

**ANALISA PENGGUNAAN *RELIEFF FEATURE SELECTION* DAN
BACKPROPAGATION PADA PENYAKIT *HEPATOCELLULAR*
*CARCINOMA***

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



**Umi Meganinditya Wulandari
30000319410008**

**SEKOLAH PASCASARJANA
UNIVERSITAS DIPONEGORO
SEMARANG
2023**

HALAMAN PENGESAHAN

TESIS

ANALISA PENGGUNAAN *RELIEFF FEATURE SELECTION* DAN *BACKPROPAGATION* PADA PENYAKIT *HEPATOCELLULAR CARCINOMA*

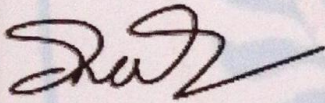
Oleh:

**Umi Meganinditya Wulandari
30000319410008**

Telah diujikan dan dinyatakan lulus ujian tesis pada tanggal 27 Juni 2023 oleh tim penguji Program Studi Magister Sistem Informasi Sekolah Pascasarjana Universitas Diponegoro.

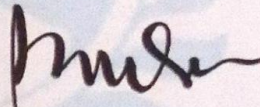
Semarang, 27 Juni 2023
Mengetahui,

Penguji I



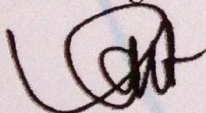
Dinar Mutiara Kusumo Nugraheni, S.T.,
M.InfoTech.(Comp)., Ph.D.
NIP 197601102009122002

Penguji II



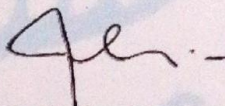
Drs. Bayu Surarso, M.Sc., Ph.D.
NIP 196311051988031001

Pembimbing I



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


Pembimbing II



Farikhin, M.Si., Ph.D
NIP 197312202000121001

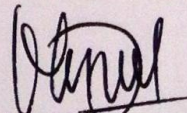
Mengetahui :

**Dekan Sekolah Pascasarjana
Universitas Diponegoro**



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

**a.n Ketua Program Studi
Sekretaris Program Studi
Magister Sistem Informasi**



Dr. Oky Dwi Nurhayati, S.T., M.T
NIP 197910022009122001

**PERNYATAAN PERSETUJUAN
PUBLIKASI TESIS UNTUK KEPENTINGAN AKADEMIS**

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

Nama : Umi Meganinditya Wulandari
NIM : 30000319410008
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 *RELIEFF FEATURE SELECTION*
DAN *BACKPROPAGATION* PADA PENYAKIT
*HEPATOCELLULAR CARCINOMA***

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 : 21 Juni 2023
Yang menyatakan

SEKOLAH PASCASARJANA



Umi Meganinditya Wulandari
30000319410008

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, 21 Juni 2023



Umi Meganinditya Wulandari



SEKOLAH PASCASARJANA

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 *Relief Feature Selection* dan *Backpropagation* pada Penyakit *Hepatocellular Carcinoma*”. 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. Selesaiannya 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 Farikhin, M.Si., 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. Bapak Drs. Bayu Surarso, M.Sc., Ph.D. selaku Ketua Program Studi Magister Sistem Informasi Universitas Diponegoro Semarang.
5. Ibu Dr. Oky Dwi Nurhayati, S.T., M.T. selaku Sekretaris Program Studi Magister Sistem Informasi Universitas Diponegoro Semarang.
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

