

**PERBANDINGAN ANALISIS DISKRIMINAN LINIER DAN REGRESI
LOGISTIK MULTINOMIAL DALAM KLASIFIKASI PENYAKIT
JANTUNG PADA DATA MULTIKELAS**

SKRIPSI

Untuk memenuhi sebagai persyaratan

Mencapai derajat Sarjana S-1

Program Studi Matematika



Diajukan Oleh

Nurul Hanin Azzahra Supriyono

22106010031

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN SKRIPSI



Universitas Islam Negeri Sunan Kalijaga



FM-UINSK-BM-05-03/RO

SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR

Hal : Persetujuan Skripsi / Tugas Akhir

Lamp : -

Kepada

Yth. Dekan Fakultas Sains dan Teknologi

UIN Sunan Kalijaga Yogyakarta

di Yogyakarta

Bismillah 'alalamin ur. ur.

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka kami selaku pembimbing berpendapat bahwa skripsi Saudara:

Nama : Nurul Hanin Azzahra Supriyoni

NIM : 22106010031

Judul Skripsi : Perbandingan Analisis Diskriminan Linier dan Regresi Logistik Multinomial dalam Klasifikasi Penyakit Jantung Pada Data Multikelas

sudah dapat diajukan kembali kepada Program Studi Matematika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu dalam Program Studi Matematika.

Dengan ini kami menghampagkan skripsi/tugas akhir Saudara tersebut di atas dapat segera dimunaqsyahkan. Atas perhatiannya kami ucapkan terima kasih.

Bismillah 'alalamin ur. ur.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

Yogyakarta, 15 April 2026

Pembimbing,

Dr. Epha Diah Sugandi, S.Si., M.Sc.

NIP. 197509122008012015

HALAMAN PENGESAHAN



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Marsda Adisucipto Telp. (0274) 540971 Fax. (0274) 519739 Yogyakarta 55281

PENGESAHAN TUGAS AKHIR

Nomor : B-912/Un.02/DST/PP.00.9/05/2026

Tugas Akhir dengan judul : Perbandingan Analisis Diskriminan linier dan Regresi Logistik Multinomial dalam Klasifikasi Penyakit Jantung Pada Data Multikelas

yang dipersiapkan dan disusun oleh:

Nama : NURUL HANIN AZZAHRA SUPRIYONO
Nomor Induk Mahasiswa : 22106010031
Telah diujikan pada : Jumat, 24 April 2026
Nilai ujian Tugas Akhir : A

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR



Ketua Sidang
Dr. Epha Diana Supandi, S.Si., M.Sc.
SIGNED

Valid ID: 6a2646cbe8f81



Penguji I
Muhamad Rashid Hilmi, S.Si., M.Sc.
SIGNED

Valid ID: 6a2646cbe8f81



Penguji II
Prof. Dr. Muhammad Wakid Musthofa,
S.Si., M.Si.
SIGNED

Valid ID: 6a2646cbe8f81



Yogyakarta, 24 April 2026
UIN Sunan Kalijaga
Dekan Fakultas Sains dan Teknologi
Prof. Dr. Dra. Hj. Khum Wardani, M.Si.
SIGNED

Valid ID: 6a2646cbe8f81

SUNAN KALIJAGA UNIVERSITY
YOGYAKARTA

HALAMAN PERNYATAAN KEASLIAN

SURAT PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini:

Nama : Nurul Hanin Azzahra Supriyono
NIM : 22106010031
Program Studi : Matematika
Fakultas : Sains dan Teknologi

Dengan ini menyatakan bahwa isi skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar sarjana di suatu Perguruan Tinggi dan sesungguhnya skripsi ini merupakan hasil pekerjaan penulis sendiri sepanjang pengetahuan penulis, bukan duplikasi atau saduran dari karya orang lain kecuali bagian tertentu yang penulis ambil sebagai bahan acuan. Apabila terbukti pernyataan ini tidak benar, sepenuhnya menjadi tanggung jawab penulis.

Yogyakarta, 15 April 2026



Nurul Hanin Azzahra Supriyono

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

MOTTO

Maka sesungguhnya bersama kesulitan ada kemudahan, sesungguhnya bersama kesulitan ada kemudahan. Maka apabila engkau telah selesai (dari sesuatu urusan), tetaplah bekerja keras (untuk urusan yang lain), dan hanya kepada Tuhanmulah engkau berharap. (Qs. Al-Insyirah [94]: 5-8)



HALAMAN PERSEMBAHAN

Dengan mengucapkan syukur, *Alhamdulillahilladzi bini'matihi tatimmush sholihat.*

Skripsi ini saya persembahkan kepada:

Orang-orang yang paling berharga dalam hidup saya yaitu kedua orang tua tercinta, Bapak Ujang Supriyana dan Ibu Susanti, yang senantiasa memberikan doa, kasih sayang, serta dukungan terbaik dalam setiap langkah saya.

Kepada kedua kakak saya, Miftahul Jannah dan Aulia Chairunnisa, serta kedua adik saya, Nabilah Nurainiyyah dan Ibnu Dzakir, terimakasih atas doa, semangat dan kebersamaan yang selalu menguatkan.

Serta kepada semua pihak yang telah memberikan dukungan, bantuan, dan doa yang tidak dapat saya sebutkan satu per satu.

Dan kepada kampus tercinta, Universitas Islam Negeri Sunan Kalijaga Yogyakarta, yang telah menjadi tempat saya menimba ilmu dan bertumbuh.



KATA PENGANTAR

Bismillahirrahmanirrahim

Puji syukur penulis panjatkan kehadirat Allah Subhanahu wa Ta'ala atas segala rahmat dan karunia-Nya yang telah diberikan, berupa keimanan, kekuatan, kesabaran, kelancaran, serta keselamatan, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Perbandingan Analisis Diskriminan Linier dan Regresi Logistik Multinomial dalam Klasifikasi Penyakit Jantung Pada Data Multikelas” dengan baik. Sholawat serta salam senantiasa tercurah kepada junjungan Nabi Muhammad Shallallahu ‘Alaihi Wasallam yang diharapkan syafa’atnya di yaumul qiyamah kelak.

Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Matematika pada Fakultas Sains dan Teknologi Universitas Islam Negeri Sunan Kalijaga Yogyakarta. Penulis menyadari bahwa dalam proses penyusunan skripsi ini terdapat berbagai tantangan dan hambatan, namun berkat bimbingan, dukungan, serta bantuan dari berbagai pihak, skripsi ini akhirnya dapat terselesaikan.

Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih kepada:

1. Bapak Prof. Noorhaidi, S.Ag., M.A., M.Phil., Ph.D., selaku Rektor Universitas Islam Negeri Sunan Kalijaga Yogyakarta.
2. Ibu Prof. Dr. Dra. Hj. Khurul Wardati, M.Si., selaku Dekan Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
3. Ibu Dr. Epha Diana Supandi, S.Si., M.Sc., selaku Ketua Program Studi Matematika Fakultas dan Teknologi UIN Sunan Kalijaga Yogyakarta.
4. Ibu Dr. Epha Diana Supandi, S.Si., M.Sc., selaku dosen pembimbing akademik penulis yang telah memberikan arahan dan bimbingan selama masa perkuliahan.
5. Seluruh dosen dan staf Fakultas Sains dan Teknologi yang telah memberikan ilmu, pengalaman serta pelayanan terbaik selama menempuh pendidikan hingga terselesaikannya skripsi ini.

