

THE IMPACT OF DUMMY VARIABLE ON NAÏVE BAYES METHOD IN GYNECOLOGY DISEASE'S CLASSIFICATION CASE

Rahmawati Danu Kusuma
Informatika, Fakultas MIPA
Universitas Sebelas Maret

Jl. Ir. Sutami No. 36 A Surakarta
rahmadanukusuma@student.uns.ac.id

Ristu Saptono
Informatika, Fakultas MIPA
Universitas Sebelas Maret

Jl. Ir. Sutami No. 36 A Surakarta
ristu.saptono@staff.uns.ac.id

Afrizal Doewes
Informatika, Fakultas MIPA
Universitas Sebelas Maret

Jl. Ir. Sutami No. 36 A Surakarta
afrizal.doewes@staff.uns.ac.id

ABSTRACT

One of the diseases that often occur in women is gynecology disease. Some kind of gynecology's disease include pelvic inflammatory disease, uterine myoma, cervical cancer, and ovarian cancer. Basically most of the causes of these deaths could be prevented, but because of many factors such as lack of doctors and gynecologists, the cost of consulting, and the lack of information regarding the health of pregnant women become the reasons why gynecology disease failed to be detected earlier. This research will go into the impact of using dummy variable on naïve bayes method in gynecology disease's classification case.

The steps included in this research are data collection, analysis and design, implementation using PHP, and also testing.

Testing use Confusion Matrix which performed five times with changing partition as training data and test data. In experiment using preliminary data, Naïve Bayes's accuracy is 88% (with 18 parameters) and get 88.88% (with 24 parameters) when using dummy data. Through this research, best method on this gynecology disease's classification is using Naïve Bayes with dummy data which all possible value has already become binary.

Keywords : *backward elimination, dummy variable, Gynecology Disease, PHP, Naïve Bayes.*

1. PENDAHULUAN

Salah satu penyakit yang sering terjadi pada kaum wanita adalah penyakit kandungan. Beberapa jenis penyakit kandungan tersebut antara lain radang panggul, mioma uteri, kanker serviks, dan kanker ovarium. Pada dasarnya sebagian besar penyebab kematian tersebut dapat dicegah, namun karena faktor seperti keterbatasan dokter maupun ahli kandungan, biaya konsultasi yang tidak murah dan kurangnya informasi mengenai kesehatan pada ibu hamil, menyebabkan keempat penyakit tersebut menjadi penyakit kandungan yang sering dijumpai pada wanita [1].

Oleh karena itu, perlu dibuat suatu aplikasi yang dapat membantu mendiagnosa jenis penyakit kandungan tersebut sehingga masyarakat tidak harus pergi ke dokter ahli kandungan apabila ingin mengetahui apakah mereka terkena penyakit kandungan atau tidak.

Beberapa penelitian yang pernah dilakukan terkait dengan penelitian ini adalah aplikasi diagnosa kanker kandungan dengan menggunakan metode *Naïve Bayes*. Pada penelitian ini, penulis membangun sebuah aplikasi

untuk mendiagnosa kanker kandungan menggunakan algoritma *Naïve Bayes*. Penelitian ini menggunakan 19 kriteria gejala yang kemudian akan menghitung nilai kemungkinan masing-masing gejala dengan mengacu data training yang ada dengan menggunakan metode *Naïve Bayes* [2]. Penelitian lain dilakukan dengan mengajukan dua pendekatan berbeda untuk penyelesaian kasus diagnosa penyakit kandungan, yaitu pendekatan probabilistik dengan *Naïve Bayesian Classifier* dan jaringan syaraf tiruan *Learning Vector Quantization*. Dari hasil uji coba dapat ditunjukkan *Naïve Bayesian Classifier* dengan *Laplacian Smoothing* merupakan metode paling baik dalam menyelesaikan kasus klasifikasi penyakit kandungan [3].

Pada penelitian ini penulis akan melakukan analisis pengaruh *dummy variable* dan *backward elimination Chi Square* pada metode *Naïve Bayes* dengan studi kasus klasifikasi penyakit kandungan dengan melihat perbandingan akurasi penggunaan *dummy variable* pada perhitungan *Naïve Bayes*.

2. LANDASAN TEORI

2.1 Dummy Variable

Dalam statistik dan ekonometrik, khususnya dalam analisis regresi, variabel *dummy* (juga dikenal sebagai variabel indikator, variabel desain, indikator *Boolean*, variabel kategori, variabel biner, atau variabel kualitatif [4] adalah salah satu variabel yang mengambil nilai 0 atau 1 untuk menunjukkan ada atau tidaknya beberapa efek kategori yang diharapkan dapat menggeser hasilnya [5].

