



**KOMBINASI SELEKSI FITUR BERBASIS FILTER DAN WRAPPER  
MENGUNAKAN NAIVE BAYES PADA KLASIFIKASI PENYAKIT  
JANTUNG**

**Skripsi**

**Untuk Memenuhi Persyaratan  
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

**Oleh**

**SITI ROZIANA AZIZAH**

**NIM 1911016320006**

**PROGRAM STUDI S-1 ILMU KOMPUTER  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMBUNG MANGKURAT  
BANJARBARU  
JUNI 2023**

# SKRIPSI

## KOMBINASI SELEKSI FITUR BERBASIS *FILTER* DAN *WRAPPER* MENGUNAKAN *NAIVE BAYES* PADA KLASIFIKASI PENYAKIT JANTUNG

Oleh:

**SITI ROZIANA AZIZAH**

**NIM. 1911016320006**

Telah dipertahankan di depan Dosen Penguji pada tanggal 19 Juni 2023.

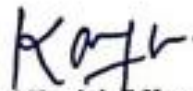
Susunan Dosen Penguji:

**Pembimbing I**



Rudy Hertanto, S.Kom., M.Kom.  
NIP. 198809252022031003

**Dosen Penguji I**



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

**Pembimbing II**



Andi Farnadi, S.Si., M.T  
NIP. 1973072520008011006

**Dosen Penguji II**



Irwan Budiman, S.T., M.Kom  
NIP. 197703252008121001



Jember, 19 Juni 2023

Direktor Program Studi Ilmu Komputer

Irwan Budiman, S.T., M.Kom  
NIP. 197703252008121001

## HALAMAN 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 juga tidak terdapat karya dan pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan diterbitkan dalam Daftar Pustaka.

Banjarbaru, 19 Juni 2023

Yang Menyatakan,



Siti Roziana Azizah  
NIM. 1911016320006

## ABSTRAK

### **KOMBINASI SELEKSI FITUR BERBASIS FILTER DAN WRAPPER MENGGUNAKAN NAIVE BAYES PADA KLASIFIKASI PENYAKIT JANTUNG**

(Oleh : Siti Roziana Azizah; Pembimbing: Rudy Herteno, S.Kom., M.Kom dan Andi Farmadi, S.Si., M.T.; 2023; 118 halaman)

Penyakit jantung menjadi salah satu penyebab utama kematian bersama dengan penyakit lainnya. Dalam bidang teknologi, data mining dapat digunakan untuk mendiagnosa suatu penyakit yang bersumber dari data rekam medis pasien. Pada klasifikasi dataset medis, *Naive Bayes* merupakan salah satu metode terbaik yang digunakan. Tujuan dari penelitian ini adalah untuk mengetahui perbandingan hasil akurasi dari *Naive Bayes* menggunakan beberapa seleksi fitur yaitu *Forward Selection*, *Backward Elimination*, kombinasi *union* hasil seleksi fitur *Forward Selection* dan *Backward Elimination*, *Information Gain*, *Gain Ratio*, dan kombinasi *union* hasil seleksi fitur *Information Gain* dengan *Gain Ratio*. Data yang digunakan dalam penelitian ini adalah data penyakit jantung yang didapatkan dari *UCI Machine Learning Repository*. Dari implementasi pemodelan yang akan dilakukan menghasilkan nilai akurasi tertinggi sebesar 91.80% pada algoritma *Naive Bayes* dengan kombinasi *union* hasil seleksi fitur *Information Gain* dan *Gain Ratio* menggunakan perbandingan data latih dan data uji 80:20. Sedangkan akurasi *Naive Bayes* dengan kombinasi *union* hasil seleksi fitur *Forward Selection* dan *Backward Elimination* hanya memiliki nilai akurasi sebesar 83.61%.