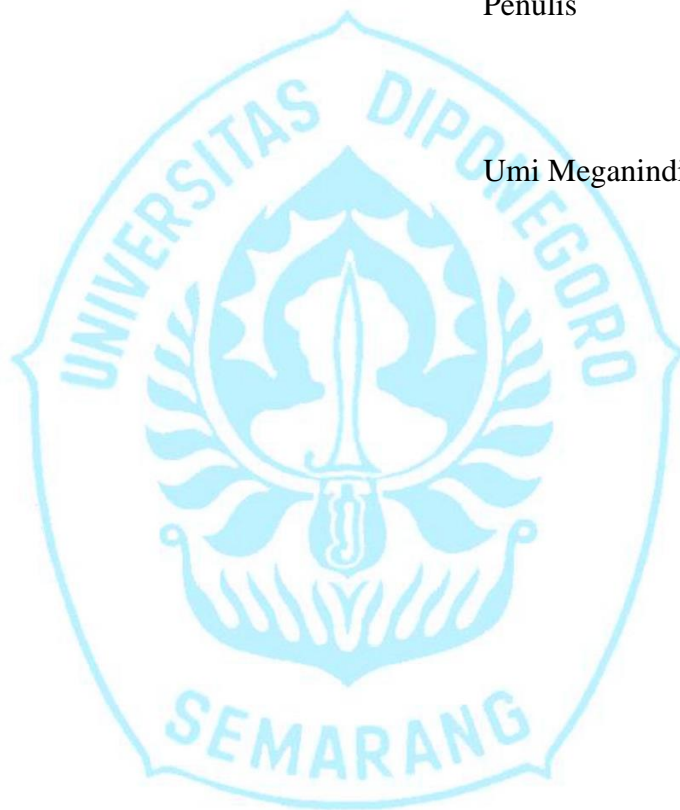
ilmu pengetahuan yang dimiliki penulis dalam penyusunan Tesis, sehingga penulis mengharapkan sekiranya ada kritik dan saran yang bersifat membangun demi kesempurnaan penelitian di masa mendatang.

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

Semarang, 21 Juni 2023

Penulis

Umi Meganinditya Wulandari



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	xv
<i>Abstract</i>	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan Penelitian	3
1.3 Manfaat Penelitian	4
BAB II LANDASAN TEORI	5
2.1 Tinjauan Pustaka	5
2.2 Dasar Teori.....	8
2.2.1 <i>Hepatocellular Carcinoma (HCC)</i>	8
2.2.2 <i>Risk Factor HCC</i>	8
2.2.3 <i>Feature Selection</i>	9
2.2.4 <i>ReliefF</i>	10
2.2.5 Jaringan Syaraf Tiruan	12
2.2.6 <i>Backpropagation</i>	12
2.2.7 <i>Confusion Matrix</i>	14
BAB III METODE PENELITIAN	16
3.1 Bahan dan Alat Penelitian.....	16
3.1.1 Bahan Penelitian.....	16

3.1.2 Alat Penelitian.....	19
3.2 Prosedur Penelitian.....	19
3.2.1 Kerangka Sistem Informasi.....	19
3.2.2 Desain Penelitian.....	21
3.3 <i>Preprocessing Dataset</i>	22
3.3.1 <i>Data Cleaning</i>	22
3.3.2 <i>Data Transformation</i>	29
BAB IV HASIL PENELITIAN DAN PEMBAHASAN.....	31
4.1 Hasil Penelitian	31
4.1.1 Skenario Pengujian.....	31
4.1.2 Implementasi Sistem.....	32
4.1.3 Implementasi <i>Feature Selection ReliefF</i>	36
4.2 Pembahasan.....	37
4.2.1 Evaluasi Model.....	37
BAB V KESIMPULAN DAN SARAN.....	42
5.1 Kesimpulan	42
5.2 Saran	42
DAFTAR PUSTAKA	43
DAFTAR LAMPIRAN.....	47

SEKOLAH PASCASARJANA

Daftar Gambar

Gambar 2.1 Lapisan <i>Backpropagation Neural Network</i>	13
Gambar 3.1 Kerangka Sistem Informasi.....	20
Gambar 3.2 Desain Penelitian.....	21
Gambar 4.1 Halaman Utama Sistem.....	32
Gambar 4.2 Halaman Preprocessing Data	33
Gambar 4.3 Hasil dari Button Process Data.....	33
Gambar 4.4 Tampilan <i>Button Result</i>	34
Gambar 4.5 Tampilan Halaman Hasil.....	35
Gambar 4.6 Tampilan Hasil <i>Button Start</i>	35
Gambar 4.7 Tampilan Hasil Pemrosesan.....	36
Gambar 4.8 Perbandingan <i>Confusion Matrix</i>	40