2.2 Naïve Bayes

Bayes merupakan teknik prediksi berbasis probabilistic sederhana atau aturan *Bayes* dengan asumsi independen (ketidaktergantungan) yang kuat (naïf). Dengan kata lain, dalam *Naïve Bayes*, model yang digunakan adalah "model fitur independen [6]. Teorema *Naïve Bayes* dapat dinyatakan dalam persamaan berikut:

$$P(X_k|Y) = \frac{P(Y|X_k)}{\sum_i P(Y|X_i)} \dots\dots\dots(2.2)$$

Dimana, keadaan *Posterior* (Probabilitas X_k di dalam Y) dapat dihitung dari keadaan prior (Probabilitas Y di dalam X_k dibagi dengan jumlah dari semua probabilitas Y di dalam semua X_i).

2.3 Laplacian Smoothing

Untuk mengatasi nilai probabilitas kondisional pada *Naïve Bayes Classifier* yang dapat saja bernilai 0,

digunakan teknik *smoothing*. Cara yang digunakan pada teknik *Laplacian Smoothing* adalah dengan cara menambahkan angka 1 pada perhitungan *Likelihood*

$$P(F_i|C) = \frac{1+n(F_i,C)}{|W|+n(C)} \dots\dots\dots(2.3)$$

dimana $n(F_i,C)$ adalah jumlah *term* F_i yang ditemukan di seluruh data pelatihan dengan kategori C , $n(C)$ adalah jumlah *term* di seluruh data pelatihan dengan kategori C , dan $|W|$ adalah jumlah seluruh *term* dari seluruh data pelatihan [7].

2.4 Confusion Matrix

Confusion matrix adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Rumus ini melakukan perhitungan dengan empat keluaran yaitu recall, presisi, akurasi, dan error rate[8].

2.5 Penyakit Kandungan

Penyakit kandungan merupakan penyakit berbahaya yang menyerang kaum wanita. Penyakit kandungan biasanya menyerang organ reproduksi wanita dan berpotensi menyebabkan kemandulan. Jenis penyakit kandungan sendiri beraneka ragam, diantaranya adalah radang panggul, *mioma uteri*, kanker *serviks*, dan kanker *ovarium*.

Radang panggul atau *Pelvic Inflammatory Disease* (PID) adalah infeksi traktus genital atas yang merupakan salah satu komplikasi dari infeksi menular seksual. Gejala penyakit radang panggul berupa nyeri perut bagian bawah, temperatur oral lebih dari 38°C, keluar cairan dari vagina, pendarahan tidak teratur, sakit kepala, lesu, nyeri berhubungan seksual dan nyeri buang air kecil [9].

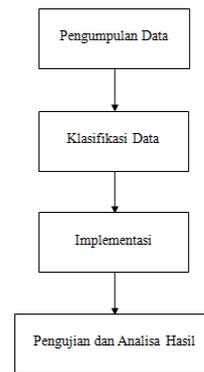
Mioma Uteri adalah tumor jinak otot polos yang terdiri dari sel-sel jaringan otot polos, jaringan fibroid dan kolagen [10]. Gejala yang terjadi dapat digolongkan seperti berikut: pendarahan abnormal, nyeri yang disertai nekrosis setempat dan peradangan, serta gejala tanda penekanan [11].

Kanker *serviks* merupakan kelainan yang terjadi pada sel-sel serviks yang berkembang dengan cepat dan tidak terkontrol. Gejala kanker serviks meliputi pendarahan abnormal, pendarahan haid abnormal, nyeri panggul, nyeri ketika berhubungan seksual, keputihan, nyeri buang air kecil, anemia, nyeri punggung bagian bawah, penurunan nafsu makan dan penurunan berat badan drastis.

Kanker ovarium adalah kanker yang bermula pada indung telur (ovarium) wanita. Gejala kanker ovarium adalah sebagai berikut [9][12][13]: nyeri ketika berhubungan seksual, kembung, sulit buang air besar, sering buang air kecil, nafsu makan menurun, cepat lelah, anemia, nyeri panggul dan nyeri punggung bagian bawah.

3. METODOLOGI

Langkah-langkah metode dalam penelitian ini diuraikan ke dalam skema tahapan penelitian agar memperlihatkan proses yang jelas dan sistematis seperti ditunjukkan pada Gambar 3.1



Gambar 3.1 Metodologi

3.1. Pengumpulan Data

Dalam penelitian ini, digunakan 18 gejala klinis yang dirasakan oleh pasien. Data diperoleh dari penelitian terkait sebelumnya yaitu penelitian oleh Prabhawaningrum yang terdiri dari 125 data rekam medik pasien RSUD Dr. Moewardi Solo. Tabel 3.1 merupakan gejala klinis yang dijadikan sebagai parameter dalam klasifikasi penyakit kandungan [14].