**Kata Kunci:** Penyakit Jantung, Klasifikasi, *Naive Bayes*, Seleksi Fitur, Kombinasi *Union*

## **ABSTRACT**

### **COMBINATIONS OF FEATURE SELECTION BASED ON FILTER AND WRAPPER USING NAIVE BAYES IN HEART DISEASE CLASSIFICATION**

*(By : Siti Roziana Azizah; Supervisor: Rudy Herteno, S.Kom., M.Kom dan Andi Farmadi, S.Si., M.T.; 2023; 118 page)*

*Heart disease is one of the leading causes of death along with other diseases. In the field of technology, data mining can be used to diagnose a disease sourced from patient medical record data. In the classification of medical datasets, Naive Bayes is one of the best methods used. The purpose of this study is to determine the comparison of the accuracy results of Naive Bayes using several feature selections, namely Forward Selection, Backward Elimination, a combination of union of Forward Selection and Backward Elimination feature selection results, Information Gain, Gain Ratio, and a combination of union of Information Gain feature selection results with Gain Ratio. The data used in this research is heart disease data obtained from the UCI Machine Learning Repository. From the implementation of modeling that will be carried out, the highest accuracy value is 91.80% in the Naive Bayes algorithm with a combination of union of Information Gain and Gain Ratio feature selection results using a ratio of training data and test data of 80:20. While the accuracy of Naive Bayes with a combination of union selection results of Forward Selection and Backward Elimination features only has an accuracy value of 83.61%.*

**Keywords:** *Heart Disease, Classification, Naive Bayes, Feature Selection, Union Combination*

## PRAKATA

Puji syukur penulis panjatkan ke Tuhan kita Yang Maha Esa karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul *“KOMBINASI SELEKSI FITUR BERBASIS FILTER DAN WRAPPER MENGGUNAKAN NAIVE BAYES PADA KLASIFIKASI PENYAKIT JANTUNG”* untuk memenuhi syarat dalam menyelesaikan pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat.

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 :

1. Keluarga yang selalu memberikan bantuan, semangat, doa dan dukungan dalam proses penyelesaian skripsi ini
2. Bapak Rudy Herteno S.Kom., Kom.selaku dosen pembimbing utama yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
3. Bapak Andi Farmadi, S.Si., M.T. selaku dosen pembimbing pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
4. Bapak Irwan Budiman S.T., M.Kom selaku Ketua Program Studi Ilmu Komputer FMIPA ULM, atas bantuan dan izin beliau skripsi ini dapat diselesaikan.
5. Seluruh Dosen dan staff Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
6. Teman-teman dan sahabat-sahabat keluarga Ilmu Komputer angkatan 2019 yang memberikan dukungan dan selalu mengingatkan serta mendoakan dalam proses mengerjakan skripsi.
7. Ucapan terima kasih kepada Almira Syahadati Arsyah, Raisa Amalia, Nurmolika, Astina, Dina, Ajwa, Fadhila dan teman teman yang memberikan dukungan dalam proses mengerjakan skripsi.
8. Semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam penyelesaian skripsi ini.

Akhir kata penulis menyadari sepenuhnya bahwa penulisan ini jauh dari sempurna, namun penulis mengharapkan bantuan serupa 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.

## DAFTAR ISI

HALAMAN JUDUL .....	i
HALAMAN PENGESAHAN .....	ii
HALAMAN PERNYATAAN .....	iii
ABSTRAK .....	iv
<i>ABSTRACT</i> .....	v
PRAKATA .....	vi
DAFTAR ISI .....	viii
DAFTAR TABEL .....	x
DAFTAR GAMBAR .....	xiv
DAFTAR LAMPIRAN .....	xv
<b>BAB I PENDAHULUAN</b>	
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	3
1.3 Tujuan .....	3
1.4 Manfaat Penelitian .....	3
1.5 Batasan Masalah .....	4
<b>BAB II TINJAUAN PUSTAKA</b>	
2.1 Kajian Terdahulu .....	5
2.2 Penyakit Jantung .....	8
2.3 Data Mining .....	9
2.4 <i>Missing Value</i> .....	11
2.5 Normalisasi Data .....	11
2.6 Seleksi Fitur .....	12
2.7 <i>Forward Selection</i> .....	12
2.8 <i>Backward Elimination</i> .....	13
2.9 <i>Information Gain</i> .....	14
2.10 <i>Gain Ratio</i> .....	15
2.11 Kombinasi Seleksi Fitur .....	15
2.12 Split Data .....	16
2.13 <i>Naive Bayes</i> .....	16



