

**REDUKSI *IMBALANCED DATA* DIAGNOSA HIPERTENSI
DENGAN *TOMEK LINKS* PADA REGRESI LOGISTIK**

SKRIPSI

**OLEH:
PUTRI AULIA FACHREZA
NIM. 210601110043**



**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG
2025**

**REDUKSI *IMBALANCED DATA* DIAGNOSA HIPERTENSI
DENGAN *TOMEK LINKS* PADA REGRESI LOGISTIK**

SKRIPSI

**Diajukan Kepada
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
untuk Memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Matematika (S.Mat)**

**Oleh
PUTRI AULIA FACHREZA
NIM. 210601110043**

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG
2025**

**REDUKSI *IMBALANCED DATA* DIAGNOSA HIPERTENSI
DENGAN *TOMEK LINKS* PADA REGRESI LOGISTIK**

SKRIPSI

**Oleh
Putri Aulia Fachreza
NIM. 210601110043**

Telah Disetujui Untuk Diuji

Malang, 22 Mei 2025

Dosen Pembimbing I



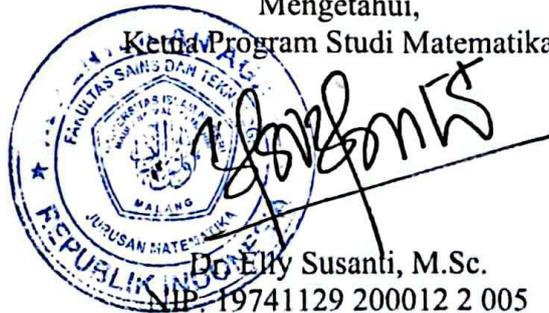
Ria Dhea Layla Nur Karisma, M.Si.
NIPPPK. 19900709 202321 2 037

Dosen Pembimbing II



Erna Herawati, M.Pd.
NIPPPK. 19760723 202321 2 006

Mengetahui,
Ketua Program Studi Matematika



Dr. Elly Susanti, M.Sc.
NIP. 19741129 200012 2 005

**REDUKSI *IMBALANCED DATA* DIAGNOSA HIPERTENSI
DENGAN *TOMEK LINKS* PADA REGRESI LOGISTIK**

SKRIPSI

Oleh
Putri Aulia Fachreza
NIM. 210601110043

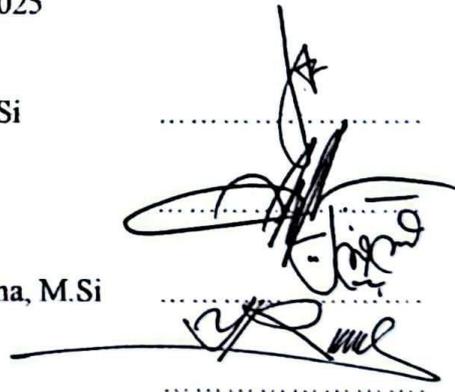
Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
untuk Memperoleh Gelar Sarjana Matematika (S.Mat)
Tanggal 16 Juni 2025

Ketua Penguji : Prof. Dr. Hj. Sri Harini, M.Si

Anggota Penguji 1 : Abdul Aziz, M.Si

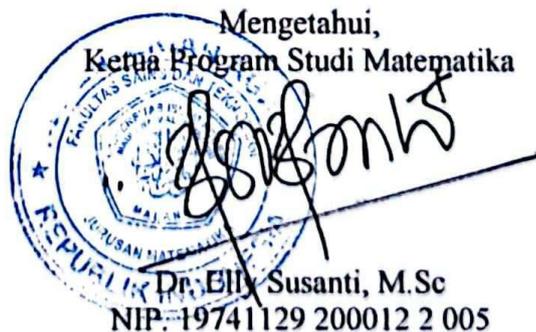
Anggota Penguji 2 : Ria Dhea Layla Nur Karisma, M.Si

Anggota Penguji 3 : Erna Herawati, M.Pd



Handwritten signatures of the examiners: Prof. Dr. Hj. Sri Harini, M.Si; Abdul Aziz, M.Si; Ria Dhea Layla Nur Karisma, M.Si; and Erna Herawati, M.Pd.

Mengetahui,
Ketua Program Studi Matematika



Official stamp of the Faculty of Science and Technology, Department of Mathematics, and the signature of Dr. Elly Susanti, M.Sc.

Dr. Elly Susanti, M.Sc
NIP. 19741129 200012 2 005

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Putri Aulia Fachreza
NIM : 210601110043
Program Studi : Matematika
Fakultas : Sains dan Teknologi
Judul Skripsi : Reduksi *Imbalanced Data* Diagnosa Hipertensi dengan
Tomek Links pada Regresi Logistik

Menyatakan dengan sebenarnya bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya sendiri, bukan merupakan pengambilan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan dan pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka. Apabila di kemudian hari terbukti atau dapat dibuktikan skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 16 Juni 2025

Yang membuat pernyataan,



Putri Aulia Fachreza

NIM. 210601110043

MOTO

“Even if the world ends tomorrow, cherish every second today.”

PERSEMBAHAN

Skripsi ini penulis persembahkan dengan penuh rasa terima kasih kepada: Mama dan Ayah tersayang yang selalu memberikan doa, dukungan, nasihat bijaksana, serta dorongan semangat yang tak pernah putus demi kesuksesan penulis. Adik tersayang yang senantiasa memberikan doa dan motivasi sehingga penulis selalu semangat dalam menjalani proses ini. Terima kasih kepada sahabat-sahabat dan teman-teman yang telah memberikan bantuan, semangat, serta inspirasi yang sangat berarti dalam menyelesaikan skripsi ini. Semoga segala doa dan dukungan tersebut menjadi sumber kekuatan penulis hingga mencapai tujuan yang diimpikan.

KATA PENGANTAR

Assalamu 'alaikum Warahmatullahi Wabarakatuh

Segala puji dan syukur bagi Allah SWT atas rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan penyusunan skripsi ini sebagai salah satu syarat untuk memperoleh gelar sarjana pada Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Penulisan dan penyusunan skripsi ini, penulis mendapatkan bimbingan, arahan, dukungan, bantuan, dan doa dari berbagai pihak. Oleh karena itu, penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. H. M. Zainuddin, M.A., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Ibu Prof. Dr. Sri Harini, M.Si., selaku dekan Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang sekaligus menjadi ketua penguji dalam ujian skripsi yang telah memberikan arahan serta ilmu yang bermanfaat.
3. Ibu Dr. Elly Susanti, M.Sc., selaku ketua Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Ibu Ria Dhea Layla Nur Kharisma, M.Si., selaku dosen pembimbing I yang telah meluangkan waktu, tenaga, dan pikiran untuk memberikan banyak arahan, nasihat, dan motivasi selama proses penyusunan skripsi.
5. Ibu Erna Herawati, M.Pd., selaku dosen pembimbing II yang telah meluangkan waktu, tenaga, dan pikiran untuk memberikan ilmunya selama proses penyusunan skripsi.
6. Bapak Abdul Aziz, M.Si., selaku penguji I dalam ujian skripsi yang telah memberikan arahan, bimbingan, serta ilmu yang bermanfaat.
7. Kedua orang tua tercinta, yang senantiasa memberikan doa, dukungan moral, dukungan materi, dan kasih sayang yang tidak terhingga.
8. Bapak Dr. Fachrur Rozi, M.Si., selaku dosen wali yang senantiasa memberikan bimbingan, arahan, dan ilmu yang bermanfaat.

9. Bapak/Ibu dosen dan segenap sivitas akademika Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang yang telah memberikan berbagai ilmu dan pengalaman yang berharga.
 10. Adik dan seluruh keluarga besar tercinta yang selalu memberikan dukungan dan doa kepada penulis.
 11. Sahabat-sahabat terbaik yang selalu membantu, menemani, dan memberikan semangat serta motivasi selama proses penyusunan skripsi.
 12. Teman-teman seperjuangan di Program Studi Matematika angkatan 2021 khususnya teman-teman konsorsium statistika yang berjuang bersama-sama.
- Semoga Allah melimpahkan taufik dan hidayah-Nya kepada kita semua. Penulis berharap semoga skripsi ini dapat memberikan manfaat bagi penulis dan pembaca serta dapat menjadi sumbangsih bagi perkembangan ilmu pengetahuan, khususnya di bidang Matematika.

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Malang, 16 Juni 2025

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGAJUAN	ii
HALAMAN PERSETUJUAN	iii
PERNYATAAN KEASLIAN TULISAN	v
MOTO	vi
PERSEMBAHAN.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
DAFTAR SIMBOL	xiv
DAFTAR LAMPIRAN	xv
ABSTRAK	xvi
ABSTRACT.....	xvii
مستخلص البحث.....	xviii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	7
1.3 Tujuan Penelitian.....	8
1.4 Manfaat Penelitian.....	8
1.5 Batasan Masalah.....	9
BAB II KAJIAN TEORI	10
2.1 Teori Pendukung.....	10
2.1.1 <i>Imbalanced Data</i>	10
2.1.2 Regresi Logistik.....	13
2.1.3 <i>Maximum Likelihood Estimator</i>	15
2.1.4 Pengujian Parameter	22
2.1.5 Uji Kesesuaian Model	25
2.1.6 Interpretasi Koefisien Model.....	26
2.1.7 Ketepatan Klasifikasi.....	27
2.1.8 Uji Multikolinearitas	28
2.1.9 Hipertensi	30
2.2 Kajian Integrasi Hipertensi dengan Al-Qu’ran dan Hadits.....	34
2.3 Kajian Hipertensi Dengan Regresi Logistik dan <i>Tomek Links</i>	41
BAB III METODE PENELITIAN	44
3.1 Jenis Penelitian	44
3.2 Data dan Sumber Data.....	44
3.3 Teknik Pengumpulan Data	45
3.4 Teknik Analisis Data	45
3.5 Diagram Alir Penelitian	48
BAB IV PEMBAHASAN.....	49
4.1 Statistika Deskriptif.....	49
4.1.1 Deskripsi Jenis Kelamin	50
4.1.2 Deskripsi Komsumsi Rokok.....	52
4.1.3 Deskripsi Aktivitas Fisik	54

4.1.4	Deskripsi Konsumsi Gula Berlebih	56
4.1.5	Deskripsi Konsumsi Garam Berlebih	58
4.1.6	Deskripsi Lemak Berlebih	60
4.1.7	Deskripsi Konsumsi Buah Dan Sayur	62
4.1.8	Deskripsi Konsumsi Alkohol	63
4.1.9	Deskripsi Usia	65
4.2	Pembagian Data <i>Training</i> dan <i>Testing</i>	66
4.3	Penerapan <i>Tomek Links</i>	67
4.4	Model Awal Regresi Logistik	68
4.5	<i>Maximum Likelihood Estimation</i>	69
4.6	Uji Multikolinearitas	74
4.7	Uji Signifikasi Parameter	76
4.7.1	Uji Simultan.....	76
4.7.2	Uji Parsial	77
4.8	Model Regresi Logistik.....	82
4.9	Uji Kesesuaian Model	83
4.10	Interpretasi Koefisien Parameter Model	84
4.11	Ketepatan Klasifikasi Model.....	86
4.12	Kajian Penelitian Dalam Prespektif Islam	89
BAB V	PENUTUP.....	93
5.1	Kesimpulan.....	93
5.2	Saran.....	94
DAFTAR PUSTAKA	95
LAMPIRAN	100
RIWAYAT HIDUP	117

DAFTAR TABEL

Tabel 2.1	Data Ilustrasi <i>Tomek Links</i>	11
Tabel 2.2	<i>Confusion Matrix</i>	27
Tabel 3.1	Variabel Penelitian	44
Tabel 3.2	Lanjutan Variabel Penelitian.....	45
Tabel 4.1	Jenis Kelamin Berdasarkan Diagnosa Hipertensi	51
Tabel 4.2	Konsumsi Rokok Berdasarkan Diagnosa Hipertensi	53
Tabel 4.3	Aktivitas Fisik Berdasarkan Diagnosa Hipertensi	55
Tabel 4.4	Konsumsi Gula Berdasarkan Diagnosa Hipertensi.....	57
Tabel 4.5	Konsumsi Garam Berlebih Berdasarkan Diagnosa Hipertensi	59
Tabel 4.6	Lemak Berlebih Berdasarkan Diagnosa Hipertensi	61
Tabel 4.7	Konsumsi Buah dan Sayur Berdasarkan Diagnosa Hipertensi	62
Tabel 4.8	Konsumsi Alkohol Berdasarkan Diagnosa Hipertensi.....	64
Tabel 4.9	Usia Berdasarkan Diagnosa Hipertensi.....	65
Tabel 4.10	Usia Berdasarkan Diagnosa Hipertensi Lanjutan	66
Tabel 4.11	Pembagian Data <i>Training</i> dan <i>Testing</i>	67
Tabel 4.12	Estimasi Parameter	73
Tabel 4.13	Nilai VIF Setiap Variabel Independen	75
Tabel 4.14	Uji Signifikansi Simultan	77
Tabel 4.15	Uji Signifikansi Parsial	78
Tabel 4.16	Uji Signifikansi Parsial Kedua	80
Tabel 4.17	Uji Signifikansi Simultan Kedua	81
Tabel 4.18	Uji Kesesuaian Model	84
Tabel 4.19	Nilai <i>Odds Ratio</i>	85
Tabel 4.20	<i>Confusion Matrix</i> Hasil Klasifikasi Model	87

DAFTAR GAMBAR

Gambar 2.1	Ilustrasi <i>Tomek Links</i>	10
Gambar 3.1	Diagram Alir Penelitian	48
Gambar 4.1	Diagnosa Hipertensi	49
Gambar 4.2	Jenis Kelamin	50
Gambar 4.3	Konsumsi Rokok.....	52
Gambar 4.4	Aktivitas Fisik	54
Gambar 4.5	Konsumsi Gula Berlebih	56
Gambar 4.6	Konsumsi Garam Berlebih	58
Gambar 4.7	Lemak Berlebih	60
Gambar 4.8	Konsumsi Buah dan Sayur	62
Gambar 4.9	Konsumsi Alkohol	64
Gambar 4.10	Perbandingan Sebelum dan Sesudah Proses <i>Tomek Links</i>	68
Gambar 4.11	Grafik Kekonvergenan Iterasi	73

DAFTAR SIMBOL

x_1	:	Variabel Jenis Kelamin
x_2	:	Variabel Konsumsi Rokok
x_3	:	Variabel Aktivitas Fisik
x_4	:	Variabel Konsumsi Gula Berlebih
x_5	:	Variabel Konsumsi Garam Berlebih
x_6	:	Variabel Lemak Berlebih
x_7	:	Variabel Konsumsi Buah dan Sayur
x_8	:	Variabel Konsumsi Alkohol
x_9	:	Variabel Tekanan Sistolik
x_{10}	:	Variabel Tekanan Diastolik
x_{11}	:	Variabel Usia
Y	:	Variabel dependen: 1 (Hipertensi), 0 (Tidak Hipertensi)
y_i	:	Nilai aktual variabel dependen ke- i
X_i	:	Vektor variabel independen pada observasi ke- i
x_{ij}	:	Nilai variabel independen ke- j pada observasi ke- i
$\pi(X_i)$:	Probabilitas sukses untuk observasi ke- i
β_j	:	Estimasi parameter atau koefisien regresi logistik untuk variabel independen ke- j
β	:	Vektor semua estimasi parameter regresi (koefisien)
$g(X_i)$:	Fungsi logit dari X_i
\ln	:	Logaritma natural
$L(\beta)$:	Fungsi <i>log-likelihood</i>
$\ell(\beta, y_i)$:	Fungsi <i>likelihood</i>
q^t	:	Vektor kemiringan (<i>gradient</i>) pada iterasi ke- t
$H(\beta^t)$:	Matriks <i>Hessian</i> pada iterasi ke- t
X^T	:	Transpos dari matriks X
V	:	Matriks diagonal dengan elemen $\pi(X_i)(1 - \pi(X_i))$
W	:	Statistik uji <i>Wald</i>
$SE(\hat{\beta}_j)$:	<i>Standar error</i> dari $\hat{\beta}_j$
ε	:	Nilai konvergensi iterasi
ψ	:	Nilai <i>Odds Ratio</i>
G	:	Statistik uji <i>Likelihood Ratio Test</i>
D	:	<i>Deviance</i> untuk uji kesesuaian model
χ^2	:	Distribusi <i>Chi-square</i>
R^2	:	Koefisien determinasi
VIF	:	<i>Variance Inflation Factor</i> untuk uji multikolinearitas
SS_{total}	:	Jumlah kuadrat total
SS_{error}	:	Jumlah kuadrat <i>error</i>
n	:	Jumlah total observasi
$n_{11}, n_{12},$ n_{21}, n_{22}	:	Elemen-elemen <i>confusion matrix</i>
$APER$:	<i>Apparent Error Rate</i> , proporsi klasifikasi salah

DAFTAR LAMPIRAN

Lampiran 1. Data Diagnosa Hipertensi.....	100
Lampiran 2. Data Diagnosa Hipertensi Setelah Pengkodean.....	101
Lampiran 3. <i>Syntax Tomek Links</i> dan Regresi Logistik.....	102
Lampiran 4. Hasil Perhitungan Manual Estimasi Parameter	110
Lampiran 5. Output Iterasi <i>Newton-Raphson</i> dengan R Studio.....	116
Lampiran 6. Output Iterasi <i>Newton-Raphson</i> dengan R Studio Lanjutan.....	116

ABSTRAK

Fachreza, Putri Aulia. 2025. **Reduksi *Imbalanced Data* Diagnosa Hipertensi dengan *Tomek Links* pada Regresi Logistik**. Skripsi. Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Ria Dhea Layla Nur Karisma, M.Si. (II) Erna Herawati, M.Pd.

Kata Kunci: Klasifikasi; *Imbalanced Data*; *Tomek Links*; Regresi Logistik; Hipertensi.

Masalah *imbalanced data* seringkali menghambat akurasi dalam proses klasifikasi, terutama dalam kasus diagnosis hipertensi, di mana jumlah data kelas minoritas jauh lebih sedikit dibandingkan kelas mayoritas. Penelitian ini bertujuan untuk membangun model regresi logistik yang akurat dengan mengatasi ketidakseimbangan data menggunakan metode *Tomek Links*. Metode ini bekerja dengan menghapus pasangan data terdekat dari kelas berbeda untuk mereduksi *noise* dan memperbaiki distribusi data. Setelah dilakukan *undersampling* dengan *Tomek Links*, model regresi logistik dibentuk dengan pendekatan *Maximum Likelihood Estimation* melalui metode iteratif *Newton-Raphson*. Evaluasi model dilakukan melalui pengujian multikolinearitas, uji signifikansi parameter, uji kesesuaian model, dan pengukuran ketepatan klasifikasi berdasarkan nilai *Apparent Error Rate* (APER). Hasil penelitian menunjukkan bahwa variabel jenis kelamin, konsumsi gula berlebih, lemak berlebih, dan usia secara signifikan memengaruhi kemungkinan seseorang menderita hipertensi. Model akhir menghasilkan tingkat akurasi sebesar 89,5%. Penelitian ini menunjukkan bahwa kombinasi metode *Tomek Links* dan regresi logistik dapat menjadi pendekatan efektif dalam menangani *imbalanced data* pada diagnosa hipertensi.

ABSTRACT

Fachreza, Putri Aulia. 2025. **Reduction of Imbalanced Hypertension Diagnosis Data with Tomek Links in Logistic Regression**. Thesis. Department of Mathematics, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Advisor: (I) Ria Dhea Layla Nur Karisma, M.Si. (II) Erna Herawati, M.Pd.

Keywords: Classification; Imbalanced Data; Tomek Links; Logistic Regression; Hypertension.

The problem of imbalanced data often hampers accuracy in the classification process, especially in the case of hypertension diagnosis, where the amount of minority class data is much less than the majority class. This study aims to build an accurate logistic regression model by overcoming data imbalance using the Tomek Links method. This method works by removing the closest pair of data from different classes to reduce noise and improve data distribution. After undersampling with Tomek Links, a logistic regression model is formed using the Maximum Likelihood Estimation approach through the Newton-Raphson iterative method. Model evaluation was conducted through multicollinearity testing, parameter significance testing, model fit testing, and measurement of classification accuracy based on Apparent Error Rate (APER) values. The results showed that the variables of gender, excess sugar consumption, excess fat, and age significantly influenced the likelihood of a person suffering from hypertension. The final model produced an accuracy rate of 89,5%. This study shows that the combination of Tomek Links method and logistic regression can be an effective approach in handling imbalanced data in hypertension diagnosis.

مستخلص البحث

فاخريزا، بوتري أوليا. ٢٠٢٥. تقليل بيانات تشخيص ارتفاع ضغط الدم غير المتوازن باستخدام روابط توميك في الانحدار اللوجستي. البحث العلمي. قسم الرياضيات، كلية العلوم والتكنولوجيا جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج.
املشرفة: (١) ريا ديا ليلي نور حرهما، الماجستير في العلوم. (٢) إيرنا هيراواتي، الماجستير في التعليم اللغة العربية.

الكلمات الأساسية: التصنيف؛ البيانات غير المتوازنة؛ روابط توميك؛ الانحدار اللوجستي؛ ارتفاع ضغط الدم

إن مشكلة البيانات غير المتوازنة غالباً ما تعيق الدقة في عملية التصنيف، وخاصة في حالة تشخيص ارتفاع ضغط الدم، حيث يكون عدد المصابين (فئة الأقلية) أقل بكثير من عدد غير المصابين (فئة الأغلبية). هدفت هذه الدراسة إلى بناء نموذج انحدار لوجستي ثنائي دقيق من خلال التغلب على اختلال توازن البيانات باستخدام طريقة *Tomek Links*. عملت هذه الطريقة عن طريق إزالة أقرب زوج من البيانات من فئات مختلفة لتقليل الضوضاء وتحسين توزيع البيانات. بعد أخذ العينات الناقصة باستخدام روابط توميك، تم تشكيل نموذج الانحدار اللوجستي باستخدام نهج تقدير الاحتمالية القصوى من خلال طريقة نيوتن-رافسون التكرارية. تم إجراء تقييم النموذج من خلال اختبار التعدد الخطي، واختبار أهمية المعلمات، واختبار ملاءمة النموذج، وقياس دقة التصنيف بناءً على قيم معدل الخطأ الظاهري (APER). وأظهرت النتائج أن متغيرات الضغط الانقباضي والضغط الانبساطي أثرت بشكل كبير على احتمالية إصابة الشخص بارتفاع ضغط الدم. حقق النموذج النهائي معدل دقة قدره ٨٩,٥%، مما يشير إلى تحسن أداء التصنيف بعد تقليل البيانات باستخدام *Tomek Links*. ظهرت هذه الدراسة أن الجمع بين روابط توميك وطريقة الانحدار اللوجستي يمكن أن يكون نهجاً فعالاً في التعامل مع البيانات غير المتوازنة في تشخيص ارتفاع ضغط الدم.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Matematika memiliki peran signifikan di era modern yang didukung oleh kemajuan teknologi dan informasi ini. Matematika merupakan ilmu dasar atau fondasi dari berbagai disiplin ilmu. Kemajuan dalam bidang teknologi seperti data sains (*science data*), data besar (*big data*), pembelajaran mesin (*mechine learning*), *deep learning*, dan kecerdasan buatan (*artificial intelligence*) semakin memperkuat peran matematika dalam inovasi dan perkembangan ilmu pengetahuan. Data yang dihasilkan dari kemajuan teknologi tersebut sangatlah masif, kompleks, dan beragam sehingga membutuhkan pendekatan matematis untuk dianalisis, dimodelkan, dan diolah menjadi informasi yang bermakna yang dapat digunakan untuk menarik kesimpulan dan mengambil keputusan. Analisis dan pengolahan data dapat membawa perubahan signifikan dalam berbagai bidang, seperti pada bidang ekonomi, bisnis, kesehatan, ataupun penelitian ilmiah.

