tersebut seperti data pribadi pasien (nama, umur, jenis kelamin, dsbnya) serta atribut-atribut yang dinyatakan pada tabel 1 di atas beserta level kelas penyakit demam berdarah yang dinyatakan diidap oleh pasien. Data dari nilai atribut pada tabel 1 selanjutnya digunakan sebagai data latih untuk sistem yang dibangun. Data latih ini sebelum disimpan ke dalam basis data perlu diproses terlebih dahulu.



Gambar 3.1. Activity Diagram Sistem Klasifikasi Demam Berdarah Dengue

Proses yang dilakukan sebelum menyimpan data latih ke basis data meliputi pengisian nilai yang hilang (*missing values*), penanganan duplikasi (*duplicate record*), penghapusan *outlier* serta normalisasi data. Pengisian nilai yang hilang untuk atribut yang bersifat kualitatif digunakan metode *most common feature value* dan atribut kuantitatif ditangani dengan metode *mean substitution*. Duplikasi data ditangani dengan cara penghapusan salah satu data yang memiliki nilai yang sama untuk semua atributnya. *Outlier* ditangani dengan metode *interquartile range*. Setelah melalui proses-proses tersebut data yang dihasilkan harus melalui tahap normalisasi sebelum disimpan ke dalam basis data untuk digunakan sebagai acuan pada proses klasifikasi.

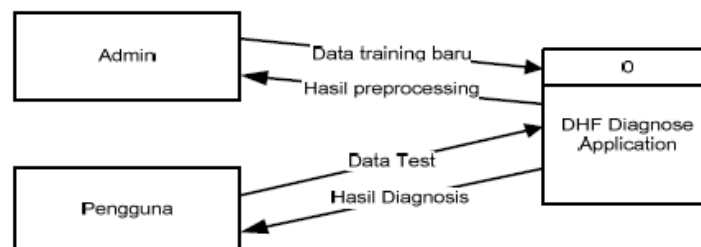
Seorang pengguna yang ingin mengetahui apakah dirinya mengidap penyakit DBD harus memasukkan semua informasi yang diperlukan. Khusus nilai data uji dari pengguna yang bertipe kontinyu akan dilakukan proses normalisasi terlebih dahulu sebelum dilakukan klasifikasi Naive Bayes. Masing-masing data uji akan direpresentasikan sebagai *n-dimensional feature vector*. Dari *n* data uji, akan diketahui masing-masing data uji akan merujuk ke kelas Y_i (misal kelas demam berdarah *dengue* derajat I) atau Y_j (misal kelas demam berdarah *dengue* derajat II) dan Y_k (misal kelas demam berdarah *dengue* derajat III). Dilakukan perhitungan *prior probability* untuk masing-masing kelas berdasarkan atas data latih yang telah didapat sebelumnya.

Selanjutnya dilakukan perhitungan probabilitas kelas bersyarat untuk masing-masing atribut. Dalam melakukan perhitungan tersebut, harus diperhatikan masing-masing atribut yang ada, apakah atribut berjenis biner, kategorikal atau kontinyu. Jika atribut berjenis kontinyu, harus digunakan persamaan rata-rata dan *covariance* dalam menentukan probabilitas kelas bersyarat.

Dari probabilitas kelas bersyarat yang telah didapat, kemudian ditentukan *posterior probability* untuk masing-masing kelas. Nilai *posterior probability* untuk masing-masing kelas dari data sampel akan dibandingkan dengan persamaan maksimalisasi, jika *posterior probability* untuk sampel data kelas Y_i lebih besar daripada *posterior probability* untuk sampel data kelas Y_j dan Y_k maka sampel data tergolongkan ke kelas Y_i , begitu pula. Perhitungan ini juga dilakukan untuk kelas-kelas selanjutnya.