6. Bapak Ujang Supriyana dan Ibu Susanti, selaku orang tua tersayang yang telah membesarkan penulis dengan penuh kasih sayang, keikhlasan, serta pengorbanan yang tiada henti, dan senantiasa memberikan doa, dukungan serta motivasi dalam setiap langkah penulis.
7. Miftahul Jannah, Aulia Chairunnisa dan Hermansyah selaku kakak penulis serta, Nabilah Nurainiyyah dan Ibnu Dzakir selaku adik tersayang yang selalu memberikan semangat, dukungan, dan doa kepada penulis. Serta Assyifa Sabiqah, keponakan penulis yang selalu menjadi sumber kebahagiaan dan penyemangat bagi penulis.
8. Alda Sonia, Alvira Lutfiah, Alya Hafizhah serta keluarga besar penulis yang selalu memberikan doa, dukungan, serta kebahagiaan dalam setiap perjalanan penulis.
9. Radwa Adila Tsabita sebagai teman yang selalu kebersamai, mendengarkan, memberikan saran, serta menjadi tempat berbagi cerita bagi penulis.
10. Ahmad Mushthofa Kamal yang kehadirannya tidak pernah diduga oleh penulis, namun selalu memberikan bantuan dalam berbagai hal, membuka wawasan berpikir penulis, serta menghadirkan dukungan dan energi positif kepada penulis.
11. Riska Fajry Rahmadhanie dan Putri Fadhillah sebagai teman dekat selama perkuliahan yang menemani penulis dalam suka dan duka, berbagi cerita, serta memberikan dukungan moral yang tulus selama proses perkuliahan hingga penyusunan skripsi ini.
12. Alifiya Aziza sebagai teman yang senantiasa memberikan dukungan, bantuan, serta afirmasi positif selama masa perkuliahan, sehingga penulis merasa memiliki teman seperjuangan dan terus termotivasi untuk berkembang serta melangkah maju bersama.
13. Carissa Fisa Azzahra sebagai teman yang selalu memberikan bantuan dan kontribusi positif, teman berdiskusi dan bertukar pikiran dalam proses penyusunan tugas akhir penulis.

14. Alya Azzahra sebagai teman yang kehadirannya selalu membawa energi positif, keceriaan, serta menjadi *mood booster* bagi penulis.
15. Intan Khairunnisa, Cattleya Zahra dan Niken Anggraini sebagai teman penulis sejak sekolah dasar yang senantiasa memberikan semangat dan masukan kepada penulis.
16. Teman-teman KKN Karangnongko, yang telah memberikan pengalaman berharga serta kenangan yang tidak terlupakan bagi penulis.
17. Teman-teman angkatan 2022 yang telah membersamai perjalanan perkuliahan selama empat tahun ini.
18. Nurul Hanin Azzahra Supriyono, yaitu diri penulis sendiri, terima kasih atas segala perjuangan, ketekunan, dan kesabaran hingga mampu bertahan dan sampai pada titik ini. Terima kasih telah berani melangkah, menghadapi berbagai tantangan, serta tetap berusaha menyelesaikan tanggung jawab dengan sebaik mungkin. Semoga segala usaha yang telah dilakukan menjadi langkah awal menuju hal-hal baik di masa depan.
19. Semua pihak yang tidak dapat penulis sebutkan satu per satu, terima kasih atas segala bantuan, doa, dan dukungan yang telah diberikan kepada penulis dalam proses penyusunan skripsi ini.

Saya berharap semoga Allah Subhanahu wa Ta'ala mengaruniakan rahmat dan hidayah-Nya kepada mereka semua. Semoga skripsi ini dapat memberikan manfaat dan dapat menambah wawasan keilmuannya, *Aamiin Allahumma Aamiin*.

Yogyakarta, 09 April 2026

Penulis,

Nurul Hanin Azzahra Supriyono

DAFTAR ISI

HALAMAN PERSETUJUAN SKRIPSI	i
HALAMAN PENGESAHAN	ii
HALAMAN PERNYATAAN KEASLIAN	iii
MOTTO	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI	ix
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN	xv
DAFTAR SIMBOL	xvi
INTISARI	xvii
ABSTRAC	xviii
BAB I	1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Batasan Masalah	6
1.3 Rumusan Masalah	6
1.4 Tujuan Penelitian	7
1.5 Manfaat Penelitian	7
1.6 Tinjauan Pustaka	7
1.7 Sistematika Penulisan	12
BAB II	14
LANDASAN TEORI	14
2.1 Statistika Inferensia	14
2.2 Matriks	15
2.2.1 Jenis-Jenis Matriks	15
2.2.2 Operasi Matriks	19
2.3 Ekspektasi, Variansi dan Kovariansi	22
2.3.1 Ekspektasi	22
2.3.2 Variansi	23
2.3.3 Kovariansi	24

2. 4	Nilai Eigen dan Vektor Eigen.....	25
2. 5	Teori Probabilitas	26
2.5.1	Probabilitas	28
2.5.2	Probabilitas bersyarat.....	29
2. 6	Variabel Acak.....	30
2.6.1	Variabel Acak Diskrit.....	30
2.6.2	Variabel Acak Kontinu	30
2. 7	Distribusi Probabilitas.....	31
2.7.1	Distribusi Probabilitas Diskrit.....	31
2.7.2	Distribusi Probabilitas Kontinu.....	32
2. 8	Distribusi Binomial.....	32
2. 9	Distribusi Normal	34
2. 10	Distribusi Normal Multivariat.....	35
2. 11	Estimasi Parameter.....	36
2. 12	Analisis Multivariat	36
2. 13	Struktur Data Multivariat.....	37
2. 14	Variabel Kategorik	38
2. 15	Vektor Rata-rata.....	39
2. 16	Matriks Kovarians.....	40
2. 17	Analisis Diskriminan	41
2.17.1	Estimasi Diskriminan Linier	42
2. 18	Analisis Regresi.....	44
2. 19	Regresi Logistik.....	45
2.19.1	Regresi Logistik Multinomial.....	47
2. 20	Estimasi Regresi Logistik	47
2. 21	Penyakit Jantung	50
2.22.1	Usia.....	51
2.22.2	Kolesterol.....	51
2.22.3	Detak Jantung dan Segmen ST	51
BAB III	53
METODOLOGI PENELITIAN	53
3. 1	Jenis Penelitian.....	53
3. 2	Variabel Penelitian	53

3.3	Metode Pengumpulan Data.....	54
3.4	Alur Analisis Data.....	55
BAB IV.....		58
PEMBAHASAN.....		58
4.1.	Model Diskriminan Linier.....	58
4.1.1	Uji Asumsi Diskriminan Linier.....	59
4.1.2	Estimasi dan Pembentukan Fungsi Diskriminan.....	64
4.1.3	Uji Signifikansi Fungsi Diskriminan (Wilk's Lambda).....	66
4.1.4	Prosedur Pengklasifikasian.....	67
4.2.	Model Regresi Logistik Multinomial.....	68
4.2.1.	Uji Multikolinieritas (VIF).....	69
4.2.2.	Estimasi Regresi Logistik Multinomial.....	71
4.2.3.	Uji Signifikansi Serentak (<i>Likelihood Ratio Test</i>).....	73
4.2.4.	Uji Signifikansi Parsial (Uji Wald).....	74
4.3.	Perbandingan Ketepatan Klasifikasi dan Nilai Akurasi.....	75
BAB V.....		78
STUDI KASUS.....		78
5.1.	Hasil Uji Asumsi Metode.....	80
5.2.1	Uji Normalitas Multivariat.....	80
5.2.2	Uji Homogenitas Matriks Varians-Kovarians.....	81
5.2.3	Uji Kesamaan Vektor Rata-Rata.....	83
5.2.4	Deteksi Outlier.....	83
5.2.5	Mendeteksi Multikolinieritas.....	84
5.2.	Uji Asumsi Setelah Penghapusan Outlier.....	85
5.2.1	Uji Normalitas Multivariat.....	85
5.2.2	Uji Homogenitas Matriks Varians-Kovarians.....	86
5.2.3	Uji Kesamaan Vektor Rata-Rata.....	88
5.2.4	Deteksi Outlier.....	89
5.2.5	Mendeteksi Multikolinieritas.....	89
5.3.	Hasil Penelitian Analisis Diskriminan Linier.....	90
5.3.1	Membentuk Fungsi Diskriminan.....	90
5.3.2	Pengujian Signifikansi Fungsi Diskriminan.....	93
5.3.3	Pengklasifikasian.....	94