Analisis data membantu dalam mempersiapkan data yang berkualitas untuk data mining. Data mining merupakan bagian dari analisis data yang lebih luas. Data mining digunakan untuk mengekstraksi pola, hubungan, atau informasi yang penting dan bermakna dari sekumpulan data besar menggunakan *mechine learning*, metode statistik, dan basis data. Teknik yang digunakan juga tergantung pada jenis data dan tujuan analisis. Salah satu teknik analisis dalam data mining adalah klasifikasi. Klasifikasi memiliki peran penting dalam pengelompokan data. Klasifikasi dapat mengelompokkan data ke dalam kategori atau kelas yang telah ditentukan

berdasarkan karakteristik tertentu. Klasifikasi juga digunakan untuk menentukan suatu model, pola, atau fungsi tertentu yang dapat mengklasifikasikan data baru ke salah satu kelas yang telah ditentukan. Pengklasifikasian data terdiri dari dua jenis yaitu klasifikasi data ke dalam dua kelas (biner) dan klasifikasi data ke lebih dari dua kelas (multikelas).

Berbagai metode umum yang sering dimanfaatkan pada data mining dalam klasifikasi data, di antaranya termasuk *logistic regression*, *Support Vector Machine* (SVM), *K-Nearest Neighbors* (K-NN), *decision tree*, *naïve bayes*, *random forest*, dan berbagai metode lainnya. Berdasarkan berbagai metode klasifikasi yang ada, regresi logistik adalah salah satu metode yang paling populer karena model yang dihasilkan dapat direpresentasikan dengan jelas dan ringkas. Penggunaan analisis regresi logistik adalah untuk menemukan keterkaitan antara variabel dependen berskala ordinal atau nominal dengan memiliki kategori sebanyak dua (dikotomus) atau memiliki kategori lebih dari dua (polikotomus) dengan beberapa variabel independen (Agresti, 2002). Kelebihan regresi logistik yaitu dapat digunakan pada data yang memiliki berbagai jenis variabel dependen, termasuk biner, multinomial, dan ordinal.

Regresi logistik biner disebut dengan regresi logistik dengan variabel dependennya memiliki dua kategori (Hosmer & Lemeshow, 2000). Regresi logistik biner memiliki keunggulan yaitu konsep yang sederhana dan fleksibel atau dapat bekerja dengan variabel independen kategori maupun kontinu (Antipov & Elenea, 2009). Penelitian dengan menggunakan regresi logistik biner dilakukan oleh Ramandhani dkk., (2017) yaitu penggunaan metode *bootstrap aggregating* pada regresi logistik biner untuk menganalisis ketepatan klasifikasi kesejahteraan rumah

tangga di Kota Pati dan menghasilkan tingkat akurasi sebesar 79,87%. Selain itu, Wan dkk., (2019) juga melakukan penelitian dengan judul “*Modeling Provision of Disaster Mutual Assistance by Electricity Utilities Using Logistik Regression*”. Penelitian tersebut menghasilkan tingkat akurasi sebesar 75,2%. Akan tetapi, regresi logistik biner akan lebih sensitif terhadap *underfitting* atau *overfitting* apabila data tidak seimbang (Antipov & Elenea, 2009).

Permasalahan dalam klasifikasi yang biasanya terjadi salah satunya adalah tidak seimbangnya jumlah kelas dalam variabel dependen. Struktur ketidakseimbangan data ini biasa disebut dengan *imbalanced data*. *Imbalanced data* terjadi ketika ditemukannya kesenjangan antar kelas, dimana terdapat kelas dengan jumlah yang lebih banyak dari kelas lain. Kelas data dengan jumlah lebih besar yaitu kelas mayoritas, sementara untuk kelas minoritas memiliki jumlah yang lebih sedikit (Pratiwi, 2018). Nilai akurasi yang tinggi seringkali tercapai dalam klasifikasi tanpa adanya penanganan untuk data yang tidak seimbang karena hanya memprioritaskan data mayoritas, sedangkan data minoritas sering dianggap sebagai data yang jarang atau tidak disengaja. Selain itu, data yang tidak seimbang dapat mempengaruhi efektivitas model dan tingkat akurasi dalam klasifikasi (Qiong, 2016). Pendekatan untuk mengatasi masalah data yang tidak seimbang dapat dilakukan teknik *sampling*.

Teknik *sampling* yang dapat digunakan untuk mereduksi data pada *imbalanced data* adalah *Tomek Links*. *Tomek Links* bekerja dengan cara mengurangi jumlah sampel dari kelas mayoritas. Kelebihan dari teknik *Tomek Links* adalah mengurangi jumlah kelas mayoritas dengan menghilangkan *noise*. Penelitian terdahulu dengan menggunakan metode *Tomek Link* untuk mengatasi data yang

tidak seimbang telah dilakukan oleh Dewi dkk., (2023) yaitu analisis sentimen *review* pada aplikasi Satu Sehat *Mobile* dengan metode SVM. Pada penelitian tersebut, dengan menggunakan metode *Tomek Links*, nilai akurasi meningkat sebesar 0,489%. Kumalasanti dan Aprilianti (2024) juga telah melakukan penelitian dengan metode *Tomek Links* yaitu analisis sentimen ulasan pada aplikasi kalender Bali menggunakan KNN. Penggunaan *Tomek Links* meningkatkan nilai akurasi sebesar 1,2% pada data sebelum pandemi dan sebesar 0,6% pada data setelah pandemi. Metode *Tomek Links* dapat digunakan untuk mereduksi data klasifikasi diagnosa hipertensi yang tidak seimbang.

Laporan *World Health Organization* (WHO) mengenai hipertensi pada tahun 2023 memprediksi bahwa secara global jumlah masyarakat yang menderita hipertensi meningkat kurang lebih dua kali lipat dalam tiga dekade terakhir. Pada tahun 1990, masyarakat yang menderita hipertensi mencapai 650 juta dan pada tahun 2019 menjadi 1,3 miliar. Menurut Kementerian Kesehatan Republik Indonesia (KEMENKES RI), hipertensi sendiri merupakan permasalahan medis dan kesehatan masyarakat yang sangat krusial baik pada tingkat global, regional, maupun nasional. Berdasarkan Survei Kesehatan Indonesia (SKI) 2023 dalam (Kemenkes RI, 2024), hipertensi termasuk dalam penyakit tidak menular (PTM) yang memiliki tingkat prevalensi tinggi karena memiliki *burden of disease* yang cukup tinggi serta perannya sebagai faktor risiko utama penyakit komplikasi jantung, pembuluh darah, bahkan kematian.

Hipertensi dapat muncul saat tekanan dalam pembuluh darah terlampaui tinggi. Tekanan darah diukur dengan menggunakan dua angka yaitu tekanan sistolik dan diastolik. Tekanan sistolik mengindikasikan seberapa besar tekanan yang

dihasilkan dalam pembuluh darah saat jantung sedang memompa darah atau berkontraksi. Sedangkan tekanan diastolik menunjukkan tekanan yang terjadi dalam pembuluh darah saat jantung sedang beristirahat di antara dua detak jantung. Pada fase ini, jantung mengisi ulang darah dan bersiap untuk memompa lagi. Dengan kata lain, penentuan status hipertensi didasarkan pada kenaikan tekanan darah sistolik yang mencapai lebih dari 140mmHg dan tekanan pada darah diastolik yang mencapai lebih dari 90mmHg.

Menurut WHO, faktor yang meningkatkan risiko hipertensi diantaranya usia, genetika, obesitas, kurang aktivitas fisik, konsumsi garam serta gula berlebih, merokok, dan konsumsi alkohol. Allah SWT telah berfirman dalam Al-Baqarah ayat 168 sebagai berikut:

يَا أَيُّهَا النَّاسُ كُلُوا مِمَّا فِي الْأَرْضِ حَلَالًا طَيِّبًا وَلَا تَتَّبِعُوا خُطُوَاتِ الشَّيْطَانِ إِنَّهُ لَكُمْ عَدُوٌّ مُّبِينٌ ﴿١٦٨﴾

Artinya: *“Wahai manusia, makanlah sebagian (makanan) yang halal dan baik di bumi, dan janganlah mengikuti jejak langkah setan, karena sesungguhnya ia adalah musuh yang nyata bagimu.”*

Berdasarkan QS. Al-Baqarah ayat 168, Allah SWT menghendaki untuk makan dan minum yang halal dan baik. Sebagai seorang mukmin, pentingnya memilih dan mengonsumsi makanan serta minuman yang halal dan baik untuk dikonsumsi merupakan upaya yang dapat mencegah dan menangani terjadinya hipertensi. Makanan dan minuman yang halal tidak semuanya secara langsung dapat dianggap baik. Kehalalan makanan dan minuman tidak selalu sesuai untuk setiap orang sehingga tergantung pada kondisi orang tersebut. Beberapa makanan dan minuman halal dan baik bagi seseorang dengan kondisi kesehatan tertentu,

tetapi kurang baik bagi orang lain meskipun halal. Sebagian makanan yang terlihat baik namun tidak mengandung kandungan gizi yang seimbang. Oleh karena itu, makanan dan minuman yang ditekankan adalah yang tidak hanya halal menurut syariat, tetapi juga baik. Makanan dan minuman yang halal dan baik akan bermanfaat, sehat, dan tidak membahayakan tubuh.

Menurut Tafsir as-Sa'di, Allah SWT memerintah orang-orang mukmin maupun yang kafir untuk mengonsumsi makanan yang ada di bumi seperti sayur-sayuran, buah-buahan, biji-bijian, dan hewan yang halal (Tafsir web, 2024). Dalam kaitannya dengan hipertensi, prinsip tersebut sejalan dengan anjuran untuk memiliki pola makan sehat. Upaya dalam rangka menjalankan pola makan yang sehat dapat berupa mengonsumsi sayur dan buah-buahan serta menghindari konsumsi gula, garam, dan lemak jenuh berlebihan. Makanan yang halal dan baik pasti mengandung nutrisi dan gizi seimbang yang dibutuhkan oleh tubuh. Gula dan garam adalah halal, akan tetapi tidak sesuai dengan prinsip baik apabila dikonsumsi secara berlebihan karena dapat berdampak buruk pada tubuh dan dapat menyebabkan hipertensi. Ayat tersebut juga menekankan untuk tidak terpengaruh terhadap godaan syaitan yang dapat mengarahkan untuk berperilaku buruk, seperti mengonsumsi makanan atau minuman haram dan tidak sehat, merokok, dan mengonsumsi alkohol yang apabila dikonsumsi dalam jangka waktu yang panjang akan memberikan pengaruh buruk pada kondisi tubuh.

Penelitian mengenai hipertensi telah dilakukan oleh Misna dkk., (2018) dengan metode regresi logistik biner dengan tingkat akurasi sebesar 77,6%. Sahila dkk., (2024) juga melakukan penelitian yaitu analisis klasifikasi menggunakan regresi logistik biner dan algoritma *naïve bayes* pada penyakit hipertensi. Penelitian

tersebut menghasilkan tingkat akurasi sebesar 93,33% dengan metode regresi logistik biner dan 63,64% dengan metode *naïve bayes*. Hal tersebut menunjukkan bahwa metode regresi logistik biner lebih baik dalam mengklasifikasikan hipertensi. Berdasarkan hasil skrining *May Measurement Month* (MMM) pada tahun 2018 di Indonesia, sekitar 47,6% masyarakat menyadari bahwa telah mengidap hipertensi (Turana dkk., 2018). Dengan demikian, perlu adanya upaya untuk meningkatkan kesadaran masyarakat akan status hipertensinya sebagai langkah awal upaya pencegahan timbulnya penyakit komplikasi seperti penyakit jantung, pembuluh darah, dan yang lainnya. Deteksi dini hipertensi sangat penting untuk memulai pengobatan seawal mungkin guna mencapai kondisi yang terkendali dan mencegah kemungkinan komplikasi lebih lanjut. Berdasarkan latar belakang yang ada, maka dilakukan penelitian mengenai reduksi *imbalanced data* diagnosa hipertensi. Penelitian ini memanfaatkan metode statistik yaitu regresi logistik dalam klasifikasi dan metode *Tomek Links* untuk mengatasi masalah *imbalanced data*.

1.2 Rumusan Masalah

Berdasarkan penjabaran latar belakang, rumusan masalah pada penelitian ini adalah:

1. Bagaimana model regresi logistik pada reduksi *imbalanced data* diagnosa hipertensi dengan *Tomek Links*?
2. Bagaimana tingkat akurasi model regresi logistik pada reduksi *imbalanced data* diagnosa hipertensi dengan *Tomek Links*?

1.3 Tujuan Penelitian

Adapun dengan merujuk pada rumusan masalah, tujuan penelitian ini adalah:

1. Mendapatkan model regresi logistik pada reduksi *imbalanced data* diagnosa hipertensi dengan *Tomek Links*.
2. Mengetahui tingkat akurasi model regresi logistik pada reduksi *imbalanced data* diagnosa hipertensi dengan *Tomek Links*.

1.4 Manfaat Penelitian

Manfaat penelitian yang diharapkan adalah:

1. Bagi Penulis
Meningkatkan dan menambah pengetahuan serta wawasan mengenai reduksi *imbalanced data* diagnosa hipertensi dengan metode *Tomek Links* pada regresi logistik.
2. Bagi Program Studi Matematika
Penelitian dapat memberikan kontribusi pada pengembangan ilmu pengetahuan yang bermanfaat bagi pembaca atau mahasiswa sebagai referensi mengenai implementasi metode regresi logistik dan *Tomek Links*.
3. Bagi Dinas Kesehatan
Memberikan informasi mengenai diagnosa penderita hipertensi sehingga dapat melakukan perencanaan dan penanganan yang efektif untuk membantu para penderita hipertensi.

1.5 Batasan Masalah

Adapun batasan dari penelitian ini adalah:

1. Data yang digunakan adalah data diagnosa hipertensi tahun 2022-2023.
2. Metode estimasi parameter yang digunakan adalah *Maximum Likelihood Estimation* (MLE) secara *Newton-Raphson*.
3. Perhitungan tingkat akurasi model berdasarkan nilai APER.

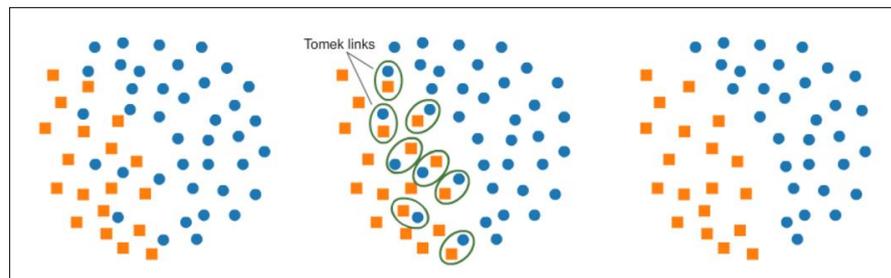
BAB II

KAJIAN TEORI

2.1 Teori Pendukung

2.1.1 *Imbalanced Data*

Imbalanced data terjadi ketika satu di antara kelas memiliki total data yang jauh lebih banyak daripada total data dari kelas yang lain. Kelas dengan jumlah lebih besar disebut kelas mayoritas dan kelas minoritas dengan jumlah lebih kecil. alternatif cara untuk menanggulangi masalah *imbalanced data* yaitu dapat dilakukan pendekatan *sampling* pada data. Dengan melakukan pendekatan *sampling*, diharapkan masalah *imbalanced data* akan semakin berkurang atau semakin kecil sehingga dapat menghasilkan hasil klasifikasi yang akurat (Solberg & Solberg, 1996). Metode *sampling* yang dapat digunakan diantaranya adalah *Tomek Links*. Ilustrasi metode *Tomek Links* disajikan pada gambar (2.1) (Choirunnisa, 2019).



Gambar 2.1 Ilustrasi *Tomek Links*

Tomek Links bertujuan untuk membersihkan data, dimana sampel dari kelas minoritas akan dihilangkan karena dianggap sebagai *noise*. Proses kerja teknik *Tomek Links* terhenti jika sampel *noise* sudah tidak ada lagi yang dapat ditemukan

pada data kelas mayoritas. Teknik ini digunakan untuk menemukan sampel data yang diinginkan dari kelas mayoritas yang memiliki jarak *euclidean* terendah dengan data kelas minoritas, kemudian menghapusnya.

Tetangga yang paling terdekat diperoleh mengacu pada jarak *euclidean* antara sepasang data. Perhitungan jarak *euclidean* $d(x,y)$ diperoleh dengan persamaan 2.1 sebagai berikut:

$$d(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \quad (2.1)$$

Teknik *Tomek Links* mampu dijelaskan seperti yang berikut ini: Jika disediakan sampel sebanyak dua yang masing-masing berasal dari kelas yang tidak sama, dengan rentang antara x dan y dapat dikatakan sebagai $d(x,y)$. Pasangan sampel (x,y) dikenal dengan *Tomek Links* apabila tidak ditemukan sampel z lain yang memiliki jarak yang lebih dekat dengan x daripada y atau jarak yang lebih dekat dengan y daripada x . Hal tersebut bisa disebut juga harus memenuhi $d(x,z)$ kurang dari $d(x,y)$ atau $d(y,z)$ kurang dari $d(y,x)$ (Batista, Bazzan, & Monard, 2003). Ilustrasi data metode *Tomek Links* disajikan pada tabel (2.1) (Sain, 2015).

Tabel 2.1 Data Ilustrasi *Tomek Links*

No	X_1	X_2	Y
1	2	2	0
2	3	6	0
3	4	2	0
4	6	5	0
5	1	2	1
6	1	4	1
7	3	1	1
8	3	3	1
9	3	4	1
10	4	4	1
11	5	1	1
12	5	3	1
13	5	6	1
14	6	2	1
15	6	4	1
16	2	3	0
17	2,5	2,5	0
18	2	1,5	0

Berdasarkan tabel (2.1) diketahui bahwa jumlah data kelas mayoritas adalah 11 ($Y = 1$), sedangkan jumlah data kelas minoritas adalah 7 ($Y = 0$). Diperkirakan titik yang saling berdekatan adalah data ke - 8, ke - 9, dan data ke - 17, kemudian titik tersebut akan dilakukan perhitungan menggunakan jarak *Euclidean* dengan persamaan (2.1) sebagai berikut:

Data ke - 8 dan ke - 17:

$$d\left(\begin{bmatrix} 3 \\ 3 \end{bmatrix}, \begin{bmatrix} 2,5 \\ 2,5 \end{bmatrix}\right) = \sqrt{(3 - 2,5)^2 + (3 - 2,5)^2} = \sqrt{0,5} \approx 0,707$$

Data ke - 8 dan ke - 9:

$$d\left(\begin{bmatrix} 3 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ 4 \end{bmatrix}\right) = \sqrt{(3 - 3)^2 + (3 - 4)^2} = \sqrt{1} = 1$$

Data ke - 9 dan ke - 17:

$$d\left(\begin{bmatrix} 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 2,5 \\ 2,5 \end{bmatrix}\right) = \sqrt{(3 - 2,5)^2 + (4 - 2,5)^2} = \sqrt{2,5} \approx 1,58$$

Sehingga diketahui bahwa $d\left(\begin{bmatrix} 3 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ 4 \end{bmatrix}\right) = 1 > d\left(\begin{bmatrix} 3 \\ 3 \end{bmatrix}, \begin{bmatrix} 2,5 \\ 2,5 \end{bmatrix}\right) = 0,707$

atau $d\left(\begin{bmatrix} 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 2,5 \\ 2,5 \end{bmatrix}\right) = 1,58 > d\left(\begin{bmatrix} 3 \\ 3 \end{bmatrix}, \begin{bmatrix} 2,5 \\ 2,5 \end{bmatrix}\right) = 0,707$. Dengan demikian dapat disimpulkan bahwa titik y_8 dan y_{17} dianggap sebagai kejadian *Tomek Links* sebab keduanya tidak mencapai kriteria dari pengertian *Tomek Links* sedemikian hingga titik-titik tersebut akan dihilangkan atau dihapus seterusnya untuk memudahkan dalam pengamatan terhadap data mayoritas lainnya. *Tomek Links* akan menghentikan sistem dan prosesnya jika tidak ada lagi sampel *noise* yang ditemukan pada kelas mayoritas.

2.1.2 Regresi Logistik

Metode analisis yang dimanfaatkan untuk menemukan keterkaitan antara variabel dependen yang memiliki skala nominal atau ordinal dengan kategori berjumlah dua (dikotomus) atau dengan kategori berjumlah lebih dari dua (polikotomus) dengan beberapa variabel independen adalah regresi logistik (Hosmer & Lemeshow, 2000). Metode ini bertujuan untuk memahami bagaimana variabel independen memiliki keterkaitan dengan memprediksi kemungkinan atau probabilitas terjadinya suatu kejadian pada variabel dependen yang bersifat kategorikal. Model regresi logistik terbagi menjadi beberapa jenis, yaitu regresi logistik biner, ordinal, dan multinomial.

Regresi logistik biner merupakan model statistik yang diperuntukkan untuk mengkaji hubungan antara sejumlah variabel independen dengan satu variabel dependen kategorik. Variabel dependen yang dimaksud memiliki sifat dikotomus atau biner, yaitu terbagi dalam dua kategori. Sedangkan variabel independennya dapat berupa satu atau lebih variabel dengan skala data rasio, interval, nominal, atau ordinal. Variabel dependen tersebut memiliki dua kategori nilai, yaitu 1 untuk menunjukkan adanya suatu karakteristik dan 0 untuk menunjukkan ketiadaan karakteristik tersebut. Variabel dependen biner dinyatakan dengan notasi Y . Dengan demikian, variabel dependen Y yang bernilai 1 ditunjukkan untuk kejadian “sukses” dan 0 ditunjukkan untuk kejadian “gagal”. Distribusi probabilitas dari variabel Y untuk $y = 1$ adalah $P(Y = 1|X = x) = \pi(x)$ dan $y = 0$ adalah $P(Y = 0|X = x) = 1 - \pi(x)$ (Agresti, 2007).

Dalam regresi logistik, model yang mencakup banyak variabel independen disebut sebagai model multivariabel. Jika data hasil pengamatan atau observasi

sebanyak p variabel independen, maka variabel independennya adalah $x_1, x_2, x_3, \dots, x_p$. Model regresi logistik multivariabel dengan p variabel independen dijelaskan dalam persamaan (2.2) di bawah ini (Hosmer & Lemeshow, 2000):

$$\pi(X_i) = \frac{\exp(\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi})}{1 + \exp(\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi})} \quad (2.2)$$

dengan keterangan nilai $i = 1, 2, \dots, N$ dan $\pi(X_i)$ merupakan peluang sukses dengan nilai probabilitas $0 \leq \pi(X_i) \leq 1$. Persamaan (2.2) kemudian diubah menggunakan transformasi logit untuk menyederhanakan prediksi atau estimasi parameter regresi sebagai berikut:

$$\pi(X_i) = \frac{\exp(\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi})}{1 + \exp(\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi})}$$

$$[\pi(X_i)][1 + \exp(\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi})] = \exp(\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi})$$

$$\pi(X_i) + \pi(X_i)\exp(\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi}) = \exp(\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi})$$

$$\pi(X_i) = \exp(\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi}) - \pi(x_i)\exp(\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi})$$

$$= [1 - \pi(x_i)]\exp(\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi})$$

$$\frac{\pi(X_i)}{1 - \pi(X_i)} = \exp(\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi})$$

$$\ln \left[\frac{\pi(X_i)}{1 - \pi(X_i)} \right] = \ln [\exp(\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi})]$$

$$= \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi}$$

$$g(X_i) = \ln \left[\frac{\pi(X_i)}{1 - \pi(X_i)} \right] = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi} \quad (2.3)$$

$$\begin{aligned}
&= \sum_{j=0}^p \beta_j x_{ij} \\
&= X_i \beta
\end{aligned}$$

dengan $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \dots, \beta_p)$ dan $X_i = (1, x_{i1}, x_{i2}, x_{i3}, \dots, x_{ip})$ adalah nilai-nilai variabel independen ke - 1 sampai p dari observasi atau pengamatan ke - i dan $g(X_i)$ merupakan model logit. Model regresi logistik dapat diekspresikan dalam bentuk persamaan (2.4) seperti yang terlihat di bawah ini:

$$\pi(X_i) = \frac{\exp(g(X_i))}{1 + \exp(g(X_i))} \quad (2.4)$$

2.1.3 *Maximum Likelihood Estimator*

Estimasi parameter pada regresi logistik diterapkan dengan metode *Maximum Likelihood Estimator* (MLE) (Hosmer & Lemeshow, 2000). Salah satu keunggulannya adalah kemampuannya untuk digunakan pada model yang tidak linier, seperti regresi logistik. Metode MLE memberikan nilai untuk parameter yang tidak dikenal yang mengoptimalkan probabilitas untuk mendapatkan kumpulan data yang diobservasi. Penerapan metode ini diawali dengan membuat fungsi *likelihood*. Fungsi ini menggambarkan seberapa besar kemungkinan data yang diamati akan muncul jika memilih nilai tertentu untuk parameter yang tidak diketahui. Nilai dari parameter-parameter yang memaksimalkan atau membuat fungsi *likelihood* mencapai angka tertinggi disebut dengan MLE. Dengan demikian, hasil penaksiran parameter menggunakan MLE cenderung mendekati nilai parameter yang sebenarnya, sehingga menghasilkan estimasi yang lebih akurat (Hosmer & Lemeshow, 2000).