Tabel 3.1 Daftar Gejala Penyakit

No.	Gejala	Nilai
1	Anemia	Tidak
		Ya
2	Nyeri Haid	Tidak
		Nyeri Setelah Haid
		Nyeri Ketika Haid
3	Susah Hamil	Tidak
		Ya
4	Benjolan di perut	Tidak
		Ya
5	Pendarahan	Tidak
		Pendarahan Menstruasi Abnormal
		Pendarahan Tiba-tiba
6	Nyeri Berhubungan Seksual	Tidak
		Ya
7	Cepat Lelah	Tidak
		Ya
8	Penurunan Berat Badan	Tidak
		Ya
9	Nyeri Panggul	Tidak
		Ya
10	Gangguan Pencernaan	Tidak
		Sembelit
		Diare
11	Nyeri Perut	Tidak
		Nyeri pada Rongga Perut
		Nyeri Perut Bagian Bawah
		Nyeri Perut Bagian Bawah dan Punggung
12	Nyeri Punggung	Tidak
		Ya

13	Penurunan Nafsu Makan	Tidak
		Ya
14	Demam	Tidak
		Ya
15	Sakit Kepala	Tidak
		Ya
16	Kembung	Tidak
		Ya
17	Keputihan	Tidak
		Ya
18	Gangguan BAK	Tidak
		Sering BAK
		Nyeri BAK
		Nyeri dan Sering BAK

Berdasar pada gejala yang telah dijabarkan pada tabel di atas, dapat diklasifikasikan menjadi 5 kategori yang dapat dilihat pada Tabel 3.2 berikut ini:

Tabel 3.2 Daftar Klasifikasi Penyakit

No.	Kategori Penyakit
1	Tidak Termasuk 4 Jenis Penyakit Kandungan
2	Radang Panggul
3	Mioma Uteri
4	Kanker Serviks
5	Kanker Ovarium

3.2. Analisis dan Perancangan

Pada penelitian ini akan diteliti mengenai pengaruh *dummy variable* dan *backward elimination Chi Square* pada metode *Naïve Bayes*. Untuk itu penulis merencanakan untuk membuat dua skenario percobaan yang berbeda-beda penerapan metodenya. Dua skenario yang akan dibuat adalah seperti pada Tabel 3.3 berikut :

Tabel 3.3 Skenario Hitungan

Nama Skenario	Dummy Variable	Naïve Bayes
Skenario Satu	-	√
Skenario Dua	√	√

Skenario pertama adalah percobaan data awal yang mana ada 18 gejala klinis dengan metode *Naïve Bayes*. Skenario yang kedua adalah percobaan menggunakan data yang telah di dummy yaitu dengan 24 gejala yang telah dijabarkan dengan metode *Naïve Bayes*. Sehingga dengan adanya skenario tersebut nantinya akan dibandingkan akurasi dari masing-masing skenario.

3.3. Implementasi

Penelitian ini dilakukan dengan menggunakan komputer yang memiliki spesifikasi sebagai berikut: prosesor Intel core i3 2,20 GHz, RAM 2GB, HDD 640 GB, OS Windows 7, XAMPP 1.7.3. Penulisan kode program pada penelitian ini menggunakan bahasa

pemrograman PHP (*Hypertext Preprocessor*) dan MySQL sebagai penyimpanan data.

3.4. Pengujian

Pengujian dilakukan pada aturan yang terbentuk pada proses klasifikasi untuk mengetahui prosentase keakuratan aturan. Pada setiap skenario percobaan yang telah ditentukan, dilakukan pengujian. Pengujian yang dilakukan pada penelitian ini menggunakan metode *k-fold cross validation*. Pengujian dengan metode *k-fold cross validation* yaitu dengan cara membagi data sebanyak k bagian, untuk data training dan data uji. Pengujian ini dilakukan sebanyak k kali dengan mengganti-ganti partisi yang berfungsi sebagai data training dan data uji. Penghitungan hasil klasifikasi menggunakan *precision*, *recall*, dan akurasi. Hasil klasifikasi akan ditampilkan dengan menggunakan confusion matrix dengan nilai k=5 karena data diklasifikasikan dalam 5 kelas. *Confusion Matrix* merupakan sebuah metode untuk evaluasi yang menggunakan tabel matriks. Penerapan penggunaan *confusion matrix* dalam penelitian ini dilihat di Tabel 3.4.