5.3.4	Validitas Diskriminan	97
5.4.	Hasil Penelitian Regresi Logistik Multinomial	98
5.4.1	Estimasi Parameter Regresi Logistik Multinomial	98
5.4.2	Pengujian Signifikansi Model Secara Serentak.....	100
5.4.3	Pengujian Signifikansi Secara Parsial.....	101
5.4.4	Eliminasi Variabel Usia (X1).....	103
5.4.5	Eliminasi Variabel Kolesterol (X2).....	106
5.4.6	Model Akhir Regresi Logistik Multinomial.....	108
5.4.7	Ketepatan Klasifikasi	109
5.4.8	Akurasi Ketepatan Klasifikasi.....	111
5.5.	Evaluasi Ketepatan Klasifikasi dan Nilai Akurasi	112
5.6.	Penerapan Prediksi Tingkat Keparahan Penyakit Jantung	115
BAB VI.....		117
PENUTUP		117
6.1.	Kesimpulan.....	117
6.2.	Saran.....	119
DAFTAR PUSTAKA.....		121
LAMPIRAN		124

DAFTAR TABEL

Tabel 1. 1 Perbedaan Penelitian Sebelumnya Dengan Penulis.....	9
Tabel 3. 1 Deskripsi Variabel Dependen.....	54
Tabel 3. 2 Deskripsi Variabel Penelitian.....	54
Tabel 5. 1 Hasil Statistik Deskriptif.....	79
Tabel 5. 2 Hasil Uji Normalitas Multivariat	80
Tabel 5. 3 Hasil Uji Homogenitas Matriks Varians-Kovarians	82
Tabel 5. 4 Hasil Uji Kesamaan Vektor Rata-Rata.....	83
Tabel 5. 5 Deteksi Outlier Berdasarkan Jarak Mahalanobis.....	83
Tabel 5. 6 Hasil Uji Multikolinearitas	84
Tabel 5. 7 Hasil Uji Normalitas (Tanpa Outlier)	85
Tabel 5. 8 Hasil Uji Homogenitas (Tanpa Outlier).....	87
Tabel 5. 9 Hasil Uji Kesamaan Vektor Rata-Rata (Tanpa Outlier).....	88
Tabel 5. 10 Hasil Uji Multikolinearitas (Tanpa Outlier).....	89
Tabel 5. 11 Koefisien Fungsi Diskriminan	91
Tabel 5. 12 Nilai Fungsi Diskriminan.....	92
Tabel 5. 13 Signifikansi Fungsi Diskriminan	93
Tabel 5. 14 Grup Centroid	94
Tabel 5. 15 Hasil Klasifikasi.....	95
Tabel 5. 16 Validitas Diskriminan	97
Tabel 5. 17 Estimasi Parameter Regresi Logistik Multinomial	98
Tabel 5. 18 Hasil Uji Likelihood Rasio	101
Tabel 5. 19 Hasil Uji Wald.....	102
Tabel 5. 20 Hasil Uji Likelihood Rasio (Setelah Eliminasi X_1).....	104
Tabel 5. 21 Hasil Uji Wald (Setelah Eliminasi X_1)	105
Tabel 5. 22 Hasil Uji Likelihood Rasio (Setelah Eliminasi X_2).....	106
Tabel 5. 23 Hasil Uji Wald (Setelah Eliminasi X_2)	107
Tabel 5. 24 Hasil Ketepatan Klasifikasi Regresi Logistik.....	109
Tabel 5. 25 Akurasi Ketepatan Klasifikasi	111
Tabel 5. 26 Perbandingan Akurasi Ketepatan Klasifikasi	112
Tabel 5. 27 Evaluasi Model Analisis Diskriminan	112
Tabel 5. 28 Evaluasi Model Regresi Logistik.....	114

DAFTAR GAMBAR

Gambar 3. 1 Skema Langkah Analisis Data. 55



DAFTAR LAMPIRAN

Lampiran 1 Penerapan model diskriminan linier dan regresi logistik.....	124
Lampiran 2 Data Tingkat Keparahan Penyakit Jantung.....	126
Lampiran 3 Output Analisis Diskriminan Linier dan Regresi Logistik Multinomial dengan Software R.....	127



DAFTAR SIMBOL

X	: Variabel independen
Y	: Variabel dependen
K	: Jumlah kategori/kelas
n	: Jumlah sampel
p	: Jumlah variabel independen
μ_k	: Rata-rata kelompok ke- k
Σ	: Matriks kovarians
a_i	: Bobot diskriminan variabel independen ke- i
a_0	: Intersep (diskriminan)
\hat{a}	: Estimasi parameter diskriminan
Z	: Skor diskriminan
Λ	: Wilk's Lambda
$\pi(x_i)$: Peluang
β	: Koefisien regresi
β_0	: Intersep (regresi)
$\hat{\beta}$: Estimasi parameter regresi
L	: Fungsi likelihood
W	: Statistik uji Wald
$g(x_i)$: Nilai logit untuk observasi ke- i

INTISARI

PERBANDINGAN ANALISIS DISKRIMINAN LINIER DAN REGRESI LOGISTIK MULTINOMIAL DALAM KLASIFIKASI PENYAKIT JANTUNG PADA DATA MULTIKELAS

Oleh

NURUL HANIN AZZAHRA SUPRIYONO

NIM. 22016010031

Dalam statistika inferensial, analisis diskriminan linier dan regresi logistik multinomial merupakan metode yang digunakan untuk klasifikasi data kategorik multikelas. Analisis diskriminan linier bertujuan membentuk kombinasi linier variabel independen yang mampu memaksimalkan perbedaan antar kelas dan meminimalkan variasi dalam kelas. Sementara itu, regresi logistik multinomial digunakan untuk memodelkan hubungan antara variabel dependen kategorik dengan lebih dari dua kelas dan satu atau lebih variabel independen yang bersifat numerik maupun kategorik. Penelitian ini menggunakan data sekunder mengenai tingkat keparahan penyakit jantung yang diperoleh dari Kaggle sebanyak 202 responden. Variabel dependen yang digunakan adalah tingkat keparahan penyakit jantung yang terdiri atas lima kategori yaitu 0 (normal), 1 (rendah), 2 (sedang), 3 (berat), dan 4 (sangat berat). Variabel independen dalam penelitian ini meliputi usia (X_1), kadar kolesterol (X_2), detak jantung maksimum (X_3), dan depresi segmen ST (X_4). Hasil penelitian menunjukkan bahwa akurasi klasifikasi menggunakan analisis diskriminan linier sebesar 52,54%, sedangkan regresi logistik multinomial sebesar 55,93%. Hal ini menunjukkan bahwa regresi logistik multinomial memiliki kemampuan klasifikasi yang lebih baik dibandingkan analisis diskriminan linier dalam mengelompokkan tingkat keparahan penyakit jantung, meskipun tingkat akurasi yang diperoleh masih tergolong sedang.

