



**PENERAPAN SMOTE UNTUK MENGATASI KETIDAKSEIMBANGAN  
DATA DAN SELEKSI FITUR ALGORITMA GENETIKA PADA  
KLASIFIKASI *INDIAN LIVER PATIENT DATASET (ILPD)*  
MENGUNAKAN METODE C4.5**

**SKRIPSI**

**Untuk memenuhi persyaratan  
Dalam Menyelesaikan Strara-1 Ilmu Komputer**

**Oleh**

**DIDIN WAHYU UTAMI**

**NIM 1711016320005**

**PROGRAM STUDI S-1 ILMU KOMPUTER  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMBUNG MANGKURAT  
BANJARBARU  
JUNI 2024**



**PENERAPAN SMOTE UNTUK MENGATASI KETIDAKSEIMBANGAN  
DATA DAN SELEKSI FITUR ALGORITMA GENETIKA PADA  
KLASIFIKASI *INDIAN LIVER PATIENT DATASET (ILPD)*  
MENGUNAKAN METODE C4.5**

**SKRIPSI**

**Untuk memenuhi persyaratan  
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

**Oleh**

**DIDIN WAHYU UTAMI**

**NIM 1711016320005**

**PROGRAM STUDI S-1 ILMU KOMPUTER  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMBUNG MANGKURAT  
BANJARBARU**

**JUNI 2024**

# SKRIPSI

## PENERAPAN SMOTE UNTUK MENGATASI KETIDAKSEIMBANGAN DATA DAN SELEKSI FITUR ALGORITMA GENETIKA PADA KLASIFIKASI *INDIAN LIVER PATIENT DATASET (ILPD)* MENGUNAKAN METODE C4.5

Oleh  
**DIDIN WAHYU UTAMI**  
1711016320005

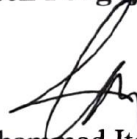
Telah dipertahankan di depan Dosen Penguji pada tanggal 27 Juni 2024,  
Susunan Dosen Penguji:

### Pembimbing I



Dwi Kartini, S.Kom., M.Kom.  
NIP. 198704212012122003

### Dosen Penguji I



Muhammad Itqan Mazdadi, S.Kom., M.Kom.  
NIP. 199006122019031013

### Pembimbing II



Fatma Indriani, S.T, M.I.T, Ph.D.  
NIP. 198404202008122004

### Dosen Penguji II



Triando Hamonangan Saragih, S.Kom., .Kom.  
NIP. 199308242019031012

Banjarbaru, 02 Juli 2024

**Koordinator Program Studi Ilmu Komputer**

  
Irwan Budiman, S.T., M.Kom.

NIP. 197703252008121001



## **PERNYATAAN**

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya jua tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Banjarbaru, 2024

Yang menyatakan,



**Didin Wahyu Utami**

**NIM. 1711016320005**

## ABSTRAK

**PENERAPAN SMOTE UNTUK MENGATASI KETIDAKSEIMBANGAN DATA DAN SELEKSI FITUR ALGORITMA GENETIKA PADA KLASIFIKASI *INDIAN LIVER PATIENT DATASET (ILPD)* MENGGUNAKAN METODE C4.5** (Oleh: Didin Wahyu Utami; Pembimbing : Dwi Kartini., S.Kom., M.Kom dan Fatma Indriani, S.T, M.I.T, Ph.D; 2024 ; 58 Halaman)

*Liver* merupakan salah satu organ penting dalam tubuh manusia yang berperan dalam proses metabolisme tubuh. Mengutip artikel dari *American Liver Foundation*, pada tahun 2020 sebanyak 51.642 orang dewasa di Amerika Serikat meninggal akibat penyakit *liver*. Data hasil tes fungsi *liver* dari laboratorium dapat digunakan untuk mendiagnosis penyakit *liver*. Klasifikasi penyakit *liver* pada pasien perlu dilakukan dengan baik karena hasilnya dapat membantu dalam diagnosis awal apakah seorang pasien mengidap penyakit *liver*. Pada penelitian ini menggunakan dataset *Indian Liver Patient Dataset (ILPD)* dari *UCI Machine Learning Repository* dengan 11 fitur dan 583 data. Metode pengklasifikasi yang digunakan yaitu Algoritma C4.5 karena metode ini merupakan salah satu metode yang dapat menganalisis data dalam jumlah yang banyak. Pada penelitian ini dilakukan metode seleksi fitur algoritma genetika dengan kombinasi parameter pc 0,1-1 dan pm 0,1-1. Selain itu, juga menggunakan metode *Synthetic Minority Oversampling Technique (SMOTE)* dengan k: 5. Berdasarkan hasil pengujian, penerapan GA dan SMOTE pada metode klasifikasi C4.5 memberikan hasil terbaik dengan nilai AUC sebesar 0.76. Hasil klasifikasi C4.5 menunjukkan nilai AUC sebesar 0,65 dan hasil seleksi fitur GA dan klasifikasi C4.5 menunjukkan nilai AUC paling besar 0,71. Berdasarkan pengujian tersebut dapat diketahui jika SMOTE pada penyeimbangan kelas dan GA pada seleksi fitur dapat meningkatkan performa klasifikasi dari metode C4.5.

