

**ANALISA PENGGUNAAN MODEL HYBRID FEATURE
SELECTION DAN JARINGAN SYARAF TIRUAN UNTUK
PENYAKIT GINJAL KRONIS**

**Tesis
untuk memenuhi sebagian persyaratan
mencapai derajat Sarjana S-2 Program Studi
Magister Sistem Informasi**



**Siti Noor Chotimah
30000319410007**

**SEKOLAH PASCASARJANA
UNIVERSITAS DIPONEGORO
SEMARANG
2021**

HALAMAN PERSETUJUAN

Tesis dengan judul :

**SISTEM DIAGNOSA AWAL PENYAKIT GINJAL KRONIS
MENGUNAKAN HYBRID FEATURE SELECTION DAN
JARINGAN SYARAF TIRUAN**

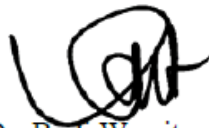
Oleh:

**Siti Noor Chotimah
30000319410007**

Telah dilakukan pembimbingan tesis dan dinyatakan layak untuk mengikuti ujian tesis pada Program Studi Magister Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro.

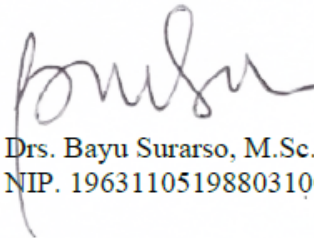
Semarang, Agustus 2021
Menyetujui,

Pembimbing I



Dr. Budi Warsito, S.Si., M.Si.
NIP: 197508241999031003

Pembimbing II



Drs. Bayu Surarso, M.Sc., Ph.D
NIP. 196311051988031001

HALAMAN PENGESAHAN

TESIS

ANALISA PENGGUNAAN MODEL HYBRID FEATURE SELECTION DAN
JARINGAN SYARAF TIRUAN UNTUK PENYAKIT GINJAL KRONIS

Oleh:
Siti Noor Chotimah
30000319410007

Telah diujikan dan dinyatakan lulus ujian tesis pada tanggal 19 Agustus 2021 oleh tim penguji
Program Studi Magister Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro.

Semarang, 25 Agustus 2021
Mengetahui,

Penguji I




Dr. Catur Edi Widodo, M.T.
NIP. 196405181992031002

Penguji II



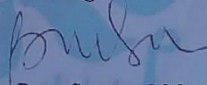
Dr. Rahmat Gemowo, M.Si.
NIP. 196511231994031003

Pembimbing I



Dr. Budi Wansito, S.Si., M.Si.
NIP. 197508241999031003

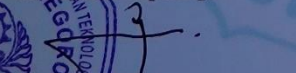
Pembimbing II



Drs. Bayu Surarso, M.Sc., Ph.D.
NIP. 196311051988031001

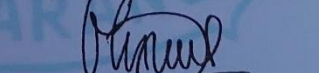
Mengetahui :

Dekan Sekolah Pascasarjana
Universitas Diponegoro



Dr. R.B. Sularto, S.H., M.Hum
NIP. 196701011991031005

Sekretaris Program Studi
Magister Sistem Informasi



Dr. Oky Dwi Nurhayati, ST., MT
NIP. 197910022009122001



**PERNYATAAN PERSETUJUAN
PUBLIKASI TESIS UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Diponegoro, saya yang bertanda tangan di bawah ini :

Nama : Siti Noor Chotimah
NIM : 30000319410007
Program Studi : Magister Sistem Informasi
Program : Sekolah Pascasarjana
Jenis Karya : Tesis

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Diponegoro Hak Bebas Royalti Noneksklusif atas karya ilmiah saya yang berjudul :

**ANALISA PENGGUNAAN MODEL HYBRID FEATURE SELECTION DAN
JARINGAN SYARAF TIRUAN UNTUK PENYAKIT GINJAL KRONIS**

beserta perangkat yang ada. Dengan Hak bebas Royalti Noneksklusif ini Program Studi Magister Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*) merawat, dan mempublikasikan tesis saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta.