Kata Kunci: Diskriminan Linier, Penyakit Jantung, Klasifikasi, Multikelas, Regresi Logistik Multinomial.

ABSTRAC

COMPARISON OF LINEAR DISCRIMINANT ANALYSIS AND MULTINOMIAL LOGISTIC REGRESSION IN MULTICLASS HEART DISEASE CLASSIFICATION

by

NURUL HANIN AZZAHRA SUPRIYONO

NIM. 22016010031

In inferential statistics, linear discriminant analysis and multinomial logistic regression are methods used for multiclass categorical data classification. Linear discriminant analysis aims to construct linear combinations of independent variables that maximize separation between classes and minimize within-class variation. Meanwhile, multinomial logistic regression is used to model the relationship between a categorical dependent variable with more than two classes and one or more independent variables, whether numerical or categorical. This study uses secondary data on the severity level of heart disease obtained from Kaggle, consisting of 202 respondents. The dependent variable is the level of heart disease severity, categorized into five groups: 0 (normal), 1 (mild), 2 (moderate), 3 (severe), and 4 (very severe). The independent variables included in this study are age (X_1), serum cholesterol (X_2), maximum heart rate (X_3), and ST segment depression (X_4). The results show that the classification accuracy using Linear Discriminant Analysis is 52.54%, while Multinomial Logistic Regression achieves an accuracy of 55.93%. This indicates that Multinomial Logistic Regression has better classification performance compared to Linear Discriminant Analysis in classifying the severity level of heart disease, although the accuracy level obtained is still considered moderate.

Keywords: Linear Discriminant Analysis, Heart Disease, Classification, Multiclass, Multinomial Logistic Regression.

BAB I

PENDAHULUAN

Pada bab ini disajikan uraian awal yang menjadi landasan penelitian. Pembahasan dalam bab pendahuluan meliputi latar belakang masalah, batasan serta perumusan masalah, tujuan dan manfaat penelitian, tinjauan pustaka, serta sistematika penulisan yang digunakan dalam penelitian ini.

1.1 Latar Belakang

Statistika memiliki keterkaitan yang sangat erat dengan matematika, karena dasarnya merupakan ilmu yang mempelajari cara mengumpulkan, mengolah, menganalisis, serta menyimpulkan data sehingga dapat memberikan informasi yang relevan dan dapat dipertanggungjawabkan. Berdasarkan penggunaannya, statistika terbagi menjadi dua jenis, yaitu statistika deskriptif dan statistika inferensial. Statistika deskriptif berfokus pada menyajikan dan mengumpulkan data, baik dalam bentuk tabel, grafik, maupun ukuran-ukuran tertentu, tanpa melakukan pengujian lebih lanjut terhadap data tersebut. Sedangkan statistika inferensial melibatkan tahapan yang lebih kompleks, seperti pengujian hipotesis, penarikan kesimpulan, dan pengambilan keputusan. Melalui statistika inferensial, data tidak hanya dijelaskan tetapi juga dapat digunakan untuk membuat prediksi atau estimasi berdasarkan hasil analisis (Supandi, 2020).

Analisis multivariat merupakan teknik statistik yang digunakan untuk mengolah dan menganalisis data dengan lebih dari dua variabel secara langsung. Secara umum, metode analisis multivariat dapat dibedakan menjadi dua kelompok, yaitu analisis dependensi dan analisis interdependensi. Analisis interdependensi digunakan jika semua variabel dianggap setara atau saling berhubungan tanpa membedakan mana yang mempengaruhi atau dipengaruhi, salah satunya melalui metode analisis kluster. Sementara itu, analisis dependensi digunakan apabila ingin melihat

pengaruh antara variabel independen (bebas) dan variabel dependen (terikat), misalnya seperti metode analisis diskriminan dan analisis regresi. Dalam konteks analisis dependensi inilah, analisis diskriminan menjadi salah satu metode yang krusial untuk dibahas lebih lanjut (Naufal et al., 2025).

Analisis diskriminan merupakan metode yang digunakan untuk mengelompokkan individu atau objek ke dalam kategori yang telah ditetapkan sebelumnya, dengan menggunakan sejumlah variabel independen (bebas) sebagai dasar klasifikasinya. Analisis diskriminan menggunakan variabel dependen (terikat) bersifat kategorik (nominal atau ordinal) dan variabel independen (bebas) berskala numerik atau metrik. Tujuan utama metode ini adalah untuk mengidentifikasi perbedaan karakteristik antar kelompok serta memprediksi klasifikasi objek ke dalam kelompok tertentu, berdasarkan sejumlah variabel independen yang berskala numerik (Savitri et al., 2021). Berdasarkan struktur data dan asumsinya, analisis diskriminan dibedakan menjadi beberapa metode, yaitu analisis diskriminan linier, analisis diskriminan kuadratik, dan analisis diskriminan regularisasi. Dalam penelitian ini, digunakan analisis diskriminan linier yang akan menghasilkan fungsi diskriminan untuk memisahkan objek secara tegas sesuai dengan kategori kelompoknya (Naufal et al., 2025).

Selain analisis diskriminan, terdapat metode alternatif yang dapat menangani variabel dependen bersifat kategorik, yaitu analisis regresi logistik. Analisis regresi logistik merupakan metode statistik yang dirancang khusus untuk memodelkan variabel dependen yang bersifat biner atau memiliki dua kategori, seperti kondisi ya atau tidak. Model ini bekerja dengan menghubungkan variabel independen terhadap probabilitas terjadinya suatu peristiwa, yang hasil estimasinya kemudian diinterpretasikan melalui nilai peluang atau *odds ratio*. Berdasarkan jenis variabel dependennya, regresi logistik dibagi menjadi tiga, yaitu regresi logistik biner, regresi logistik multinomial, dan regresi logistik ordinal

(Naufal et al., 2025). Dalam penelitian ini digunakan regresi logistik multinomial, dimana metode ini adalah perluasan dari regresi logistik biner yang digunakan ketika variabel dependen (terikat) memiliki skala pengukuran nominal dengan lebih dari dua kategori (Roflin dkk. 2023). Dalam penerapannya, meskipun tidak mensyaratkan asumsi normalitas multivariat secara ketat, model ini tetap memerlukan uji multikolinearitas untuk memastikan tidak adanya korelasi yang tinggi antar variabel bebas. Adapun teknik estimasi parameter yang menjadi standar dalam pembentukan model ini adalah metode maksimum likelihood.

Penelitian ini menggunakan analisis diskriminan linier dan regresi logistik multinomial sebagai metode klasifikasi multikelas dengan pertimbangan efisiensi estimasi parameter dan kejelasan model matematis. Pemilihan analisis diskriminan linier dibandingkan pendekatan kuadrat didasarkan pada prinsip parsimoni, di mana metode ini menggunakan matriks kovarians gabungan sehingga jumlah parameter yang diestimasi lebih sedikit, menjadikan model lebih stabil dan efisien terutama pada ukuran sampel yang terbatas (Johnson & Wichern, 2007; Rencher, 2002). Sementara itu, regresi logistik multinomial dipilih karena kemampuannya membentuk fungsi logit yang merepresentasikan hubungan antara variabel independen dan dependen secara eksplisit dalam bentuk persamaan matematis, yang selanjutnya dapat ditransformasikan menjadi probabilitas untuk setiap kategori kelas (Hosmer & Lemeshow, 2000). Keunggulan ini menjadikan regresi logistik multinomial lebih mudah diinterpretasikan karena berbasis persamaan matematis yang eksplisit (Agresti, 2002). Oleh karena itu, penggunaan kedua metode dalam penelitian ini ditujukan untuk kepentingan komparasi metode, khususnya dalam membandingkan hasil klasifikasi dan tingkat akurasi masing-masing metode pada data multikelas.

