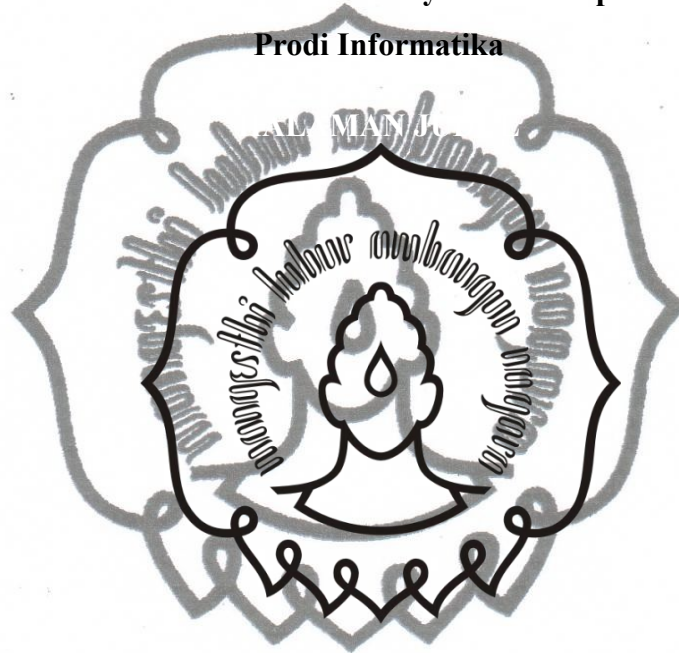


**PENGARUH *DUMMY VARIABLE* PADA METODE *NAÏVE*  
*BAYES* DALAM KASUS KLASIFIKASI PENYAKIT  
KANDUNGAN**

**SKRIPSI**

**Diajukan untuk Memenuhi Salah Satu Syarat Mencapai Gelar Strata Satu  
Prodi Informatika**



**Disusun Oleh:  
Rahmawati Danu Kusuma  
M0509058**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS SEBELAS MARET  
SURAKARTA  
2016**

*commito user*

**SKRIPSI**  
**PENGARUH *DUMMY VARIABLE* PADA METODE *NAÏVE BAYES***  
**DALAM KASUS KLASIFIKASI PENYAKIT KANDUNGAN**

**Disusun oleh:**

**Rahmawati Danu Kusuma**

**M0509058**

**ditulis dan diajukan untuk memenuhi sebagian persyaratan  
memperoleh gelar Strata Satu Program Studi Informatika**

**PROGRAM STUDI INFORMATIKA**  
**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**  
**UNIVERSITAS SEBELAS MARET**  
**SURAKARTA**  
**2016**

*commito user*

**SKRIPSI**  
**PENGARUH *DUMMY VARIABLE* PADA METODE *NAÏVE BAYES***  
**DALAM KASUS KLASIFIKASI PENYAKIT KANDUNGAN**


Disusun oleh :  
Rahmawati Danu Kusuma  
M0509058

Skripsi ini telah disetujui untuk dipertahankan di hadapan dewan penguji,  
Pada tanggal: 29 Juli 2016

Pembimbing 1,

  
**Ristu Saptono, S.Si., M.T**  
NIP. 19790210 200212 1 001

Pembimbing 2,

  
**Afrizal Doewes, S.Kom., M.Sc**  
NIP. 19850831 201212 1 004

**SKRIPSI**  
**PENGARUH DUMMY VARIABLE PADA METODE NAÏVE BAYES**  
**DALAM KASUS KLASIFIKASI PENYAKIT KANDUNGAN**

Disusun oleh :  
Rahmawati Danu Kusuma  
M0509058

Telah dipertahankan di hadapan Dewan Penguji  
Pada tanggal : 29 Juli 2016

Susunan Dewan Penguji

1. Ristu Saptono, S.Si., M.T (Ketua) (  )  
NIP. 19790210 200212 1 001
2. Afrizal Doewes, S.Kom., M.Sc (Sekretaris) (  )  
NIP. 19850831 201212 1 004
3. Sari Widya Sihwi S.Kom., M.T.I (Anggota) (  )  
NIP. 19830412 200912 2 003
4. Rini Anggrainingsih S.T.,M.T (Anggota) (  )  
NIP. 19780909 200812 2 002

