

**PENGARUH PENGGUNAAN *DUMMY VARIABLE*  
PADA KLASIFIKASI PENYAKIT KANDUNGAN DENGAN  
METODE C4.5**

**SKRIPSI**

**Diajukan untuk memenuhi salah satu syarat mencapai gelar Strata Satu  
Program Studi Informatika**



**DISUSUN OLEH:**

**MOCH SHOFIEYUDDIN**

**M0509047**

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

**2016**

**SKRIPSI**  
**PENGARUH PENGGUNAAN *DUMMY VARIABLE***  
**PADA KLASIFIKASI PENYAKIT KANDUNGAN DENGAN METODE**  
**C4.5**

**Disusun Oleh :**  
**Moch Shofieyuddin**  
**M0509047**

**Ditulis dan diajukan untuk memenuhi salah satu syarat mencapai gelar**  
**Strata Satu Program Studi Informatika**

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

SKRIPSI  
PENGARUH PENGGUNAAN *DUMMY VARIABLE*  
PADA KLASIFIKASI PENYAKIT KANDUNGAN DENGAN METODE  
C4.5

Disusun Oleh :  
Moch Shofieyuddin  
M0509047

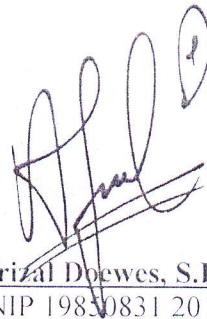
Skripsi ini telah disetujui untuk dipertahankan di hadapan dewan penguji,  
Pada tanggal : 27 Oktober 2016

Pembimbing 1,



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

Pembimbing 2,



Afrizal Doewes, S.Kom., M.Sc.  
NIP 19830831 201212 1 004



## MOTTO

*“Hard work beats talent when talent doesn’t work hard”*

**-Tim Notke-**

“Semua barang yang kasat mata itu dapat dikerjakan oleh manusia”

**-H. Hartana-**

“Semua ini hanya ilusi”

## **HALAMAN PERSEMBAHAN**

Kupersembahkan karya ini kepada:

- 1. Kedua orang tua saya (H.Hartana – Farima Mahmudah)  
(Malaikat yang dikirim tuhan untuk menjaga saya)**
- 2. Kedua Adik saya ( Mitslal Bahar & Ranu Arya Wiguna)**
- 3. Keluarga besar Bani Ilham, Bani Ichsan, dan Bani Murawi**
- 4. Teman-teman himawaloz**
- 5. Teman-teman last squad 2009**
- 6. Keluarga besar 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 Penggunaan *Dummy Variable* Pada Klasifikasi Penyakit Kandungan dengan Metode C4.5”, 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. Semua Dosen Program Studi Informatika, terima kasih atas bimbingannya selama ini
5. Keluarga yang dirumah yang selalu support untuk mengerjakan skripsi
6. Teman-teman Last Squad 2009 Pitra, Gandhi, Sambodo, Arie, Basuki, Febri, Betty, Novi, Dyah, Insan, Andy.
7. 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 PENGGUNAAN *DUMMY VARIABLE*  
PADA KLASIFIKASI PENYAKIT KANDUNGAN DENGAN METODE  
C4.5**

**MOCH SHOFIEYUDDIN**

Prodi Informatika, Fakultas MIPA, Universitas Sebelas Maret.

**ABSTRAK**

Penggunaan *dummy variable* disarankan karena pada gejala penyakit kandungan yang memiliki kemungkinan nilai yang muncul lebih dari dua (*non-binary*), ada kemungkinan tidak semua jenis kemunculannya terkait pada gejala penyakit kandungan lainnya sehingga perlu dilakukan pemecahan gejala agar nilainya menjadi *binary* dan gejala menjadi lebih spesifik. Dengan mengaplikasikan *dummy variable*, diharapkan dapat meningkatkan akurasi pada pendekatan probabilistik *Naïve Bayes Classifier*, karena asumsi independensi antar gejala penyakit terpenuhi. Selain *Naïve Bayes classifier*, *Decission Tree* juga umum digunakan dalam klasifikasi, salah satu metode *Decission Tree* adalah metode C4.5. Penelitian ini membahas tentang pengaruh penggunaan *dummy variable* dalam klasifikasi penyakit kandungan menggunakan metode C4.5. Dari hasil penelitian ini disimpulkan bahwa penggunaan *dummy variable* menghasilkan nilai rata-rata *accuracy*, *precission*, *recall*, dan *F-measure* yang stabil pada angka 87.2% pada pengujian *k-fold cross validation* dengan nilai *k* (5, 10, 15, 20, dan 25). Akan tetapi, penggunaan *dummy variable* mengurangi nilai rata – rata *accuracy*, *precission*, *recall*, dan *f-measure* secara berurutan dari 89.6%, 89.74%, 89.7%, dan 89.6% menjadi 87.2%, 87.2%, 87.2%, dan 87.2%. Selain itu, penggunaan *dummy variable* menspesifikasikan atribut gejala penyakit yang digunakan pada klasifikasi penyakit kandungan.