Berdasarkan dari kedua metode yang telah ditetapkan, penelitian ini akan berfokus pada studi kasus klasifikasi tingkat keparahan penyakit jantung koroner. Secara medis, penyakit jantung koroner terjadi akibat adanya akumulasi plak pada arteri koroner yang berfungsi menyuplai

oksigen ke otot jantung. Penyakit ini tergolong sebagai jenis penyakit kardiovaskuler yang paling umum dijumpai. Adapun penyakit kardiovaskuler sendiri didefinisikan sebagai gangguan sistematis pada jantung dan pembuluh darah, yang mencakup stroke, penyakit jantung rematik serta kondisi terkait lainnya (Ghani et al., 2016). Mengingat cakupannya yang luas dan sifatnya yang mematikan, penyakit ini tidak hanya menjadi masalah medis semata, tetapi juga telah berkembang menjadi ancaman serius bagi demografi kesehatan masyarakat luas.

Urgensi permasalahan ini terbukti secara statistik, dimana penyakit jantung masih menjadi tantangan kesehatan terbesar secara global maupun nasional. Berdasarkan laporan Organisasi Kesehatan Dunia (World Health Organization/WHO), sekitar 17,7 juta kematian setiap tahun disebabkan oleh penyakit jantung dan pembuluh darah, yang setara dengan 31% dari total penyebab kematian di seluruh dunia. Jumlah tersebut diperkirakan akan terus mengalami peningkatan dan diproyeksikan mencapai 23,3 juta kematian pada tahun 2030 (Nurmasani & Pristyanto, 2021). Tingginya signifikansi penyakit ini juga terlihat di negara maju seperti Amerika Serikat, dimana penyakit jantung tercatat sebagai penyebab utama kematian dengan jumlah hampir 2 juta kasus setiap tahunnya (Nurkholifah et al., 2023). Data dari Centers for Disease Control and Prevention (CDC) menunjukkan bahwa penyakit jantung merupakan penyebab kematian tertinggi pada pria maupun wanita dari berbagai kelompok ras dan etnis, dengan estimasi satu kematian terjadi setiap 36 detik akibat penyakit ini (Haganta Depari et al., 2022).

Kompleksitas data penyakit jantung yang melibatkan berbagai faktor risiko klinis menjadikan kasus ini relevan digunakan sebagai sarana implementasi dalam penelitian yang berfokus pada komparasi metode klasifikasi statistik. Penelitian ini membangun model klasifikasi menggunakan dataset *Cleveland Heart Disease* yang diakses melalui Kaggle, dimana dataset tersebut memiliki struktur data multikelas dengan sejumlah variabel numerik sehingga memenuhi karakteristik dan asumsi

dasar dari kedua metode klasifikasi yang digunakan, yaitu analisis diskriminan linier dan regresi logistik multinomial. Dalam konteks pemodelan klasifikasi, pemilihan variabel prediktor yang relevan menjadi faktor penting untuk menghasilkan analisis yang akurat dan dapat dibandingkan secara objektif. Oleh karena itu, penelitian ini menggunakan empat faktor risiko utama sebagai variabel independen, yaitu usia pasien, kadar kolesterol, detak jantung maksimum (*thalach*), dan depresi segmen ST (*oldpeak*), yang dipilih karena memenuhi karakteristik data yang dibutuhkan oleh kedua metode klasifikasi serta memiliki indikasi medis yang kuat terhadap kondisi iskemia jantung. Sementara itu, variabel dependen yang digunakan adalah tingkat keparahan penyakit jantung koroner yang diklasifikasikan ke dalam skala multikelas, mulai dari level 0 yang menunjukkan tidak adanya penyempitan pembuluh darah hingga level 4 yang menunjukkan penyempitan pembuluh darah dengan tingkat keparahan sangat berat. Oleh karena itu, dataset ini dinilai ideal untuk diuji melalui perbandingan metode analisis diskriminan linier dan regresi logistik multinomial guna menentukan model klasifikasi terbaik.

Penelitian ini hadir untuk mengisi kesenjangan pada studi sebelumnya dengan membandingkan performa analisis diskriminan linier dan regresi logistik multinomial dalam konteks klasifikasi data multikelas. Pemilihan kedua metode ini didasarkan pada perbedaan karakteristik statistiknya, dimana analisis diskriminan linier merupakan pendekatan parametrik yang efisien ketika asumsi normalitas multivariat terpenuhi, sedangkan regresi logistik multinomial merupakan metode parametrik yang lebih fleksibel karena tidak mensyaratkan normalitas multivariat. Oleh karena itu, penelitian ini menjadi penting untuk mengevaluasi secara empiris perbedaan hasil klasifikasi dan tingkat akurasi antara kedua metode tersebut pada data penyakit jantung multikelas. Penggunaan dataset standar Cleveland Heart Disease diharapkan dapat memberikan evaluasi yang objektif mengenai kinerja masing-masing metode dalam memodelkan tingkat risiko penyakit jantung.

Berdasarkan penjelasan di atas, penelitian dilakukan untuk mengetahui model klasifikasi terbaik antara analisis diskriminan linier dan regresi logistik multinomial, serta mengetahui faktor-faktor risiko klinis yang paling signifikan mempengaruhi tingkat keparahan penyakit jantung. Dengan demikian, diharapkan hasil penelitian ini dapat memberikan kontribusi teoritis mengenai ketepatan pemilihan metode klasifikasi statistik pada kasus data multikelas. Oleh karena itu, penulis memilih judul penelitian: “Perbandingan Analisis Diskriminan Linier dan Regresi Logistik Multinomial dalam Klasifikasi Penyakit Jantung pada Data Multikelas”.

1.2 Batasan Masalah

Dalam penulisan skripsi ini, batasan masalah yang digunakan adalah sebagai berikut:

1. Analisis diskriminan yang digunakan adalah analisis diskriminan linier.
2. Regresi logistik yang digunakan adalah model regresi logistik multinomial.
3. Penelitian ini difokuskan pada klasifikasi atau pengelompokan kategori tingkat keparahan penyakit jantung.
4. Data yang digunakan adalah dataset standar penyakit jantung (*Cleveland Heart Disease*) yang diakses melalui Kaggle.

1.3 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan di atas, maka rumusan dalam penulisan skripsi ini adalah sebagai berikut:

1. Bagaimana langkah-langkah pembentukan model analisis diskriminan linier dan regresi logistik multinomial?
2. Bagaimana model klasifikasi yang terbentuk dari penerapan metode analisis diskriminan linier dan regresi logistik multinomial pada data tingkat keparahan penyakit jantung?
3. Manakah metode yang memiliki performa dan akurasi klasifikasi terbaik antara analisis diskriminan linier dan regresi logistik multinomial?

1.4 Tujuan Penelitian

Berdasarkan permasalahan yang diajukan, tujuan penelitian dari penulisan skripsi ini adalah sebagai berikut:

1. Mempelajari metode analisis diskriminan linier dan analisis regresi logistik multinomial dalam mengklasifikasikan obeservasi ke dalam kelas yang sesuai.
2. Menerapkan metode analisis diskriminan linier dan analisis regresi logistik multinomial pada studi kasus klasifikasi tingkat keparahan penyakit jantung.
3. Membandingkan tingkat performa dan akurasi klasifikasi dalam penerapan analisis diskriminan linier dan regresi logistik multinomial.