Disahkan Oleh

Ketua Program Studi Informatika



Drs. Bambang Harjito MAPP.Sc, Ph.D  
NIP. 19621130 199103 1 002

iv

## MOTTO

“Allah tidak membebani seseorang melainkan sesuai kesanggupannya”

-Al Baqarah : 286-

*“Life is tough, but so are you”*

-English Proverb-



## HALAMAN PERSEMBAHAN

Kupersembahkan karya ini kepada:

**“Orang tuaku tercinta, bapak Maryoto dan ibu Kartini”**

**“Mba Arum Danu Kusumawati”**

**“Mas Arie Zakaria”**

**“Sahabatku Brigadir”**

**“Sahabatku VVACCC”**

**“Teman Informatika UNS khususnya angkatan 2009”**



## KATA PENGANTAR

Segala puji dan syukur penulis ucapkan kepada Allah SWT, yang hanya karena rahmat dan karunia-Nya, penulis dapat menyelesaikan penelitian dan pembuatan laporan penelitian ini dengan judul “Pengaruh *Dummy Variable* Pada Metode *Naïve Bayes* Dalam Kasus Klasifikasi Penyakit Kandungan”, yang merupakan salah satu syarat mendapatkan gelar strata satu Informatika Universitas Sebelas Maret Surakarta.

Skripsi ini tidak akan selesai tanpa adanya bantuan dari banyak pihak. Penulis ingin menyampaikan ucapan terima kasih kepada :

1. Bapak Ristu Saptono, S.Si., M.T., selaku Dosen Pembimbing I yang telah memberikan bimbingan, masukan dan pengarahan.
2. Bapak Afrizal Doewes, S.Kom., M.Sc, selaku Dosen Pembimbing II yang telah memberikan bimbingan dan pengarahan.
3. Bapak Drs. Bambang Harjito, M.Apps.Sc., Ph.D selaku Kepala Program Studi Informatika FMIPA UNS.
4. Bapak,ibu, mba Arum, dan mas Arie yang selalu memberikan dukungan, nasehat, motivasi dan doa kepada penulis.
5. Vincent, Vera, Choco, Ana, Chandra, Imas, Aprissya, Septiana, kak Ruth, mba Lidya yang selalu memberikan dukungan, doa dan motivasi dalam memperjuangkan skripsi.
6. Keluarga Besar Informatika UNS, khususnya angkatan 2009.

Semua pihak yang tidak bisa disebutkan satu-persatu yang telah memberikan bantuan dan dukungan terhadap penulis. Penulis berharap semoga skripsi ini bermanfaat.

Surakarta, 2016

Penulis



# PENGARUH *DUMMY VARIABLE* PADA METODE *NAÏVE BAYES* DALAM KASUS KLASIFIKASI PENYAKIT KANDUNGAN

RAHMAWATI DANU KUSUMA

Prodi Informatika. Fakultas MIPA. Universitas Sebelas Maret

## ABSTRAK

Salah satu penyakit yang sering terjadi pada kaum wanita adalah penyakit kandungan. Penyakit kandungan yang dimaksud dalam penelitian ini adalah radang panggul, mioma uteri, kanker serviks, dan kanker ovarium. Penelitian ini mempelajari pengaruh penggunaan *dummy variable* pada metode *naïve bayes* pada kasus klasifikasi penyakit kandungan. Penggunaan *dummy variable* untuk pemecahan gejala agar nilainya menjadi *binary* dan gejala menjadi lebih spesifik. *Laplacian smoothing* diterapkan ketika klasifikasi data menggunakan metode *Naïve Bayes* guna menghindari kegagalan klasifikasi akibat kemungkinan hadirnya nilai nol pada model probabilitas.

Langkah penelitian yang dilakukan adalah pengumpulan data, klasifikasi data, implementasi menggunakan PHP, pengujian dan analisa hasil. Dua skenario diterapkan di penelitian ini. Skenario satu merupakan klasifikasi data menggunakan 18 gejala. Sedangkan skenario dua merupakan klasifikasi data menggunakan 24 gejala yang diperoleh dari penjabaran 18 gejala awal yang telah mengalami proses *dummy*.