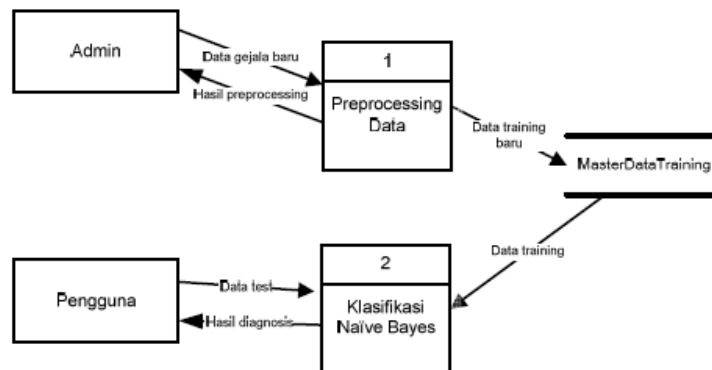
Berdasarkan nilai *posterior probability* tersebut sistem dapat mengklasifikasikan apakah pengguna mengidap DBD atau tidak. Jika ternyata pengguna mengidap DBD maka sistem akan menentukan kelas (derajat DBD I/II/III) dengan menggunakan persamaan maksimalisasi sebelumnya. Untuk kasus tidak terinfeksi demam berdarah *dengue* juga ditentukan berdasarkan nilai masing-masing kelas, yaitu apabila nilai *posterior probability* masing-masing kelas bernilai nol.

Desain alur data dari sistem yang dibangun adalah sebagai berikut:



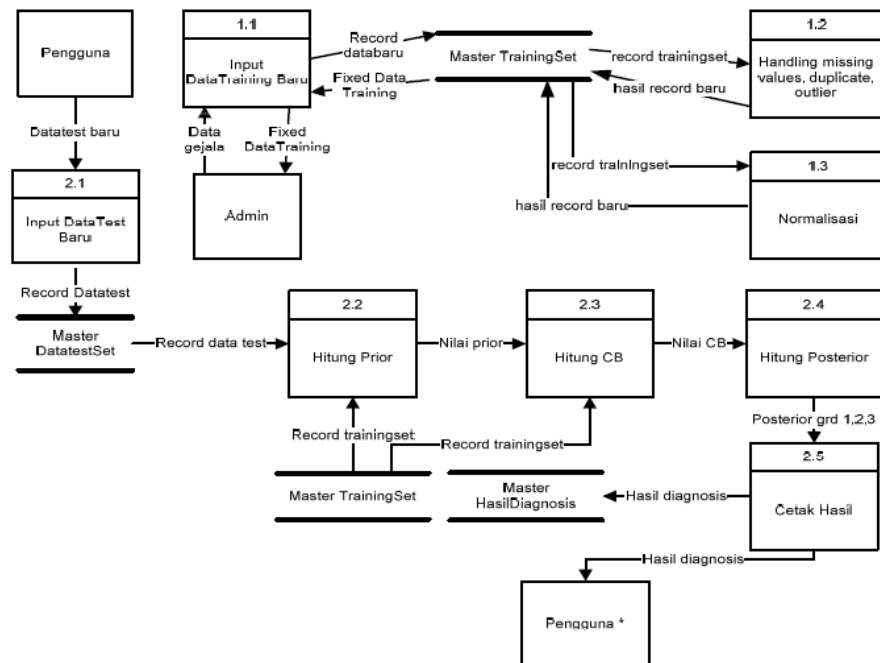
Gambar 3.2. DFD Level 0

Data masukan untuk sistem adalah berupa data latih dan data uji, sedangkan keluaran sistem kepada pengguna adalah hasil diagnosis.