Apabila terdapat N sampel, maka jumlah sukses dinyatakan sebagai y_i dengan probabilitas sukses (π_i) dan probabilitas gagal ($1 - \pi_i$). y_i mengikuti distribusi *bernoulli*. Probabilitas $\pi(x)$ menunjukkan bahwa $Y = 1$ jika diberikan x dan dilambangkan sebagai $P(Y = 1|x)$. Sedangkan probabilitas ($1 - \pi(x)$) menunjukkan bahwa $Y = 0$ jika diberikan x dan dilambangkan sebagai $P(Y = 0|x)$. Sehingga fungsi probabilitas pasangan (X_i, y_i) untuk pengamatan ke - i dinyatakan dalam persamaan (2.5) sebagai berikut (Hosmer & Lemeshow, 2000):

$$\pi(Y_i = y_i) = \pi(X_i)^{y_i} [1 - \pi(X_i)]^{(1-y_i)} \quad (2.5)$$

dengan keterangan nilai $i = 1, 2, \dots, N$. Fungsi *likelihood* untuk pasangan (X_i, y_i) dalam regresi logistik dinyatakan pada persamaan (2.6) sebagai berikut:

$$\begin{aligned} \ell(\beta, y_i) &= \prod_{i=1}^N [\pi(Y_i = y_i)] \\ &= \prod_{i=1}^N [\pi(X_i)^{y_i} [1 - \pi(X_i)]^{(1-y_i)}] \end{aligned} \quad (2.6)$$

Persamaan (2.6) kemudian dimaksimumkan dan diperoleh fungsi *log-likelihood* sebagai berikut:

$$\begin{aligned} L(\beta, y_i) &= \ln \ell(\beta, y_i) \\ &= \ln \prod_{i=1}^N [\pi(X_i)^{y_i} [1 - \pi(X_i)]^{(1-y_i)}] \\ &= \sum_{i=1}^N [y_i \ln(\pi(X_i)) + (1 - y_i) \ln(1 - \pi(X_i))] \\ &= \sum_{i=1}^N \left[y_i \ln \left(\frac{e^{g(X_i)}}{1 + e^{g(X_i)}} \right) + (1 - y_i) \ln \left(1 - \frac{e^{g(X_i)}}{1 + e^{g(X_i)}} \right) \right] \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^N \left[y_i \left[\ln(e^{g(X_i)}) - \ln(1 + e^{g(X_i)}) \right] + (1 - y_i) \ln \left(\frac{1}{1 + e^{g(X_i)}} \right) \right] \\
&= \sum_{i=1}^N \left[y_i \left[\ln(e^{g(X_i)}) - \ln(1 + e^{g(X_i)}) \right] + (1 - y_i) \left[\ln(1) - \right. \right. \\
&\quad \left. \left. \ln(1 + e^{g(X_i)}) \right] \right] \\
&= \sum_{i=1}^N \left[y_i \left[\ln(e^{g(X_i)}) - \ln(1 + e^{g(X_i)}) \right] - (1 - y_i) \left[\ln(1 + \right. \right. \\
&\quad \left. \left. e^{g(X_i)}) \right] \right] \\
&= \sum_{i=1}^N \left[y_i \ln(e^{g(X_i)}) - y_i \ln(1 + e^{g(X_i)}) - \ln(1 + e^{g(X_i)}) + \right. \\
&\quad \left. y_i \ln(1 + e^{g(X_i)}) \right] \\
&= \sum_{i=1}^N \left[y_i \ln(e^{g(X_i)}) - \ln(1 + e^{g(X_i)}) \right] \\
&= \sum_{i=1}^N \left[y_i (g(X_i)) - \ln(1 + e^{g(X_i)}) \right]
\end{aligned}$$

Berdasarkan persamaan (2.3), untuk $g(X_i) = \ln \left[\frac{\pi(X_i)}{1 - \pi(X_i)} \right]$ diperoleh fungsi

log-likelihood:

$$\begin{aligned}
L(\beta) &= \sum_{i=1}^N \left[y_i (X_i \beta) - \ln(1 + e^{(X_i \beta)}) \right] \\
&= \sum_{i=1}^N y_i (X_i \beta) - \sum_{i=1}^N \ln(1 + e^{(X_i \beta)}) \\
&= \sum_{i=1}^N y_i \left(\sum_{j=0}^p x_{ij} \beta_j \right) - \sum_{i=1}^N \ln \left(1 + e^{\left(\sum_{j=0}^p x_{ij} \beta_j \right)} \right) \quad (2.7)
\end{aligned}$$

Langkah yang dilakukan untuk memperoleh estimasi parameter adalah dengan menghitung turunan pertama dari $L(\beta)$ secara parsial terhadap β_j .

Turunan pertama tersebut digunakan untuk memperoleh elemen-elemen vektor kemiringan (g). Fungsi *log-likelihood* pada persamaan (2.7) diturunkan terhadap β_j dengan $j = 0, 1, 2, \dots, p$ dan diperoleh bentuk secara umum sebagai berikut (Afifah, 2020):

$$\begin{aligned}
 \frac{\partial L(\beta)}{\partial \beta_j} &= \frac{\partial \left(\sum_{i=1}^N y_i (\sum_{j=0}^p x_{ij} \beta_j) - \sum_{i=1}^N \ln \left(1 + e^{\left(\sum_{j=0}^p x_{ij} \beta_j \right)} \right) \right)}{\partial \beta_j} \\
 &= \sum_{i=1}^N y_i (x_{ij}) - \sum_{i=1}^N x_{ij} \left(\frac{e^{\left(\sum_{j=0}^p x_{ij} \beta_j \right)}}{1 + e^{\left(\sum_{j=0}^p x_{ij} \beta_j \right)}} \right) \\
 &= \sum_{i=1}^N y_i (x_{ij}) - \sum_{i=1}^N x_{ij} (\pi(X_i)) \\
 &= \sum_{i=1}^N x_{ij} (y_i - \pi(X_i)) \tag{2.8}
 \end{aligned}$$

Hasil yang diperoleh pada turunan pertama secara parsial terhadap β masih dalam bentuk implisit. Metode yang digunakan untuk menyelesaikan permasalahan tersebut adalah metode numerik dengan menggunakan pendekatan iterasi *Newton-Raphson*. Metode *Newton-Raphson* dapat digunakan untuk menentukan nilai dari parameter β . Metode *Newton-Raphson* bekerja melalui proses iterasi atau algoritma berulang untuk menentukan titik yang memaksimalkan fungsi $L(\beta)$ dan digunakan apabila fungsi $L(\beta)$ tidak linear (Agresti, 2007). Apabila β^0 adalah nilai awal dari β pada proses iterasi ke - t dengan ($t = 0, 1, 2, \dots$), maka dapat dimisalkan bahwa $\beta = \beta^{(t+1)}$ dan diperoleh persamaan sebagai berikut (Agresti, 1990):

$$\beta^{(t+1)} = \beta^{(t)} - (H^{(t)})^{-1} q^{(t)} \quad (2.9)$$

dengan $q^{(t)}$ adalah vektor kemiringan (*slope*) yang diperoleh dari hasil turunan parsial pertama sebagai berikut:

$$q^{(t)} = \begin{bmatrix} \frac{\partial L(\beta)}{\partial \beta_0} \\ \frac{\partial L(\beta)}{\partial \beta_1} \\ \vdots \\ \frac{\partial L(\beta)}{\partial \beta_j} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^N y_i(x_{i0}) - \sum_{i=1}^N x_{i0}(\pi(X_i)) \\ \sum_{i=1}^N y_i(x_{i1}) - \sum_{i=1}^N x_{i1}(\pi(X_i)) \\ \vdots \\ \sum_{i=1}^N y_i(x_{ij}) - \sum_{i=1}^N x_{ij}(\pi(X_i)) \end{bmatrix}$$

Sehingga pada iterasi ke - t dapat dituliskan sebagai berikut:

$$\begin{aligned} q^{(t)} &= X^T Y - X^T \pi(X_i) \\ &= X^T (Y - \pi(X_i)) \\ &= X^T W \end{aligned} \quad (2.10)$$

Bentuk matriks dari persamaan (2.10) dinyatakan sebagai berikut:

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{12} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Np} \end{bmatrix} \begin{bmatrix} y_1 - \pi(x_1) \\ y_2 - \pi(x_2) \\ \vdots \\ y_N - \pi(x_N) \end{bmatrix} = X^T W$$

Kemudian fungsi $L(\beta)$ pada persamaan (2.7) diturunkan parsial untuk kedua kalinya terhadap β_j dengan $j = 0, 1, 2, \dots, p$ dan dinyatakan secara umum sebagai berikut (Afifah, 2020):

$$\frac{\partial^2 L(\beta)}{(\partial \beta_j)^2} = \frac{\partial(\partial L(\beta))}{\partial \beta_j \partial \beta_j} = \frac{\partial(\sum_{i=1}^N x_{ij}(y_i - \pi(X_i)))}{\partial \beta_j}$$

$$\begin{aligned}
&= -\sum_{i=1}^N [x_{ij}x_{ij}\pi(X_i) - x_{ij}x_{ij}(\pi(X_i))^2] \\
&= -\sum_{i=1}^N x_{ij}x_{ij}[\pi(X_i)(1 - \pi(X_i))] \\
&= -\sum_{i=1}^N x_{ij}^2[\pi(X_i)(1 - \pi(X_i))] \tag{2.11}
\end{aligned}$$

Elemen-elemen dari turunan parsial kedua dari fungsi $L(\beta)$ akan membentuk matriks *hessian* sebagai berikut:

$$H(\beta^{(t)}) = \begin{bmatrix} \frac{\partial^2 L(\beta)}{(\partial \beta_0)^2} & \frac{\partial^2 L(\beta)}{\partial \beta_0 \partial \beta_1} & \cdots & \frac{\partial^2 L(\beta)}{\partial \beta_0 \partial \beta_p} \\ \frac{\partial^2 L(\beta)}{\partial \beta_1 \partial \beta_0} & \frac{\partial^2 L(\beta)}{\partial \beta_1 \partial \beta_1} & \cdots & \frac{\partial^2 L(\beta)}{\partial \beta_1 \partial \beta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 L(\beta)}{\partial \beta_p \partial \beta_0} & \frac{\partial^2 L(\beta)}{\partial \beta_p \partial \beta_1} & \cdots & \frac{\partial^2 L(\beta)}{(\partial \beta_p)^2} \end{bmatrix}$$

Kemudian matriks diagonal dengan elemen-elemennya merupakan nilai dari $\pi(X_i)(1 - \pi(X_i))$ dimisalkan sebagai V . Sedangkan bentuk matriks dari persamaan (2.11) dinyatakan sebagai berikut:

$$-\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{12} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Np} \end{bmatrix} \begin{bmatrix} \pi(x_1)(1 - \pi(x_1)) & 0 & \cdots & 0 \\ 0 & \pi(x_2)(1 - \pi(x_2)) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \pi(x_p)(1 - \pi(x_p)) \end{bmatrix}$$

$$\begin{bmatrix} 1 & x_{11} & \cdots & x_{N1} \\ 1 & x_{12} & \cdots & x_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1p} & \cdots & x_{Np} \end{bmatrix} = -X^T V X$$

Dengan demikian dapat dituliskan sebagai:

$$H(\beta^{(t)}) = -X^T V X \tag{2.12}$$

Sehingga didapatkan hasil estimasi parameter menggunakan proses iterasi dengan memperbarui proses iterasi terhadap β secara iteratif. Iterasi akan berhenti ketika nilai taksiran yang didapatkan telah konvergen.

$$\beta^{(t)} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \text{ dengan kriteria } \beta^{(t+1)} \cong \beta^{(t)} \text{ atau } \|\beta^{(t+1)} - \beta^{(t)}\| < \varepsilon$$

Dengan ε adalah tingkat ketelitian dalam bilangan positif, misalkan $\varepsilon = 0,0001$ dan t adalah banyaknya iterasi. Sehingga berdasarkan persamaan (2.9) estimasi parameter regresi logistik sebagai berikut:

$$\begin{aligned} \beta^{(t+1)} &= \beta^{(t)} - (H^{(t)})^{-1} q^{(t)} \\ &= \beta^{(t)} - (-X^T V^{(t)} X)^{-1} X^T W^{(t)} \\ &= \beta^{(t)} + (X^T V^{(t)} X)^{-1} X^T W^{(t)} \end{aligned} \quad (2.13)$$

Estimasi parameter pertama memerlukan $\beta^{(0)}$ atau nilai tebakan awal untuk perhitungannya, $\beta^{(0)}$ dihitung dengan menggunakan *Ordinary Least Square* (OLS). Metode tersebut hanya digunakan untuk menghitung $\beta^{(0)}$ sebagai nilai tebakan awal sebagai berikut:

$$\beta^{(0)} = (X^T X)^{-1} X^T Q \quad (2.14)$$

Dengan Q adalah vektor kolom dengan panjang N yang dihitung dari data observasi dan mempunyai elemen $Q_i = \log \frac{y_i}{n_i - y_i}$ sehingga diperoleh matriks $\beta^{(0)}$ (Fatah & Mahmood, 2016).

2.1.4 Pengujian Parameter

Pengujian parameter dijalankan guna menentukan adakah variabel-variabel independen signifikan dengan variabel dependen. Pengujian parameter terbagi menjadi dua, yaitu uji serentak atau simultan yang mengevaluasi signifikansi variabel independen keseluruhan terhadap variabel dependen dan uji parsial yang mengevaluasi signifikansi tiap-tiap variabel independen secara terpisah.

1. Uji Simultan

Likelihood ratio test adalah sebuah pengujian signifikansi yang dilakukan secara serentak. Uji ini bertujuan untuk mengidentifikasi secara menyeluruh bagaimana variabel-variabel saling berhubungan dan berpengaruh dalam suatu model. *Likelihood ratio test* digunakan dalam uji signifikansi parameter dengan cara serentak. Prinsip dasar uji *G* atau *likelihood ratio test* adalah perbandingan antara nilai observasi dari variabel dependen dengan nilai dugaan yang didapatkan dari model dengan dan tanpa variabel independen. Model tersebut diantaranya yaitu *fitted model* (model dengan variabel independen) dan *saturated model* (model tanpa variabel independen atau hanya β_0) (Hosmer & Lemeshow, 2000). Perbandingan nilai yang diobservasi dengan nilai prediksi didasarkan pada fungsi *log-likelihood* pada persamaan (2.6) dan ditulis pada persamaan (2.15) sebagai berikut:

$$D = -2 \ln \left[\frac{(\text{likelihood of the fitted model})}{(\text{likelihood of the saturated model})} \right] \quad (2.15)$$

Dengan mengoperasikan persamaan (2.6) dan persamaan (2.15) maka didapatkan *deviance* (D) pada persamaan (2.16) sebagai berikut:

$$D = -2 \sum_{i=1}^N \left[y_i \ln \left(\frac{\hat{\pi}(X_i)}{y_i} \right) + (1 - y_i) \ln \left(\frac{1 - \hat{\pi}(X_i)}{1 - y_i} \right) \right] \quad (2.16)$$

Tujuan menilai signifikansi untuk mengetahui apakah suatu variabel independen (variabel yang memprediksi atau mempengaruhi) benar-benar berkontribusi secara signifikan terhadap model, maka dilakukan perbandingan antara nilai deviasi dengan dan tanpa variabel tersebut. Pengaruh variabel independen diperoleh dengan menghitung selisih nilai deviasi saat variabel tersebut dimasukkan dan saat variabel tersebut tidak dimasukkan ke dalam model. Sehingga diperoleh persamaan (2.17) sebagai berikut:

$$G = D (\text{dengan variabel}) - D (\text{tanpa variabel}) \quad (2.17)$$

Karena nilai deviasi didasarkan pada fungsi *log-likelihood*, maka diperoleh nilai G pada persamaan (2.18) sebagai berikut:

$$G = -2 \ln \left[\frac{(\text{likelihood tanpa variabel})}{(\text{likelihood dengan variabel})} \right] \quad (2.18)$$

Hipotesis uji G (Hosmer & Lemeshow, 2000):

$$H_0: \beta_0 = \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1: \text{Minimal terdapat satu } \beta_j \neq 0 \text{ dengan } j = 0, 1, 2, \dots, p$$

Penggunaan statistik uji pada hipotesis ini ditulis pada persamaan (2.19) sebagai berikut:

$$G = -2 \ln \left(\frac{\left(\frac{n_0}{n}\right)^{n_0} \left(\frac{n_1}{n}\right)^{n_1}}{\prod_{i=1}^N [\hat{\pi}(X_i)^{y_i} (1 - \hat{\pi}(X_i))^{(1-y_i)}]} \right) \quad (2.19)$$

Keterangan:

G : Likelihood Ratio Test

n_0 : Jumlah pengamatan yang memiliki nilai $Y = 0$

n_1 : Jumlah pengamatan yang memiliki nilai $Y = 1$

n : Jumlah pengamatan

Statistik uji G mengikuti sebaran distribusi *chi-square* (χ^2) dengan dengan derajat bebas v yang merupakan jumlah variabel independen. Kriteria penolakan yang ditentukan dengan taraf signifikansi α adalah:

Tolak H_0 jika $p\text{-value} < \alpha$ atau nilai $G > \chi_{\alpha, v}^2$ artinya variabel independen secara bersama-sama berpengaruh signifikan terhadap variabel dependen.

2. Uji Parsial

Tujuan uji parsial adalah untuk menguji signifikansi koefisien β secara individual. Pengujian signifikansi secara parsial ini menggunakan metode *wald test* untuk menentukan variabel independen mana yang memiliki pengaruh signifikan terhadap probabilitas sukses. Hipotesis yang dipakai adalah sebagai berikut (Hosmer & Lemeshow, 2000):

$H_0: \beta_j = 0$ (Variabel ke - j tidak memiliki pengaruh signifikan)

$H_1: \beta_j \neq 0$ dengan $j = 0, 1, 2, \dots, p$ (Variabel ke - j memiliki pengaruh signifikan)

Penggunaan statistik uji pada hipotesis ini ditulis pada persamaan (2.20) sebagai berikut:

$$W = \left[\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \right]^2 \quad (2.20)$$

dengan $SE(\hat{\beta}_j) = \sqrt{var(\hat{\beta}_j)}$. Penggunaan uji statistik W beriringan dengan distribusi χ^2 sehingga pengujiannya dijalankan dengan mengontraskan hasil uji statistik W dan hasil dari tabel distribusi χ^2 . Kriteria penolakan yang ditentukan dengan mengacu pada taraf signifikansi α adalah:

Tolak H_0 jika $p\text{-value} < \alpha$ atau nilai $W > \chi_{\alpha, v}^2$ yang artinya variabel independen ke - j memiliki pengaruh signifikan terhadap variabel dependen.

2.1.5 Uji Kesesuaian Model

Goodness of fit dikenal sebagai uji kesesuaian model memiliki peran dalam mengevaluasi seberapa baik model regresi menjelaskan data yang diamati atau diobservasi. Uji kesesuaian model dijalankan guna menentukan apakah ditemukan ketaksamaan antara data yang diamati atau data hasil observasi dengan prediksi yang dihasilkan oleh model. Hipotesis dalam uji kesesuaian model seperti berikut:

H_0 : Model sudah sesuai (terdapat kesamaan antara hasil observasi dengan peluang perkiraan model)

H_1 : Model belum sesuai (terdapat ketaksamaan antara hasil observasi dengan peluang perkiraan model)

Penggunaan uji statistik yang digunakan ditulis dalam persamaan (2.21) sebagai berikut (Hosmer & Lemeshow, 2000):

$$D = -2 \sum_{i=1}^N \left[y_i \ln \left(\frac{\hat{\pi}(X_i)}{y_i} \right) + (1 - y_i) \ln \left(\frac{1 - \hat{\pi}(X_i)}{1 - y_i} \right) \right] \quad (2.21)$$

Kriteria penolakan yang ditentukan dengan mengacu pada taraf signifikansi α adalah tolak H_0 jika $p\text{-value} < \alpha$ atau nilai $D > \chi_{(df)}^2$, yang berarti model belum sesuai atau terdapat ketaksamaan antara hasil obeservasi dengan peluang perkiraan model.

2.1.6 Interpretasi Koefisien Model

Interpretasi koefisien dalam regresi logistik bertujuan untuk memahami bagaimana variabel independen memengaruhi variabel dependen. Dengan kata lain untuk mengetahui seberapa besar dan ke arah mana suatu perubahan pada variabel independen dapat memengaruhi kemungkinan terjadinya suatu kejadian pada variabel dependen. Dalam konteks regresi logistik, interpretasi ini umumnya dilakukan melalui nilai *odds ratio*. *Odds* merupakan rasio peluang terjadinya kejadian sukses (π) dengan kejadian gagal ($1 - \pi$) dari variabel dependen. Nilai *odds* didefinisikan pada persamaan (2.22) (Agresti, 1990):

$$Odds = \frac{\pi_j}{1 - \pi_j} \quad (2.22)$$

Odds ratio menggambarkan perbandingan peluang terjadinya suatu peristiwa pada satu kategori variabel independen dibandingkan dengan kategori referensinya. Nilai ini sangat berguna, terutama ketika variabel independen

bersifat kualitatif, karena membantu menjelaskan kecenderungan atau risiko relatif antara kelompok-kelompok yang berbeda. Model regresi logistik dengan p variabel independen, maka nilai *odds ratio* dapat dihitung untuk masing-masing variabel independen yang didefinisikan sebagai berikut:

$$\psi = \exp(\hat{\beta}_j) \quad (2.23)$$

2.1.7 Ketepatan Klasifikasi

Prosedur klasifikasi adalah penilaian untuk memeriksa seberapa baik suatu model klasifikasi dapat memprediksi kelas yang benar dari suatu data. Fungsi klasifikasi adalah mengklasifikasikan jumlah sampel yang salah dalam perhitungan nilai *Apparent Error Rate* (APER) (Johnson & Wichern, 2007). Nilai APER yang dihasilkan dari perhitungan tersebut adalah proporsi dari prediksi yang tidak tepat oleh klasifikasi (Ramandhani dkk., 2017). Perhitungan nilai APER lebih sederhana dengan menggunakan *confusion matrix*. *Confusion matrix* digunakan untuk mengidentifikasi perbedaan data asli dengan data hasil proses prediksi klasifikasi. Bentuk *confusion matrix* untuk klasifikasi 2 kelas disajikan dalam tabel 2.2 (Umaroh, 2020).