SEKOLAH PASCASARJANA

Daftar Tabel

Tabel 2.1 <i>Confusion Matrix</i>	15
Tabel 3.1 Keterangan <i>dataset</i> HCC	16
Tabel 3.2 Jumlah data hilang pada fitur	23
Tabel 3.3 Fitur dengan data hilang kurang dari 40 data.....	25
Tabel 3.4 <i>Dataset</i> dengan <i>missing value</i>	27
Tabel 3.5 <i>Dataset</i> dengan <i>value</i> “ <i>Nan</i> ”	27
Tabel 3.6 Fitur nilai modus	28
Tabel 3.7 Fitur nilai rata-rata	28
Tabel 3.8 <i>Dataset</i> dengan nilai modus dan rata-rata.....	29
Tabel 3.9 Perubahan data nominal	29
Tabel 3.10 Data ordinal bernilai khusus	30
Tabel 4.1 Fitur terbaik.....	36
Tabel 4.2 Hasil <i>confusion matrix</i> skenario pengujian 1	37
Tabel 4.3 Hasil terbaik dari skenario pengujian 1.....	38
Tabel 4.4 Hasil <i>confusion matrix</i> skenario pengujian 2.....	39
Tabel 4.5 Hasil terbaik dari skenario pengujian 2.....	39

SEKOLAH PASCASARJANA

Daftar Lampiran

Lampiran 1. <i>Ethical Clearence</i> Pengambilan Data Rekam Medis	47
Lampiran 2. Surat Ijin Penelitian di RSUD Dr. Kariadi Semarang	48



SEKOLAH PASCASARJANA

Daftar Arti Lambang dan Singkatan

DAFTAR ARTI LAMBANG

Lambang	Arti Lambang
x_k	<i>learning instance</i>
p_y	Probabilitas kelas
m	Parameter pengambilan sampel
$diff$	Fungsi diferensial
W [i]	Bobot Kualitas
x_l	Satu <i>instance</i>
y_k	Kelas sebuah <i>instance</i>
y	kelas <i>miss</i> terdekatnya
i	<i>neuron</i> ke- i ($i=1,2,3, \dots, n$) pada <i>input layer</i> (lapisan input)
j	<i>neuron</i> ke- j ($j=1,2,3, \dots, p$) pada <i>hidden layer</i>
v_{0j}	bias pada <i>input layer</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>
δ_k	informasi kesalahan
y_k	unit <i>output</i>
t_k	pola target
$w_{jk}(new)$	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
HCC	<i>Hepatocellular Carcinoma</i>
WHO	<i>World Health Organization</i>
HBV	<i>Hepatitis B Virus</i>
HCV	<i>Hepatitis C Virus</i>
FS	<i>Feature selection</i>
FN	<i>False Negative</i>
FP	<i>False Positive</i>
NCA	<i>Neighborhood Component Analysis</i>
JST	Jaringan Syaraf Tiruan
<i>RegLog</i>	<i>Logistic Regression</i>
MLP	<i>Multi-layer Perceptron</i>
SVM	<i>Support-Vector Machine</i>
DFS	<i>Disease-Free Survival</i>
RFA	<i>Radiofrequency ablation</i>
TN	<i>True Negative</i>
TP	<i>True Positive</i>
ESCC	<i>Esophageal squamous cell carcinoma</i>
BP	<i>Backpropagation</i>
ASSA	<i>Adaptive Salp Swarm Algorithm</i>
ANN	<i>Artificial Neural Network</i>
CART	<i>Classification And Regression Tree</i>
OA	<i>Overall Accuracy</i>