Tabel 3.4. Confusion Matrix

	Prediction					
	C001	C002	C003	C004	C005	
C001	TP-C001	Error	Error	Error	Error	TC001
C002	Error	TP-C002	Error	Error	Error	TC002
C003	Error	Error	TP-C003	Error	Error	TC003
C004	Error	Error	Error	TP-C004	Error	TC004
C005	Error	Error	Error	Error	TP-C005	TC005
Total	PreC001	PreC002	PreC003	PreC004	PreC005	

Keterangan:

- C001= Kelas Tidak Termasuk 4 Jenis Penyakit Kandungan
- C002 = Kelas Radang Panggul
- C003 = Kelas Mioma Uteri
- C004 = Kelas Kanker Serviks
- C005 = Kelas Kanker Ovarium
- TP = True Positive
- Prei = Jumlah prediksi dalam kelas i
- TCi = Jumlah data aktual dalam kelas i

Evaluasi dengan *confusion matrix* menghasilkan nilai *accuracy*, *precision*, dan *recall*. Nilai *accuracy* merupakan persentase jumlah record data yang diklasifikasikan secara benar oleh sebuah algoritma dapat membuat klasifikasi setelah dilakukan pengujian pada hasil klasifikasi tersebut[15]. Nilai *precision* atau dikenal juga dengan nama *confidence* merupakan proporsi jumlah kasus yang diprediksi positif yang juga positif benar pada data yang sebenarnya. Sedangkan nilai dari *recall* merupakan proporsi jumlah kasus positif yang sebenarnya yang diprediksi positif secara benar [16].

True Positive adalah jumlah *record* positif yang diklasifikasikan sebagai positif, kemudian masukkan data uji. Setelah data uji dimasukkan ke dalam *confusion matrix*, hitung nilai-nilai yang telah dimasukkan tersebut untuk dihitung jumlah *Precision (P)*, *Recall (R)*, dan *Accuracy* [17].

Perhitungan untuk akurasi dapat dikalkulasi sebagai berikut:

$$Precision (PC_i) = \frac{TP - C_i}{PreC_i}; \dots\dots\dots (3.1)$$

$$Recall (RC_i) = \frac{TPC_i}{TC_i}; \dots\dots\dots (3.2)$$

$$Akurasi = \frac{\sum TP - C_i}{Jumlah\ Data} \times 100\% \dots\dots\dots (3.3)$$

4. HASIL DAN PEMBAHASAN

4.1. Deskripsi dan Pemodelan Data

4.1.1 Penentuan Kelas

Pada penelitian ini, terdapat empat penyakit kandungan yang akan diklasifikasikan, yaitu radang panggul, mioma uteri, kanker serviks, dan kanker ovarium. kemudian juga terdapat kelas yan merupakan kelas untuk pasien yang tidak termasuk pada empat jenis penyakit kandungan yang telah disebutkan. Selanjutnya, penulis memasukkan lima kelas tersebut ke dalam aplikasi yang telah dibuat dimana masing-masing kelas tersebut akan memiliki ID yang unik (berbeda-beda). Dapat dilihat pada Tabel 4.1 lima kelas yang telah dimasukkan pada sistem beserta ID masing-masing kelas.

Tabel 4.1 Kategori Penyakit dalam Sistem

ID Penyakit	Kategori Penyakit
C001	Tidak Termasuk 4 Jenis Penyakit
C002	Radang Panggul
C003	Mioma Uteri
C004	Kanker Serviks
C005	Kanker Ovarium

4.1.2 Penentuan Parameter

Penelitian ini menggunakan 18 gejala klinis awal yang dapat digunakan untuk mendeteksi empat penyakit kandungan yang telah dijabarkan sebelumnya yang kemudian akan menjadi 24 gejala klinis setelah dilakukan proses *dummy variable*. Penulis memasukkan 18 gejala klinis tersebut pada aplikasi yang telah dibuat, kemudian masing-masing gejala tersebut akan memiliki ID yang unik (berbeda-beda). Masing-masing gejala diberikan rentang nilai dari 0 sampai 1 tergantung pada kasus di setiap gejala. Pemberian nilai berdasarkan tingkat parahnya kasus, semakin parah kasus semakin besar nilainya. Tabel 4.2 adalah gejala yang akan dimasukkan pada sistem yang diambil dari penelitian [14].

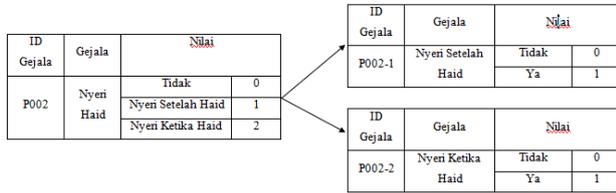
Tabel 4.2 Gejala Dalam Sistem

ID Gejala	Gejala	Nilai	
P001	Anemia	Tidak	0
		Ya	1
P002	Nyeri Haid	Tidak	0
		Nyeri Setelah Haid	1
		Nyeri Ketika Haid	2