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) jika menggunakan data yang telah *didummy*. Sehingga berdasarkan hasil dari penelitian, akurasi klasifikasi data menggunakan *dummy variable* lebih tinggi daripada akurasi klasifikasi data tanpa penggunaan *dummy variable*.

**Kata kunci :** *Confusion Matrix, Dummy variable, PHP, Penyakit kandungan, Naïve Bayes, Laplacian Smoothing.*



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

**RAHMAWATI DANU KUSUMA**

Departement of Informatics.Mathematics and Natural Science Faculty.  
Sebelas Maret University

## ABSTRACT

One of the diseases that often occur in women is gynecology disease. In this research, gynecology diseases are limited to pelvic inflammatory, uterine myoma, cervical cancer, and ovarium cancer. This research study the impact of using dummy variable on naïve bayes method in gynecology disease's classification case. The use of dummy variable aimed for splitting the symptoms so that the value become binary and the symptoms become more spesific. Laplacian smoothing is applied when classification process using Naïve Bayes in order to avoid failure due to zero-probability possible presence.

The steps included in this research are data collection, data classification, implementation using PHP, testing and result analysis. Two scenarios are applied in this research. First scenario is data classification using 18 symptoms. While second scenario is data classification using 24 symptoms which were derived from elaboration of 18 initial symptoms that have been through the dummy process.

Testing use Confusion Matrix which is 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 variables. Through this research, the accuracy of data classification using dummy variables is higher than the accuracy of data classification without using dummy variables.

**Keyword** : *Confusion Matrix, Dummy variable, Gynecology diseases, PHP, Naïve Bayes, Laplacian Smoothing.*

## DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGAJUAN.....	ii
HALAMAN PERSETUJUAN.....	<b>Error! Bookmark not defined.</b> i
HALAMAN PENGESAHAN.....	<b>Error! Bookmark not defined.</b>
MOTTO.....	iv
HALAMAN PERSEMBAHAN .....	vii
KATA PENGANTAR.....	viii
ABSTRAK .....	viii
ABSTRACT.....	ixx
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR.....	xiii
DAFTAR LAMPIRAN.....	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	4
1.6 Sistematika Penulisan.....	4
BAB II TINJAUAN PUSTAKA.....	6
2.1. Dasar Teori.....	6
2.1.1 <i>Dummy Variable</i> .....	6
2.1.2 <i>Naïve Bayes</i> .....	7
2.1.3 <i>Laplacian Smoothing</i> .....	8
2.1.4 <i>Confusion Matrix</i> .....	8
2.1.5 Penyakit Kandungan.....	9
2.2 Penelitian Terkait.....	11
2.3 Kerangka Pemikiran.....	13
BAB III METODOLOGI.....	16