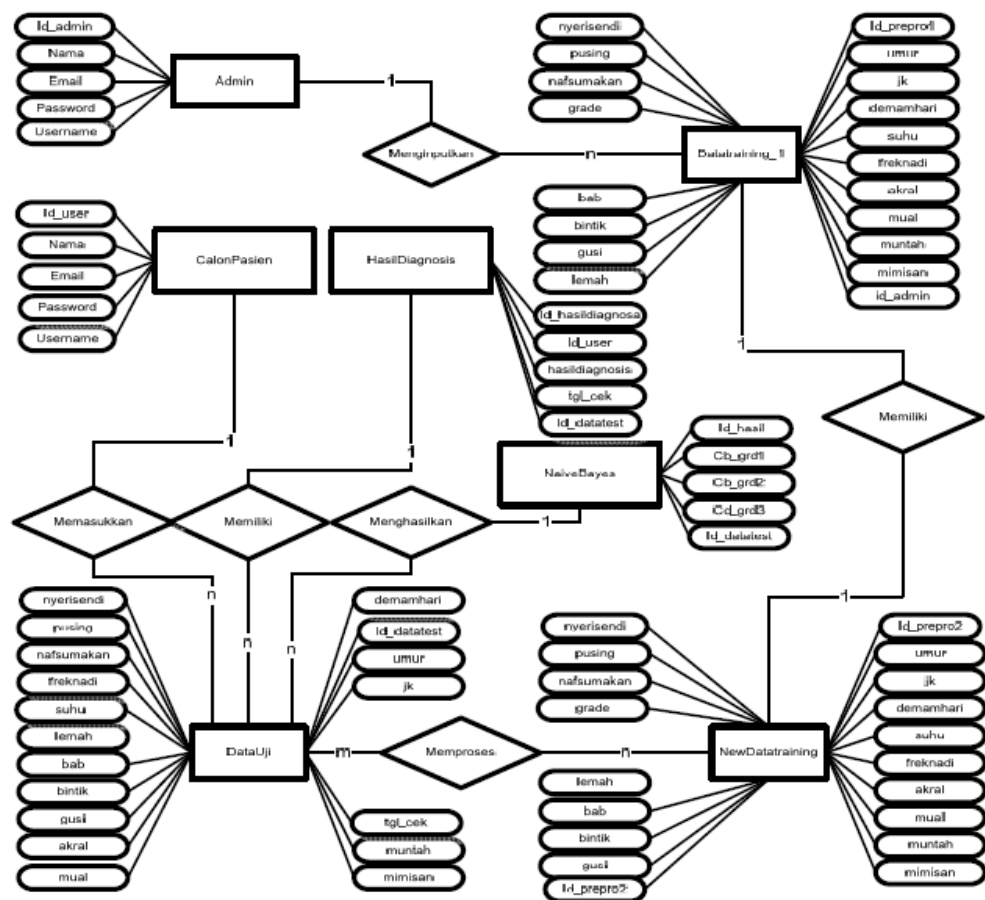
Gambar 3.3. DFD Level 1

Detail dari alur data digambarkan pada DFD level 2 sebagai berikut:



Gambar 3.4. DFD Level 2

Desain basis data yang digunakan:



Gambar 3.5. ER Diagram

3.3 Implementasi Program

Data mentah yang berasal dari rekam medik pasien akan melalui tahap pemrosesan awal, dimana data dalam data latih dianalisa dengan empat proses inti, yaitu penanganan nilai yang hilang (*handling missing values*), mengatasi duplikasi data, menghilangkan *outlier* dan mentransformasi nilai-nilai dalam suatu atribut yang sifatnya kontinyu.

Action :	Handling Missing Values		Remove Duplicate		Remove Outlier		Normalization		Back		Halaman : 1 2 3 4 5 >>>						
id_pasien	Umur Pasien	Jk Pasien	Demam Hari	Suhu Tubuh	Frek. nadi	Akral	Mual	Muntah	Nyeri Sendi	Pusing	Nafsu Makan	Lemah	Bintik Merah	Gusi Berdarah	Mimisan	Bab	Grade
1	6	pria	4	36.8	80	hangat	iya	tidak	tidak	tidak	normal	tidak	tidak	tidak	tidak	normal	1
4	8	pria	4	38.8	120	hangat	iya	iya	tidak	tidak	berkurang	iya	tidak	tidak	iya	normal	2
7	12	perempuan	4	37.5	100	hangat	tidak	tidak	tidak	normal	normal	iya	iya	tidak	tidak	normal	2
194	27	perempuan	4	37	80	hangat	tidak	tidak	tidak	iya	normal	tidak	iya	tidak	tidak	tidak normal	2
195	9	perempuan	4	37.8	100	hangat	iya	iya	tidak	iya	normal	tidak	tidak	tidak	iya	tidak normal	2
196	4	pria	5	37	100	hangat	tidak	tidak	tidak	tidak	normal	tidak	tidak	tidak	iya	normal	2
12	22	perempuan	5	38.3	88	hangat	iya	tidak	iya	iya	normal	tidak	tidak	tidak	tidak	normal	1
17	3	pria	2	37.7	100	hangat	iya	tidak	tidak	tidak	normal	tidak	tidak	tidak	tidak	normal	1
18	27	pria	5	36	88	hangat	iya	tidak	tidak	tidak	berkurang	iya	tidak	tidak	tidak	normal	1
19	30	perempuan	4	30	104	dingin	tidak	iya	tidak	tidak	berkurang	tidak	tidak	tidak	tidak	normal	3
20	11	pria	3	37	100	dingin	tidak	iya	tidak	tidak	normal	iya	tidak	tidak	tidak	tidak normal	3
21	12	perempuan	3	39.5	120	dingin	tidak	iya	tidak	tidak	normal	tidak	tidak	tidak	tidak	normal	3
22	30	perempuan	4	37.5	104	dingin	iya	iya	tidak	tidak	normal	tidak	tidak	tidak	tidak	normal	3
24	56	pria	3	36	96	hangat	tidak	tidak	tidak	iya	berkurang	iya	tidak	tidak	tidak	normal	1
193	15	perempuan	5	38	102	hangat	tidak	tidak	tidak	iya	normal	iya	iya	tidak	tidak	normal	2
34	12	pria	3	38	108	hangat	iya	iya	tidak	tidak	normal	iya	tidak	tidak	tidak	normal	1
93	46	pria	3	37	88	hangat	tidak	tidak	iya	iya	normal	tidak	tidak	iya	tidak	normal	2
191	14	pria	2	39.4	100	hangat	tidak	iya	tidak	tidak	normal	tidak	tidak	tidak	tidak	normal	1
192	10	pria	4	38	100	hangat	iya	iya	tidak	iya	normal	tidak	tidak	tidak	tidak	normal	1
42	46	perempuan	6	36.7	96	hangat	tidak	tidak	tidak	iya	berkurang	iya	tidak	tidak	tidak	normal	1