P003	Susah Hamil	Tidak	0
		Ya	1
P004	Benjolan di Perut	Tidak	0
		Ya	1
P005	Pendarahan	Tidak	0
		Pendarahan Menstruasi Abnormal	1
		Pendarahan Tiba-tiba	2
P006	Nyeri Berhubungan Seksual	Tidak	0
		Ya	1
P007	Cepat Lelah	Tidak	0
		Ya	1
P008	Penurunan Berat Badan	Tidak	0
		Ya	1
P009	Nyeri Panggul	Tidak	0
		Ya	1
P010	Gangguan Pencernaan	Tidak	0
		Sembelit	1
		Diare	2
P011	Nyeri Perut	Tidak	0
		Nyeri Pada Rongga Perut	1
		Nyeri Perut Bagian Bawah	2
		Nyeri Perut Bagian Bawah dan Pinggul	3
P012	Nyeri Punggung	Tidak	0
		Ya	1
P013	Penurunan Nafsu Makan	Tidak	0
		Ya	1
P014	Demam	Tidak	0
		Ya	1
P015	Sakit Kepala	Tidak	0
		Ya	1
P016	Kembung	Tidak	0
		Ya	1
P017	Keputihan	Tidak	0
		Ya	1
P018	Gangguan BAK	Tidak	0
		Sering BAK	1
		Nyeri BAK	2
		Nyeri dan Sering BAK	3

Beberapa gejala di atas memiliki *possible value* lebih dari dua (*non-binary*) sehingga perlu dilakukan pemecahan agar nilainya menjadi *binary* dengan menggunakan *dummy variable*. Proses *dummy* dilakukan pada semua parameter yang memiliki *possible value* tidak

binary, yaitu P002, P005, P010,P011, dan P018. Salah satu contohnya adalah P002 yang memiliki nilai 2 sehingga dianggap tidak *binary*. Proses *dummy* untuk P002 dapat dilihat pada Gambar 4.1 dimana P002 mengalami penjabaran menjadi P002-1 dan P002-2. Contoh lain adalah P005 dapat dilihat pada Gambar 4.2. Sedangkan Tabel 4.3 adalah uraian gejala klinis yang telah di-*dummy*.



Gambar 4.1 Proses *Dummy* P002

Tabel 4.3 Gejala yang di-*Dummy*

ID Gejala	Gejala	Nilai	
		Tidak	Ya
P001	Anemia	Tidak	0
		Ya	1
P002-1	Nyeri Setelah Haid	Tidak	0
		Ya	1
P002-2	Nyeri Ketika Haid	Tidak	0
		Ya	1
P003	Susah Hamil	Tidak	0
		Ya	1
P004	Benjolan di Perut	Tidak	0
		Ya	1
P005-1	Pendarahan Menstruasi Abnormal	Tidak	0
		Ya	1
P005-2	Pendarahan Tiba-Tiba	Tidak	0
		Ya	1
P006	Nyeri Berhubungan Seksual	Tidak	0
		Ya	1
P007	Cepat Lelah	Tidak	0
		Ya	1
P008	Penurunan Berat Badan	Tidak	0
		Ya	1
P009	Nyeri Panggul	Tidak	0
		Ya	1
P010-1	Sembelit	Tidak	0
		Ya	1
P010-2	Diare	Tidak	0
		Ya	1
P011-1	Nyeri Pada Rongga Perut	Tidak	0
		Ya	1
P011-2	Nyeri Perut Bagian Bawah	Tidak	0
		Ya	1
P011-3	Nyeri Perut Bagian Bawah dan Pinggul	Tidak	0
		Ya	1
P012	Nyeri Punggung	Tidak	0

		Ya	1
P013	Penurunan Nafsu Makan	Tidak	0
		Ya	1
P014	Demam	Tidak	0
		Ya	1
P015	Sakit Kepala	Tidak	0
		Ya	1
P016	Kembung	Tidak	0
		Ya	1
P017	Keputihan	Tidak	0
		Ya	1
P018-1	Sering BAK	Tidak	0
		Ya	1
P018-2	Nyeri BAK	Tidak	0
		Ya	1

4.1.3 Cross Tabulation

Analisis *Cross tabulation* merupakan analisis deskriptif yang digunakan untuk mengetahui hubungan dari dua variabel. Penjelasan lebih rinci mengenai jumlah inputan, kelas, dan masing-masing parameter untuk data awal dapat dilihat pada Tabel 4.4 dan Tabel 4.5 untuk data yang telah mengalami proses *dummy*.