**Kata kunci:** *Liver*, Klasifikasi, C4.5, SMOTE, Algoritma Genetika

## ABSTRACT

**PENERAPAN SMOTE UNTUK MENGATASI KETIDAKSEIMBANGAN DATA DAN SELEKSI FITUR ALGORITMA GENETIKA PADA KLASIFIKASI *INDIAN LIVER PATIENT DATASET (ILPD)* MENGGUNAKAN METODE C4.5** (By: Didin Wahyu Utami; Pembimbing : Dwi Kartini., S.Kom., M.Kom dan Fatma Indriani, S.T, M.I.T, Ph.D; 2024 ; 58 Pages)

The liver is an important organ in the human body which plays a role in the body's metabolic processes. Quoting an article from the American Liver Foundation, in 2020 as many as 51,642 adults in the United States died from liver disease. Liver function test results data from the laboratory can be used to diagnose liver disease. Classification of liver disease in patients needs to be done properly because the results can help in the initial diagnosis of whether a patient has liver disease. This study uses the Indian Liver Patient Dataset (ILPD) dataset from the UCI Machine Learning Repository with 11 features and 583 data. The classifier method used is the C4.5 Algorithm because this method is a method that can analyze large amounts of data. In this research, a genetic algorithm feature selection method was carried out with a combination of parameters  $pc$  0.1-1 and  $pm$  0.1-1. Apart from that, it also uses the Synthetic Minority Oversampling Technique (SMOTE) method with  $k$ : 5. Based on the test results, the application of GA and SMOTE to the C4.5 classification method gives the best results with an AUC value of 0.76. The results of the C4.5 classification show an AUC value of 0.65 and the results of GA feature selection and C4.5 classification show an AUC value of 0.71. Based on this test, it can be seen that SMOTE in class balancing and GA in feature selection can improve the classification performance of the C4.5 method.

**Keyword:** *Liver*, Klasifikasi, C4.5, SMOTE, Algoritma Genetika

## PRAKATA

Assalamualaikum warahmatullahi wabarakatuh.

Puji syukur penulis panjatkan ke hadirat Allah SWT karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul “Penerapan Smote Untuk Mengatasi Ketidakseimbangan Data Dan Seleksi Fitur Algoritma Genetika Pada Klasifikasi *Indian Liver Patient Dataset (ILPD)* Menggunakan Metode C4.5”, untuk memenuhi syarat dalam menyelesaikan pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat. Tak lupa pula penulis panjatkan shalawat dan salam ke hadirat Rasulullah Muhammad SAW beserta para sahabat, keluarga, dan pengikut beliau hingga *yaumul qiyamah..*

Pada lembar ini penulis ingin menyampaikan ucapan terimakasih kepada pihak-pihak yang sangat mendukung penulis dalam pembuatan dan penyusunan skripsi ini, adapun yang dimaksud adalah sebagai berikut :Allah SWT,

1. Keluarga terutama Abah dan Mama yang senantiasa memberikan doa dan dukungan agar dapat menyelesaikan skripsi dengan baik.
2. Ibu Dwi Kartini, S.Kom., M.Kom selaku dosen pembimbing utama dan Dosen Pembimbing akademik sekaligus koordinator panitia skripsi yang turut serta membantu dan meluangkan waktu untuk membantu menyelesaikan skripsi ini.
3. Ibu Fatma Indriani, S.T, M.I.T, Ph.D selaku dosen pembimbing pendamping yang turut serta membantu dan meluangkan waktu untuk membantu menyelesaikan skripsi ini.
4. Bapak Irwan Budiman, S.T., M.Kom selaku Koordinator Program Studi Ilmu Komputer FMIPA ULM, atas bantuan dan izin beliau skripsi ini dapat diselesaikan.
5. Seluruh Dosen dan staf Program Studi Ilmu Komputer FMIPA UNLAM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
6. Maria ulfah, S.Kom. dan Najwa khalisa ramadlana teman satu rumah yang tinggal bersama saya dan membantu saya tetap waras selama mengerjakan

skripsi ini

7. Teman yang selalu mau menjadi teman diskusi serta membantu menjawab setiap pertanyaan saya dan membantu saya mencari solusi akan permasalahan yang saya dapatkan selama mengerjakan skripsi.
8. Teman – teman Ilmu Komputer 2017 yang masuk jurusan ini bersama saya, terimakasih atas tangis dan tawanya.
9. Ada banyak pihak yang tidak dapat disebutkan satu persatu, kepada semua pihak tersebut saya juga ingin mengucapkan terimakasih yang sebanyak – banyaknya karena sudah menjadi teman ataupun musuh selama menjalani perkuliahan ini.