Tabel 2.2 *Confusion Matrix*

Hasil Observasi	Prediksi		Total
	y_1	y_2	
y_1	n_{11}	n_{12}	n_1
y_2	n_{21}	n_{22}	n_2
Total	n_1	n_2	n

Keterangan:

n_{11} : Banyak subjek dari y_1 benar diklasifikasikan sebagai y_1

n_{12} : Banyak subjek dari y_1 benar diklasifikasikan sebagai y_2

n_{21} : Banyak subjek dari y_2 benar diklasifikasikan sebagai y_1

n_{22} : Banyak subjek dari y_2 benar diklasifikasikan sebagai y_2

Perhitungan ketepatan klasifikasi sebagai berikut:

1. Rumus untuk menghitung nilai APER untuk menghitung tingkat kesalahan klasifikasi adalah sebagai berikut:

$$\text{APER} = \left(\frac{n_{12} + n_{21}}{n} \right) \times 100 \quad (2.24)$$

2. Rumus untuk menghitung tingkat ketepatan klasifikasi adalah sebagai berikut:

$$1 - \text{APER} = \left(\frac{n_{11} + n_{22}}{n} \right) \times 100 \quad (2.25)$$

2.1.8 Uji Multikolinearitas

Uji multikolinear bertujuan untuk mengidentifikasi adanya hubungan linear antara variabel-variabel independen. Sebuah model yang efektif seharusnya tidak menunjukkan korelasi yang kuat antara variabel independennya. Multikolinearitas dapat mengakibatkan bias dalam estimasi parameter dari model (Sungkono & Nugrahaningsih, 2017). Artinya ketika variabel independen saling berkorelasi kuat, maka untuk membedakan kontribusi masing-masing variabel terhadap respon yang diukur akan menjadi sulit. Hal tersebut akan membuat estimasi parameter menjadi bias atau tidak akurat. Tujuan dari deteksi multikolinieritas adalah untuk mengidentifikasi keberadaan hubungan linear antara variabel independen.

Sebuah model yang efektif sebaiknya tidak memiliki keterkaitan atau hubungan yang kuat antara variabel independennya. Salah diantara metode guna mengidentifikasi multikolinieritas adalah dengan melihat pada nilai *Variance Inflation Factor* (VIF). Penilaian seberapa kuat variabel independen dipengaruhi oleh korelasi dengan variabel lain dalam suatu model dapat dilakukan dengan memanfaatkan nilai VIF. Nilai VIF dirumuskan seperti berikut (Hocking, 2003):

$$VIF = \frac{1}{Tolerance} \quad (2.26)$$

dengan,

$$\begin{aligned} Tolerance &= \frac{1}{(1 - R^2)} \\ R^2 &= 1 - \frac{SS_{error}}{SS_{total}} \\ &= \frac{SS_{total} - SS_{error}}{SS_{total}} \\ &= \frac{SS_{regresi}}{SS_{total}} \end{aligned}$$

Dengan keterangan R^2 merupakan koefisien determinan. Hipotesis yang digunakan dalam pengujian ini adalah:

H_0 : Tidak terjadi pelanggaran multikolinieritas

H_1 : Terjadi pelanggaran multikolinieritas

Kriteria penolakan yang ditentukan dengan mengacu pada taraf signifikansi α adalah:

Tolak H_0 jika $p\text{-value} < \alpha$ atau nilai $VIF \geq 10$ yang artinya terjadi pelanggaran multikolinieritas.

2.1.9 Hipertensi

Hipertensi merupakan salah satu penyakit serius dan kronis yang menjadi penyebab kematian dan disabilitas di penjuru dunia. Hipertensi terjadi ketika tekanan darah diastolik melebihi angka 90 mmHg dan tekanan darah sistolik melebihi angka 140 mmHg (WHO, 2023). Hipertensi juga sering kali disebut dengan *the silent killer* karena gejalanya sering tidak terdeteksi oleh penderita sehingga menyebabkan komplikasi yang serius pada tubuh. Hipertensi dapat meningkatkan risiko penyakit yang menyerang jantung, otak, ginjal, dan penyakit lainnya.

Laporan WHO pada tahun 2023 mengenai hipertensi memprediksi banyaknya orang dewasa yang menderita hipertensi meningkat kurang lebih dua kali lipat dalam tiga dekade terakhir. Secara global hampir setiap 1 dari 3 orang dewasa menderita hipertensi dan prevalensi hipertensi diperkirakan akan terus meningkat secara signifikan dengan proyeksi sekitar 29% orang dewasa pada tahun 2025. Dampak bagi kesehatan dari terjadinya peningkatan tren hipertensi menyebabkan kurang lebih 10,8 juta kematian setiap tahun (WHO, 2024). Menurut SKI tahun 2023, hipertensi menempati peringkat keempat sebagai faktor penyebab kematian di Indonesia dengan persentase sebesar 10,2% terhadap total kematian (Kemenkes RI, 2024).

Hipertensi terbagi menjadi dua berdasarkan penyebabnya. Hipertensi esensial atau primer merupakan jenis hipertensi yang penyebab pastinya belum diketahui secara jelas. Meskipun penyebabnya tidak diketahui secara pasti, terdapat sejumlah faktor yang memicu kondisi tersebut, seperti faktor genetik atau riwayat keluarga, gaya hidup tidak sehat, obesitas, dan lainnya. Sedangkan

hipertensi sekunder merupakan jenis hipertensi yang disebabkan oleh adanya kondisi kesehatan tertentu sebagai penyebab utamanya. Berbeda dengan hipertensi essensial yang penyebabnya tidak jelas, hipertensi sekunder biasanya dapat ditelusuri ke gangguan yang spesifik, seperti gangguan hormon, cacat bawaan, efek samping obat-obatan tertentu, dan lainnya.

Faktor yang meningkatkan risiko terkena hipertensi adalah penambahan usia, faktor genetik, kelebihan berat badan, tidak aktif secara fisik, mengonsumsi alkohol, merokok, dan pola hidup yang tidak sehat. Faktor risiko hipertensi yang dapat diubah diantaranya adalah pola makan yang tidak sehat, kurangnya aktivitas fisik, obesitas, merokok, dan mengonsumsi alkohol. Pola makan yang tidak sehat dapat berupa konsumsi gula berlebih, konsumsi garam berlebih, makan dengan makanan yang tinggi lemak, dan kurang konsumsi buah serta sayur. Sedangkan faktor risiko yang tidak dapat diubah adalah usia dan faktor genetik atau riwayat keluarga. Berikut merupakan faktor-faktor yang dapat meningkatkan risiko menderita hipertensi (WHO, 2023):

1. Jenis Kelamin

Jenis kelamin merupakan salah satu faktor yang memengaruhi risiko hipertensi. Laki-laki cenderung memiliki risiko lebih tinggi mengalami hipertensi pada usia muda hingga paruh baya dibandingkan wanita. Namun, setelah masa menopause, wanita mengalami peningkatan risiko karena penurunan kadar hormon estrogen, yang sebelumnya berperan sebagai pelindung alami terhadap tekanan darah tinggi.

2. Konsumsi Rokok

Merokok merupakan faktor risiko utama penyakit kardiovaskular, termasuk hipertensi. Kandungan nikotin dalam rokok menyebabkan penyempitan pembuluh darah, meningkatkan denyut jantung, serta merusak dinding arteri yang dapat mempercepat terbentuknya plak yang menyumbat aliran darah.

3. Kurangnya Aktivitas Fisik

Kurangnya aktivitas fisik dapat berkontribusi pada peningkatan tekanan darah. Aktivitas fisik secara teratur membantu menjaga berat badan ideal dan meningkatkan kesehatan jantung, yang keduanya penting dalam pencegahan hipertensi. Aktivitas seperti jalan cepat, bersepeda, atau olahraga ringan bisa membantu menjaga tekanan darah tetap normal.

4. Konsumsi Gula Berlebih

Asupan gula yang berlebihan, terutama dari minuman manis dan makanan olahan, dapat meningkatkan risiko hipertensi baik secara langsung maupun tidak langsung. Konsumsi gula yang berlebih akan berpengaruh pada kadar gula dalam darah. Sehingga, konsumsi gula berlebih dapat menyebabkan peningkatan berat badan, resistensi insulin, dan peradangan sistemik yang semuanya berkontribusi pada naiknya tekanan darah.

5. Konsumsi Garam Berlebih

Konsumsi garam yang berlebihan dapat menyebabkan tubuh menahan air, yang pada akhirnya meningkatkan tekanan darah. Konsumsi garam yang berlebih juga akan berpengaruh pada kadar garam dalam darah. WHO

merekomendasikan untuk membatasi asupan garam guna mengurangi risiko hipertensi.

6. Lemak Berlebih

Konsumsi lemak, terutama lemak jenuh secara berlebihan dapat meningkatkan kadar kolesterol jahat dalam darah dan mempercepat proses penyempitan pembuluh darah. Konsumsi lemak yang berlebih akan berpengaruh pada kadar lemak dalam darah. Kondisi ini memicu peningkatan tekanan darah karena jantung harus bekerja lebih keras untuk memompa darah melalui pembuluh yang menyempit.

7. Konsumsi Sayur dan Buah

Sayur dan buah merupakan sumber serat serta berbagai vitamin yang bermanfaat untuk menjaga tekanan darah tetap stabil. Kurangnya konsumsi sayur dan buah dikaitkan dengan meningkatnya risiko hipertensi serta penyakit kronis lainnya.

8. Konsumsi Alkohol

Mengonsumsi alkohol dalam jumlah besar secara rutin dapat meningkatkan tekanan darah. Dalam jangka panjang, alkohol juga bisa merusak organ-organ penting seperti hati dan jantung. Tidak ada tingkat konsumsi alkohol yang sepenuhnya aman bagi kesehatan, dan jika dikonsumsi, jumlahnya harus sangat dibatasi.

9. Usia

Seiring bertambahnya usia, risiko seseorang mengalami hipertensi meningkat. Hal ini disebabkan oleh perubahan fisiologis seperti penurunan elastisitas pembuluh darah dan fungsi ginjal yang menurun. WHO

mencatat bahwa prevalensi hipertensi meningkat pada individu berusia di atas 30-79 tahun.

Berdasarkan KEMENKES RI tahun 2024, sekitar 90-95% penyebab utama kematian di dunia adalah hipertensi esensial. Gejala yang ditimbulkan tidak terlihat jelas menyebabkan sebagian besar orang yang menderita tekanan darah tinggi tidak menyadarinya. Hal tersebut yang menyebabkan hipertensi sulit terdeteksi tanpa adanya pemeriksaan. Penderita hipertensi esensial tidak merasakan sakit atau tanda-tanda yang dapat menunjukkan secara langsung bahwa sedang mengalami hipertensi. Penderita juga dapat merasa sehat namun tekanan darah sedang tinggi, yang pada akhirnya dapat menyebabkan serangan jantung, stroke, kerusakan ginjal yang serius, bahkan kematian. Inilah alasan mengapa hipertensi sering disebut sebagai "*the silent killer*" atau penyakit yang mematikan secara diam-diam. Pada penderita hipertensi esensial, gejala jarang muncul sampai tekanan darahnya sudah sangat tinggi dan membahayakan nyawa sampai tekanan darahnya sudah terlalu tinggi dan mengancam nyawa. Dengan demikian penting untuk meningkatkan kesadaran masyarakat mengenai bahaya hipertensi, melakukan pemeriksaan tekanan darah secara rutin, dan menerapkan gaya hidup sehat untuk mencegah hipertensi.

2.2 Kajian Integrasi Hipertensi dengan Al-Qu'ran dan Hadits

Hipertensi merupakan permasalahan kesehatan yang berhubungan dengan darah, pembuluh darah, dan sirkulasi darah. Seorang mukmin yang sehat memiliki sirkulasi darah yang normal dan stabil. Hipertensi terjadi ketika sirkulasi darah tidak normal atau mengalami tekanan yang tinggi. Apabila seorang mukmin

menderita hipertensi, maka kegiatan sosial dan ibadah sehari-harinya akan terganggu. Dalam Al-Qur'an, istilah medis seperti "hipertensi" atau "tekanan darah tinggi" tidak disebutkan secara langsung, karena istilah tersebut merupakan istilah modern yang muncul dari perkembangan ilmu kedokteran. Namun, Islam mendorong umatnya untuk menjaga kesehatan, keseimbangan, dan menghindari hal-hal yang merusak tubuh dan jiwa. Upaya yang dapat dilakukan untuk menjaga kesehatan tubuh diantaranya adalah dengan mengonsumsi makanan yang halal dan baik.

Allah SWT berfirman dalam QS. Al-Baqarah ayat 168 mengenai perintah untuk mengonsumsi makanan dan minuman yang halal dan baik serta larangan untuk mengikuti jejak setan. Dengan mengonsumsi makanan dan minuman yang halal dan baik, tubuh juga akan terhindar dari berbagai jenis penyakit, termasuk hipertensi. Wajib bagi umat Islam untuk fokus pada apapun yang jelas diperbolehkan oleh Allah SWT karena itu pasti aman dan membawa berkah serta menghindari apapun yang jelas dilarang Allah SWT karena itu akan merugikan diri sendiri. Apapun yang jelas diperbolehkan oleh Allah SWT adalah halal dan yang dilarang adalah haram. Dalam kaitannya dengan hipertensi, mengonsumsi makanan yang halal dan menghindari yang haram dapat terhindar dari risiko hipertensi.

Makanan yang halal dan baik tidak mengandung zat atau bahan yang dapat merusak dan membahayakan kesehatan tubuh. Sedangkan makanan dan minuman yang haram pasti mengandung bahan atau zat yang dapat merusak dan membahayakan kesehatan tubuh. Salah satu hal atau kebiasaan buruk yang dapat membahayakan tubuh adalah merokok dan mengonsumsi alkohol. Merokok dan mengonsumsi alkohol juga menjadi faktor penyebab hipertensi. Mengonsumsi

alkohol jelas merupakan perbuatan haram karena tidak hanya dapat merusak tubuh tetapi juga akal seseorang. Oleh karena itu, menghindari dan tidak mengonsumsi alkohol bukan hanya mematuhi perintah Allah SWT, tetapi juga merupakan langkah yang bijak untuk menjaga kondisi tubuh agar tetap sehat dan terhindar dari risiko hipertensi.

Bentuk upaya untuk menjaga kondisi tubuh seorang mukmin agar sehat dan terhindar dari risiko hipertensi adalah dengan cara menjaga pola makan. Pola makan yang sehat dan seimbang dilakukan dengan tidak makan atau minum secara berlebihan dan dengan mengonsumsi makanan serta minuman yang baik dan halal. Allah SWT berfirman dalam QS. Al-A'raf ayat 31:

يٰۤاٰدَمُ خُذْ زِينَتَكَ عِنْدَ كُلِّ مَسْجِدٍ وَكُلْ وَاشْرَبْ وَلَا تُسْرِفْ ۗ اِنَّهٗ لَا يُحِبُّ الْمُسْرِفِيْنَ ﴿٣١﴾

Artinya: *“Hai anak Adam, pakailah pakaianmu yang indah di setiap (memasuki) mesjid, makan dan minumlah, dan janganlah berlebih-lebihan. Sesungguhnya Allah tidak menyukai orang-orang yang berlebih-lebihan.”*

Berdasarkan ayat tersebut Allah SWT tidak menyukai orang-orang yang berlebihan dalam segala hal, baik perilaku, pakaian, minum, maupun makan. Apabila seorang mukmin makan atau minum secara berlebihan, maka akan menyebabkan obesitas maupun mendatangkan penyakit lainnya. Mengonsumsi berlebih makanan yang mengandung tinggi gula, garam, dan lemak juga tidak baik untuk kesehatan tubuh. Obesitas serta mengonsumsi gula dan garam berlebih merupakan faktor yang dapat menyebabkan seorang mukmin menderita hipertensi. Dengan demikian, Allah SWT menganjurkan seorang mukmin untuk menjaga pola makan yang seimbang dan tidak berlebihan agar terhindar dari hipertensi.

Menurut Fathul Karim Mukhtashar Tafsir Al-Qu'ran Al-Adzhim, Ibnu Jarir berkata tentang firman Allah SWT “*Sesungguhnya Allah tidak menyukai orang-orang yang berlebih-lebihan*” bahwa berlebih-lebihan yang dimaksud adalah melampaui batas-batas yang telah Allah SWT tetapkan dalam perkara yang halal atau haram (Tafsir web, 2024). Allah SWT tidak menyukai orang-orang yang menghalalkan sesuatu yang sudah jelas diharamkan oleh Allah SWT sebagai sesuatu yang boleh dilakukan dan mengharamkan sesuatu yang dihalalkan oleh Allah SWT sebagai hal yang terlarang. Makan merupakan sesuatu yang halal, akan tetapi jika mengonsumsi makanan yang mengandung gula, garam, dan lemak secara berlebihan akan melampaui batas kebutuhan tubuh dan dapat meningkatkan risiko terkena hipertensi. Dengan kata lain, Allah SWT menyukai orang-orang yang mematuhi segala perintah, hukum, dan aturan-Nya sebagaimana adanya tanpa menambah atau mengurangi.

Dalam agama Islam, segala aspek kehidupan telah diatur oleh Allah SWT dengan sangat jelas dan proporsional atau sesuai porsi masing-masing. Selain itu, Islam juga menekankan pentingnya melakukan sesuatu dengan memperhatikan adab atau tata kramanya. Adab tidak hanya sekedar aturan yang perlu ditaati, tetapi juga merupakan bagian dari ibadah kepada Allah SWT. Adab merupakan sesuatu yang diajarkan dan dicontohkan langsung oleh Rasulullah Muhammad SAW, sehingga menjadi pedoman bagi umat Islam. Salah satu adab dalam makan dan minum adalah duduk. Nabi Muhammad SAW menyerukan kepada umat Islam untuk makan dan minum dilakukan dengan duduk, sebagai mana yang dijelaskan dalam hadist yang diriwayatkan oleh Muslim nomor 2024 sebagai berikut:

وَعَنْ أَبِي سَعِيدٍ أَنَّ النَّبِيَّ صَلَّى اللَّهُ عَلَيْهِ وَآلِهِ وَسَلَّمَ نَهَى عَنِ الشُّرْبِ قَائِمًا (رَوَاهُ أَحْمَدُ وَمُسْلِمٌ)

Artinya: “Dari Abu Said bahwa Nabi Muhammad SAW melarang minum sambil berdiri” (HR Muslim).

Berdasarkan sudut pandang medis, tata cara seseorang makan dan minum dapat mempengaruhi seberapa baik tubuh dapat menyerap nutrisi. Makan dan minum dengan duduk dapat membantu tubuh agar tetap sehat, terhindar dari penyakit, dan mendukung fungsi pencernaan yang optimal. Proses pencernaan makanan dalam tubuh lebih optimal apabila dilakukan dengan cara duduk. Ketika proses pencernaan berjalan lancar, aliran darah ke organ pencernaan dapat berlangsung normal tanpa ada tekanan berlebihan. Sebaliknya, apabila seseorang makan dan minum dengan berdiri, makanan akan masuk dengan cepat ke lambung sehingga dapat memicu tekanan berlebihan sementara pada pembuluh darah (Manembu dkk., 2015). Hal tersebut dapat memperburuk risiko hipertensi, terutama jika dilakukan secara terus. Dengan demikian, makan dan minum dengan duduk dapat membantu menstabilkan tekanan darah dan meningkatkan efisiensi proses pencernaan.

Faktor penyebab hipertensi selain dilihat dari segi fisik juga dapat dilihat dari segi psikologi. Keadaan *psychosociological* yang mencakup faktor psikologi (seperti stres, kecemasan, atau depresi) dan sosial (seperti hubungan dengan orang lain atau lingkungan) dapat mempengaruhi aktivitas pada pusat otak dalam mengatur tekanan darah (Taufiq dkk., 2020). Allah SWT berfirman dalam QS. An-Nisa’ ayat 32:

وَلَا تَتَمَنَّوْا مَا فَضَّلَ اللَّهُ بِهِ بَعْضَكُمْ عَلَى بَعْضٍ ۗ لِلرِّجَالِ نَصِيبٌ مِّمَّا كَسَبُوا ۗ وَلِلنِّسَاءِ نَصِيبٌ مِّمَّا كَسَبْنَ ۗ وَسَأَلُوا اللَّهَ مِنْ فَضْلِهِ ۗ إِنَّ اللَّهَ كَانَ بِكُلِّ شَيْءٍ عَلِيمًا ﴿٣٢﴾

Artinya: “Dan janganlah kamu iri hati terhadap apa yang dikaruniakan Allah kepada sebahagian kamu lebih banyak dari sebahagian yang lain. (Karena) bagi orang laki-laki ada bahagian dari pada apa yang mereka usahakan, dan bagi para wanita (pun) ada bahagian dari apa yang mereka usahakan, dan mohonlah kepada Allah sebagian dari karunia-Nya. Sesungguhnya Allah Maha Mengetahui segala sesuatu.”

Berdasarkan surat An-Nisa’ ayat 32 tersebut, Allah SWT memerintahkan seorang mukmin untuk tidak iri hati terhadap karunia Allah. Iri hati adalah penyakit hati yang merasa tidak senang terhadap kebahagiaan atau kesuksesan orang lain. Seseorang yang iri hati cenderung memiliki emosi yang tidak stabil. Emosi yang tidak stabil tersebut sering kali memicu ketegangan, kecemasan, bahkan perasaan marah. Tekanan emosi, ketegangan, dan kecemasan yang tinggi dan berlebihan dapat mengakibatkan stres. Stres merupakan salah satu penyebab tekanan darah meningkat (Prasetyorini, 2012). Dengan demikian, apabila seseorang mengalami perasaan iri hati terus menerus, maka tubuh akan mengalami stres. Hal tersebut akan menyebabkan tekanan darah menjadi naik sehingga dapat meningkatkan risiko hipertensi.

Upaya untuk membantu menjaga tekanan darah tetap normal dan terhindar dari risiko hipertensi, seorang mukmin harus mengelola hati agar lebih lapang dan tenang untuk mengendalikan emosi dengan baik. Ketenangan hati dapat dicapai dengan cara berdzikir atau mengingat Allah SWT seperti firman-Nya dalam QS. Ar-Rad ayat 28:

الَّذِينَ آمَنُوا وَتَطْمَئِنُّ قُلُوبُهُمْ بِذِكْرِ اللَّهِ أَلَا بِذِكْرِ اللَّهِ تَطْمَئِنُّ الْقُلُوبُ ﴿٢٨﴾

Artinya: “(yaitu) orang-orang yang beriman dan hati mereka menjadi tenteram dengan mengingat Allah. Ingatlah, hanya dengan mengingat Allah-lah hati menjadi tenteram.”

Berdasarkan ayat tersebut, dengan berdzikir dan mengingat Allah SWT hati akan menjadi tenteram. Hati yang tenteram juga akan membuat pikiran menjadi tenang dan berfikir positif. Hati dan pikiran yang tenang dapat mengurangi dan bahkan menghilangkan stres, kecemasan, dan ketegangan. Apabila seseorang memiliki hati yang lapang dan tenang pasti dapat berfikir positif dan dapat mengatur emosinya. Dengan demikian, berdzikir atau mengingat Allah SWT akan berdampak positif pada kesehatan fisik termasuk menjaga tekanan darah agar tetap stabil dan dalam batas normal.

قَدْ أَفْلَحَ الْمُؤْمِنُونَ ﴿١﴾ الَّذِينَ هُمْ فِي صَلَاتِهِمْ خَاشِعُونَ ﴿٢﴾ وَالَّذِينَ هُمْ عَنِ اللَّغْوِ مُعْرِضُونَ ﴿٣﴾

Artinya: “Sungguh, beruntunglah orang-orang mukmin. (Yaitu) orang-orang yang khusyuk dalam shalatnya, orang-orang yang meninggalkan (perbuatan dan perkataan) yang tidak berguna.”

Berdasarkan ayat tersebut, Allah SWT memerintahkan umat Islam untuk khusyuk dalam shalatnya. Gerakan dalam shalat seperti berdiri, rukuk, sujud, dan duduk merupakan bentuk aktivitas fisik ringan yang dapat membantu melancarkan aliran darah ke seluruh tubuh. Sujud, misalnya, menempatkan posisi kepala lebih rendah dari jantung, sehingga memudahkan darah mengalir ke otak dan meningkatkan fungsi saraf. Aktivitas ini bila dilakukan secara rutin dapat menurunkan tekanan darah, menstabilkan detak jantung, serta mengurangi risiko penyakit jantung. Selain itu, shalat yang dilakukan dengan khusyuk juga membantu

tubuh menjadi lebih rileks dan tenang, yang berdampak langsung pada sistem peredaran darah.