2.14	<i>Confusion Matrix</i> .....	18
<b>BAB III METODE PENELITIAN</b>		
3.1	Bahan Penelitian .....	20
3.2	Alat Penelitian.....	20
3.3	Variabel Penelitian.....	20
3.4	Prosedur Penelitian .....	21
<b>BAB IV HASIL DAN PEMBAHASAN</b>		
4.1	Hasil .....	25
4.1.1	Pengumpulan Dataset.....	25
4.1.2	Preprocessing Data.....	26
4.1.3	Pembagian Data Latih dan Data Uji .....	54
4.1.4	Klasifikasi .....	55
4.1.5	Evaluasi.....	100
4.2	Pembahasan.....	104
<b>BAB V PENUTUP</b>		
5.1	Kesimpulan .....	114
5.2	Saran .....	114
DAFTAR PUSTAKA		
LAMPIRAN		

## DAFTAR TABEL

Tabel 1. Keaslian Penelitian.....	6
Tabel 2. Rancangan Penelitian .....	8
Tabel 3. Confusion Matrix .....	18
Tabel 4. Dataset Penyakit Jantung .....	22
Tabel 5. Deskripsi Atribut Dataset .....	22
Tabel 6. Dataset Penyakit Jantung .....	25
Tabel 7. Distribusi Data Kelas Target .....	26
Tabel 8. Missing Value.....	27
Tabel 9. Hasil Penanganan Missing Value .....	27
Tabel 10. Hasil Normalisasi Data .....	29
Tabel 11. Atribut Dataset.....	29
Tabel 12. Hasil Nilai F iterasi pertama .....	30
Tabel 13. Atribut Hasil Iterasi Pertama.....	31
Tabel 14. Kombinasi Variabel Yang Akan Dilakukan Pengujian.....	32
Tabel 15. Hasil Nilai F iterasi kedua.....	33
Tabel 16. Atribut Hasil Seleksi Fitur Forward Selection .....	34
Tabel 17. Kombinasi Variabel Iterasi Pertama.....	35
Tabel 18. Hasil Fparsial Iterasi Pertama .....	35
Tabel 19. Atribut Hasil Iterasi Pertama.....	37
Tabel 20. Kombinasi Variabel Iterasi Kedua.....	37
Tabel 21. Hasil Nilai Fparsial iterasi kedua .....	38
Tabel 22. Atribut Hasil Seleksi Fitur Bakcward Elimination.....	39
Tabel 23. Atribut Hasil Kombinasi Seleksi Fitur Forward Selection Dan Backward Elimination.....	40
Tabel 24. Data Atribut Kontinu.....	41
Tabel 25. Data Atribut Kategorik.....	41
Tabel 26. Hasil Mengurutkan Data Atribut Kontinu.....	42
Tabel 27. Hasil Perhitungan Split Point Atribut Kontinu.....	42
Tabel 28. Hasil Deskritisasi Data Atribut Kontinu .....	43