Akhir kata penulis menyadari sepenuhnya bahwa penulisan ini jauh dari sempurna, namun penulis mengharapkan bantuan berupa saran dan kritik yang membangun dari semua pihak demi kesempurnaan dan mutu penulisan skripsi ini.

Semoga tulisan ini dapat bermanfaat bagi ilmu pengetahuan dan pembaca khususnya serta mendapat keridhaan Allah SWT.

Banjarbaru, Juni 2024



Didin Wahyu Utami

## DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	ii
HALAMAN PERNYATAAN .....	iii
ABSTRAK .....	iv
ABSTRACT.....	v
PRAKATA.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR .....	xiii
DAFTAR LAMPIRAN.....	xiv
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	2
1.3. Batasan Masalah.....	3
1.4. Tujuan Penelitian.....	3
1.5. Manfaat penelitian.....	3
1.6. Sistematika penulisan.....	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Kajian Terdahulu.....	5
2.2 Keaslian Penelitian.....	6
2.3 Landasan Teori.....	8
2.3.1 Definisi Penyakit <i>Liver</i> .....	8

2.3.2 Data Mining.....	8
2.3.3 Klasifikasi.....	9
2.3.4 Algoritma C4.5.....	9
2.3.5 Seleksi Fitur.....	11
2.3.6 Algoritma Genetika.....	12
2.3.7 SMOTE.....	14
2.3.8 Evaluasi Kinerja Algoritma.....	15
<b>BAB III METODE PENELITIAN.....</b>	<b>17</b>
3.1. Alat Penelitian.....	17
3.2. Bahan Penelitian.....	17
3.3. Prosedur Penelitian.....	17
<b>BAB IV HASIL DAN PEMBAHASAN.....</b>	<b>20</b>
4.1. HASIL.....	20
4.1.1. Pengumpulan Dataset.....	20
4.1.2. <i>Preprocessing</i> Data.....	21
4.1.3. Pembagian Dataset.....	24
4.1.4. Klasifikasi Menggunakan Metode C4.5.....	24
4.1.5. Seleksi Fitur Algoritma Genetika - Klasifikasi C4.5.....	25
4.1.6. SMOTE - Seleksi Fitur Algoritma Genetika - Klasifikasi C4.5.....	37
4.1.7. Evaluasi.....	48
4.2. Pembahasan.....	52
<b>BAB V PENUTUP.....</b>	<b>58</b>
5.1. Kesimpulan.....	58
5.2. Saran.....	58
<b>DAFTAR PUSTAKA.....</b>	<b>59</b>

LAMPIRAN..... 63

## DAFTAR TABEL

Tabel 1 Keaslian Penelitian.....	6
Tabel 2 Rancangan Penelitian.....	7
Tabel 3 Model Optimasi Algoritma Genetika.....	12
Tabel 4 Oversampling dengan SMOTE.....	14
Tabel 5 Confussion Matrix .....	15
Tabel 6 Deskripsi Dataset .....	20
Tabel 7. Jumlah Data di Setiap Label Kelas .....	20
Tabel 8 Sample dataset.....	21
Tabel 9 Pengecekan Jumlah Missing value Pada Setiap Fitur .....	22
Tabel 10 Perhitungan min-max normalization.....	23
Tabel 11 Sampel Data Hasil Min-max normalization.....	23
Tabel 12 Pembagian Data Training dan Testing.....	24
Tabel 13 Hasil Evaluasi Klasifikasi C4.5 .....	25
Tabel 14 Pengujian Nilai Populasi.....	25
Tabel 15 Pengujian Nilai Generasi .....	26
Tabel 16 Parameter Algoritma Genetika.....	26
Tabel 17 Proses Inisialisasi Populasi dan Fitness Value .....	27
Tabel 18 Persilangan Crossover dengan Titik Potong Berbeda.....	28
Tabel 19 Individu Baru .....	29
Tabel 20 individu Baru (Offspring) .....	29
Tabel 21 Individu yang Didapatkan.....	30
Tabel 22 Rangking Setiap Individu Terbaik .....	31
Tabel 23 Hasil GA Pc 0.1 dan Pm 0.1-1 .....	32
Tabel 24 Hasil GA Pc 0.2 dan Pm 0.1-1 .....	33
Tabel 25 Hasil GA Pc 0.3 dan Pm 0.1-1 .....	33
Tabel 26 Hasil GA Pc 0.4 dan Pm 0.1-1 .....	33
Tabel 27 Hasil GA Pc 0.5 dan Pm 0.1-1 .....	34
Tabel 28 Hasil GA Pc 0.6 dan Pm 0.1-1 .....	34