Gambar 3.6. Tampilan Tahap Pemrosesan Awal

Hasil Preprocessing Normalization		Halaman : 1 2 3 4 5 6 7 8 >>>		Home	Logout
ID_Pasien	Demam Hari	Suhu Tubuh	Frekuensi Nadi	Grade	
1	0.428571428571	0.686868686869	0.166666666667	1	
4	0.428571428571	0.888888888889	0.583333333333	2	
7	0.428571428571	0.757575757576	0.375	2	
194	0.428571428571	0.707070707071	0.166666666667	2	
195	0.428571428571	0.787878787879	0.375	2	
196	0.571428571429	0.707070707071	0.375	2	
12	0.571428571429	0.838383838384	0.25	1	
17	0.142857142857	0.777777777778	0.375	1	
18	0.571428571429	0.606060606061	0.25	1	
19	0.428571428571	0	0.416666666667	3	
20	0.285714285714	0.707070707071	0.375	3	
21	0.285714285714	0.959595959596	0.583333333333	3	
22	0.428571428571	0.757575757576	0.416666666667	3	
24	0.285714285714	0.606060606061	0.333333333333	1	
193	0.571428571429	0.808080808081	0.395833333333	2	
34	0.285714285714	0.808080808081	0.458333333333	1	
93	0.285714285714	0.707070707071	0.25	2	
191	0.142857142857	0.949494949495	0.375	1	
192	0.428571428571	0.808080808081	0.375	1	

Gambar 3.7. Tampilan Normalisasi Data

Info Hasil Diagnosis

Anda suspect DBD Grade 1 hari ke 4
Segera periksakan diri Anda ke dokter terdekat untuk memastikan diagnosis ini.

Rincian :
Grade 1 : 64.01%, Grade 2 : 35.99%, Grade 3 : 0%

Home Admin Change Pas

Search

Go

DHF Diagnose Application

- Tentang DBD >>
- Cegah DBD >>
- Diagnosa Diri Anda
- Berita Tentang DBD

Umur

JK

Demam hari

Suhu tubuh 39

Frekuensi nadi 120 /menit

Mual Iya

Muntah Tidak

Nyeri sendi Tidak

Gusi berdarah Tidak

Mimisan Tidak

Akral Hangat

BAB Normal

Nafsu makan Berkurang

Hapus Proses

DEMOEDITION

HD User : 11/Date : 31 January 2010

User Login

Login As : 1

Diagnosa : Belum Diagnosa

Last Date : 29 December 2010

Logout

Gambar3.8. Tampilan Hasil Diagnosis

Setelah data uji pengguna melewati proses validasi, data uji akan melalui proses normalisasi data terlebih dahulu, dimana data uji pengguna seperti suhu tubuh, frekuensi nadi dan demam hari akan ditransformasi menjadi nilai dengan kisaran 0.0 hingga 1.0. Kemudian data uji yang telah di-normalisasi ditentukan kelasnya dengan menggunakan klasifikasi naïve bayes.