Dibuat di : Semarang
Pada tanggal : 10 Agustus 2021
Yang menyatakan



Siti Noor Chotimah
NIM. 30000319410007

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam tesis ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar akademik di suatu perguruan tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Semarang, 10 Agustus 2021



Siti Noor Chotimah

KATA PENGANTAR

Dengan menyebut nama Allah SWT yang Maha Pengasih dan Maha Penyayang, penulis memanjatkan puja dan puji syukur atas limpahan rahmat, hidayah dan karunia-Nya kepada penulis sehingga dapat menyelesaikan Tesis dengan judul “Analisa Penggunaan Model Hybrid Feature Selection dan Jaringan Syaraf Tiruan untuk Penyakit Ginjal Kronis”. Tesis ini disusun sebagai pemenuhan salah satu syarat menyelesaikan studi serta memperoleh gelar Magister Komputer (M.Kom) pada Program Studi Magister Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro Semarang.

Tesis ini telah penulis usahakan dengan sebaik dan semaksimal mungkin. Selesainya penulisan Tesis ini tak lepas dari adanya bantuan dari berbagai pihak, baik materi maupun teknis. Oleh karena itu, pada kesempatan ini penulis ingin mengucapkan terima kasih kepada:

1. Bapak Dr. Budi Warsito, S.Si., M.Si selaku pembimbing I yang telah senantiasa meluangkan banyak waktu untuk memberikan bimbingan, memberikan masukan, dan motivasi kepada penulis selama penyusunan Tesis ini.
2. Bapak Drs. Bayu Surarso, M.Sc., Ph.D selaku pembimbing II yang telah senantiasa meluangkan banyak waktu untuk memberikan bimbingan dan juga memberikan masukan, dan motivasi kepada penulis selama penyusunan Tesis ini.
3. Bapak Dr. R.B. Sularto, S.H., M.Hum selaku Dekan Sekolah Pascasarjana Universitas Diponegoro Semarang.
4. Ibu Dr. Oky Dwi Nurhayati, S.T., M.T selaku Sekretaris Program Studi Magister Sistem Informasi Universitas Diponegoro Semarang.
5. Dokter Andri, suster kepala ruangan dan segenap jajaran lain dari Rumah Sakit TK. IV 04.07.03 dr.Asmir Salatiga yang telah bersedia membantu untuk memberikan data dari rekam medis sebagai data penelitian untuk penyusunan Tesis ini.

6. Dan semua pihak yang tidak bisa saya sebutkan satu persatu.

Dengan segala kerendahan hati, penulis sepenuhnya menyadari masih terdapat banyak kekurangan dari segi kualitas maupun kuantitas dan juga dari segi ilmu pengetahuan yang dimiliki penulis dalam penyusunan Tesis, sehingga penulis mengharapkan sekiranya ada kritik dan saran yang bersifat membangun demi kesempurnaan penelitian di masa mendatang.

Akhir kata, apabila sekiranya masih terdapat kata-kata yang kurang berkenan, penulis mengucapkan permohonan maaf yang sebesar-besarnya.