Tabel 4.4 *Cross Tabulation* Data Awal

Parameter	Nilai	Kelas				
		C001	C002	C003	C004	C005
Anemia	Tidak	17	25	11	0	0
	Ya	8	1	13	26	24
Nyeri Haid	Tidak	14	17	11	16	17
	Nyeri Setelah Haid	0	2	4	5	1
Susah Hamil	Nyeri Ketika Haid	11	7	9	5	6
	Tidak	24	22	19	26	18
Benjolan di Perut	Ya	1	4	5	0	6
	Tidak	24	24	0	25	11
Pendarahan	Ya	1	2	24	1	13
	Tidak	23	5	0	1	0
	Pendarahan Menstruasi Abnormal	1	16	5	0	0
Nyeri Berhubungan Seksual	Pendarahan Tiba-tiba	1	5	19	25	24
	Tidak	22	8	17	14	18
Cepat Lelah	Ya	3	18	7	12	6
	Tidak	15	14	14	5	3
Penurunan Berat Badan	Ya	10	12	10	21	21
	Tidak	21	24	10	7	12
Nyeri Panggul	Ya	4	2	14	19	12
	Tidak	24	17	21	3	10

	Ya	1	9	3	23	14
Gangguan Pencernaan	Tidak	12	19	11	23	10
	Sembelit	7	2	13	3	12
	Diare	6	5	0	0	2
Nyeri Perut	Tidak	22	1	0	0	0
	Nyeri Pada Rongga Perut	2	0	0	4	0
	Nyeri Perut Bagian Bawah	1	19	11	13	14
	Nyeri Perut Bagian Bawah dan Punggul	0	6	13	9	10
Nyeri Punggung	Tidak	23	11	23	17	14
	Ya	2	15	1	9	10
Penurunan Nafsu Makan	Tidak	18	19	24	5	7
	Ya	7	7	0	21	17
Demam	Tidak	22	4	14	21	15
	Ya	3	22	10	5	9
Sakit Kepala	Tidak	9	20	7	2	2
	Ya	16	6	17	24	22
Kembung	Tidak	20	25	24	23	8
	Ya	5	1	0	3	16
Keputihan	Tidak	21	12	13	0	22
	Ya	4	14	11	26	2
Gangguan BAK	Tidak	18	11	15	0	4
	Sering BAK	5	0	2	0	7
	Nyeri BAK	2	14	3	23	10
	Nyeri dan Sering BAK	0	1	4	3	3

Pendarahan Menstruasi Abnormal	Tidak	24	10	19	26	24
	Ya	1	16	5	0	0
Pendarahan Tiba-Tiba	Tidak	24	21	5	1	0
	Ya	1	5	19	25	24
Nyeri Berhubungan Seksual	Tidak	22	8	17	12	18
	Ya	3	18	7	12	6
Cepat Lelah	Tidak	15	14	14	15	3
	Ya	10	12	10	21	21
Penurunan Berat Badan	Tidak	21	24	10	7	12
	Ya	4	2	14	19	12
Nyeri Punggul	Tidak	24	17	21	3	10
	Ya	1	9	3	23	14
Sembelit	Tidak	18	24	11	23	12
	Ya	7	2	13	3	12
Diare	Tidak	19	21	24	26	22
	Ya	6	5	0	0	2
Nyeri Pada Rongga Perut	Tidak	23	26	24	22	24
	Ya	2	0	0	4	0
Nyeri Perut Bagian Bawah	Tidak	24	7	13	13	10
	Ya	1	19	11	13	14
Nyeri Perut Bagian Bawah dan Punggul	Tidak	25	20	11	17	14
	Ya	0	6	13	9	10
Nyeri Punggung	Tidak	23	11	23	17	14
	Ya	2	15	1	9	10
Penurunan Nafsu Makan	Tidak	18	19	24	5	7
	Ya	7	7	0	21	17
Demam	Tidak	22	4	14	21	15
	Ya	3	22	10	5	9
Sakit Kepala	Tidak	9	20	7	2	2
	Ya	16	6	17	24	22
Kembung	Tidak	20	25	24	23	8
	Ya	5	1	0	3	16
Keputihan	Tidak	21	12	13	0	22
	Ya	4	14	11	26	2
Sering BAK	Tidak	20	25	18	23	14
	Ya	5	1	6	3	10
Nyeri BAK	Tidak	23	11	17	0	11
	Ya	2	15	7	26	13

Tabel 4.5 Cross Tabulation Data Dummy

Parameter	Nilai	Kelas				
		C001	C002	C003	C004	C005
Anemia	Tidak	17	25	11	0	0
	Ya	8	1	13	26	24
Nyeri Setelah Haid	Tidak	25	24	20	21	23
	Ya	0	2	4	5	1
Nyeri Ketika Haid	Tidak	14	19	15	21	18
	Ya	11	7	9	5	6
Susah Hamil	Tidak	24	22	19	26	18
	Ya	1	4	5	0	6
Benjolan di Perut	Tidak	24	24	0	25	11
	Ya	1	2	24	1	13

4.2. Algoritma Percobaan Sistem

4.2.1 Skenario Satu

Skenario satu adalah percobaan data awal menggunakan 18 gejala dengan metode *Naïve Bayes*. Pada Tabel 4.6 tersaji contoh data training.