3.4 Uji Coba Sistem

Uji coba dibagi menjadi dua tahap. Tahap I dilakukan untuk melihat tingkat akurasi sistem dalam menentukan kelas data uji termasuk DBD atau tidak. Tahap II dilakukan untuk melihat akurasi sistem dalam menentukan level dari data uji yang telah dinyatakan terinfeksi DBD. Pada masing-masing tahap dilakukan uji coba sebanyak tiga kali dengan jumlah data yang berbeda. Uji coba 1 dengan 50 data, uji coba 2 dengan 90 data dan terakhir dengan 141 data.

Tabel 3.2. Hasil Uji Coba Tahap I

Pengujian Ke	Jumlah Data Latih	DBD		Akurasi	Error
		Benar	Salah		
1	50	50	0	100	0
2	90	89	1	98,89	1,11
3	141	140	1	99,29	0,71

Dari tabel 2 didapatkan rata-rata tingkat akurasi dari sistem sebesar 99,39%. Dapat disimpulkan bahwa sistem dapat mengklasifikasikan apakah seorang pengguna terinfeksi DBD atau tidak.

Tabel 3.3. Hasil Uji Coba Tahap II

Pengujian ke	Jumlah Data Latih	Grade I		Grade II		Grade III		Akurasi	Error
		Benar	Salah	Benar	Salah	Benar	Salah		
1	40	9	8	15	2	13	3	74	26
2	90	20	10	32	3	19	6	78,88	21,12
3	141	34	19	50	3	29	6	80,14	19,86

Dari tabel 3 didapatkan rata-rata tingkat akurasi sistem dalam menentukan level (grade) penyakit DBD yang diidap seorang pengguna adalah 77,67%. Adanya penurunan dalam tingkat akurasi yang dihasilkan dapat disebabkan oleh adanya keterkaitan/hubungan antara atribut-atribut yang digunakan. Salah satu syarat algoritma Naïve Bayes adalah tidak adanya hubungan antar masing-masing atribut. Dapat dilihat pula bahwa tingkat akurasi berbanding lurus dengan jumlah data latih yang digunakan. Semakin banyak data latih yang digunakan maka semakin baik sistem dalam menentukan level dari penyakit DBD pengguna.

IV. KESIMPULAN

Kesimpulan yang dapat diambil dari penelitian yang dilakukan adalah sebagai berikut:

1. Sistem klasifikasi DBD menggunakan algoritma Naïve Bayes yang dibangun memiliki rata-rata tingkat akurasi yang tinggi (99,39%) dalam menentukan apakah data uji pengguna terinfeksi atau tidak.
2. Rata-rata akurasi sistem menurun menjadi 77,67% ketika digunakan untuk menentukan level penyakit DBD yang diidap pengguna. Hal ini dapat disebabkan karena adanya keterkaitan antar atribut yang digunakan dalam klasifikasi.

V. DAFTAR PUSTAKA

- [1] Anumalla, Kalyani. 2007. “*Data Preprocessing Management System*”. Faculty of The University of Akron.
- [2] Cao, J., Panetta, R., Yue,S., and Steyaert, A. 2003. “A Naive Bayes Model To Predict Coupling Between Seven Transmembrane Domain Receptors And G-Proteins”. *Department of Molecular Sciences*. Vol. 19 no. 2.
- [3] Kohonen, J., Talikota, S., Corander, J., and Auvinen, P. 2008. “A Naïve Bayes Classifier for Protein Function Prediction”. In *Silico Biology* 9.
- [4] Kotsiantis, S. B., Kanellopoulos , D., and Pintelas, P. E. 2006. “Data Preprocessing for Supervised Learning”. *International Journal Of Computer Science*, Volume 1 Number 2.
- [5] Nahar, J., Yi-Ping, P.C. 2007. “Kernel-Based Naive Bayes Classifier For Breast Cancer Prediction. *Journal of Biological Systems*”, Vol. 15 No. 1 (2007) 17–25. Australia.
- [6] Pang-Ning Tan, Michael S., and Vipin K. 2006. “*Introduction to Data Mining*”. Pearson Education Inc. Boston USA.
- [7] World Health Organization. 1997. “*Dengue Haemorrhagic Fever: Diagnosis, Treatment, Prevention and Control*”. World Health Organization, Second Edition, Genewa.
- [8] Yang, Y. and Webb, G. 2005. “Discretization for Naïve-Bayes Learning:Managing Discretization Bias and Variance”. *Springer Science Bussines Media*, 74:39- 74.