Sementara itu, dzikir yang dilakukan secara teratur, terutama dalam suasana tenang, terbukti secara ilmiah mampu menurunkan tingkat stres dan kecemasan. Saat seseorang berdzikir, tubuh merespons dengan mengaktifkan sistem saraf parasimpatik yang memberikan efek relaksasi, menurunkan tekanan darah, dan memperlancar aliran darah ke organ-organ penting. Secara keseluruhan, sholat dan dzikir merupakan bentuk ibadah yang memberikan efek sinergis bagi kesehatan jasmani dan rohani, serta mendukung kinerja sistem peredaran darah secara optimal. Sholat dan dzikir memiliki manfaat yang luar biasa tidak hanya dari sisi spiritual, tetapi juga bagi kesehatan fisik, khususnya untuk menstabilkan tekanan darah.

2.3 Kajian Hipertensi Dengan Regresi Logistik dan *Tomek Links*

Penelitian penerapan metode regresi logistik dan *Tomek Links* pada *imbalance data* diagnosa hipertensi memiliki sejumlah rangkaian tahapan yang didasarkan pada teori-teori pendukung yang berkaitan. Dalam penelitian ini, beberapa langkah sistematis dilakukan untuk menganalisis dan mengolah data diagnosa hipertensi. Tahap pertama adalah pengumpulan data yang dilaksanakan untuk memperoleh informasi mengenai variabel dependen (diagnosa hipertensi) dan variabel independen (faktor-faktor yang mempengaruhi hipertensi) yang relevan. Setelah data terkumpul, langkah berikutnya adalah menggambarkan karakteristik data secara visual dengan analisis statistika deskriptif, seperti distribusi dan ukuran pemusatan data.

Tahap selanjutnya adalah data dibagi menjadi dua bagian yaitu data *training* dan data *testing* dengan pembagian acak yang proporsional. Data *training* digunakan untuk melatih model, sementara data *testing* digunakan untuk menguji performa model. Pembagian ini dilakukan dengan perbandingan 70% untuk data *training* dan 30% untuk data *testing*. Metode *Tomek Links* kemudian diterapkan pada data *training* untuk mereduksi data diagnosa hipertensi, yaitu dengan menghapus data *noise* pada kelas mayoritas. Teknik ini digunakan untuk menemukan sampel data yang diinginkan dari kelas mayoritas yang memiliki jarak *euclidean* terendah dengan data kelas minoritas, kemudian menghapusnya. Data yang telah diproses kemudian dianalisis lebih lanjut dengan regresi logistik.

Tahap pertama dalam regresi logistik adalah membentuk model awal regresi logistik dengan melibatkan semua variabel independen, kemudian estimasi parameter menggunakan metode MLE secara *Newton-Raphson*, yang bertujuan untuk mencari nilai parameter model yang memaksimalkan kemungkinan data yang diamati berdasarkan persamaan (2.13). Kemudian melakukan uji multikolinearitas untuk mengetahui hubungan atau korelasi yang sangat kuat antara variabel-variabel independen. Uji multikolinearitas ditentukan berdasarkan nilai VIF pada persamaan (2.26). Selanjutnya, dilakukan uji signifikansi parameter untuk mengkaji peran masing-masing variabel independen terhadap model. Uji simultan dilakukan menggunakan uji G untuk mengetahui apakah semua variabel independen signifikan secara bersama-sama, sementara uji parsial dilakukan dengan uji *wald* untuk menilai signifikansi masing-masing variabel secara individu. Uji G dihitung berdasarkan persamaan (2.18) dan uji *wald* dihitung berdasarkan persamaan (2.20).

Setelah parameter model dievaluasi dan dianggap signifikan, langkah selanjutnya adalah membuat model regresi logistik (Afifah, 2020).

Model yang diperoleh kemudian diuji kesesuaiannya dengan tujuan untuk memastikan bahwa model yang dibangun dapat menggambarkan data dengan baik dan sesuai. Uji kesesuaian model dihitung dengan nilai *deviance* berdasarkan persamaan (2.21). Apabila model tidak sesuai, maka dilakukan pemilihan model terbaik berdasarkan nilai *Akaike Information Criterion* (AIC) (Pratama, 2018). Model yang telah sesuai kemudian dilakukan interpretasi parameter. Langkah terakhir adalah evaluasi yang dilakukan untuk mengukur ketepatan model dalam melakukan klasifikasi diagnosis hipertensi, menggunakan *confusion matrix* yang mengukur kinerja model dan menghitung tingkat akurasi ketepatan model. Perhitungan untuk mengetahui kinerja model regresi logistik dalam mengklasifikasikan data diagnosa hipertensi dengan berdasarkan persamaan (2.24) yaitu menghitung tingkat kesalahan model dalam klasifikasi dan persamaan (2.25) yaitu menghitung tingkat ketepatan model dalam klasifikasi.

BAB III

METODE PENELITIAN

3.1 Jenis Penelitian

Jenis penelitian yang diterakan yaitu penelitian kuantitatif. Data yang diterapkan pada penelitian ini adalah data kuantitatif berupa data diagnosa hipertensi. Data kuantitatif tersebut kemudian dianalisis dengan menggunakan metode regresi logistik dan reduksi *imbalanced data* dengan metode *Tomek Links*.

3.2 Data dan Sumber Data

Penelitian ini menerapkan data sekunder yang merupakan jenis pengambilan data melalui sumber yang sudah tersedia. Adapun data sekunder yang digunakan dalam penelitian ini diperoleh dari Puskesmas Mojolangu, Kecamatan Lowokwaru, Kota Malang. Variabel yang diterapkan pada penelitian ini mencakup variabel dependen (Y) dan variabel independen (X). Variabel dependen yang digunakan dalam penelitian ini adalah diagnosa hipertensi berskala nominal yang dikategorikan menjadi 2 yaitu: menderita hipertensi (1) dan tidak menderita hipertensi (0) dan. Berikut merupakan variabel-variabel yang digunakan:

Tabel 3.1 Variabel Penelitian

No	Variabel		Skala	Keterangan
1	Diagnosa Hipertensi	Y	Nominal	1: Hipertensi 0: Tidak Hipertensi
2	Jenis Kelamin	x_1	Nominal	1: Laki-laki 0: Perempuan
3	Konsumsi Rokok	x_2	Nominal	1: Ya 0: Tidak
4	Aktivitas Fisik	x_3	Nominal	1: Ya 0: Tidak

Tabel 3.2 Lanjutan Variabel Penelitian

5	Konsumsi Gula Berlebih	x_4	Nominal	1: Ya 0: Tidak
6	Konsumsi Garam Berlebih	x_5	Nominal	1: Ya 0: Tidak
7	Lemak Berlebih	x_6	Nominal	1: Ya 0: Tidak
8	Konsumsi Buah dan Sayur	x_7	Nominal	1: Ya 0: Tidak
9	Konsumsi Alkohol	x_8	Nominal	1: Ya 0: Tidak
10	Usia	x_9	Nominal	1: Ya 0: Tidak

3.3 Teknik Pengumpulan Data

Penelitian ini menggunakan teknik dokumentasi dalam teknik pengumpulan datanya. Teknik dokumentasi merupakan metode pengumpulan data dengan mencari informasi dari sumber yang tertulis maupun lisan. Data yang dikumpulkan bisa berupa surat, pernyataan tertulis, pengumuman, data dari situs web, dan lain-lain (Nilamsari, 2014).

3.4 Teknik Analisis Data

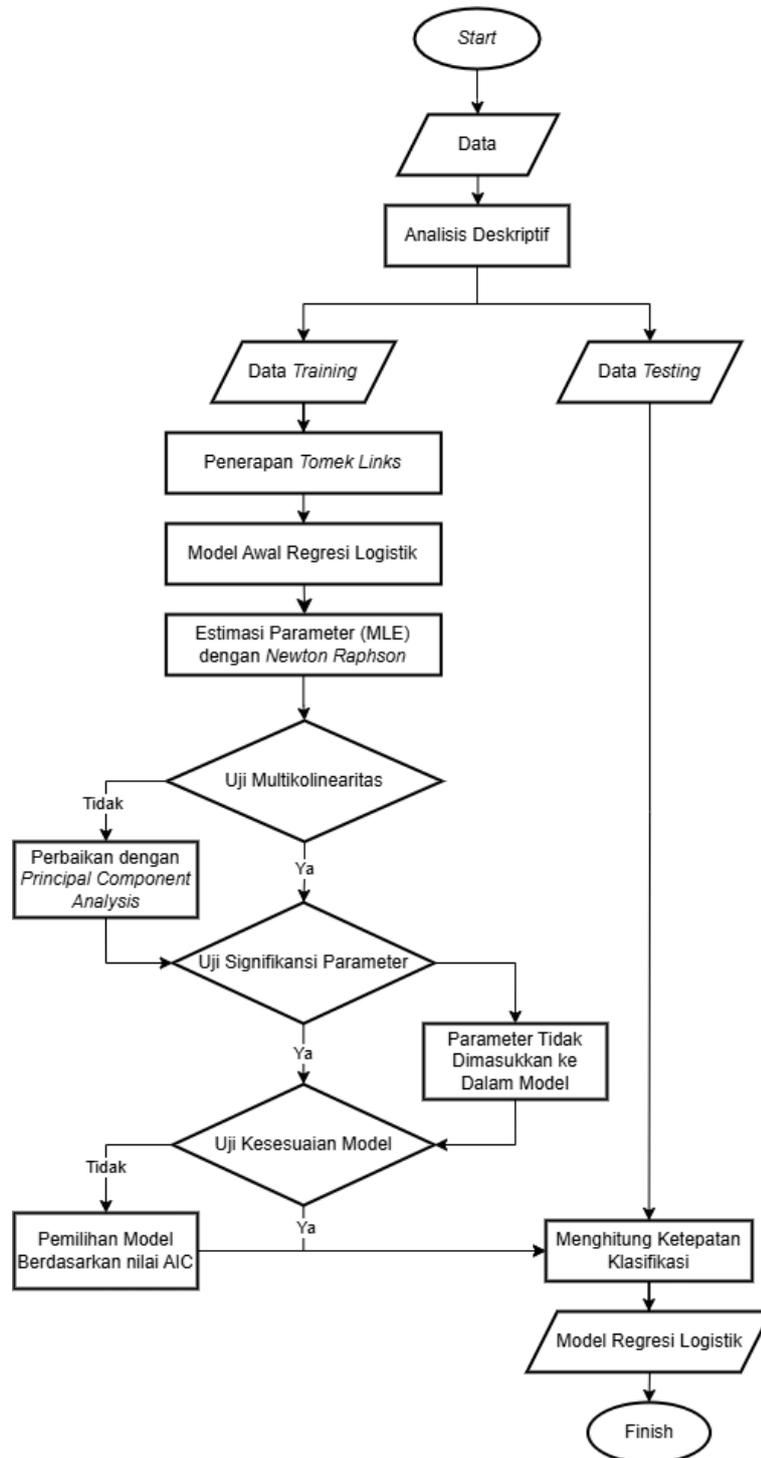
Adapun teknik analisis data yang dilakukan dalam penelitian ini memiliki beberapa langkah. Langkah-langkah yang dilakukan sebagai berikut:

1. Melakukan pengumpulan data diagnosa hipertensi yang meliputi variabel dependen dan variabel independen.
2. Melakukan analisis statistika deskriptif untuk mengetahui visualisasi karakteristik data diagnosa hipertensi.
3. Melakukan pembagian data dengan dua bagian sehingga menjadi data *training* dan data *testing*. Pembagian tersebut dilakukan secara acak dengan perbandingan data *training* sebesar 70% dan data *testing* sebesar 30%.

4. Menerapkan metode *Tomek Links* pada data *training* untuk mereduksi data dengan menghapus data dari kelas mayoritas berdasarkan jarak *euclidean* pada persamaan (2.1).
5. Melakukan pengolahan dan pengujian data dengan menggunakan regresi logistik:
 - 1) Membentuk model awal regresi logistik dengan melibatkan semua variabel independen berdasarkan persamaan (2.4)
 - 2) Melakukan estimasi parameter dengan menggunakan metode MLE dengan metode numerik *Newton-Raphson* berdasarkan persamaan (2.13).
 - 3) Melakukan uji multikolinearitas dengan berdasarkan nilai VIF pada persamaan (2.26).
 - 4) Melakukan pengujian signifikansi parameter yaitu uji simultan yang dilakukan dengan uji *G* dengan berdasarkan persamaan (2.18) dan uji parsial yang dilakukan dengan uji *wald* dengan berdasarkan persamaan (2.20). Apabila parameter yang tidak memenuhi signifikansi, maka parameter tidak dimasukkan dalam pembentukan model.
 - 5) Membentuk model regresi logistik berdasarkan variabel yang signifikan berdasarkan persamaan (2.4).
 - 6) Melakukan pengujian kesesuaian model dengan nilai *deviance* berdasarkan persamaan (2.21).
 - 7) Menginterpretasikan koefisien parameter model dengan berdasarkan nilai *odds ratio* pada persamaan (2.23)

6. Melakukan evaluasi ketepatan klasifikasi regresi logistik dengan *confusion matrix* dan menghitung tingkat akurasi ketepatan model dalam klasifikasi berdasarkan persamaan (2.24) dan (2.25).

3.5 Diagram Alir Penelitian



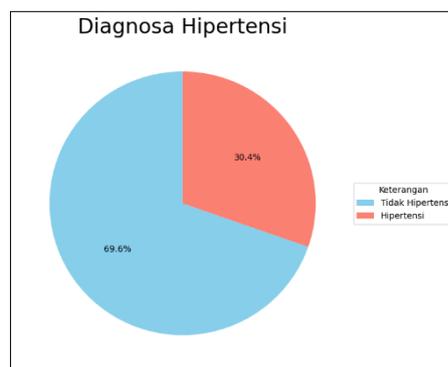
Gambar 3.1 Diagram Alir Penelitian

BAB IV

PEMBAHASAN

4.1 Statistika Deskriptif

Penelitian ini menggunakan data diagnosa hipertensi tahun 2022 hingga 2023 yang diperoleh dari Puskesmas Mojolangu, Kecamatan Lowokwaru, Kota Malang. Dataset terdiri dari 2.387 pasien dengan informasi yang dikumpulkan meliputi 9 variabel independen yaitu jenis kelamin, konsumsi rokok, aktivitas fisik, konsumsi gula berlebih, konsumsi garam berlebih, lemak berlebih, konsumsi buah dan sayur, konsumsi alkohol, dan usia. Sementara itu, variabel dependen dalam data ini adalah diagnosa hipertensi yang memiliki kategori hipertensi dan tidak hipertensi. Analisis statistika deskriptif dilakukan pada data diagnosa hipertensi untuk memberikan visualisasi karakteristik data dan faktor-faktor yang mempengaruhi diagnosa hipertensi. Visualisasi dari diagnosa hipertensi di Puskesmas Mojolangu disajikan dalam Gambar 4.1.



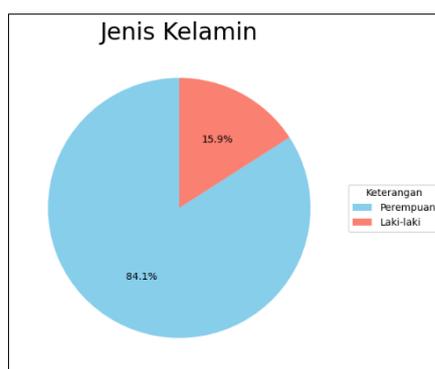
Gambar 4.1 Diagnosa Hipertensi

Berdasarkan Gambar 4.1, persentase pasien yang tidak hipertensi lebih banyak yaitu sebesar 69,6% atau sebanyak 1.662 pasien dalam klasifikasi tidak

hipertensi. Sedangkan persentase pasien yang memiliki hipertensi sebesar 30,4% atau sebanyak 725 pasien diklasifikasikan hipertensi. Hal tersebut menunjukkan bahwa sekitar 1 dari 3 pasien dalam sampel memiliki hipertensi. Meskipun mayoritas pasien tidak memiliki hipertensi, persentase penderita hipertensi yang mencapai lebih dari 30% menandakan adanya masalah kesehatan yang serius. Hal tersebut yang dapat menjadi dasar untuk melakukan analisis lebih lanjut untuk mengidentifikasi faktor-faktor risiko yang berpengaruh terhadap tekanan darah pada kejadian hipertensi, seperti jenis kelamin, konsumsi rokok, aktivitas fisik, konsumsi gula berlebih, konsumsi garam berlebih, lemak berlebih, konsumsi buah dan sayur, konsumsi alkohol, dan usia.

4.1.1 Deskripsi Jenis Kelamin

Jenis kelamin merupakan salah satu faktor yang dapat mempengaruhi hipertensi. Deskripsi tentang jenis kelamin disajikan dalam Gambar 4.2.



Gambar 4.2 Jenis Kelamin

Berdasarkan Gambar 4.2, menunjukkan bahwa pasien yang berjenis kelamin perempuan lebih banyak dibandingkan dengan pasien laki-laki. Tercatat dari 2.387 pasien, sebesar 84,1% atau sebanyak 2.008 pasien yang berjenis

kelamin perempuan. Sedangkan sebesar 15,9% atau sebanyak 379 pasien berjenis kelamin laki-laki. Deskripsi jenis kelamin pasien berdasarkan diagnosa hipertensi disajikan pada Tabel 4.1.

Tabel 4.1 Jenis Kelamin Berdasarkan Diagnosa Hipertensi

		Diagnosa Hipertensi		Total
		Hipertensi	Tidak Hipertensi	
Jenis Kelamin	Laki-laki	89 (23,5%)	290 (76,5%)	379
	Perempuan	636 (31,7%)	1372 (68,3%)	2008

Tabel 4.1 menunjukkan mengenai jenis kelamin pasien berdasarkan diagnosa hipertensi. Berdasarkan 725 diagnosa hipertensi, terdapat 89 pasien berjenis kelamin laki-laki dan 636 pasien berjenis kelamin perempuan. Sedangkan pada diagnosa tidak hipertensi terdapat 290 pasien berjenis kelamin laki-laki dan 1.372 pasien berjenis kelamin perempuan. Jika ditinjau lebih lanjut berdasarkan status hipertensi, dari 379 laki-laki, sebanyak 89 orang (23,5%) menderita hipertensi, sedangkan dari 2.008 perempuan, sebanyak 636 orang (31,7%) terdiagnosa hipertensi. Artinya, secara absolut jumlah penderita hipertensi lebih banyak pada perempuan, hal tersebut juga dikarenakan jumlah total perempuan memang jauh lebih tinggi.

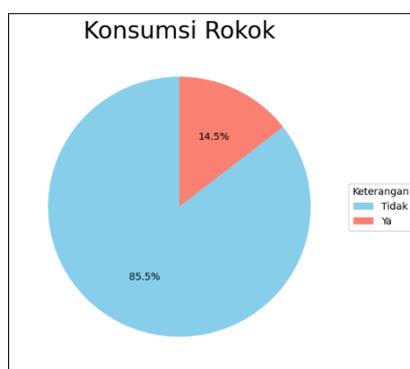
Tingginya prevalensi hipertensi pada perempuan mungkin dipengaruhi oleh faktor-faktor risiko lain yang juga diamati dalam penelitian, seperti konsumsi garam, gula, atau faktor yang lainnya. Perempuan cenderung memiliki risiko lebih tinggi terkena hipertensi setelah memasuki masa menopause. Hal tersebut disebabkan oleh penurunan kadar hormon estrogen yang memiliki peran protektif terhadap sistem peredaran darah. Selain itu, perempuan cenderung lebih responsif terhadap gejala-gejala kesehatan yang muncul dalam tubuhnya. Ketika mulai

merasakan sakit atau ketidaknyamanan, perempuan lebih mungkin untuk segera memeriksakan diri ke fasilitas kesehatan seperti rumah sakit atau klinik.

Kebiasaan tersebut memungkinkan deteksi dini terhadap berbagai kondisi medis, yang pada akhirnya meningkatkan peluang untuk mendapatkan perawatan yang efektif dan mencegah komplikasi lebih lanjut. Karena itulah, secara statistik, perempuan sering kali memiliki angka harapan hidup yang lebih tinggi dibandingkan laki-laki. Tingkat kunjungan medis yang lebih tinggi serta kepatuhan terhadap anjuran pengobatan turut menjadi faktor pendukung mengapa perempuan cenderung memiliki umur yang lebih panjang.

4.1.2 Deskripsi Konsumsi Rokok

Mengonsumsi rokok merupakan salah satu faktor yang dapat mempengaruhi hipertensi. Variabel konsumsi rokok dikategorikan menjadi 2 yaitu pasien yang mengonsumsi rokok dan pasien yang tidak mengonsumsi rokok. Deskripsi tentang konsumsi rokok disajikan dalam Gambar 4.3.



Gambar 4.3 Konsumsi Rokok

Berdasarkan Gambar 4.3, menunjukkan bahwa pasien yang tidak mengonsumsi rokok lebih banyak dibandingkan dengan pasien yang

mengonsumsi rokok. Tercatat dari 2.387 pasien, sebesar 85,5% atau sebanyak 2.042 pasien yang tidak mengonsumsi rokok. Sedangkan sisanya yaitu sebesar 14,5% atau sebanyak 345 merupakan pasien yang mengonsumsi rokok. Deskripsi konsumsi rokok pasien berdasarkan diagnosa hipertensi disajikan pada Tabel 4.2.

Tabel 4.2 Konsumsi Rokok Berdasarkan Diagnosa Hipertensi

Konsumsi Rokok	Diagnosa Hipertensi				Total
	Hipertensi		Tidak Hipertensi		
Ya	Perempuan	Laki-laki	Perempuan	Laki-laki	345
	65	38	129	113	
	103 (29,8%)		242 (70,2%)		
Tidak	Perempuan	Laki-laki	Perempuan	Laki-laki	2042
	571	51	1243	177	
	622 (30,5%)		1420 (69,5%)		

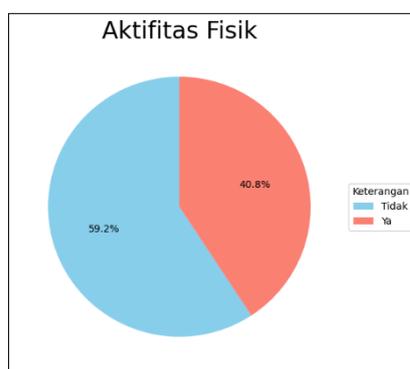
Tabel 4.2 menunjukkan mengenai konsumsi rokok pasien berdasarkan diagnosa hipertensi. Berdasarkan 725 diagnosa hipertensi, terdapat 103 pasien yang mengonsumsi rokok dan 622 pasien tidak mengonsumsi rokok. Sedangkan pada diagnosa tidak hipertensi terdapat 242 pasien yang mengonsumsi rokok dan 1.420 pasien tidak mengonsumsi rokok. Jika dilihat berdasarkan status hipertensi, dari 345 pasien yang mengonsumsi rokok, 29,8% terdiagnosa hipertensi, sedangkan dari 2.042 pasien tidak mengonsumsi rokok, 30,5% mengalami hipertensi. Artinya, secara proporsional, persentase penderita hipertensi pada kelompok yang mengonsumsi rokok lebih rendah dibandingkan kelompok yang tidak mengonsumsi rokok. Temuan ini tampak bertentangan dengan literatur medis yang umumnya menyatakan bahwa kebiasaan merokok merupakan faktor risiko kuat terhadap hipertensi.

Salah satu kemungkinan penjelasannya adalah bahwa kelompok yang tidak mengonsumsi rokok dalam sampel ini didominasi oleh perempuan, yang

berdasarkan data sebelumnya (Tabel 4.1). Karena merokok lebih umum pada laki-laki, maka kelompok yang tidak mengonsumsi rokok cenderung berisi lebih banyak perempuan, yang dalam sampel ini memang lebih banyak jumlahnya. Selain itu, faktor-faktor lain juga dapat berkontribusi terhadap terjadinya hipertensi, terlepas dari status merokok. Dengan demikian, meskipun secara umum merokok merupakan faktor risiko hipertensi, dalam konteks data ini tingginya prevalensi hipertensi pada kelompok yang tidak mengonsumsi rokok tidak serta-merta menandakan bahwa merokok lebih aman.

4.1.3 Deskripsi Aktivitas Fisik

Aktivitas fisik yang dilakukan seseorang dalam kegiatan sehari-hari merupakan salah satu faktor yang dapat mempengaruhi hipertensi. Variabel aktivitas fisik dikategorikan menjadi 2 yaitu pasien dengan aktivitas fisik dan pasien yang tidak melakukan aktivitas fisik. Deskripsi tentang aktivitas fisik pasien disajikan dalam Gambar 4.4.