Semarang, 10 Agustus 2021

Penulis



Siti Noor Chotimah

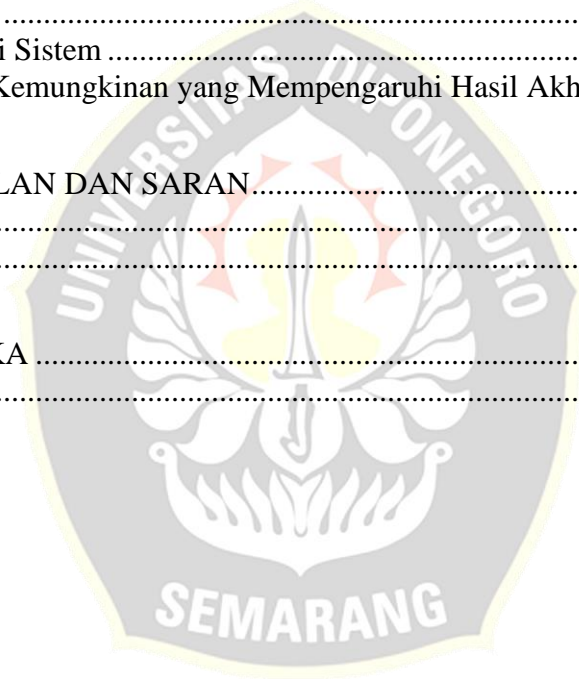


Sekolah Pascasarjana

DAFTAR ISI

	Halaman
Halaman Judul.....	i
Halaman Pengesahan	ii
Halaman Pernyataan Persetujuan Publikasi	iii
Halaman Pernyataan.....	iv
Kata Pengantar	v
Daftar Isi.....	vii
Daftar Gambar.....	ix
Daftar Tabel	x
Daftar Lampiran	xi
Daftar Arti Lambang dan Singkatan	xii
Abstrak	xiv
Abstract	xv
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Tujuan Penelitian.....	3
1.3. Manfaat Penelitian.....	4
BAB II TINJAUAN PUSTAKA DAN DASAR TEORI.....	5
2.1. Tinjauan Pustaka	5
2.2. Dasar Teori	8
2.2.1. Penyakit Ginjal Kronis	8
2.2.2. <i>Feature Selection</i>	10
2.2.3. <i>Information Gain</i>	11
2.2.4. <i>Sequential Feature Selection</i>	12
2.2.5. Jaringan Syaraf Tiruan	13
2.2.6. Algoritma <i>Backpropagation</i>	14
2.2.7. Parameter Evaluasi	18
BAB III METODE PENELITIAN.....	21
3.1. Bahan dan Alat Penelitian	21
3.1.1. Bahan Penelitian.....	21
3.1.2. Alat Penelitian	22
3.2. Prosedur Penelitian.....	23
3.2.1. Kerangka Sistem Informasi.....	24
3.2.2. Desain Penelitian.....	26
3.3. <i>Preprocessing Dataset</i>	27

3.3.1. <i>Data Cleaning</i>	28
3.3.2. <i>Data Transformation</i>	29
3.4. Desain Antarmuka Sistem	34
BAB IV HASIL PENELITIAN DAN PEMBAHASAN	38
4.1. Hasil Penelitian	38
4.1.1. Implementasi Sistem	38
4.1.2. Implementasi <i>Feature Selection Information gain</i>	47
4.1.3. Implementasi <i>Feature Selection Sequential Backward Feature Selection</i>	48
4.1.4. Implementasi Jaringan Syaraf Tiruan <i>Backpropagation</i>	49
4.2. Pembahasan	51
4.2.1. Evaluasi Sistem	51
4.2.2. Faktor Kemungkinan yang Mempengaruhi Hasil Akhir.....	54
BAB V KESIMPULAN DAN SARAN.....	55
5.1. Kesimpulan.....	55
5.2. Saran.....	56
DAFTAR PUSTAKA	57
LAMPIRAN.....	61



Sekolah Pascasarjana

DAFTAR GAMBAR

	Halaman
Gambar 2. 1 JST sederhana (Larose, 2005)	14
Gambar 3. 1 Kerangka Sistem Informasi	24
Gambar 3. 2 Blok Diagram Penelitian	26
Gambar 3. 3 Tampilan Awal Sistem	34
Gambar 3. 4 Tampilan Menu Feature Selection	34
Gambar 3. 5 Tampilan proses JST	35
Gambar 3. 6 Tampilan drop-down hidden neuron	36
Gambar 3. 7 Tampilan NNtraintool	37
Gambar 4. 1 Tampilan mulai dari sistem	38
Gambar 4. 2 Tampilan proses feature selection pada sistem	39
Gambar 4. 3 Tampilan hasil Information gain	40
Gambar 4. 4 Tampilan hasil Backward selection.....	41
Gambar 4. 5 Tampilan menu JST pada sistem.....	42
Gambar 4. 6 Menu hidden neuron	43
Gambar 4. 7 nntraintool default	44
Gambar 4. 8 Result default neuron	45
Gambar 4. 9 nntraintool determined	46
Gambar 4. 10 Result determined neuron.....	47
Gambar 4. 11 Conf matrix default	51
Gambar 4. 12 Conf matrix determined	52