1.5 Manfaat Penelitian

Dari hasil penulisan skripsi ini, manfaat yang diharapkan adalah sebagai berikut:

1. Bagi penulis, penelitian ini dapat menambah pengetahuan mengenai proses dan hasil pemodelan analisis diskriminan linier dan regresi logistik multinomial, serta memperluas wawasan terkait penerapan matematika dalam bidang statistika multivariat.
2. Bagi lembaga, hasil penelitian ini dapat menjadi tambahan pemikiran keilmuan dalam bidang statistika serta menjadi tolak ukur terhadap kemampuan mahasiswa dalam menetapkan teori yang telah diperoleh.
3. Bagi pembaca, penelitian ini dapat dijadikan referensi bagi penelitian selanjutnya, sekaligus menambah wawasan dan memberikan gambaran mengenai langkah-langkah dari pemodelan metode terkait.

1.6 Tinjauan Pustaka

Tinjauan pustaka yang relevan dengan tema yang digunakan dalam penulisan skripsi ini adalah sebagai berikut:

1. Penelitian yang berjudul “Perbandingan Model Analisis Diskriminan dengan Multiple Discriminant Analysis dan Regresi Logistik Ordinal” yang ditulis oleh Hesty Eko Setyowati mahasiswa Program Studi

Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Sunan Kalijaga Yogyakarta. Penelitian ini bertujuan untuk membandingkan model analisis diskriminan berganda dan regresi logistik ordinal dalam menganalisis faktor-faktor yang memengaruhi Indeks Pembangunan Manusia (IPM) di Jawa Tengah tahun 2022, serta mengidentifikasi faktor-faktor yang berpengaruh signifikan.

2. Penelitian yang berjudul "Perbandingan Metode Analisis Diskriminan Linier dan Regresi Logistik Multinomial Untuk Klasifikasi Pada Data Multikelas" yang ditulis oleh Muhammad Tahta Rizqi Yuandri, mahasiswa Universitas Gadjah Mada. Penelitian ini bertujuan untuk membandingkan performa metode regresi logistik multinomial dan analisis diskriminan linier dalam mengklasifikasikan jenis penyakit obesitas, di mana hasilnya menunjukkan bahwa regresi logistik multinomial memiliki akurasi dan *f1-score* yang lebih baik dibandingkan analisis diskriminan linier.
3. Penelitian yang berjudul "Klasifikasi Penyakit Jantung Menggunakan Algoritma Analisis Diskriminan Linier" yang ditulis oleh Ibnu Rashad, R Rizal Isnanto, dan Catur Edi Widodo dari Universitas Diponegoro. Penelitian ini bertujuan untuk menerapkan algoritma *Linier Discriminant Analysis* (LDA) dalam mendeteksi penyakit jantung dengan biaya rendah, serta membandingkan hasil akurasi penggunaan 2 label target (sakit/tidak) dengan 5 label target (tingkatan stadium).
4. Penelitian yang berjudul "Estimasi Parameter Regresi Logistik Multinomial Menggunakan Metode Maksimum Likelihood" yang ditulis oleh Karina Sagita, mahasiswa Program Studi Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Sunan Kalijaga Yogyakarta. Penelitian ini bertujuan untuk mengestimasi parameter model regresi logistik multinomial pada kasus penyakit Tuberkulosis di RSUD Temanggung, serta menentukan faktor-faktor yang berpengaruh signifikan terhadap penyakit tersebut, yaitu jenis kelamin dan sesak nafas.

5. Penelitian yang berjudul “Perbandingan Analisis Diskriminan Linier dan Regresi Logistik Multinomial dalam Klasifikasi Penyakit Jantung pada Data Multikelas” yang ditulis oleh Nurul Hanin Azzahra Supriyono mahasiswa Program Studi Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Sunan Kalijaga Yogyakarta. Penelitian ini mencari metode klasifikasi terbaik terhadap data multikelas menggunakan analisis diskriminan linier dan regresi logistik multinomial.

Tabel 1. 1 Perbedaan Penelitian Sebelumnya Dengan Penulis

No	Judul Penelitian	Metode	Studi Kasus	Hasil Penelitian
1.	Perbandingan Model Analisis Diskriminan dengan Multiple Discriminant Analysis dan Regresi Logistik Ordinal (Hesty Eko S)	<i>Multiple discriminant analysis</i> dan regresi logistik ordinal.	Faktor-faktor yang mempengaruhi Indeks Pembangunan Manusia di Provinsi Jawa Tengah tahun 2022.	Kedua metode menghasilkan ketepatan klasifikasi yang sama baiknya, yaitu sebesar 77,14%. Faktor yang signifikan berpengaruh adalah rata-rata lama sekolah.
2.	Perbandingan Metode Analisis Diskriminan Linier dan Regresi Logistik	Analisis diskriminan linier dan regresi logistik multinomial.	Klasifikasi jenis penyakit obesitas.	Akurasi metode regresi logistik multinomial (93.04%) lebih baik dibandingkan analisis

No	Judul Penelitian	Metode	Studi Kasus	Hasil Penelitian
	Multinomial Untuk Klasifikasi Pada Data Multikelas (M. Tahta Rizqi Y)			diskriminan linier (88.61%). metode regresi logistik multinomial direkomendasikan untuk data ini.
3.	Klasifikasi Penyakit Jantung Menggunakan Algoritma Analisis Diskriminan Linier (Ibnu Rashad dkk)	Analisis diskriminan linier.	Klasifikasi Penyakit Jantung (Risiko Tinggi vs Risiko Rendah).	Klasifikasi LDA dengan 2 kelas target menghasilkan akurasi lebih baik (81,22%) dibandingkan dengan menggunakan 5 kelas target (59,38%).
4.	Estimasi Parameter Regresi Logistik Multinomial Menggunakan Metode Maksimum Likelihood	Regresi logistik multinomial dengan metode maksimum likelihood.	Faktor-faktor yang mempengaruhi penyakit tuberkulosis rawat inap di RSUD Temanggung 2018.	Variabel independen yang berpengaruh signifikan adalah jenis kelamin dan sesak nafas dengan ketepatan klasifikasi model sebesar 56,6%.

No	Judul Penelitian	Metode	Studi Kasus	Hasil Penelitian
	(Karina Sagita)			
5.	Perbandingan Analisis Diskriminan Linier dan Regresi Logistik Multinomial dalam Klasifikasi Penyakit Jantung pada Data Multikelas (Nurul Hanin)	Analisis diskriminan linier dan regresi logistik multinomial	Klasifikasi Penyakit Jantung pada Data Multikelas.	Perbandingan ketepatan klasifikasi dan nilai akurasi antara analisis diskriminan linier dan regresi logistik multinomial.

Berdasarkan telaah terhadap penelitian-penelitian sebelumnya, dapat disimpulkan bahwa sebagian besar studi masih menempatkan studi kasus sebagai fokus utama analisis, baik dengan menggunakan satu metode klasifikasi maupun perbandingan metode untuk memperoleh kesimpulan substantif terkait faktor-faktor yang berpengaruh pada kasus yang diteliti. Sementara itu, penelitian ini memiliki perbedaan mendasar pada orientasi metodologis, dimana analisis diskriminan linier dan regresi logistik multinomial diperlakukan sebagai objek utama komparasi, sedangkan kasus penyakit jantung multikelas digunakan sebagai media implementasi untuk menguji kinerja kedua metode pada karakteristik data yang sama. Dengan demikian, penelitian ini berkontribusi dalam memberikan gambaran empiris mengenai perbandingan akurasi dan ketepatan klasifikasi kedua metode

statistik tersebut, khususnya pada data multikelas, yang belum secara eksplisit menjadi fokus utama pada penelitian-penelitian sebelumnya.