DT	<i>Decision tree</i>
LD	<i>Linear Discriminant</i>
kNN	<i>K-nearest neighborhood</i>
SD	<i>Subspace Discriminant</i>
ResExLBP	<i>Residual Exemplar Local Binary Pattern</i>
LOOCV	<i>Leave one out cross-validation</i>
CV	<i>Cross-validation</i>
FDM	<i>Fault diagnosis model</i>
CDR	<i>Correct diagnosis rates</i>
PD	<i>Parkinson's disease</i>
SWEDD	<i>Scans without evidence of dopaminergic deficit</i>
BPNN	<i>Backpropagation Neural Network</i>

SEKOLAH PASCASARJANA

Abstrak

Penelitian ini dilakukan untuk mengetahui seberapa baik tingkat performa *ReliefF* dalam pemilihan fitur pada data penyakit HCC. Penelitian ini dilatarbelakangi karena *dataset* memiliki fitur-fitur yang berpengaruh dan sedikit berpengaruh. Fitur yang memiliki banyak pengaruh merupakan faktor risiko penyakit HCC. Salah satu cara untuk menanganinya dengan memanfaatkan teknologi komputer dalam pemilihan fitur pada rekam medis dengan informasi yang paling berpengaruh pada HCC. Pada penelitian ini, model yang diusulkan adalah penggunaan *ReliefF* pada tahap seleksi fitur dan *Backpropagation* pada tahap klasifikasi. Pada tahap seleksi fitur ini terdapat dua langkah yaitu perhitungan bobot dan proses reduksi fitur. Pada langkah perhitungan bobot fitur, setiap fitur diberi bobot dan fitur yang dihasilkan akan diproses dalam proses reduksi fitur. Hasil perhitungan bobot fitur akan menghasilkan rangking dari nilai tertinggi hingga nilai terendah yang kemudian akan dikurangi dengan rangking fitur. Fitur terbaik yang telah dihasilkan akan dijadikan *input* untuk tahap kedua yaitu tahap klasifikasi menggunakan *Backpropagation*. Berdasarkan hasil penelitian, ditemukan 10 fitur yang dianggap terbaik dari 39 fitur yang digunakan. Akurasi dihasilkan oleh metode *ReliefF*+BPNN sebesar 80%. Hasil perbandingan menunjukkan bahwa metode pemilihan fitur *ReliefF* dapat memberikan peningkatan pada hasil akurasi dan hasil *recall*. Hasil penelitian ini menunjukkan bahwa metode yang diusulkan berhasil. Namun, cara ini masih memerlukan pertimbangan dari tenaga medis karena bukan merupakan hasil akhir bagi pasien.

Kata Kunci : *Feature Selection*, *ReliefF*, klasifikasi, *Backpropagation*, *Hepatocellular Carcinoma*, *Survival*

SEMARANG
SEKOLAH PASCASARJANA

Abstract

This research was conducted to find out how well ReliefF performs in feature selection and implementation on unbalanced data on HCC disease data. The background of this research is because the dataset has influential and slightly influential features. Features that have multiple influences are risk factors for HCC disease. One way to handle this is by utilizing computer technology in selecting features in medical records with the most influential information on HCC. In this study, the model proposed is the use of ReliefF at the feature selection stage and Backpropagation at the classification stage. At this feature selection stage, there are two steps, namely the calculation of weights and the process of reducing features. In the feature weight calculation step, each feature is given a weight and the resulting features will be processed in the feature reduction process. The results of the feature weight calculation will produce a ranking from the highest value to the lowest value which will then be reduced by the ranking features. The best features that have been produced will be used as input for the second stage, namely the classification stage using Backpropagation. Based on the research results, 10 features were found that were considered the best out of the 39 features used. The accuracy produced by the ReliefF+BPNN method is 80%. The comparison results show that the ReliefF feature selection method can provide an increase in the results of the accuracy of data imbalances. The results of this study indicate that the proposed method is successful. However, this method still requires consideration from medical personnel because it is not the final result for the patient.

Keywords: Feature Selection, ReliefF, Classification, Backpropagation, Hepatocellular Carcinoma, Survival



SEKOLAH PASCASARJANA