Sekolah Pascasarjana

DAFTAR TABEL

	Halaman
Tabel 2. 1 Confusion matrix.....	18
Tabel 3. 1 Keterangan dataset CKD.....	22
Tabel 3. 2 Dataset dengan missing value	28
Tabel 3. 3 Dataset tanpa missing value	28
Tabel 3. 4 Konversi data ordinal	29
Tabel 3. 5 Deskripsi dataset anemia.....	29
Tabel 3. 6 Deskripsi dataset mual, muntah	30
Tabel 3. 7 Deskripsi dataset sesak nafas	30
Tabel 3. 8 Deskripsi dataset edema.....	30
Tabel 3. 9 Deskripsi dataset dehidrasi	30
Tabel 3. 10 Deskripsi dataset pucat.....	30
Tabel 3. 11 Deskripsi dataset nadi	30
Tabel 3. 12 Deskripsi dataset nafsu makan	31
Tabel 3. 13 Deskripsi dataset gender	31
Tabel 3. 14 Deskripsi dataset penyakit lain	31
Tabel 3. 15 Deskripsi dataset riwayat jatuh	31
Tabel 3. 16 Deskripsi dataset tingkat kesadaran	31
Tabel 3. 17 Perhitungan interval	32
Tabel 3. 18 Fitur Age	32
Tabel 3. 19 Fitur Bp	32
Tabel 3. 20 Fitur Wei	33
Tabel 3. 21 Dataset bernilai khusus	33
Tabel 4. 1 Peningkatan Hasil Information gain	48
Tabel 4. 2 Fitur Terbaik dari SBS	49
Tabel 4. 3 Pembagian <i>dataset</i> dalam toolbox	50
Tabel 4. 4 Perbandingan hasil JST	50
Tabel 4. 5 Perhitungan conf matrix default.....	52
Tabel 4. 6 Perhitungan conf matrix determined.....	53

DAFTAR LAMPIRAN

	Halaman
Lampiran 1 Surat Keterangan Pelaksanaan Penelitian.....	61
Lampiran 2 Contoh form rekam medis	62



Sekolah Pascasarjana

DAFTAR ARTI LAMBANG DAN SINGKATAN

DAFTAR ARTI LAMBANG

Lambang	Arti Lambang
c	Jumlah nilai pada atribut target (jumlah kelas)
p_i	Rasio antara jumlah sampel kelas i dengan jumlah semua sampel himpunan data
A	Atribut
v	Suatu nilai yang mungkin untuk atribut A
$Values(A)$	Himpunan nilai yang mungkin untuk atribut A
$ S_v $	Jumlah sampel untuk nilai v
$ S $	Jumlah seluruh sampel nilai data
$Entropy(S_v)$	<i>Entropy</i> untuk sampel-sampel yang memiliki nilai v
i	<i>Neuron ke-i pada lapisan input</i>
j	<i>Neuron ke-j pada hidden layer</i>
v_{0j}	Bias pada lapisan <i>input</i>
x_i	Bobot pada i
v_{ij}	Nilai <i>input</i> pada i ke j
$f(z_{in_j})$	Nilai aktivasi pada <i>hidden layer</i> ke <i>output</i>
exp	Bilangan natural
w_{0k}	Bias pada <i>hidden layer</i>
z_j	Hasil nilai fungsi aktivasi dari <i>hidden layer</i>
w_{jk}	Bobot pada j ke k pada <i>output</i>
$f(y_{in_k})$	Nilai aktivasi pada <i>hidden layer</i> ke <i>output</i>
δ_k	Informasi kesalahan / nilai galat
y_k	Unit <i>output</i>
t_k	Pola target
α	Laju pembelajaran
Δw_{0k}	Besar perubahan untuk ditambahkan pada bobot lama yang diperbarui
δ_j	Informasi kesalahan / nilai galat
$w_{jk}(new)$	Bobot baru antara <i>input layer</i> dan <i>hidden layer</i> yang akan dicari
$w_{jk}(old)$	Bobot lama yang diperbaharui
$v_{ij}(new)$	Bobot baru antara <i>hidden layer</i> dan <i>output layer</i> yang akan dicari
$v_{ij}(old)$	Bobot lama yang diperbaharui