Tabel 29. Hasil Perhitungan Nilai Entropy Data Atribut Kontinu .....	44
Tabel 30. Hasil information Gain Data Atribut Kontinu.....	45
Tabel 31. Hasil Entropy Tiap Parameter Data Kategorik.....	46
Tabel 32. Hasil Nilai Information Gain Data Kategorik .....	47
Tabel 33. Atribut Hasil Seleksi Fitur Information Gain.....	48
Tabel 34. Hasil Split Info Atribut Kontinu.....	49
Tabel 35. Hasil Gain Ratio Atribut Kontinu .....	50
Tabel 36. Hasil Split Info Atribut Kategorik.....	51
Tabel 37. Hasil Gain Ratio Atribut Kategorik.....	51
Tabel 38. Hasil Seleksi Fitur Gain Ratio.....	52
Tabel 39. Hasil Seleksi Fitur Information Gain .....	53
Tabel 40. Hasil Seleksi Fitur Gain Ratio.....	53
Tabel 41. Hasil Kombinasi Seleksi Fitur Information Gain dan Gain Ratio .....	54
Tabel 42. Data Latih Pemodelan Naive Bayes dengan Forward Selection.....	55
Tabel 43. Data Uji Pemodelan Naive Bayes dengan Forward Selection .....	56
Tabel 44. Hasil Mean dan Standar Deviasi Atribut Kontinu.....	57
Tabel 45. Hasil Probabilitas Setiap Parameter Atribut Kategorik.....	59
Tabel 46. Hasil Distribusi Gaussian Atribut Kontinu .....	60
Tabel 47. Hasil Probabilitas Akhir Setiap Kelas .....	61
Tabel 48. Hasil Probabilitas Akhir .....	62
Tabel 49. Hasil Normalisasi .....	62
Tabel 50. Hasil Prediksi Naive Bayes dengan Forward Selection.....	63
Tabel 51. Data Latih Pemodelan Naive bayes dengan Backward Elimination.....	64
Tabel 52. Data Uji Pemodelan Naive Bayes dengan Backward Elimination .....	64
Tabel 53. Hasil Mean dan Standar Deviasi Data Kontinu .....	65
Tabel 54. Hasil Probabiitas Setiap Parameter Atribut Kategorik.....	67
Tabel 55. Hasil Distribusi Gaussin Atribut Kontinu .....	68
Tabel 56. Hasil Probabilitas Akhir Setiap Kelas .....	69
Tabel 57. Hasil Probabilitas Akhir .....	70
Tabel 58. Hasil Normalisasi .....	71
Tabel 59. Hasil Prediksi Naive Bayes dengan Backward Elimination .....	71

Tabel 60. Data Latih Pemodelan Naive Bayes dengan Kombinasi Forward Selection dan Backward Elimination.....	72
Tabel 61. Data Uji Pemodelan Naive Bayes dengan Kombinasi Forward Selection dan Backward Elimination .....	72
Tabel 62. Hasil Mean dan Standar Deviasi Atribut Kontinu.....	73
Tabel 63. Hasil Probabilitas Setiap Parameter Atribut Kategorik.....	75
Tabel 64. Hasil Distribusi Gaussian Atribut Kontinu .....	77
Tabel 65. Hasil Probabilitas Akhir Setiap Kelas .....	78
Tabel 66. Hasil Probabilitas Akhir .....	78
Tabel 67. Hasil Normalisasi .....	79
Tabel 68. Hasil Prediksi Naive Bayes dengan Kombinasi Forward Selection dan Backward Elimination .....	79
Tabel 69. Data Latih Pemodelan Naive Bayes dengan Information Gain .....	80
Tabel 70. Data Uji Pemodelan Naive Bayes dengan Information Gain.....	81
Tabel 71. Hasil Probabilitas Setiap Parameter .....	83
Tabel 72. Hasil Probabilitas Akhir Setiap Kelas .....	84
Tabel 73. Hasil Probabilitas Akhir .....	84
Tabel 74. Hasil Normalisasi .....	85
Tabel 75. Hasil Prediksi Naive Bayes dengan Information Gain .....	85
Tabel 76. Data Latih Pemodelan Naive Bayes dengan Gain Ratio.....	86
Tabel 77. Data Uji Pemodelan Naive Bayes dengan Gain Ratio .....	86
Tabel 78. Hasil Mean dan Standar Deviasi Atribut Kontinu.....	88
Tabel 79. Hasil Ditsribusi Gaussian Data Kontinu .....	90
Tabel 80. Hasil Probabilitas Akhir Setiap Kelas .....	90
Tabel 81. Hasil Probabilitas Akhir .....	91
Tabel 82. Hasil Normalisasi .....	92
Tabel 83. Hasil Prediksi Naive Bayes dengan Gain Ratio .....	92
Tabel 84. Data Latih Pemodelan Naive Bayes dengan Kombinasi Information Gain dan Gain Ratio .....	93
Tabel 85. Data Uji Pemodelan Naive Bayes dengan Information Gain dan Gain Ratio .....	93