Gambar 4.4 Aktivitas Fisik

Berdasarkan Gambar 4.4, menunjukkan bahwa pasien yang tidak melakukan aktivitas fisik lebih banyak dibandingkan dengan pasien yang

melakukan aktivitas fisik. Tercatat dari 2.387 pasien, sebesar 59,2% atau sebanyak 1.414 pasien yang tidak melakukan aktivitas fisik. Sedangkan sisanya yaitu sebesar 40,8% atau sebanyak 973 merupakan pasien yang melakukan aktivitas fisik. Deskripsi aktivitas fisik pasien berdasarkan diagnosa hipertensi disajikan pada Tabel 4.3.

Tabel 4.3 Aktivitas Fisik Berdasarkan Diagnosa Hipertensi

Aktivitas Fisik	Diagnosa Hipertensi				Total
	Hipertensi		Tidak Hipertensi		
Ya	Perempuan	Laki-laki	Perempuan	Laki-laki	973
	259	52	507	155	
	311 (32%)		622 (68%)		
Tidak	Perempuan	Laki-laki	Perempuan	Laki-laki	1414
	377	37	865	135	
	414 (29,3%)		1000 (70,7%)		

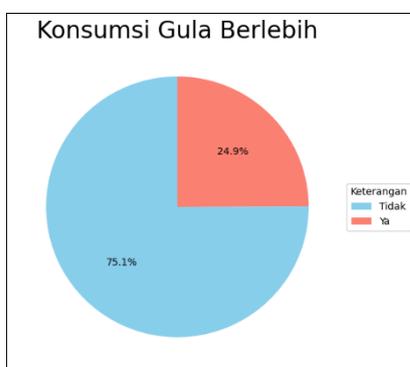
Tabel 4.3 menunjukkan mengenai aktivitas fisik pasien berdasarkan diagnosa hipertensi. Berdasarkan 725 diagnosa hipertensi, terdapat 311 pasien yang melakukan aktivitas fisik dan 414 pasien tidak melakukan aktivitas fisik. Sedangkan pada diagnosa tidak hipertensi terdapat 622 pasien yang melakukan aktivitas fisik dan 1.000 pasien tidak melakukan aktivitas fisik. Secara numerik, jumlah penderita hipertensi lebih banyak terdapat pada kelompok yang tidak melakukan aktivitas fisik (414 kasus vs. 311 kasus).

Temuan tersebut tampak tidak sejalan dengan banyak literatur kesehatan yang menyatakan bahwa aktivitas fisik secara rutin dapat menurunkan risiko hipertensi. Hal tersebut mungkin disebabkan oleh pengaruh faktor lain. Pasien yang aktif secara fisik akan tetapi tidak mengatur pola makan seperti, konsumsi gula berlebih, konsumsi garam berlebih, dan lainnya. Di sisi lain, sebagian pasien yang tidak aktif secara fisik bisa saja memiliki tekanan darah normal karena faktor

lain, seperti tidak merokok atau mengatur pola makan dengan baik. Dengan demikian, hasil ini menunjukkan bahwa aktivitas fisik bukan satu-satunya faktor yang menentukan diagnosa hipertensi, melainkan merupakan bagian dari kombinasi perilaku dan faktor risiko lainnya.

4.1.4 Deskripsi Konsumsi Gula Berlebih

Jumlah gula yang dikonsumsi seseorang merupakan salah satu faktor yang dapat mempengaruhi hipertensi. Hal tersebut konsumsi gula berlebih akan berpengaruh pada kadar gula dalam darah. Variabel konsumsi gula dikategorikan menjadi 2 yaitu pasien dengan konsumsi gula berlebih dan pasien yang tidak mengonsumsi gula berlebih. Deskripsi tentang konsumsi gula pasien disajikan dalam Gambar 4.5.



Gambar 4.5 Konsumsi Gula Berlebih

Berdasarkan Gambar 4.5, menunjukkan bahwa pasien yang tidak berlebihan dalam mengonsumsi gula lebih banyak dibandingkan dengan pasien yang mengonsumsi gula berlebih. Tercatat dari 2.387 pasien, sebesar 75,1% atau sebanyak 1.793 pasien yang tidak mengonsumsi gula berlebih. Sedangkan sisanya yaitu sebesar 24,9% atau sebanyak 594 merupakan pasien yang mengonsumsi

gula berlebih. Deskripsi konsumsi gula pasien berdasarkan diagnosa hipertensi disajikan pada Tabel 4.4.

Tabel 4.4 Konsumsi Gula Berdasarkan Diagnosa Hipertensi

Konsumsi Gula	Diagnosa Hipertensi				Total
	Hipertensi		Tidak Hipertensi		
Ya	Perempuan	Laki-laki	Perempuan	Laki-laki	594
	214	15	296	69	
	229 (38,5%)		365 (61,5%)		
Tidak	Perempuan	Laki-laki	Perempuan	Laki-laki	1793
	422	74	1076	221	
	496 (27,7%)		1297 (72,3%)		

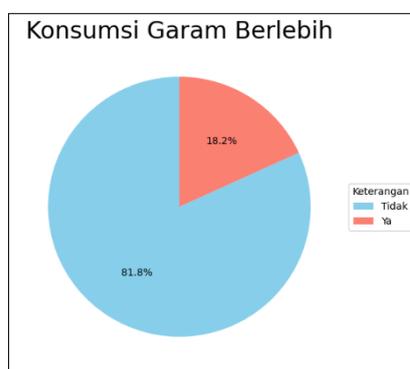
Tabel 4.4 menunjukkan mengenai konsumsi gula pasien berdasarkan diagnosa hipertensi. Berdasarkan 725 diagnosa hipertensi, terdapat 229 pasien yang mengonsumsi gula berlebih dan 496 pasien tidak mengonsumsi gula berlebih. Sedangkan pada diagnosa tidak hipertensi terdapat 365 pasien dengan konsumsi gula berlebih dan 1.297 pasien tidak mengonsumsi gula berlebih. Meskipun secara teori konsumsi gula berlebih merupakan salah satu faktor risiko hipertensi, namun berdasarkan hasil yang disajikan Tabel 4.4, justru jumlah absolut penderita hipertensi lebih banyak berasal dari kelompok yang tidak mengonsumsi gula secara berlebihan. Jika ditinjau secara proporsional, prevalensi hipertensi justru lebih tinggi pada kelompok yang mengonsumsi gula berlebih yaitu 38,5% (229 dari 594) dibandingkan kelompok yang tidak mengonsumsi gula berlebih yaitu 27,7% (496 dari 1.793).

Secara proporsional, prevalensi hipertensi lebih tinggi pada kelompok yang mengonsumsi gula secara berlebihan, meskipun secara jumlah absolut lebih banyak penderita hipertensi berasal dari kelompok yang tidak mengonsumsi gula berlebih. Hal ini konsisten dengan temuan dalam literatur yang menyatakan

bahwa asupan gula berlebih, terutama dalam bentuk minuman manis dan makanan olahan, dapat meningkatkan risiko hipertensi. Selain itu, jumlah penderita hipertensi yang lebih tinggi pada kelompok yang tidak mengonsumsi gula berlebih tidak serta-merta menunjukkan bahwa konsumsi gula tidak berpengaruh terhadap hipertensi. Beberapa pasien mungkin mengalami hipertensi karena faktor genetik, usia, pola makan, atau faktor lain yang tidak terkait langsung dengan konsumsi gula.

4.1.5 Deskripsi Konsumsi Garam Berlebih

Jumlah garam yang dikonsumsi seseorang juga merupakan salah satu faktor yang dapat mempengaruhi hipertensi. Hal tersebut konsumsi garam berlebih akan berpengaruh pada kadar garam dalam darah. Variabel konsumsi garam dikategorikan menjadi 2 yaitu pasien dengan konsumsi garam berlebih dan pasien yang tidak mengonsumsi garam berlebih. Deskripsi tentang konsumsi garam pasien disajikan dalam Gambar 4.6.



Gambar 4.6 Konsumsi Garam Berlebih

Berdasarkan Gambar 4.6 menunjukkan bahwa pasien yang tidak berlebihan dalam mengonsumsi garam lebih banyak dibandingkan dengan pasien

yang mengonsumsi garam berlebih. Tercatat dari 2.387 pasien, sebesar 81,8% atau sebanyak 1.953 pasien yang tidak mengonsumsi garam berlebih. Sedangkan sisanya yaitu sebesar 18,2% atau sebanyak 434 merupakan pasien yang mengonsumsi garam berlebih. Deskripsi konsumsi garam pasien berdasarkan diagnosa hipertensi disajikan pada Tabel 4.5.

Tabel 4.5 Konsumsi Garam Berlebih Berdasarkan Diagnosa Hipertensi

Konsumsi Garam	Diagnosa Hipertensi				Total
	Hipertensi		Tidak Hipertensi		
Ya	Perempuan	Laki-laki	Perempuan	Laki-laki	434
	158	21	209	46	
	179 (41,2%)		255 (58,8%)		
Tidak	Perempuan	Laki-laki	Perempuan	Laki-laki	1953
	478	68	1163	244	
	546 (28%)		1407 (72%)		

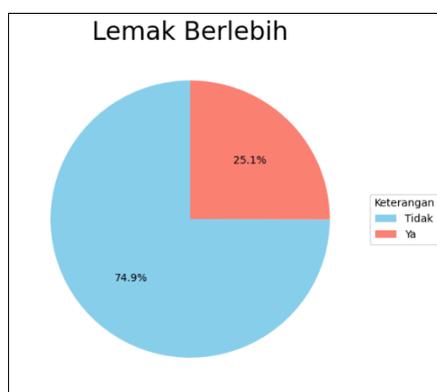
Tabel 4.5 menunjukkan mengenai konsumsi garam pasien berdasarkan diagnosa hipertensi. Berdasarkan 725 diagnosa hipertensi, terdapat 179 pasien yang mengonsumsi garam berlebih dan 546 pasien tidak mengonsumsi garam berlebih. Sedangkan pada diagnosa tidak hipertensi terdapat 255 pasien dengan konsumsi garam berlebih dan 1.407 pasien tidak mengonsumsi garam berlebih. Jika ditinjau secara proporsional, prevalensi hipertensi justru lebih tinggi pada kelompok yang mengonsumsi garam berlebih yaitu 41,2% (179 dari 434) dibandingkan kelompok yang tidak yaitu 28% (546 dari 1.953).

Jumlah penderita hipertensi yang lebih tinggi pada kelompok yang tidak mengonsumsi garam berlebih tidak berarti garam bukan faktor risiko hipertensi. Konsumsi garam berlebih tetap menjadi variabel independen penting yang berkontribusi terhadap kejadian hipertensi, meskipun hasil deskriptif secara jumlah absolut tampak sebaliknya karena faktor-faktor lain yang mungkin

dimiliki pasien yang tidak mengonsumsi garam berlebih. Sebagian pasien yang tidak mengonsumsi garam berlebih bisa saja memiliki mengonsumsi rokok, kurang aktivitas fisik, atau faktor lain yang membuat pasien tetap mengalami hipertensi meskipun konsumsi garam rendah.

4.1.6 Deskripsi Lemak Berlebih

Kadar lemak berlebih dalam darah merupakan salah satu faktor yang dapat mempengaruhi hipertensi. Hal tersebut konsumsi lemak berlebih akan berpengaruh pada kadar lemak dalam darah. Variabel lemak berlebih dikategorikan menjadi 2, yaitu pasien dengan kadar lemak berlebih dan pasien yang tidak memiliki kadar lemak berlebih. Deskripsi lemak berlebih pasien berdasarkan diagnosa hipertensi disajikan pada Tabel 4.6.



Gambar 4.7 Lemak Berlebih

Berdasarkan Gambar 4.7, menunjukkan bahwa pasien yang tidak berlebihan dalam mengonsumsi lemak lebih banyak dibandingkan dengan pasien yang mengonsumsi lemak berlebih. Tercatat dari 2.387 pasien, sebesar 74,9% atau sebanyak 1.789 pasien yang tidak mengonsumsi lemak berlebih. Sedangkan

sisanya yaitu sebesar 25,1% atau sebanyak 598 merupakan pasien yang mengonsumsi lemak berlebih.

Tabel 4.6 Lemak Berlebih Berdasarkan Diagnosa Hipertensi

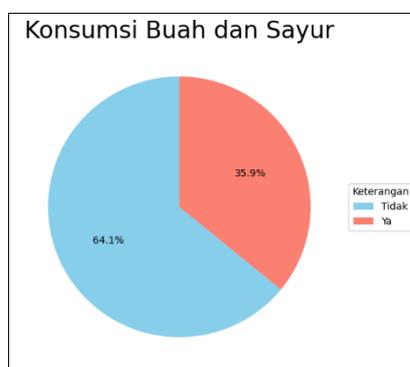
Lemak Berlebih	Diagnosa Hipertensi				Total
	Hipertensi		Tidak Hipertensi		
Ya	Perempuan	Laki-laki	Perempuan	Laki-laki	598
	199	19	333	47	
	218 (36,5%)		380 (63,5%)		
Tidak	Perempuan	Laki-laki	Perempuan	Laki-laki	1789
	437	70	1039	243	
	507 (28,3%)		1282 (71,7%)		

Tabel 4.6 menunjukkan mengenai kadar lemak dalam darah pasien berdasarkan diagnosa hipertensi. Berdasarkan 725 diagnosa hipertensi, terdapat 218 pasien yang memiliki kadar lemak dalam darah yang berlebih dan 507 pasien tidak memiliki kadar lemak berlebih dalam darah. Sedangkan pada diagnosa tidak hipertensi terdapat 380 pasien memiliki kadar lemak dalam darah yang berlebih dan 1.282 pasien tidak memiliki kadar lemak berlebih dalam darah.

Jumlah penderita hipertensi yang lebih tinggi pada kelompok yang tidak memiliki kadar lemak berlebih tidak berarti konsumsi lemak tidak berpengaruh terhadap hipertensi. Secara proporsional, risiko hipertensi lebih tinggi pada kelompok yang konsumsi lemak berlebih, yaitu 36,5% (218 dari 598), sedangkan pada kelompok yang tidak konsumsi lemak berlebih hanya 28,3% (507 dari 1.789). Hal ini mengindikasikan bahwa lemak berlebih tetap menjadi faktor risiko penting terhadap hipertensi, meskipun jumlah absolut penderita lebih banyak pada kelompok yang tidak memiliki kadar lemak berlebih. Selain itu, beberapa pasien yang tidak memiliki kadar lemak berlebih tetap bisa terkena hipertensi akibat faktor-faktor lain.

4.1.7 Deskripsi Konsumsi Buah Dan Sayur

Jumlah buah dan sayur yang dikonsumsi seseorang merupakan salah satu faktor yang dapat mempengaruhi hipertensi. Variabel konsumsi buah dan sayur dikategorikan menjadi 2 yaitu pasien dengan konsumsi buah dan sayur dan pasien yang tidak mengonsumsi buah dan sayur. Deskripsi tentang konsumsi buah dan sayur disajikan dalam Gambar 4.8. Deskripsi konsumsi buah dan sayur berdasarkan diagnosa hipertensi disajikan pada Tabel 4.7.



Gambar 4.8 Konsumsi Buah dan Sayur

Tabel 4.7 Konsumsi Buah dan Sayur Berdasarkan Diagnosa Hipertensi

Konsumsi Buah dan Sayur	Diagnosa Hipertensi				Total
	Hipertensi		Tidak Hipertensi		
Ya	Perempuan	Laki-laki	Perempuan	Laki-laki	858
	252	26	492	88	
	278 (32,4%)		580 (67,6%)		
Tidak	Perempuan	Laki-laki	Perempuan	Laki-laki	1529
	384	63	880	202	
	447 (29,2%)		1082 (70,8%)		

Berdasarkan Gambar 4.8 menunjukkan bahwa pasien yang tidak mengonsumsi buah dan sayur lebih banyak dibandingkan dengan pasien yang mengonsumsi buah dan sayur. Tercatat dari 2.387 pasien, sebesar 64,1% atau sebanyak 1.529 pasien yang tidak mengonsumsi buah dan sayur. Sedangkan

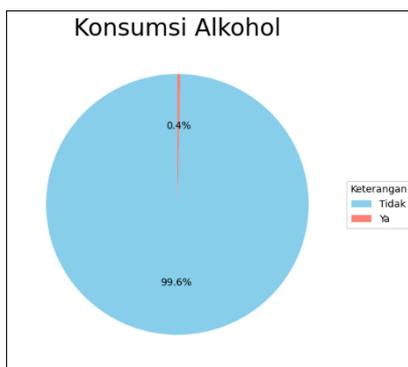
sisanya yaitu sebesar 35,9% atau sebanyak 858 merupakan pasien yang mengonsumsi buah dan sayur.

Tabel 4.7 menunjukkan mengenai konsumsi buah dan sayur berdasarkan diagnosa hipertensi. Berdasarkan 725 diagnosa hipertensi, terdapat 278 pasien yang mengonsumsi buah dan sayur dan 447 pasien tidak mengonsumsi buah dan sayur. Sedangkan pada diagnosa tidak hipertensi terdapat 580 pasien dengan konsumsi buah dan sayur dan 1.082 pasien tidak mengonsumsi buah dan sayur. Jumlah pasien yang tidak mengonsumsi buah dan sayur dan tidak hipertensi memang terlihat lebih besar (1.082 orang) dibandingkan dengan yang hipertensi (447 orang). Fenomena di mana pasien yang tidak makan buah dan sayur justru tidak terkena hipertensi tidak berarti bahwa buah dan sayur menyebabkan hipertensi, melainkan karena faktor risiko lainnya.

4.1.8 Deskripsi Konsumsi Alkohol

Konsumsi alkohol merupakan salah satu faktor yang dapat mempengaruhi hipertensi. Variabel konsumsi alkohol dikategorikan menjadi 2 yaitu pasien mengonsumsi alkohol dan pasien yang tidak mengonsumsi alkohol. Deskripsi tentang konsumsi alkohol disajikan dalam Gambar 4.9.

Berdasarkan Gambar 4.9 menunjukkan bahwa pasien yang tidak mengonsumsi alkohol lebih banyak dibandingkan dengan pasien yang mengonsumsi alkohol. Tercatat dari 2.387 pasien, sebesar 99,6% atau sebanyak 2.378 pasien yang tidak mengonsumsi alkohol. Sedangkan pasien yang mengonsumsi alkohol sebesar 0,4% atau sebanyak 9. Deskripsi konsumsi alkohol berdasarkan diagnosa hipertensi disajikan pada Tabel 4.8.



Gambar 4.9 Konsumsi Alkohol

Tabel 4.8 Konsumsi Alkohol Berdasarkan Diagnosa Hipertensi

Konsumsi Alkohol	Diagnosa Hipertensi				Total
	Hipertensi		Tidak Hipertensi		
Ya	Perempuan	Laki-laki	Perempuan	Laki-laki	9
	3	1	2	3	
	4 (44,4%)		5 (55,5%)		
Tidak	Perempuan	Laki-laki	Perempuan	Laki-laki	2378
	633	88	1370	287	
	721 (30,3%)		1657 (69,7%)		

Tabel 4.8 menunjukkan mengenai konsumsi alkohol berdasarkan diagnosa hipertensi. Berdasarkan 725 diagnosa hipertensi, terdapat 4 pasien yang mengonsumsi alkohol dan 721 pasien tidak mengonsumsi alkohol. Sedangkan pada diagnosa tidak hipertensi terdapat 5 pasien dengan konsumsi alkohol dan 1.657 pasien tidak mengonsumsi alkohol. Meskipun secara absolut jumlah penderita hipertensi lebih besar pada kelompok yang tidak mengonsumsi alkohol, hal ini wajar karena kelompok tersebut jauh lebih besar secara jumlah total.

Jika dilihat dari proporsi, justru kelompok yang mengonsumsi alkohol memiliki persentase hipertensi yang lebih tinggi dibandingkan kelompok yang tidak mengonsumsi alkohol. Secara proposional, kelompok yang tidak mengonsumsi alkohol yaitu, 30,3% (721 dari 2.378), sementara dari kelompok yang mengonsumsi alkohol yaitu, 44,4% (4 dari 9). Sebagian besar pasien adalah

tidak mengonsumsi alkohol (99,6%), sehingga secara alami jumlah penderita hipertensi dalam kelompok ini lebih tinggi secara absolut. Selain itu, bisa saja pasien yang tidak mengonsumsi alkohol kemungkinan memiliki faktor risiko lain yang lebih dominan.

4.1.9 Deskripsi Usia

Usia juga merupakan faktor yang sangat berpengaruh terhadap hipertensi. Deskripsi tentang usia pasien berdasarkan diagnosa hipertensi disajikan pada Tabel 4.9.

Tabel 4.9 Usia Berdasarkan Diagnosa Hipertensi

		<i>Mean</i>	<i>Median</i>	<i>Minimum</i>	<i>Maximum</i>
Diagnosa	Hipertensi	55	56	4	103
	Tidak Hipertensi	46	45	1	89

Tabel 4.9 menunjukkan mengenai usia pasien berdasarkan diagnosa hipertensi. Berdasarkan 725 diagnosa hipertensi, rata-rata pasien memiliki usia 55 tahun, dengan median atau nilai tengah 56 tahun. Selain itu, usia minimal pasien adalah 4 tahun dan maksimal 103 tahun. Sedangkan pada 1.662 diagnosa tidak hipertensi, rata-rata pasien memiliki usia 46 tahun dengan median atau nilai tengah 45 tahun. Usia minimal pasien pada diagnosa tidak hipertensi pada angka 1 tahun dan maksimal pada angka 89 tahun.

Data pada variabel tekanan diastolik kemudian diubah menjadi kategorik yaitu 0 untuk tidak hipertensi dan 1 untuk hipertensi. Berdasarkan WHO (2023), seseorang berisiko terkena hipertensi apabila berusia 30-79 tahun. Dengan demikian, data dengan usia lebih dari sama dengan 30 tahun akan dikodekan

sebagai 1, sedangkan data dengan usia kurang dari 30 tahun dikodekan sebagai 0. Setelah dilakukan pengkodean, pasien yang dikategorikan hipertensi berjumlah 2.001 dan pasien yang dikategorikan tidak hipertensi berjumlah 386.

Tabel 4.10 Usia Berdasarkan Diagnosa Hipertensi Lanjutan

Usia	Diagnosa Hipertensi				Total
	Hipertensi		Tidak Hipertensi		
Ya	Perempuan	Laki-laki	Perempuan	Laki-laki	2001
	614	79	1089	219	
	693 (34,6%)		1308 (65,4%)		
Tidak	Perempuan	Laki-laki	Perempuan	Laki-laki	386
	22	10	283	71	
	32 (8,3%)		354 (91,7%)		

Tabel 4.10 menunjukkan bahwa berdasarkan 725 diagnosa hipertensi, terdapat 693 pasien termasuk dalam usia rentan terkena hipertensi dan 32 pasien tidak termasuk dalam usia rentan terkena hipertensi. Sedangkan pada 1.662 diagnosa tidak hipertensi terdapat 1.308 pasien yang termasuk dalam usia rentan terkena hipertensi dan 354 pasien tidak termasuk dalam usia rentan terkena hipertensi. Dengan demikian, untuk memahami lebih jauh faktor penyebab hipertensi dalam sampel ini, perlu dilakukan analisis seperti regresi logistik, agar dapat mengetahui kontribusi masing-masing variabel independen terhadap tekanan darah pada kejadian hipertensi.

4.2 Pembagian Data *Training* dan *Testing*

Pembagian data *training* dan *testing* dilakukan dengan rasio 70% dan 30% pada data *training*. Pembagian data dengan rasio 70% dan 30% dipilih karena memiliki tingkat akurasi yang lebih tinggi dari rasio lainnya. Hasil pembagian data dengan menggunakan *software* R Studio disajikan pada Tabel 4.11.