Tabel 4.6 Contoh Data Dengan 18 Gejala

Kelas	Amneia	Nyeri Haid	Susah Hamil	Benjolan di Perut	Perubahan Pendarahan	Cepat Lelah	Penurunan BB	Nyeri Panggul	Gangguan Penebaran	Nyeri Perut	Nyeri Punggung	Demam	Sakit Kepala	Kembung	Keputihan	Gangguan BAK
C001	0	2	0	0	0	0	1	0	0	1	0	0	0	1	0	0
C002	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C003	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C004	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C005	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Dari data di atas, dilakukan percobaan menggunakan metode *Naïve Bayes* dengan cara:

1. Menghitung frekuensi kemunculan nilai inputan di masing-masing parameter untuk setiap kelas. Hitung semua parameter sesuai dengan inputan dan ulang percobaan untuk kelas lain, yaitu C002, C003, C004, dan C005.
2. Menambahkan nilai positif yang kecil pada setiap frekuensi kemunculan (*smoothing*). Nilai positif kecil yang diambil adalah satu. Sehingga setiap hasil frekuensi mendapat tambahan nilai 1.
3. Membagi masing-masing frekuensi kemunculan dengan total data masing-masing kelas.
4. Mengalikan semua hasil bagi masing-masing parameter sesuai nilai inputan parameter dan berdasarkan kelas
5. Mencari hasil maksimal diantara perkalian parameter berdasarkan kelas
Hasil maksimal tersebut dianggap sebagai hasil klasifikasi percobaan

4.2.2 Skenario Dua

Alur untuk skenario dua sama dengan alur skenario satu. Pembedanya adalah skenario dua menggunakan data *dummy* dengan 24 parameter.

4.3 Pengembangan Sistem

Tampilan awal sistem akan menyuguhkan opsi skenario percobaan dapat dilihat pada Gambar 4.3. Pilihan penggunaan data antara data murni (data awal) atau data *dummy*.



Gambar 4.3 Tampilan Awal Sistem

Tampilan hasil percobaan akan mencakup proses uji dan menampilkan hasil akurasi seperti yang terlihat pada Gambar 4.4 untuk penggunaan data awal dan Gambar 4.5 untuk penggunaan data *dummy*.

-	C001	C002	C003	C004	C005
C001	22	2	0	0	1
C002	3	21	2	0	0
C003	1	0	23	0	0
C004	2	0	0	23	1
C005	1	1	1	0	21

Presisi = 88.625591050629%
Recall = 88.112820512821%
Akurasi = 88%

Gambar 4.4 Hasil Percobaan Naïve Bayes Menggunakan Data Awal

-	C001	C002	C003	C004	C005
C001	21	3	0	0	1
C002	3	22	1	0	0
C003	0	1	23	0	0
C004	2	0	0	23	1
C005	1	0	1	0	22

Presisi = 89.211965811966%
Recall = 88.915384615385%
Akurasi = 88.8%

Gambar 4.5 Hasil Percobaan Naïve Bayes Menggunakan Data Dummy

4.4 Pengujian

4.4.1 Pengujian Skenario Satu

Hasil confusion matrix untuk uji skenario satu seperti pada Tabel 4.6

Tabel 4.6 Hasil Penjumlahan Confusion Matrix Skenario Satu

		Prediction					
		C001	C002	C003	C004	C005	
Actual Class	C001	22	2	0	0	1	25
	C002	3	21	2	0	0	25
	C003	1	0	23	0	0	25
	C004	2	0	0	23	1	25
	C005	1	1	1	0	21	25
	Total	29	24	26	23	23	

Dari tabel di atas maka nilai *precision* adalah 88.624%, *recall* 88%, dan nilai akurasi adalah 88%

4.4.2 Pengujian Skenario Dua

Hasil confusion matrix untuk uji skenario dua seperti pada Tabel 4.7

Tabel 4.7 Hasil Penjumlahan Confusion Matrix Skenario Dua

		Prediction					
		C001	C002	C003	C004	C005	
Actual Class	C001	21	3	0	0	1	25
	C002	3	22	1	0	0	26
	C003	0	1	23	0	0	24
	C004	2	0	0	23	1	26
	C005	1	0	1	0	22	24
	Total	29	24	26	23	23	

Total	27	26	25	23	24	
-------	----	----	----	----	----	--

Dari tabel di atas maka nilai *precision* adalah 89.21%, *recall* 88.90%, dan nilai akurasi adalah 88.88%.