Tabel 86. Hasil Mean dan Standar Deviasi Data Kontinu .....	95
Tabel 87. Hasil Probabilitas Setiap Parameter Atribut Kategorik.....	96
Tabel 88. Hasil Distribusi Gaussian Atribut Kontinu .....	97
Tabel 89. Hasil Probabilitas Akhir Setiap Kelas .....	98
Tabel 90. Hasil Probabilitas Akhir .....	99
Tabel 91. Hasil Normalisasi .....	99
Tabel 92. Hasil Prediksi Naive Bayes dengan Kombinasi Information Gain dan Gain Ratio .....	100
Tabel 93. Confusion Matrix Naive Bayes dengan Forward Selection.....	101
Tabel 94. Confusion Matrix Naive Bayes dengan Backward Elimination .....	101
Tabel 95. Confusion Matrix Naive Bayes dengan Kombinasi Forward Selection dan Backward Elimination .....	102
Tabel 96. Confusion Matrix Naive Bayes dengan Information Gain .....	102
Tabel 97. Confusion Matrix Naive Bayes dengan Gain Ratio .....	103
Tabel 98. Confusion Matrix Naive Bayes dengan Kombinasi Information Gain dan Gain Ratio .....	104
Tabel 99. Jumlah Atribut Hasil Seleksi Fitur .....	109

## DAFTAR GAMBAR

Gambar 1. Tahapan Data Mining .....	10
Gambar 2. Kombinasi Seleksi Fitur pendekatan union dan intersection .....	16
Gambar 3. Alur Penelitian.....	21
Gambar 4. Bar Chart Distribusi Kelas Target .....	26
Gambar 5. Kurva Hasil Uji Hipotesis Atribut Thal .....	31
Gambar 6. Kurva Hasil Uji Hipotesis Atribut ca .....	34
Gambar 7. Kurva Hasil Uji Hipotesis Atribut Age .....	36
Gambar 8. Kurva Hasil Uji Hipotesis Atribut Chol.....	39
Gambar 9. Bar Chart Split Data 80:20.....	54
Gambar 10. Chart Hasil Akurasi Seleksi Fitur.....	110

## DAFTAR LAMPIRAN

Lampiran 1 Source Code Import Dataset.....	115
Lampiran 2 Source Code Mengisi Missing Value dengan Nilai Modus.....	115
Lampiran 3 Source Code Normalisasi Atribut Kontinu dengan Min-Max Scaler...	115
Lampiran 4 Source Code Seleksi Fitur Forward Selection.....	115
Lampiran 5 Source Code Naive Bayes dengan Forward Selection .....	116
Lampiran 6 Source Code Seleksi Fitur Backward Elimination .....	116
Lampiran 7 Source Code Naive Bayes dengan Backward Elimination .....	117
Lampiran 8 Naive Bayes dengan Kombinasi Forward Selection dan Backward Elimination.....	117
Lampiran 9 Source Code Seleksi Fitur Information Gain .....	118
Lampiran 10 Source Code Naive Bayes dengan Information Gain.....	122
Lampiran 11 Source Code Naive Bayes dengan Gain Ratio .....	123
Lampiran 12 Naive Bayes dengan Kombinasi Forward Selection dan Backward Elimination.....	123