Tabel 4.11 Pembagian Data *Training* dan *Testing*

		Pembagian Data		Total
		<i>Testing</i> (30%)	<i>Training</i> (70%)	
Diagnosa	Hipertensi	226	499	725
	Tidak Hipertensi	491	1171	1662
Total		717	1670	2387

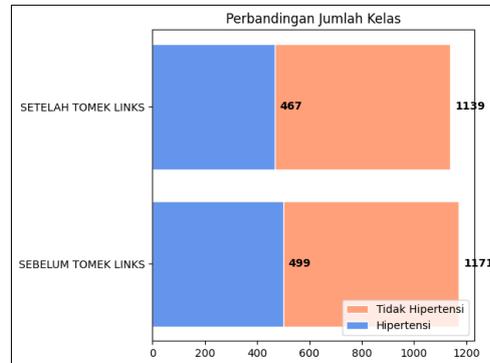
Tabel 4.11 menunjukkan pembagian data untuk proses *training* dan *testing* terhadap data diagnosis hipertensi. Data keseluruhan berjumlah 2.387 sampel, yang kemudian dibagi menjadi dua bagian yaitu 70% untuk data *training* (sebanyak 1.670 sampel) dan 30% untuk data *testing* (sebanyak 717 sampel). Berdasarkan total 725 sampel pada diagnosa hipertensi, sebanyak 499 digunakan untuk *training* dan 226 digunakan untuk *testing*. Sementara itu, dari 1.662 sampel pada diagnosa tidak hipertensi, 1.171 digunakan untuk *training* dan 491 untuk *testing*.

4.3 Penerapan *Tomek Links*

Tomek Links digunakan untuk mereduksi data yang tidak seimbang. *Tomek Links* bekerja dengan cara mengurangi data pada kelas mayoritas. Total data diagnosa hipertensi berjumlah 2.387 yang terbagi menjadi 2 kategori yaitu 725 merupakan kategori hipertensi dan 1.662 adalah kategori tidak hipertensi. Penerapan *Tomek Links* dilakukan pada data *training*. Data *training* terdiri dari 499 kategori hipertensi dan 1.171 kategori tidak hipertensi. Perbedaan data kelas mayoritas dan minoritas pada data sesudah dan sebelum dilakukan proses *Tomek Links* disajikan pada Gambar 4.10.

Gambar 4.10 menunjukkan bahwa pada data *training* kategori tidak hipertensi, data mengalami pengurangan. Data sebelumnya berjumlah 1.171 dan setelah dilakukan proses *Tomek Links* berkurang menjadi 1.139. Sedangkan pada

data kategori hipertensi data berkurang dari 499 data menjadi 467 data. Sehingga jumlah total data *training* setelah *Tomek Links* adalah 1.606 data.



Gambar 4.10 Perbandingan Sebelum dan Sesudah Proses *Tomek Links*

4.4 Model Awal Regresi Logistik

Model awal regresi logistik dibentuk dengan melibatkan seluruh variabel independen yang tersedia dalam penelitian. Model awal ini bertujuan untuk memberikan gambaran umum mengenai pengaruh masing-masing variabel independen terhadap variabel dependen secara simultan. Penyusunan model dilakukan tanpa terlebih dahulu melakukan proses seleksi variabel, sehingga seluruh variabel independen dimasukkan ke dalam model untuk kemudian dianalisis kontribusinya. Berdasarkan persamaan (2.4) diperoleh model regresi logistik sebagai berikut:

$$\begin{aligned} \pi(X) &= \frac{\exp(g(X))}{1 + \exp(g(X))} \\ &= \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \dots + \beta_9 x_9)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \dots + \beta_9 x_9)} \end{aligned}$$

4.5 Maximum Likelihood Estimation

Metode MLE merupakan metode yang digunakan untuk mengestimasi parameter-parameter dalam regresi logistik biner. Metode MLE bertujuan untuk menunjukkan nilai estimasi β dengan memaksimalkan fungsi *likelihood*-nya. Akan tetapi, karena bentuk fungsi *likelihood* dalam regresi logistik tidak memiliki solusi yang sederhana, maka metode numerik *Newton-Raphson* digunakan untuk memperoleh estimasi parameter β secara iteratif yang terdapat pada persamaan (2.13). Penentuan nilai awal estimasi awal ($\beta^{(0)}$) dengan menggunakan OLS. Perhitungan dilakukan dengan variabel independen dan variabel dependen dari data diagnosa hipertensi yang sudah dikategorikan pada Lampiran 2. Perhitungan manual nilai awal estimasi parameter ($\beta^{(0)}$) secara OLS pada persamaan (2.14) sebagai berikut:

$$X = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 \\ \vdots & \vdots \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, Y = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

$$\beta^{(0)} = (X^T X)^{-1} X^T Y$$

$$= \left[\begin{bmatrix} 1 & 1 & \dots & 1 \\ 0 & 0 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 0 \end{bmatrix} \right]^{-1} \left[\begin{bmatrix} 1 & 1 & \dots & 1 \\ 0 & 0 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right]$$

$$= \begin{bmatrix} 0,11284 \\ 0,10299 \\ \vdots \\ 0,81396 \end{bmatrix}$$

Kemudian, nilai awal tersebut digunakan untuk proses iterasi *Newton-Raphson* berdasarkan persamaan (2.13). Perhitungan manual untuk estimasi parameter pada iterasi ke-1 sebagai berikut:

$$g(x_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_9 x_{9i}$$

Untuk observasi ke $- i = 1$

$$\begin{aligned} \pi(x_1) &= \frac{\exp(g(x_1))}{1 + \exp(g(x_1))} \\ &= \frac{\exp(0,11284 + 0,10299.0 + 0,01362.0 + \dots + 0,0220.0)}{1 + \exp(0,11284 + 0,10299.0 + 0,01362.0 + \dots + 0,0220.0)} \\ &= \frac{\exp(0,10963)}{1 + \exp(0,10963)} \\ &= 0,52738 \end{aligned}$$

Perhitungan nilai peluang kejadian dilakukan dengan cara yang sama untuk observasi ke $- 2, 3, \dots, 1606$ sehingga diperoleh matriks berukuran 1606×1

$$\pi(x)^{(0)} = \begin{bmatrix} 0,52738 \\ 0,53778 \\ \vdots \\ 0,49945 \end{bmatrix}$$

Kemudian menghitung nilai dari $\pi(x)^{(0)}(1 - \pi(x)^{(0)})$. Hasil dari $\pi(x)^{(0)}(1 - \pi(x)^{(0)})$ kemudian diubah menjadi matriks diagonal $V^{(0)}$ berukuran 1606×1606 .

Untuk observasi ke $- i = 1$

$$\begin{aligned} \pi(x_1)^{(0)}(1 - \pi(x_1)^{(0)}) &= 0,52738(1 - 0,52738) \\ &= 0,24925 \end{aligned}$$

Perhitungan nilai peluang tidak terjadi kejadian dilakukan dengan cara yang sama untuk observasi ke - 2, 3, ..., 1606 sehingga diperoleh:

$$\pi(x)^{(0)}(1 - \pi(x)^{(0)}) = \begin{bmatrix} 0,24925 \\ 0,24857 \\ \vdots \\ 0,25000 \end{bmatrix}$$

$$V^{(0)} = \begin{bmatrix} 0,24925 & 0 & \cdots & 0 \\ 0 & 0,24857 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0,25000 \end{bmatrix}$$

Kemudian menghitung matriks *hessian* berdasarkan persamaan (2.12) dan diinverskan:

$$(X^T V^{(0)} X)^{-1} =$$

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \cdot \begin{bmatrix} 0,24925 & 0 & \cdots & 0 \\ 0 & 0,24857 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0,25000 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 0 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} 384,193 & 60,484 & \cdots & 65,204 \\ 60,484 & 60,484 & \cdots & 9,271 \\ \vdots & \vdots & \ddots & \vdots \\ 65,204 & 9,271 & \cdots & 65,204 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} 0,00788 & -0,00242 & \cdots & -0,0285 \\ -0,00242 & 0,02282 & \cdots & 0,00062 \\ \vdots & \vdots & \ddots & \vdots \\ -0,00285 & 0,00062 & \cdots & 0,01865 \end{bmatrix}$$

Menghitung $Y - \pi(x)^{(0)}$ atau $W^{(0)}$ untuk membuat vektor kemiringan berdasarkan persamaan (2.10):

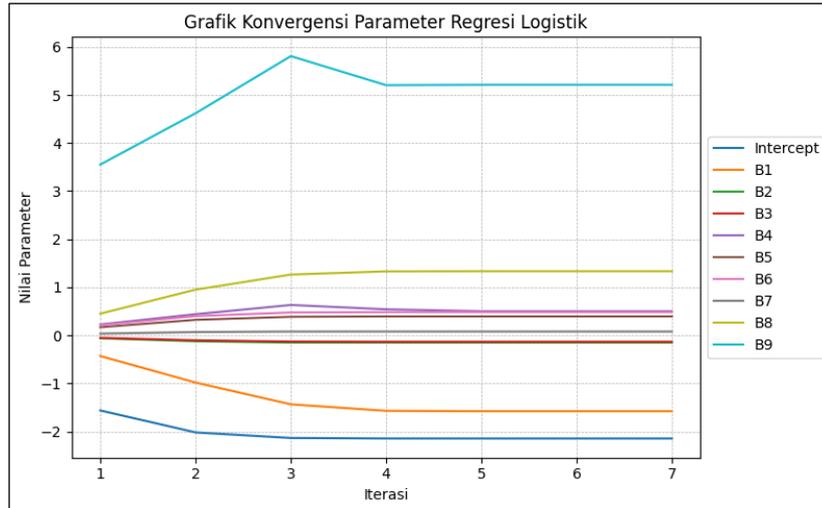
$$X^T W^{(0)} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} - \begin{bmatrix} 0,52738 \\ 0,53778 \\ \vdots \\ 0,49945 \end{bmatrix}$$

$$\begin{aligned}
&= \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \cdot \begin{bmatrix} -0,52738 \\ -0,53778 \\ \vdots \\ -0,49945 \end{bmatrix} \\
&= \begin{bmatrix} -447,344 \\ -91,741 \\ \vdots \\ 73,573 \end{bmatrix}
\end{aligned}$$

Sehingga dengan demikian hasil estimasi parameter $\beta^{(1)}$ dengan proses iterasi *Newton-Raphson* berdasarkan persamaan (2.13) sebagai berikut:

$$\begin{aligned}
\beta^{(1)} &= \beta^{(0)} + (X^T V^{(0)} X)^{-1} X^T W^{(0)} \\
&= \begin{bmatrix} 0,11284 \\ 0,10299 \\ \vdots \\ 0,81396 \end{bmatrix} + \begin{bmatrix} 0,00788 & -0,00242 & \cdots & -0,0285 \\ -0,00242 & 0,02282 & \cdots & 0,00062 \\ \vdots & \vdots & \ddots & \vdots \\ -0,00285 & 0,00062 & \cdots & 0,01865 \end{bmatrix} \cdot \begin{bmatrix} -447,344 \\ -91,741 \\ \vdots \\ 73,573 \end{bmatrix} \\
&= \begin{bmatrix} 0,11284 \\ 0,10299 \\ \vdots \\ 0,81396 \end{bmatrix} + \begin{bmatrix} -1,67284 \\ -0,32549 \\ \vdots \\ 2,73744 \end{bmatrix} \\
&= \begin{bmatrix} -1,56000 \\ -0,42848 \\ \vdots \\ 3,55140 \end{bmatrix}
\end{aligned}$$

Proses estimasi parameter pada iterasi selanjutnya dilakukan menggunakan langkah yang sama. Proses akan berhenti apabila nilai taksiran telah konvergen seperti yang terlampir pada Lampiran 5. Grafik kekonvergenan proses iterasi estimasi parameter disajikan pada Gambar 4.11. Hasil dari estimasi parameter dengan *software* R Studio disajikan dalam Tabel 4.12.



Gambar 4.11 Grafik Kekonvergenan Iterasi

Tabel 4.12 Estimasi Parameter

Variabel	β_i	Estimasi Parameter
<i>Intercept</i>	β_0	-2,143
x_1	β_1	-1,576
x_2	β_2	-0,150
x_3	β_3	-0,131
x_4	β_4	0,543
x_5	β_5	0,395
x_6	β_6	0,485
x_7	β_7	0,080
x_8	β_8	1,334
x_9	β_9	5,210

Tabel 4.12 menunjukkan hasil estimasi parameter model regresi logistik setelah proses iteratif menggunakan metode *newton-raphson*. Dengan tingkat konvergensi $\varepsilon = 0,000001$, proses iterasi untuk semua parameter telah konvergen pada iterasi ke-7 sudah sangat kecil, dengan tingkat konvergensi sebesar $\varepsilon = 0,0000000007$.

4.6 Uji Multikolinearitas

Uji multikolinear bertujuan untuk mengidentifikasi adanya hubungan linear antara variabel-variabel independen. Salah satu metode yang digunakan untuk mengidentifikasi multikolinieritas adalah dengan berdasarkan nilai VIF.

1. Hipotesis dalam pengujian ini adalah:

H_0 : Tidak terjadi pelanggaran multikolinearitas

H_1 : Terjadi pelanggaran multikolinearitas

2. Taraf signifikansi:

Pengujian menggunakan taraf signifikansi $\alpha = 5\%$ (0,05)

3. Kriteria penolakan:

Tolak H_0 jika $p\text{-value} < \alpha$ atau nilai $VIF \geq 10$ yang artinya terjadi pelanggaran multikolinearitas. Nilai VIF untuk setiap variabel independen disajikan pada Tabel 4.13. Perhitungan manual nilai VIF berdasarkan persamaan (2.25) sebagai berikut:

Perhitungan untuk variabel x_1 :

$$\begin{aligned} R_1^2 &= \frac{SS_{total} - SS_{error}}{SS_{total}} \\ &= \frac{210,394 - 180,455}{210,394} \\ &= \frac{29,939}{210,394} \\ &= 0,142 \end{aligned}$$

$$\begin{aligned} Tolerance &= \frac{1}{1 - R^2} \\ &= \frac{1}{1 - 0,142} \end{aligned}$$

$$\begin{aligned}
 &= 0,858 \\
 \text{VIF} &= \frac{1}{\text{Tolerance}} \\
 &= \frac{1}{0,858} \\
 &= 1,166
 \end{aligned}$$

Dengan cara yang sama, untuk variabel x_2, x_3, \dots, x_9 diperoleh hasil sebagaimana disajikan pada Tabel 4.13.

Tabel 4.13 Nilai VIF Setiap Variabel Independen

Variabel	VIF	Keputusan
x_1	1,166	Terima H_0
x_2	1,139	Terima H_0
x_3	1,037	Terima H_0
x_4	1,043	Terima H_0
x_5	1,087	Terima H_0
x_6	1,047	Terima H_0
x_7	1,053	Terima H_0
x_8	1,017	Terima H_0
x_9	1,012	Terima H_0

Tabel 4.13 menunjukkan nilai VIF untuk masing-masing variabel. Berdasarkan data dalam tabel, seluruh variabel yang memiliki nilai VIF kurang dari atau berada jauh di bawah ambang batas umum sebesar 10. Nilai VIF tertinggi adalah 1,166 yang dimiliki oleh variabel x_1 , sedangkan nilai VIF terendah adalah 1,012 pada variabel x_9 . Dengan demikian, menunjukkan bahwa keputusan terima H_0 yang berarti tidak terdapat masalah multikolinearitas yang signifikan dalam model atau tidak terdapat korelasi tinggi antar variabel.

4.7 Uji Signifikansi Parameter

Pengujian parameter terbagi menjadi dua, yaitu uji serentak atau simultan yang mengevaluasi signifikansi variabel independen keseluruhan terhadap variabel dependen dan uji parsial yang mengevaluasi signifikansi tiap-tiap variabel independen secara terpisah.

4.7.1 Uji Simultan

Uji simultan dilakukan untuk mengidentifikasi secara keseluruhan bagaimana variabel-variabel saling berhubungan dan berpengaruh dalam suatu model. Uji yang digunakan adalah *likelihood ratio test* atau uji G .

1. Hipotesis uji G :

$$H_0: \beta_0 = \beta_1 = \beta_2 = \dots = \beta_9 = 0$$

$$H_1: \text{Minimal terdapat satu } \beta_j \neq 0 \text{ dengan } j = 0, 1, 2, \dots, 9$$

2. Taraf signifikansi:

Pengujian menggunakan taraf signifikansi $\alpha = 5\%$ (0,05)

3. Kriteria penolakan:

Tolak H_0 jika $p\text{-value} < \alpha$ atau nilai $G > \chi_{\alpha, v}^2$ artinya variabel independen secara bersama-sama berpengaruh signifikan terhadap variabel dependen.

4. Statistika uji:

Hasil dari uji signifikansi parameter secara simultan dengan *software R Studio* disajikan dalam Tabel 4.14. Perhitungan manual uji G berdasarkan persamaan (2.17) sebagai berikut:

$$\begin{aligned} G &= 2(L_{full} - L_{null}) \\ &= 2(1936,363 - 1034,961) \\ &= 901,402 \end{aligned}$$

Tabel 4.14 Uji Signifikansi Simultan

	G	Df	P-Value	χ^2
Model	901,402	9	0,000	16,919

Tabel 4.14 menunjukkan hasil uji signifikansi parameter secara simultan menghasilkan $p\text{-value} = 0,000 < \alpha = 0,05$. Nilai statistik uji G yang diperoleh sebesar 901,402 dengan derajat kebebasan sebanyak 9, sesuai dengan jumlah variabel independen yang diuji. Karena $p\text{-value} = 0,000 < \alpha = 0,05$ dan nilai uji $G = 901,402 > \chi^2 = 16,919$, maka keputusannya adalah tolak H_0 yang berarti variabel independen yaitu jenis kelamin, konsumsi rokok, aktivitas fisik, konsumsi gula berlebih, konsumsi garam berlebih, lemak berlebih, konsumsi buah dan sayur, konsumsi alkohol, dan usia secara bersama-sama berpengaruh signifikan terhadap diagnosa hipertensi.

4.7.2 Uji Parsial

Uji simultan dilakukan untuk menguji signifikansi secara parsial variabel independen terhadap model. Uji yang digunakan adalah *wald test*.

1. Hipotesis yang dipakai adalah sebagai berikut:

$H_0: \beta_j = 0$ (Variabel ke - j tidak memiliki pengaruh signifikan)

$H_1: \beta_j \neq 0$ dengan $j = 0, 1, 2, \dots, 9$ (Variabel ke - j memiliki pengaruh signifikan)

2. Taraf signifikansi:

Pengujian menggunakan taraf signifikansi $\alpha = 5\% (0,05)$

3. Kriteria penolakan:

Tolak H_0 jika $p\text{-value} < \alpha$ atau nilai $W > \chi_{0,05,1}^2 = 3,841$, yang artinya variabel independen ke- j memiliki pengaruh signifikan terhadap variabel dependen.

4. Statistika uji:

Hasil dari uji signifikansi parameter secara parsial dengan *software* R Studio disajikan dalam Tabel 4.15. Perhitungan nilai *wald* berdasarkan persamaan (2.19) sebagai berikut:

$$\begin{aligned} W_0 &= \left[\frac{\hat{\beta}_0}{SE(\hat{\beta}_0)} \right]^2 \\ &= \left[\frac{-2,143}{0,144} \right]^2 \\ &= 220,352 \end{aligned}$$

Dengan cara yang sama, untuk variabel $x_1, x_2, x_3, \dots, x_9$ diperoleh hasil sebagaimana disajikan pada Tabel 4.15.

Tabel 4.15 Uji Signifikansi Parsial

Parameter	Estimasi	Std. Error	P-Value	Wald	Keputusan
β_0	-2,143	0,144	0,000	220,352	Tolak H_0
$\beta_1(1)$	-1,576	0,341	0,000	21,347	Tolak H_0
$\beta_2(1)$	-0,150	0,286	0,601	0,273	Terima H_0
$\beta_3(1)$	-0,131	0,172	0,447	0,579	Terima H_0
$\beta_4(1)$	0,543	0,181	0,003	9,019	Tolak H_0
$\beta_5(1)$	0,395	0,211	0,061	3,500	Terima H_0
$\beta_6(1)$	0,485	0,180	0,007	7,279	Tolak H_0
$\beta_7(1)$	0,080	0,175	0,646	0,210	Terima H_0
$\beta_8(1)$	1,334	1,113	0,231	1,435	Terima H_0
$\beta_9(1)$	5,210	0,298	0,000	305,397	Tolak H_0

Tabel 4.15 menunjukkan bahwa hasil uji signifikansi parsial memperoleh 4 variabel independen yang berpengaruh terhadap diagnosa hipertensi. Variabel independen tersebut diantaranya adalah x_1 (jenis kelamin), x_4 (konsumsi gula berlebih), x_6 (lemak berlebih), dan x_9 (usia). Pada variabel x_1 menghasilkan $p\text{-value} = 0,000 < \alpha = 0,05$. Dengan demikian keputusannya adalah tolak H_0 yang artinya variabel independen jenis kelamin memiliki pengaruh signifikan terhadap diagnosa hipertensi. Variabel x_4 diperoleh nilai $p\text{-value} = 0,003 < \alpha = 0,05$. Dengan demikian keputusannya adalah tolak H_0 yang artinya variabel independen konsumsi gula berlebih memiliki pengaruh signifikan terhadap diagnosa hipertensi.

Variabel x_6 diperoleh nilai $p\text{-value} = 0,007 < \alpha = 0,05$. Dengan demikian keputusannya adalah tolak H_0 yang artinya variabel independen lemak berlebih memiliki pengaruh signifikan terhadap diagnosa hipertensi. Sedangkan variabel x_9 diperoleh nilai $p\text{-value} = 0,000 < \alpha = 0,05$. Dengan demikian keputusannya adalah tolak H_0 yang artinya variabel independen usia memiliki pengaruh signifikan terhadap diagnosa hipertensi.

Sedangkan variabel lainnya seperti x_2 (konsumsi rokok), x_3 (aktivitas fisik), x_5 (konsumsi garam berlebih), x_7 (konsumsi sayur dan buah), dan x_8 (konsumsi alkohol) menghasilkan nilai $p\text{-value}$ yang lebih besar dari α . Dengan demikian, keputusannya adalah terima H_0 yang berarti variabel-variabel tersebut tidak memiliki pengaruh secara parsial terhadap diagnosa hipertensi.

Berdasarkan uji signifikansi parsial diketahui bahwa terdapat 4 variabel yang signifikan dan terdapat 5 variabel independen yang tidak signifikan. Hal tersebut menunjukkan bahwa perlu dilakukan uji signifikansi parsial yang kedua

dengan hanya melibatkan 4 variabel yang signifikan. Hasil uji signifikansi parsial kedua dengan *software* R Studio disajikan pada Tabel 4.16.

Tabel 4.16 Uji Signifikansi Parsial Kedua

Parameter	Estimasi	Std. Error	P-Value	Wald	Keputusan
β_0	-2,130	0,119	0,000	319,636	Tolak H_0
$\beta_1(1)$	-1,608	0,323	0,000	24,775	Tolak H_0
$\beta_4(1)$	0,594	0,178	0,001	11,085	Tolak H_0
$\beta_6(1)$	0,552	0,175	0,002	9,904	Tolak H_0
$\beta_9(1)$	5,189	0,296	0,000	308,141	Tolak H_0

Berdasarkan Tabel 4.16 menunjukkan hasil uji signifikansi parsial kedua yang hanya menggunakan 4 variabel independen yaitu x_1 (jenis kelamin), x_4 (konsumsi gula berlebih), x_6 (lemak berlebih), dan x_9 (usia). Variabel x_1 menghasilkan $p\text{-value} = 0,000 < \alpha = 0,05$ dan $W = 24,775 > \chi_{0,05,1}^2 = 3,841$, sehingga keputusannya adalah tolak H_0 yang artinya variabel independen jenis kelamin memiliki pengaruh signifikan terhadap diagnosa hipertensi. Variabel x_4 menghasilkan $p\text{-value} = 0,001 < \alpha = 0,05$ dan $W = 11,085 > \chi_{0,05,1}^2 = 3,841$, sehingga keputusannya adalah tolak H_0 yang artinya variabel independen konsumsi gula berlebih memiliki pengaruh signifikan terhadap diagnosa hipertensi.