**Kata kunci:** *Dummy Variable*, Metode C4.5, Penyakit kandungan.



## **C4.5 THE EFFECT OF USING DUMMY VARIABLE ON CLASSIFICATION OF WOMB DISEASE WITH C4.5 METHOD**

**MOCH SHOFIEYUDDIN**

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

### **ABSTRACT**

The use of dummy variables is recommended because the symptoms of the womb disease compounds that have the possible values that appear more than two (non-binary), there is a possibility that not all types of occurrence related to the disease symptoms as other content that needs to be done solving the symptoms so that the value to binary and symptoms become more specific. By applying the dummy variable, is expected to improve the accuracy of the probabilistic approach Naïve Bayes classifier, because the assumption of independency between the symptoms of the disease are met. Besides Naïve Bayes classifier, Decission Tree is also commonly used in classification, one of Decission Tree method is C4.5. This study discusses the effect of the use of dummy variables in the womb disease classification using C4.5. From the results of this study concluded that the use of dummy variables to produce an average value accuracy, precission, recall, and F-measure which remained stable at 87.2% in testing k-fold cross validation with value of k (5, 10, 15, 20, and 25). However, the use of dummy variables reduces the average value of accuracy, precission, recall, and F-measure sequentially from 89.6%, 89.74%, 89.7%, and 89.6% to 87.2%, 87.2%, 87.2% and 87.2%. Besides, the use of dummy variables to specify the attributes of disease symptoms used in the classification of disease womb.

*Keywords: Dummy Variable, C4.5 method, Womb Disease.*

## DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGAJUAN.....	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN.....	iv
MOTTO .....	v
HALAMAN PERSEMBAHAN .....	vi
KATA PENGANTAR .....	vii
ABSTRAK .....	viii
ABSTRACT.....	ix
DAFTAR ISI.....	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR .....	xiii
DAFTAR LAMPIRAN.....	xiv
BAB I.....	1
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	2
1.3. Batasan Masalah.....	2
1.4. Tujuan Penelitian.....	2
1.5. Manfaat Penelitian.....	2
1.6. Sistematika Penulisan.....	3
BAB II.....	4
2.1. Dasar Teori .....	4
2.1.1. Decission Tree.....	4
2.1.2. C4.5.....	4
2.1.3. Dummy Variable.....	6
2.1.4. Penyakit Kandungan .....	7
2.2. Penelitian Terkait .....	9
2.3. Rencana Penelitian .....	10
BAB III .....	12
3.1. Pengumpulan Data .....	13
3.2. Pemodelan data.....	13
3.3. Implementasi model dalam aplikasi .....	13
3.4. Pengujian .....	13

3.5. Analisa Hasil Pengujian .....	15
BAB IV .....	16
4.1. Deskripsi Data .....	16
4.2. Pemodelan Data.....	19
4.3. Tabulasi Silang .....	20
4.4. Klasifikasi C4.5 .....	25
4.5. Implementasi .....	34
4.6. Pengujian .....	36
4.7. Analisa Hasil Pengujian .....	41
BAB V.....	43
5.1. Kesimpulan.....	43
5.2. Saran .....	43
DAFTAR PUSTAKA .....	44

## DAFTAR TABEL

<b>Tabel 2.1.</b> Penelitian Sebelumnya .....	10
<b>Tabel 3. 1.</b> Confusion matrix hasil klasifikasi .....	14
<b>Tabel 4. 1.</b> Daftar kelas penyakit .....	16
<b>Tabel 4. 2.</b> Daftar gejala penyakit .....	16
<b>Tabel 4. 3.</b> Hasil aplikasi Dummy Variable pada gejala penyakit .....	19
<b>Tabel 4. 4.</b> Tabulasi silang gejala penyakit dan kelas pada data asli .....	20
<b>Tabel 4. 5.</b> Tabulasi silang gejala penyakit dan kelas pada data dummy .....	23
<b>Tabel 4. 6.</b> Nilai gain ratio pada masing-masing atribut pada node 1 (root) .....	28
<b>Tabel 4. 7.</b> Nilai gain ratio pada masing-masing atribut pada node 1.1 .....	31
<b>Tabel 4. 8.</b> Atribut gejala yang dominan .....	38
<b>Tabel 4. 9.</b> Confusion matrix data asli pada 5-fold cross validation .....	39
<b>Tabel 4. 10.</b> Confusion matrix data dummy pada 5-fold cross validation .....	40
<b>Tabel 4. 11.</b> Hasil pengujian k-fold cross validation pada data asli dan data dummy .....	41

## DAFTAR GAMBAR

Gambar 2.2. Contoh penggunaan <i>dummy variable</i> .....	7
Gambar 3.1. Diagram alir metodologi penelitian.....	12
Gambar 3.2. Aplikasi <i>dummy variable</i> pada gejala penyakit.....	13
Gambar 4.1. Pohon keputusan sementara untuk node akar (root) .....	29
Gambar 4.2. Pohon keputusan sementara untuk node 1.1 .....	32
Gambar 4.3. Pohon keputusan data asli tanpa <i>pruning</i> (pemangkasan) .....	33
Gambar 4.4. Pohon keputusan data asli setelah <i>pruning</i> (pemangkasan) .....	33
Gambar 4.5. Pohon keputusan data dummy setelah <i>pruning</i> (pemangkasan) .....	34
Gambar 4.6. Tampilan awal sistem.....	35
Gambar 4.7. Tampilan klasifikasi dan hasil pengujian .....	36
Gambar 4.8. Pohon keputusan data asli .....	37
Gambar 4.9. Pohon keputusan data dummy.....	38

## DAFTAR LAMPIRAN

LAMPIRAN I .....	46
LAMPIRAN II .....	57
LAMPIRAN III.....	63