1.7 Sistematika Penulisan

Sistematika penulisan yang digunakan dalam skripsi ini adalah sebagai berikut:

1. BAB I PENDAHULUAN

Dalam pendahuluan akan dijelaskan mengenai latar belakang dan perumusan masalah yang disajikan dalam bentuk uraian kronologis untuk menggambarkan konteks penelitian. Selain itu, pada bab ini juga dijelaskan batasan masalah yang digunakan sebagai ruang lingkup penelitian, rumusan masalah yang berisi pertanyaan-pertanyaan penelitian, serta tujuan penelitian yang menjelaskan capaian yang ingin diperoleh. Bab ini juga memuat manfaat penelitian, baik secara teoritis maupun praktis, guna memberikan kontribusi bagi penelitian selanjutnya. Selain itu, disajikan pula tinjauan pustaka yang berisi ulasan singkat dari penelitian-penelitian terdahulu yang relevan dengan tema penelitian ini.

2. BAB II LANDASAN TEORI

Bab ini memuat definisi, konsep, sifat, serta teori dasar dan teori penunjang yang berkaitan dengan analisis diskriminan linier dan regresi logistik multinomial. Teori-teori tersebut digunakan sebagai dasar dalam memahami serta mendukung analisis yang dilakukan dalam penelitian ini.

3. BAB III METODE PENELITIAN

Bagian ini menjelaskan secara rinci setiap tahapan penelitian yang dilakukan, meliputi sumber data, metode pengumpulan data, variabel penelitian, metode analisis data, alat pengolahan data, serta alur dan langkah-langkah analisis yang digunakan dalam penelitian. Dengan demikian, metodologi yang dipaparkan dapat memberikan gambaran sistematis mengenai proses penelitian dari awal hingga akhir.

4. BAB IV PEMBAHASAN

Bab ini berisi implementasi konsep analisis diskriminan linier dan regresi logistik multinomial termasuk tahapan analisis, algoritma yang digunakan, serta contoh penerapannya pada kasus penelitian. Pembahasan dilakukan untuk menilai sejauh mana kedua metode dapat memberikan hasil klasifikasi yang akurat dan relevan dengan data yang dianalisis.

5. BAB V STUDI KASUS

Pada bab ini dibahas penerapan metode analisis diskriminan linier dan regresi logistik multinomial dalam proses klasifikasi faktor-faktor risiko yang mempengaruhi tingkat keparahan penyakit jantung. Analisis ini bertujuan untuk mengetahui sejauh mana kedua metode dapat memberikan hasil klasifikasi yang optimal, serta membandingkan kinerja keduanya dalam menentukan model yang paling sesuai dengan karakteristik data.

6. BAB VI PENUTUP

Bab ini memuat kesimpulan yang diperoleh berdasarkan hasil analisis yang telah dilakukan. Kesimpulan disusun untuk menjawab rumusan masalah yang telah dirumuskan pada bab sebelumnya. Selain itu, bab ini juga menyajikan saran-saran yang dapat dijadikan pertimbangan bagi penelitian selanjutnya maupun penerapan metode yang digunakan dalam penelitian ini.

BAB VI

PENUTUP

Bagian ini memuat kesimpulan yang diperoleh berdasarkan hasil analisis serta saran untuk penelitian selanjutnya.

6.1. Kesimpulan

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan, maka dapat diperoleh beberapa kesimpulan sebagai berikut:

1. Prosedur Analisis Diskriminan Linier dan Regresi Logistik Multinomial

a. Analisis Diskriminan Linier

Tahapan dalam analisis diskriminan linier meliputi:

1) Melakukan pengujian asumsi analisis diskriminan

- Uji normalitas multivariat
- Uji homogenitas matrik varians-kovarians
- Uji kesamaan vektor rata-rata
- Deteksi outlier

2) Pembentukan fungsi diskriminan

$$Z_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p$$

3) Menguji signifikansi dari fungsi diskriminan dengan Wilk's Lambda

$$\Lambda = \frac{|W|}{|W+B|}$$

4) Melakukan pengklasifikasian

- Menghitung skor diskriminan
- Menentukan grup centroid

$$Z_k = \frac{\sum_{i=1}^{N_k} LD_{ik}}{N_k}$$

- Penentuan keanggotaan kelompok

$$D_k = \sqrt{(LD_1 - C_{1k})^2 + \dots + (LD_m - C_{mk})^2}$$

b. Regresi Logistik Multinomial

Langkah-langkah dalam regresi logistik multinomial adalah:

- 1) Melakukan uji asumsi multikolinearitas
- 2) Estimasi parameter regresi logistik multinomial

Fungsi logit yang terbentuk

$$g(x_i) = \beta_{k0} + \sum_{j=1}^p \beta_{kj}x_{ij}$$

Sedangkan fungsi peluangnya adalah:

$$\pi_k(x_i) = \frac{\exp(\beta_{k0} + \sum_{j=1}^p \beta_{kj}x_{ij})}{1 + \sum_{h=1}^{K-1} \exp(\beta_{h0} + \sum_{j=1}^p \beta_{hj}x_{ij})}$$

- 3) Pengujian signifikansi serentak dengan Uji Likelihood Rasio

$$G^2 = -2(\ln L_0 - \ln L_1)$$

- 4) Pengujian signifikansi parsial dengan Uji Wald

$$W = \left[\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \right]^2$$

c. Evaluasi Prediksi Ketepatan Klasifikasi

- 1) $Akurasi = \frac{\sum_{i=1}^K n_{ii}}{n} \times 100\%$

- 2) $precision = \frac{TP}{TP+FP}$

- 3) $recall = \frac{TP}{TP+FN}$

- 4) $F1 = 2 \times \frac{precision \times recall}{precision + recall}$

2. Model Klasifikasi yang Terbentuk

a. Model Analisis Diskriminan Linier

Fungsi diskriminan yang diperoleh:

$$Z_1 = -0,05643X_1 + 0,23082X_2 - 0,78467X_3 + 0,78062X_4$$

$$Z_2 = -0,29425X_1 - 0,91203X_2 + 0,14731X_3 + 0,41706X_4$$

$$Z_3 = 0,17864X_1 + 0,39333X_2 + 0,80448X_3 + 0,66483X_4$$

$$Z_4 = -1,03960X_1 + 0,48402X_2 - 0,29156X_3 + 0,01294X_4$$

b. Model Regresi Logistik Multinomial

Model logit yang diperoleh:

$$g_1(x_i) = -0,41461 - 0,71095X_3 + 0,36859X_4$$

$$g_2(x_i) = -1,13895 - 1,31622X_3 + 1,02893X_4$$

$$g_3(x_i) = -0,95695 - 1,09334X_3 + 1,09874X_4$$

$$g_4(x_i) = -2,72781 - 1,25675X_3 + 1,60509X_4$$

3. Metode Terbaik

Berdasarkan hasil evaluasi ketepatan klasifikasi, analisis diskriminan linier memperoleh tingkat akurasi sebesar 52,54%, sedangkan regresi logistik multinomial sebesar 55,93%. Dengan demikian, dapat disimpulkan bahwa metode regresi logistik multinomial memiliki performa dan kinerja yang lebih baik dibandingkan analisis diskriminan linier dalam mengklasifikasikan tingkat keparahan penyakit jantung pada penelitian ini, meskipun tingkat akurasinya masih berada pada kategori sedang.