DAFTAR SINGKATAN

Singkatan	Kepanjangan Singkatan
ANN	<i>Artificial Neural Network</i>
BPANN	<i>Back-Propagation Artificial Neural Network</i>
CFS	<i>Correlation based Feature Selection</i>
CKD	<i>Chronic Kidney Disease</i>
DE	<i>Differential Evolution</i>
DLBCL	<i>Diffuse Large B-cell Lymphoma</i>
DR	<i>Discriminant Relevance</i>
FE	<i>Feature Extraction</i>
FN	<i>False Negatives</i>
FP	<i>False Positives</i>
GA	<i>Genetic Algorithm</i>
GR	<i>Gain Ratio</i>
HMANN	<i>Heterogeneous Modified Artificial Neural Network</i>
IG	<i>Information Gain</i>
IoMT	<i>Internet of Medical Things</i>
JST	<i>Jaringan Syaraf Tiruan</i>
LRS	<i>Plus-L Minus-R Selection</i>
MLP	<i>Multilayer Perceptron</i>
MSE	<i>Mean Square Error</i>
NN	<i>Neural Network</i>
PNN	<i>Probabilistic Neural Network</i>
RBF	<i>Radial Basis Function</i>
RLF	<i>Relief F</i>
SBS	<i>Sequential Backward Feature Selection</i>
SFS	<i>Sequential Forward Feature Selection</i>
SVM	<i>Support Vector Machine</i>
TN	<i>True Negatives</i>
TP	<i>True Positives</i>

ANALISA PENGGUNAAN MODEL HYBRID FEATURE SELECTION DAN JARINGAN SYARAF TIRUAN UNTUK PENYAKIT GINJAL KRONIS

ABSTRAK

Jumlah pasien penderita penyakit ginjal kronis (PGK) yang terus mengalami peningkatan membuat penyakit ini menjadi salah satu penyakit yang termasuk dalam permasalahan kesehatan publik secara global. Ada banyak faktor yang bisa dikategorikan sebagai diagnosa untuk penyakit ginjal kronis. Penelitian ini bertujuan untuk memilih diagnosa penyakit ginjal kronis yang dianggap memiliki lebih banyak pengaruh diantara diagnosa lainnya. Terdapat tiga tahapan yang dilakukan dalam penelitian ini. Tahapan pertama yaitu memberi peringkat pada fitur diagnosa dari peringkat yang tertinggi hingga terendah dengan menggunakan *Information Gain* (IG). Tahapan kedua yaitu mengeliminasi fitur yang dianggap memiliki sedikit pengaruh jika dibandingkan dengan fitur lainnya dengan menggunakan *Sequential Backward Feature Selection* (SBS). Tahapan ketiga yaitu menggunakan fitur-fitur terpilih sebagai *input* pada algoritma klasifikasi Jaringan Syaraf Tiruan (JST) *Backpropagation*. Berdasarkan hasil penelitian, dengan menggunakan *hybrid feature selection* antara IG dan SBS, terpilih 15 fitur yang dianggap terbaik dari total 18 fitur dengan hasil akurasi sebesar 90% untuk penelitian dengan menggunakan *default hidden neuron* dan 86% untuk penelitian dengan menggunakan *determined hidden neuron*.

Kata kunci: diagnosa awal penyakit ginjal kronis, hybrid feature selection, information gain, sequential backward feature selection, jaringan syaraf tiruan

SEMARANG
Sekolah Pascasarjana

ANALYSIS OF USING HYBRID FEATURE SELECTION AND ARTIFICIAL NEURAL NETWORK MODEL FOR CHRONIC KIDNEY DISEASE

ABSTRACT

The number of patients with chronic kidney disease (CKD) which continues to increase makes this disease one of the diseases that are included in global public health problems. There are many factors that can be categorized as a diagnosis for chronic kidney disease. This study aims to select the diagnosis of chronic kidney disease that is considered to have more influence among other diagnoses. This study has three steps. The first step is to rank the diagnostic features from the highest to the lowest using Information Gain (IG). The second step is to eliminate features that are considered to have little influence when compared to other features by using Sequential Backward Feature Selection (SBS). The third step is using the selected features as input to the Backpropagation Artificial Neural Network (ANN) classification algorithm. Based on the results of the study, using hybrid feature selection between IG and SBS, 15 features were selected that were considered the best from a total of 18 features with an accuracy of 90% for studies using default hidden neurons and 86% for studies using determined hidden neurons.

Keywords: chronic kidney disease early diagnosis, hybrid feature selection, information gain, sequential backward feature selection, ANN



Sekolah Pascasarjana