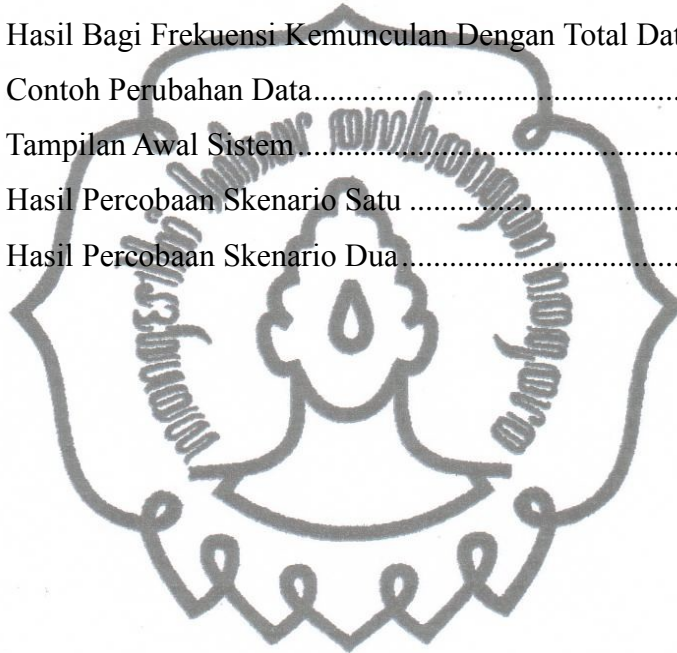
3.1. Pengumpulan Data.....	16
3.2. Klasifikasi Data .....	17
3.3. Implementasi.....	17
3.4. Pengujian dan Analisa Hasil .....	18
BAB IV HASIL DAN PEMBAHASAN.....	20
4.1 Deskripsi dan Pemodelan Data.....	20
4.1.1 Penentuan Kelas.....	20
4.1.2 Penentuan Parameter.....	20
4.1.3 <i>Cross Tabulation</i> .....	25
4.2 Algoritma Percobaan Sistem .....	30
4.2.1 Implementasi Skenario Satu.....	30
4.2.2 Implementasi Skenario Dua.....	35
4.3. Pengembangan Sistem.....	41
4.4 Pengujian .....	44
4.4.1 Pengujian Skenario Satu.....	44
4.4.2 Pengujian Skenario Dua .....	46
4.5 Pembahasan .....	48
BAB V PENUTUP.....	49
5.1 Kesimpulan.....	49
5.2 Saran.....	49
DAFTAR PUSTAKA .....	50

## DAFTAR TABEL

Tabel 2.1 Tabel <i>Confusion Matrix</i> .....	9
Tabel 2.2 Matriks Penelitian .....	14
Tabel 3.1 Daftar Klasifikasi Penyakit .....	17
Tabel 3.2 Penerapan <i>Confusion Matrix</i> .....	18
Tabel 4.1 Kategori Penyakit dalam Sistem.....	20
Tabel 4.2 Gejala Dalam Sistem.....	21
Tabel 4.3 Gejala yang di- <i>Dummy</i> .....	24
Tabel 4.4 <i>Cross Tabulation</i> Data Awal.....	26
Tabel 4.5 <i>Cross Tabulation</i> Data <i>Dummy</i> .....	28
Tabel 4.6 Contoh Data.....	30
Tabel 4.7 Contoh Data ( <i>Dummy</i> ) .....	37
Tabel 4.8 Hasil Frekuensi Kemunculan Nilai Per Parameter ( <i>Dummy</i> ).....	38
Tabel 4.9 Hasil <i>Smoothing</i> Frekuensi Kemunculan ( <i>Dummy</i> ).....	39
Tabel 4.10 Hasil Bagi Frekuensi Kemunculan Dengan Total Data Tiap Kelas ( <i>Dummy</i> ) .....	40
Tabel 4.11 Hasil Penjumlahan <i>Confusion Matrix</i> Skenario Satu.....	45
Tabel 4.12 Hasil Penjumlahan <i>Confusion Matrix</i> Skenario Dua .....	47
Tabel 4.13 Hasil <i>Precision</i> , <i>Recall</i> , dan Akurasi.....	48

**DAFTAR GAMBAR**

Gambar 3.1 Metodologi Penelitian .....	16
Gambar 4.1 Proses <i>Dummy</i> P002.....	23
Gambar 4.2 Proses <i>Dummy</i> P005.....	23
Gambar 4.3 Hasil Frekuensi Kemunculan Nilai Inputan.....	32
Gambar 4.4 Hasil <i>Smoothing</i> Frekuensi Kemunculan .....	33
Gambar 4.5 Hasil Bagi Frekuensi Kemunculan Dengan Total Data Tiap Kelas ..	34
Gambar 4.6 Contoh Perubahan Data.....	36
Gambar 4.7 Tampilan Awal Sistem.....	41
Gambar 4.8 Hasil Percobaan Skenario Satu .....	42
Gambar 4.9 Hasil Percobaan Skenario Dua.....	43



### DAFTAR LAMPIRAN

Lampiran A : Data Responden .....	52
Lampiran B : Pengelompokan Data .....	62
Lampiran C : Gejala Penyakit .....	71
Lampiran D : <i>Confusion Matrix</i> Skenario Satu.....	73
Lampiran E : <i>Confusion Matrix</i> Skenario Dua .....	75