6.2. Saran

Adapun beberapa saran untuk penelitian selanjutnya adalah sebagai berikut:

1. Penelitian ini hanya membandingkan metode analisis diskriminan linier dan regresi logistik multinomial. Oleh karena itu, penelitian selanjutnya dapat menggunakan metode klasifikasi lain seperti K-Nearest Neighbors (KNN), Naïve Bayes, Support Vector Machine (SVM), maupun Classification and Regression Tree (CART) untuk memperoleh hasil klasifikasi yang lebih optimal.
2. Pada analisis diskriminan linier, penelitian ini menggunakan pendekatan direct method dalam pembentukan fungsi diskriminan. Penelitian selanjutnya dapat mempertimbangkan penggunaan metode *stepwise discriminant* untuk mengetahui variabel yang paling berpengaruh secara bertahap.

3. Pada regresi logistik multinomial, estimasi parameter dilakukan menggunakan metode maksimum likelihood. Penelitian berikutnya dapat mempertimbangkan pendekatan estimasi lain atau melakukan validasi model menggunakan teknik cross-validation agar diperoleh hasil yang lebih robust.
4. Penelitian selanjutnya disarankan untuk menambahkan variabel prediktor lain yang relevan dengan tingkat keparahan penyakit jantung guna meningkatkan kemampuan dan akurasi model klasifikasi. Selain itu, peneliti selanjutnya juga dapat menggunakan sumber data multikelas lain yang memiliki karakteristik dan struktur yang sesuai, sehingga kinerja metode analisis diskriminan linier dan regresi logistik multinomial dapat dievaluasi secara lebih komprehensif.

DAFTAR PUSTAKA

- Agresti, Alan. (2002). *Categorical Data Analysis*. Wiley-Interscience.
- Amruddin, Priyanda, R., Agustina, T. S., Ariantini, N. S., Rusmayani, N. G. A. L., Ningsih, K. P., Wulandari, S., Putranto, P., Yuniat, I., Untari, I., Mujiani, S., & Wicaksono, D. (2022). *Metodologi Penelitian Kuantitatif* (F. Sukmawati, Ed.; 1st ed.). Pradina Pustaka.
- Anna, I. D. (2017). *Matriks dan Ruang Vektor* (MNC Publishing, Ed.). Media Nusa Creative.
- Anton, H. (1987). *Aljabar Linier Elementer*. Erlangga.
- Bain, L. J., & Engelhardt, M. (1992). *Introduction to Probability and Mathematical Statistics* (2nd Edition). Duxbury Press.
- Budiyono. (2009). *Statistika untuk Penelitian* (Suyono, Ed.; 3rd ed.). UPT Penerbitan dan Percetakan UNS.
- Chamidah, N., & Lestari, B. (2022). *Analisis Regresi Nonparametrik dengan Perangkat Lunak R*. Airlangga University Press.
- Chandra, T., & Priyono. (2024). *Pengantar Statistika Inferensial* (Abd. W. H. S. S. Yoyok, Ed.; 1st ed.). PT Literasi Nusantara Abadi Grup.
- Detrano, R., Janosi, A., Steinbrunn, W., Pfisterer, M., Schmid, J.-J., Sandhu, S., Guppy, K. H., Lee, S., & Froelicher, V. (1989). International Application of a New Probability Algorithm for the Diagnosis of Coronary Artery Disease. *The American Journal of Cardiology*, *64*, 304–310.
- E. Walpole, R., H. Myers, R., L. Myers, S., & Ye, K. (2011). *Probability Statistics for Engineers & Scientists* (Deirdre Lynch, Ed.; 9th ed.). Pearson Education.
- Ghani, L., Dewi Susilawati, M., & Novriani, H. (2016). Faktor Risiko Dominan Penyakit Jantung Koroner di Indonesia. *Buletin Penelitian Kesehatan*, *44*(3), 153–164.
- Haganta Depari, D., Widiastiwi, Y., & Mega Santoni, M. (2022). Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung. *Jurnal Informatik*, *18*(3), 239–248.
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression* (2nd ed.). John Wiley & Sons, Inc.

- Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*. Pearson Prentice Hall.
- Julie, H., Apriani, M. S., & Krisnamurti, C. N. (2017). *Buku Ajar Teori Peluang* (1st ed.). Sanata Dharma University Press.
- Kurniadi, H. (2015). *Stop! Gejala Penyakit Jantung Koroner*. Relasi Inti Media.
- Kutner, M. H. ., Nachtsheim, Chris., Neter, John., & Li, William. (2005). *Applied linear statistical models*. McGraw-Hill Irwin.
- Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3), 519–530.
- Martono, T. (2017). *Aljabar Matriks untuk Metode Kuantitatif* (T. Martono & A. Pratama, Eds.). PT Penerbit IPB Press.
- Naufal, N., Apriani, F., Dr.Fajriana, Dr.Nurdin, Nurdin, K., & Hadi Siswanto, D. (2025). *Analisis Multivariat* (W. Yuliani, Ed.). Literasi Langsung Terbit.
- Nurkholifah, M., Nofiar.Am, A., & Kurnia Oktorina, F. (2023). Analisa Penyakit Jantung Menggunakan Algoritma Naïve Bayes. *Journal of System and Computer Engineering (JSCE) ISSN*, 4(1), 26.
- Nurmasani, A., & Pristyanto, Y. (2021). Algoritme Stacking untuk Klasifikasi Penyakit Jantung pada Dataset Imbalanced Class. *Jurnal Pseudocode*, VIII(1), 21–26. www.ejournal.unib.ac.id/index.php/pseudocode
- O.Sihombing, S. (2020). *Pengantar Metode Analisis Multivariat* (M. Nasrudin, Ed.). PT. Nasya Expanding Management.
- Powers, D. M. W. (2007). *Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation*. 1–24.
- Purnomo, Sutadji, E., Wargijono Utomo, Mp., Purnawirawan, O., Farich, R., Ratna Fajarwati, Mp. M., Carina, A., & Gilang, N. (2022). *Analisis Data Multivariat* (W. Nur L, Ed.). Omera Pustaka.
- Rencher, A. C. (2002). *Methods of Multivariate Analysis* (2nd Edition). John Wiley & Sons, Inc.
- Rifa'i, R. (2016). *Aljabar Matriks Dasar* (D. Novidiantoko & I. Candrawinata, Eds.). Deepublish.

- Roflin, E., Riana, F., Pariyana, E. M., & Liberty, I. A. (2023). *Regresi Logistik Biner dan Multinomial* (Moh. Nasrudin, Ed.; 1st ed.). PT Nasya Expanding Management.
- Savitri, C., Faddila, S. P., Iswari, R., Anam, C., Syah, S., Mulyani, R., Sihombing, R., Kismawadi, R., Pujianto, A., Mulyati, A., Astuti, Y., Adinugroho, W. C., Imanuddin, R., Kristia, A., Nuraini, M., & Tirtana, S. (2021). *Statistik Multivariat dalam Riset* (I. Ahmaddien, Ed.). Widina Bhakti Persada Bandung. www.penerbitwidina.com
- Supandi, E. D. (2020). *Statistika dan Terapannya* (R. Trisnadewi, Ed.; 1st ed.). PT Refika Aditama.
- Supangat, A. (2008). *Statistika: Dalam Kajian Deskriptif, Inferensia, dan Non Parametrik* (I). Kencana Prenada Media Group.
- Wahyu Wibowo, F. (2024). *Aljabar Linier dan Matriks* (F. Wahyu Wibowo, Ed.). Universitas Amikom Yogyakarta.