Tabel 29 Hasil GA Pc 0.7 dan Pm 0.1-1 .....	35
Tabel 30 Hasil GA Pc 0.8 dan Pm 0.1-1 .....	35
Tabel 31 Hasil GA Pc 0.9 dan Pm 0.1-1 .....	36
Tabel 32 Hasil GA Pc 1 dan Pm 0.1-1 .....	36
Tabel 33 Parameter SMOTE_GA .....	38
Tabel 34 Proses inisialisasi Populasi Awal dan Vitness Value.....	39
Tabel 35 Crossover dengan Titik Poton Berbeda .....	40
Tabel 36 Individu Baru .....	40
Tabel 37 Individu Baru (Offspring) .....	41
Tabel 38 Individu Yang Didapatkan.....	42
Tabel 39 Ranging Setiap Individu Terbaik .....	43
Tabel 40 Hasil GA Pc 0.1 dan Pm 0.1-1 .....	44
Tabel 41 Hasil GA Pc 0.2 dan Pm 0.1-1 .....	44
Tabel 42 Hasil GA Pc 0.3 dan Pm 0.1-1 .....	44
Tabel 43 Hasil GA Pc 0.4 dan Pm 0.1-1 .....	45
Tabel 44 Hasil GA Pc 0.1 dan Pm 0.1-1 .....	45
Tabel 45 Hasil GA Pc 0.6 dan Pm 0.1-1 .....	46
Tabel 46 Hasil GA Pc 0.7 dan Pm 0.1-1 .....	46
Tabel 47 Hasil GA Pc 0.8 dan Pm 0.1-1 .....	47
Tabel 48 Hasil GA Pc 0.9 dan Pm 0.1-1 .....	47
Tabel 49 Hasil GA Pc 1 dan Pm 0.1-1 .....	48
Tabel 50 Confussion Matrix Klasifikasi C4.5.....	48
Tabel 51 Confussion Matrix Klasifikasi GA-C4.5 .....	48
Tabel 52 Confussion Matrix Klasifikasi SMOTE-GA-C4.5.....	50
Tabel 53 perbandingan Nilai AUC dari 3 Tahapan .....	56

## DAFTAR GAMBAR

Gambar 1 Flow Chart Optimasi Algoritma Genetika .....	12
Gambar 2 Alur Penelitian.....	17
Gambar 3 Persentase Jumlah Pasien Liver dan Pasien Non Liver.....	21
Gambar 4 Jumlah Data pada Masing Masing Kelas .....	37
Gambar 5 Presentase Setiap Jumlah Kelas Data Pasca SMOTE .....	38
Gambar 6 Grafik Perbandingan Evaluasi Kinerja Klasifikasi C4.5.....	49
Gambar 7 Hasil Pohon klasifikasi C4.5 .....	47
Gambar 8 Grafik Perbandingan Evaluasi Kinerja Klasifikasi GA-C4.5.....	48
Gambar 9 Hasil Pohon Klasifikasi GA-C4.5 .....	49
Gambar 10 Grafik Perbandingan Evaluasi Kinerja Klasifikasi SMOTE-GA-C4.5..	50
Gambar 11 Hasil Pohon SMOTE-GA-C4.5 .....	51
Gambar 12 Hasil Presentase Perbandingan Nilai AUC .....	56
Gambar 13 Hasil Perbandingan Persentase Nilai AUC Pembagian Data 90:10.....	57

## DAFTAR LAMPIRAN

Lampiran 1 Dataset ILPD .....	64
Lampiran 2 Import Library .....	92
Lampiran 3 Import Data.....	92
Lampiran 4 Preprocessing .....	94
Lampiran 5 Pembagian Data.....	94
Lampiran 6 Function Algoritma Genetika .....	95
Lampiran 7 C4.5.....	103
Lampiran 8 GA-C4.5 .....	105
Lampiran 9 SMOTE-GA-C4.5.....	109