6. KESIMPULAN DAN SARAN

Penelitian ini menerapkan *dummy variable* pada metode *Naïve Bayes* sebagai algoritma klasifikasi penyakit kandungan. Dua skenario telah diterapkan di penelitian ini. Pengujian terhadap skenario tersebut menggunakan metode pengujian *Confusion Matrix* yang dilakukan sebanyak lima kali dengan mengganti-ganti partisi yang berfungsi sebagai data training dan data uji.

Pada percobaan menggunakan data awal, akurasi *Naïve Bayes* yang didapat adalah 88% (dengan 18 parameter) dan mendapat akurasi 88.88% (dengan 24 parameter).

Sehingga berdasar hasil dari penelitian, metode terbaik untuk kasus penyakit kandungan ini adalah percobaan menggunakan *Naïve Bayes* dengan menggunakan data *dummy* dimana semua *possible value* telah menjadi *binary*.

Pengembangan penelitian ini selanjutnya dapat dilakukan pada tahap pengujian dengan mengacak data training dan data uji. Pengumpulan data gejala untuk penelitian dapat melibatkan pakar secara langsung sehingga data yang digunakan dapat lebih berkualitas.

7. DAFTAR PUSTAKA

- [1] Pangestika, M.W., Irawan, B., Brianorman, Y., 2013. *Sistem Pakar Untuk Diagnosa Penyakit Ibu Hamil Berbasis Mobile*. ICT For Smart Society | Think Ecosystem Act Convergence.
- [2] Karina, N.E., Yamasari, Y., 2013. *Aplikasi Diagnosa Kanker Kandungan Dengan Menggunakan Metode Naïve Bayes*. Jurnal Manajemen Informatika. Vol 2, No 2. ISSN : 2085-3548
- [3] Nugraha, P. A., 2013. *Perbandingan Metode Probabilistik Naive Bayesian Classifier dan Jaringan Syaraf Tiruan Leaning Vector Quantization dalam Kasus Klasifikasi Penyakit Kandungan*. JURNAL ITSMART. Vol 2. No 2. ISSN : 2301-7201
- [4] Gujarati, Damodar. 2003. *Ekonometrika Dasar : Edisi Keenam*. Jakarta: Erlangga.
- [5] Draper, N. R. and Smith, H., 1998. *Dummy Variables, in Applied Regression Analysis, Third Edition*. ISBN: 9780471170822 (print). ISBN: 9781118625590 (online).
- [6] Prasetyo, E., 2012. *DATA MINING : Konsep dan Aplikasi menggunakan MATLAB*. ANDI, Yogyakarta.
- [7] Dai, W., Xue, G.-R., Yang, Q. & Yu, Y., 2007. *Transferring Naive Bayes Classifiers for Text Classification. Association for the Advancement of Artificial Intelligence*.
- [8] Doreswamy, Hemanth. K.S. 2011. *Performance Evaluation of Predictive Classifiers For Knowledge Discovery From Engineering Materials Data Sets*. eprint arXiv:1209.2501. Volume 3, No 3, March 2011.
- [8] Nugroho, T., 2010. *Buku Ajar Ginekologi untuk Mahasiswa Kebidanan*. Nuha Media, Yogyakarta.
- [9] Hardibroto, B. R., 2005. *Mioma Uteri*. *Majalah Kedokteran Nusantara*, 38, 254-259.
- [10] Khoman, JS. 2007. *Mioma-uteri*. *Cermin Dunia Kedokteran*, Edisi Khusus No. 80, 1992:60-3.
- [11] Tjhay, F., 2009. *Ancaman Penyakit Radang Panggul pada Infeksi Menular Seksual*. *Majalah Kedokteran Damianus*, 8, 105-114.
- [12] Norwitz, E., & Shorge, J., 2008. *At a glance Obsteri & Ginekologi*. Erlangga, Jakarta.
- [13] Prabhawaningrum, A. 2013. *Perbandingan Algoritma Levenberg-Marquardt dengan Backpropagation untuk Mendiagnosa Jenis Penyakit Kandungan*. Surakarta: Universitas Sebelas Maret.
- [14] Han, J., & Kamber, M. 2006. *Data Mining Concept and Tehniques*. San Fransisco: Morgan Kauffman. ISBN 13: 978-1-55860-901-3
- [15] Powers, D.M.W. 2011. *Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation*. *Journal of Machine Learning Technologies*. ISSN: 2229-3981 & ISSN: 2229-399X, Volume 2, Issue 1, 2011, pp-37-63
- [16] Andriani, A. 2013. *Sistem Pendukung Keputusan Berbasis Decision Tree Dalam Pemberian Beasiswa Studi Kasus: AMIK "BSI Yogyakarta"*. Seminar Nasional Teknologi Informasi dan Komunikasi 2013 (SENTIKA 2013). ISSN: 2089-9815.