Variabel x_6 menghasilkan $p\text{-value} = 0,002 < \alpha = 0,05$ dan $W = 9,904 > \chi_{0,05,1}^2 = 3,841$, sehingga keputusannya adalah tolak H_0 yang artinya variabel independen lemak berlebih memiliki pengaruh signifikan terhadap diagnosa hipertensi. Sedangkan untuk variabel x_9 menghasilkan $p\text{-value} = 0,000 < \alpha = 0,05$ dan $W = 308,141 > \chi_{0,05,1}^2 = 3,841$, sehingga keputusannya adalah tolak

H_0 yang artinya variabel independen usia memiliki pengaruh signifikan terhadap diagnosa hipertensi.

Keempat variabel independen yang signifikan tersebut kemudian dilakukan uji signifikansi simultan untuk mengetahui bagaimana keempat variabel independen tersebut secara bersama-sama mempengaruhi variabel diagnosa hipertensi. Hasil dari uji signifikansi secara simultan pada keempat variabel independen yang signifikan dengan *software* R Studio disajikan pada Tabel 4.17. Perhitungan manual uji G berdasarkan persamaan (2.17) sebagai berikut:

$$\begin{aligned} G &= 2(L_{full} - L_{null}) \\ &= 2(1936,363 - 1040,603) \\ &= 895,760 \end{aligned}$$

Tabel 4.17 Uji Signifikansi Simultan Kedua

	G	Df	P-Value	χ^2
Model	895,760	4	0,000	9,487

Berdasarkan Tabel 4.17 menunjukkan hasil uji signifikansi parameter secara simultan kedua menghasilkan $p\text{-value} = 0,000 < \alpha = 0,05$. Dengan demikian, keputusannya adalah tolak H_0 yang berarti keempat variabel independen yaitu variabel jenis kelamin, konsumsi gula berlebih, lemak berlebih, dan usia secara bersama-sama berpengaruh signifikan terhadap diagnosa hipertensi.

4.8 Model Regresi Logistik

Model regresi logistik diperoleh berdasarkan pengujian signifikansi parameter secara simultan dan parsial. Berdasarkan hasil pengujian pada Tabel 4.14 dan 4.15 diperoleh variabel independen yang signifikan adalah variabel x_1 (jenis kelamin), x_4 (konsumsi gula berlebih), x_6 (lemak berlebih), dan x_9 (usia). Dengan demikian dari hasil estimasi koefisien regresi pada Tabel 4.16 diperoleh model regresi logistik berdasarkan persamaan (2.4):

$$\pi(X) = \frac{\exp(-2,130 - 1,608x_1(1) + 0,594x_4(1) + 0,552x_6(1) + 5,189x_9(1))}{1 + \exp(-2,130 - 1,608x_1(1) + 0,594x_4(1) + 0,552x_6(1) + 5,189x_9(1))}$$

Model regresi logistik untuk setiap diagnosa hipertensi berdasarkan 4 variabel independen yang signifikan sebagai berikut:

1. Peluang perempuan hipertensi apabila $x_4 = 0$, $x_6 = 0$ dan $x_9 = 0$:

$$\begin{aligned}\pi(x_1 = 0) &= \frac{\exp(-2,130 - 1,608(0) + 0,594(0) + 0,552(0) + 5,189(0))}{1 + \exp(-2,130 - 1,608(0) + 0,594(0) + 0,552(0) + 5,189(0))} \\ &= 0,106\end{aligned}$$

Hasil tersebut menunjukkan bahwa ketika perempuan pada usia yang tidak rentan terkena hipertensi serta tidak mengonsumsi gula dan lemak berlebih, peluang mengalami hipertensi yaitu sekitar 10,6%.

2. Peluang perempuan hipertensi apabila $x_4 = 1$, $x_6 = 1$ dan $x_9 = 1$:

$$\begin{aligned}\pi(x_1 = 0) &= \frac{\exp(-2,130 - 1,608(0) + 0,594(1) + 0,552(1) + 5,189(1))}{1 + \exp(-2,130 - 1,608(0) + 0,594(1) + 0,552(1) + 5,189(1))} \\ &= 0,985\end{aligned}$$

Hasil tersebut menunjukkan bahwa ketika perempuan yang berusia lanjut atau pada usia yang rentan terkena hipertensi mengonsumsi gula dan lemak berlebih, peluang mengalami hipertensi sangat besar, yaitu sekitar 98,5%.

3. Peluang laki-laki hipertensi apabila $x_4 = 0$, $x_6 = 0$ dan $x_9 = 0$:

$$\begin{aligned}\pi(x_1 = 1) &= \frac{\exp(-2,130-1,608(1)+0,594(0)+0,552(0)+5,189(0))}{1+\exp(-2,130-1,608(1)+0,594(0)+0,552(0)+5,189(0))} \\ &= 0,023\end{aligned}$$

Hasil tersebut menunjukkan bahwa jika laki-laki pada usia yang tidak rentan terkena hipertensi serta tidak mengonsumsi gula dan lemak berlebih, peluang mengalami hipertensi yaitu sekitar 2,3%.

4. Peluang laki-laki hipertensi apabila $x_4 = 1$, $x_6 = 1$ dan $x_9 = 1$:

$$\begin{aligned}\pi(x_1 = 1) &= \frac{\exp(-2,130-1,608(1)+0,594(1)+0,552(1)+5,189(1))}{1+\exp(-2,130-1,608(1)+0,594(1)+0,552(1)+5,189(1))} \\ &= 0,930\end{aligned}$$

Hasil tersebut menunjukkan bahwa jika laki-laki yang berusia lanjut atau pada usia yang rentan terkena hipertensi mengonsumsi gula dan lemak berlebih, peluang mengalami hipertensi sangat besar, yaitu sekitar 93%.

4.9 Uji Kesesuaian Model

Uji kesesuaian model dijalankan guna menentukan apakah ditemukan ketaksamaan antara data yang diamati atau data hasil observasi dengan prediksi yang dihasilkan oleh model.

1. Hipotesis dalam uji kesesuaian model seperti berikut:

H_0 : Model sudah sesuai (terdapat kesamaan antara hasil observasi dengan peluang perkiraan model)

H_1 : Model belum sesuai (terdapat ketaksamaan antara hasil observasi dengan peluang perkiraan model)

2. Taraf signifikansi:

Pengujian menggunakan taraf signifikansi $\alpha = 5\%$ (0,05)

3. Statistik uji:

$$D = -2 \sum_{i=1}^N \left[y_i \ln \left(\frac{\hat{\pi}(X_i)}{y_i} \right) + (1 - y_i) \ln \left(\frac{1 - \hat{\pi}(X_i)}{1 - y_i} \right) \right]$$

4. Kriteria penolakan:

Tolak H_0 jika $p\text{-value} < \alpha$ atau nilai $D > \chi_{(df)}^2$, yang berarti model belum sesuai atau terdapat ketaksamaan antara hasil obeservasi dengan peluang perkiraan model.

Hasil dari uji kesesuaian model dengan *software* R Studio disajikan pada Tabel 4.18.

Tabel 4.18 Uji Kesesuaian Model

	<i>Deviance</i>	DF	<i>P-Value</i>	χ^2
Model	1040,603	1601	1,000	1695,2

Tabel 4.18 menunjukkan hasil uji kesesuaian model yaitu $p\text{-value} = 1,0000 > \alpha = 0,05$. Dapat dilihat juga bahwa nilai *deviance* sebesar $1040,603 < \chi_{(df)}^2 = 1695,2$. Dengan demikian, keputusannya adalah terima H_0 yang berarti model yang digunakan telah sesuai.

4.10 Interpretasi Koefisien Parameter Model

Interpretasi koefisien model regresi logistik dapat dilakukan dengan berdasarkan nilai *odds ratio*. Hasil nilai *odds ratio* dengan *software* R Studio

disajikan pada Tabel 4.19. Perhitungan manual nilai *odds ratio* berdasarkan persamaan (2.22) sebagai berikut:

Perhitungan nilai *odds ratio* untuk variabel x_1 :

$$\begin{aligned}\psi_1 &= \exp(\beta_1(1)) \\ &= \exp(-1,608) \\ &= 0,2004\end{aligned}$$

Dengan cara yang sama, untuk variabel x_4 , x_6 , dan x_9 diperoleh hasil sebagaimana disajikan pada Tabel 4.19.

Tabel 4.19 Nilai *Odds Ratio*

Variabel	Estimasi	<i>Odds Ratio</i>
<i>Intercept</i>	-2,130	0,1188
x_1	-1,608	0,2004
x_4	0,594	1,8117
x_5	0,552	1,7372
x_9	5,189	179,28

Tabel 4.19 menunjukkan hasil perhitungan nilai *odds ratio* untuk mengetahui seberapa besar pengaruh variabel independen terhadap diagnosa hipertensi di Puskesmas Mojolangu. Interpretasi koefisien parameter dari variabel jenis kelamin laki-laki dengan estimasi parameter negatif menghasilkan nilai *odds ratio* sebesar 0,2004. Hal tersebut menunjukkan bahwa pasien laki-laki memiliki risiko lebih kecil dibandingkan dengan perempuan untuk menjadi kategori hipertensi.

Variabel konsumsi gula berlebih dengan estimasi parameter 0,594 menghasilkan nilai *odds ratio* sebesar 1,8117. Hal tersebut menunjukkan bahwa pasien yang mengonsumsi atau memiliki kadar gula berlebih memiliki risiko 1,8117 kali lebih tinggi dibandingkan dengan pasien yang tidak mengonsumsi atau memiliki kadar gula berlebih untuk menjadi kategori hipertensi.

Variabel lemak berlebih dengan estimasi parameter 0,552 menghasilkan nilai *odds ratio* sebesar 1,7372. Hal tersebut menunjukkan bahwa pasien yang mengonsumsi atau memiliki kadar lemak berlebih memiliki risiko 1,7372 kali lebih tinggi dibandingkan dengan pasien yang tidak mengonsumsi atau memiliki kadar lemak berlebih untuk menjadi kategori hipertensi. Sedangkan variabel usia dengan estimasi parameter 5,189 menghasilkan nilai *odds ratio* sebesar 179,28.

Pasien yang berusia lanjut atau berada pada usia yang rentan terkena hipertensi memiliki risiko lebih tinggi dibandingkan dengan pasien yang masih muda atau tidak berada pada usia rentan untuk menjadi kategori hipertensi. Seseorang pada usia lanjut cenderung memiliki fisik yang lemah. Secara keseluruhan, hasil ini menunjukkan bahwa keempat variabel independen memiliki pengaruh yang terhadap diagnosis hipertensi di Puskesmas Mojolangu.

4.11 Ketepatan Klasifikasi Model

Ketepatan klasifikasi model digunakan untuk mengetahui seberapa layak suatu model dapat mengklasifikasikan kelas yang benar dari suatu data. Perhitungan ketepatan klasifikasi berdasarkan hasil observasi dan prediksi klasifikasi disajikan dalam bentuk *confusion matrix* pada Tabel 4.20.

Tabel 4.20 menunjukkan *confusion matrix* hasil klasifikasi model regresi logistik yang digunakan untuk memprediksi diagnosis hipertensi. Berdasarkan tabel, diketahui bahwa sebanyak 483 kasus yang sebenarnya tidak mengalami hipertensi (y_0) berhasil diklasifikasikan dengan benar sebagai tidak hipertensi, dan 159 kasus yang benar-benar hipertensi (y_1) juga berhasil diklasifikasikan secara akurat. Akan tetapi, terdapat 8 kasus hipertensi yang salah diklasifikasikan sebagai

tidak hipertensi dan terdapat 67 kasus tidak hipertensi yang diklasifikasikan sebagai hipertensi.

Tabel 4.20 *Confusion Matrix* Hasil Klasifikasi Model

Hasil Observasi	Prediksi		Total
	y_0	y_1	
y_0	483	67	550
y_1	8	159	167
Total	491	226	717

8 kasus hipertensi yang salah diklasifikasikan sebagai tidak hipertensi kemungkinan besar disebabkan oleh faktor lain pada saat pemeriksaan. Misalnya pasien sedang menjalani pengobatan antihipertensi pada saat pemeriksaan atau ketika data diambil, sehingga tekanan darah yang tercatat berada pada kisaran normal atau hampir normal. Kondisi tersebut dapat menyebabkan model gagal mendeteksi bahwa pasien sebenarnya adalah penderita hipertensi.

1. Perhitungan nilai APER untuk tingkat kesalahan model dalam klasifikasi berdasarkan persamaan (2.23) sebagai berikut:

$$\begin{aligned}
 \text{APER} &= \frac{n_{01} + n_{10}}{n} \\
 &= \frac{67 + 8}{717} \\
 &= 0,105 \\
 &= 10,5\%
 \end{aligned}$$

2. Perhitungan tingkat ketepatan model dalam klasifikasi berdasarkan persamaan (2.24) sebagai berikut:

$$1 - \text{APER} = \frac{n_{00} + n_{11}}{n}$$

$$\begin{aligned} &= \frac{483 + 159}{717} \\ &= 0,895 \\ &= 89,5\% \end{aligned}$$

Berdasarkan perhitungan nilai APER, diketahui bahwa persentase kesalahan model dalam mengklasifikasikan kelas sebesar 10,5%. Nilai ini menunjukkan bahwa tingkat kesalahan model sangat rendah, yang berarti hanya sekitar 10,5% dari total data yang diklasifikasikan secara tidak tepat. Sebaliknya hasil ketepatan model dalam mengklasifikasikan sebesar 89,5%. Hal ini menunjukkan bahwa model mampu mengklasifikasikan data dengan baik. Dengan tingkat akurasi yang tinggi dan kesalahan yang rendah, model regresi logistik yang dibangun dapat disimpulkan memiliki performa yang baik dalam memprediksi diagnosis hipertensi berdasarkan variabel-variabel yang digunakan.

Selain itu, variabel-variabel yang terbukti signifikan dalam model regresi logistik, yaitu x_1 , x_4 , x_6 , dan x_9 , menunjukkan kontribusi yang bermakna dalam membedakan antara individu yang terdiagnosis hipertensi dan yang tidak. Hasil estimasi model menunjukkan bahwa kombinasi variabel-variabel tersebut mampu meningkatkan akurasi klasifikasi hingga mencapai 89,5%. Tingkat ketepatan yang tinggi ini mengindikasikan bahwa model yang dibangun memiliki kinerja prediktif yang baik dan layak digunakan sebagai alat bantu dalam proses pengambilan keputusan berbasis data.

Model ini berpotensi untuk dimanfaatkan oleh instansi terkait, seperti Dinas Kesehatan, dalam merumuskan strategi pencegahan dan intervensi yang lebih efektif dan tepat sasaran. Selain itu, identifikasi variabel-variabel yang berpengaruh

signifikan terhadap risiko hipertensi juga dapat dijadikan dasar untuk menyusun program edukasi kesehatan masyarakat. Penyebarluasan informasi mengenai faktor-faktor risiko tersebut diharapkan dapat meningkatkan kesadaran masyarakat terhadap pentingnya upaya preventif, sehingga dapat mendorong perubahan perilaku menuju gaya hidup yang lebih sehat. Dengan demikian, model ini tidak hanya bermanfaat secara akademis, tetapi juga memiliki implikasi praktis dalam upaya penanggulangan hipertensi di tingkat populasi.

4.12 Kajian Penelitian Dalam Prespektif Islam

Berdasarkan hasil penelitian diagnosa hipertensi, diketahui bahwa tekanan sistolik dan tekanan diastolik memberikan pengaruh terhadap diagnosa hipertensi. Tekanan sistolik dan diastolik dalam tubuh yang meningkat secara tidak normal menunjukkan terganggunya sistem peredaran dan tekanan darah yang berpengaruh juga pada keseimbangan tubuh. Keseimbangan tubuh yang terganggu dapat berdampak pada kesehatan tubuh. Hal tersebut menunjukkan bahwa tekanan sistolik dan diastolik pada seseorang menjadi faktor yang perlu diperhatikan untuk menjaga kesehatan tubuh. Hal tersebut selaras dengan ajaran Islam yang tercantum dalam QS. Al-Infithar ayat 6 sampai 8 sebagai berikut:

يَا أَيُّهَا الْإِنْسَانُ مَا غَرَّبَكَ بِرَبِّكَ الْكَرِيمِ ﴿٦﴾ الَّذِي خَلَقَكَ فَسَوَّاكَ فَعَدَلَكَ ﴿٧﴾ فِي أَيِّ صُورَةٍ مَّا شَاءَ رَكَّبَكَ ﴿٨﴾

Artinya: “Wahai manusia, apakah yang telah memperdayakanmu (berbuat durhaka) terhadap Tuhanmu yang Maha Pengasih? Yang telah menciptakan kamu lalu menyempurnakan kejadianmu dan menjadikan (susunan tubuh)-mu seimbang? Dalam bentuk apa saja yang Dia kehendaki, Dia menyusun tubuhmu.”

Berdasarkan ayat tersebut Allah SWT memberikan peringatan dan teguran kepada umatnya agar tidak lalai terhadap nikmat penciptaan yang sempurna. Allah SWT telah menciptakan manusia dengan kesempurnaan bentuk, keseimbangan struktur dan sistem biologis tubuh yang teratur. Frasa “menjadikan (susunan tubuh)-mu seimbang” menunjukkan bahwa setiap sistem dalam tubuh manusia termasuk sistem peredaran dan tekanan darah telah dirancang dalam keseimbangan yang sangat presisi. Apabila sistem peredaran dan tekanan dalam darah memiliki tekanan yang tinggi dan tidak normal, kondisi tersebut mencerminkan keseimbangan tubuh yang terganggu.

Sistem peredaran dan tekanan dalam darah telah Allah SWT ciptakan dengan fungsi yang sangat teratur. Jantung memompa darah sesuai ritme tertentu agar darah dapat teralirkan keseluruh tubuh dengan stabil. Seseorang yang tidak menjaga keseimbangan tubuh, maka tubuh akan mudah untuk terpapar penyakit terutama hipertensi. Lebih dalam lagi, Allah SWT telah menyebutkan dalam frasa “telah memperdayakanmu” yang berarti bahwa manusia sering lalai akan rahmat-Nya. Dalam konteks hipertensi, banyak orang yang baru menyadari pentingnya menjaga keseimbangan dan kesehatan tubuh setelah mengalami hipertensi atau bahkan komplikasi.

Menjaga tekanan darah agar tetap seimbang merupakan bagian dari tanggung jawab spiritual dalam menghargai tubuh sebagai amanah dari Allah SWT. Menjaga tekanan darah agar tetap stabil bukan hanya tindakan medis agar terhindar dari hipertensi, tetapi juga bagian dari ibadah dan manifestasi rasa syukur atas kesempurnaan ciptaan Allah SWT. Seiring bertambahnya usia, elastisitas pembuluh darah menurun sehingga menyebabkan penyempitan pembuluh darah atau

kekakuan. Penyempitan pembuluh darah menyebabkan aliran darah terhambat. Hal tersebut akan membuat jantung memompa darah lebih kuat yang menyebabkan tekanan darah akan meningkat.

Al-Qur'an memiliki banyak ayat yang mengingatkan manusia akan fase kehidupan termasuk masa tua sebagai bagian dari perjalanan yang harus disiapkan dengan amal sholeh, kesadaran, dan kesehatan tubuh. Salah satunya terdapat dalam QS. Ar-Rum ayat 54 sebagai berikut:

اللَّهُ الَّذِي خَلَقَكُمْ مِنْ ضَعْفٍ ثُمَّ جَعَلَ مِنْ بَعْدِ ضَعْفٍ قُوَّةً ثُمَّ جَعَلَ مِنْ بَعْدِ قُوَّةٍ ضَعْفًا وَشَيْبَةً يَخْلُقُ مَا يَشَاءُ
وَهُوَ الْعَلِيمُ الْقَدِيرُ ﴿٥٤﴾

Artinya: “Allah, Dialah yang menciptakan kamu dalam kondisi lemah sejak awal, lalu menjadikanmu kuat setelah masa kelemahan itu. Kemudian, setelah masa kuat itu, Dia mengembalikannya pada kelemahan dan uban. Allah menciptakan segala sesuatu sesuai kehendak-Nya, dan Dia Maha Mengetahui serta Maha Kuasa atas segala sesuatu.”

Berdasarkan ayat tersebut, manusia diisyaratkan bahwa bertambahnya usia atau penuaan merupakan fase alamiah dalam kehidupan yang mengharuskan seseorang untuk lebih merawat tubuh terhadap kelemahan fisik, termasuk potensi terkena hipertensi. Dengan demikian, semakin tua usia seseorang, maka kondisi tubuhnya akan semakin rapuh dan rentang terhadap penyakit. Nabi Muhammad SAW menekankan pentingnya menjaga kesehatan sebelum datangnya sakit dan usia tua. Dari Ibnu ‘Abbas, Nabi Muhammad SAW sabdanya dalam hadist:

إِغْتَنِمِ خَمْسًا قَبْلَ خَمْسٍ : شَبَابَكَ قَبْلَ هَرَمِكَ وَصِحَّتَكَ قَبْلَ سَقَمِكَ وَغِنَاكَ قَبْلَ فُقْرِكَ وَفَرَاغَكَ قَبْلَ شُغْلِكَ وَحَيَاتَكَ قَبْلَ مَوْتِكَ

Artinya: *“Manfaatkan lima perkara sebelum lima perkara: masa mudamu sebelum datang masa tuamu, masa sehatmu sebelum datang masa sakitmu, masa kayamu sebelum datang masa kefakiranmu, masa luangmu sebelum datang masa sibukmu, dan hidupmu sebelum datang matimu”* (HR Al Hakim).

Hadits tersebut diriwayatkan oleh Imam Al-Hakim dalam kitabnya yang terkenal yaitu "Al-Mustadrak 'ala Al-Sahihain", dikatakan oleh Adz-Dzahabi dalam kitabnya At-Talkhish berdasarkan syarat Bukhari dan Muslim serta hadits ini dikatakan shahih oleh Syaikh Al-Albani dalam kitabnya yaitu Al-Jami' Ash-Shaghir (Supandi, Istikhori, & Suryana, 2025). Berdasarkan hadist tersebut, dapat diketahui bahwa Islam tidak memisahkan aspek fisik dan spiritual dalam memaknai kesehatan. Hadist ini menjelaskan bahwa masa muda dan sehat harus dimanfaatkan untuk membangun tubuh agar tetap sehat dan tidak menyesal di masa tua saat tubuh mulai rentan terhadap penyakit, termasuk hipertensi.

Integrasi antara ilmu kesehatan dan nilai-nilai Islam tentang bertambahnya usia terhadap hipertensi menciptakan pendekatan yang tidak hanya fokus pada aspek klinis, tetapi juga menghubungkan jasmani dan rohani. Praktiknya dalam kehidupan sehari-hari adalah kesadaran akan tanggung jawab terhadap tubuh seiring bertambahnya usia. Tubuh yang sehat memungkinkan seseorang untuk lebih optimal dalam melaksanakan ibadah, berkontribusi dalam masyarakat, dan meraih keberkahan hidup. Islam memandang kesehatan sebagai bagian integral dari keberhasilan hidup dunia dan akhirat yang harus dijaga dengan kesadaran penuh.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan rumusan masalah dan pembahasan pada penelitian ini, dapat disimpulkan sebagai berikut:

1. Model regresi logistik pada reduksi *imbalanced data* diagnosa hipertensi dengan *Tomek Links* dengan variabel signifikan adalah

$$\pi(X) = \frac{\exp(-2,130 - 1,608x_1(1) + 0,594x_4(1) + 0,552x_6(1) + 5,189x_9(1))}{1 + \exp(-2,130 - 1,608x_1(1) + 0,594x_4(1) + 0,552x_6(1) + 5,189x_9(1))}$$

Berdasarkan hasil perhitungan peluang kejadian hipertensi, dapat disimpulkan bahwa faktor jenis kelamin, usia, serta konsumsi gula dan lemak memiliki pengaruh yang signifikan terhadap risiko hipertensi. Untuk perempuan yang berada pada usia yang tidak rentan terhadap hipertensi serta tidak mengonsumsi gula dan lemak berlebihan, peluang mengalami hipertensi relatif rendah, yakni sekitar 10,6%. Namun, ketika perempuan tersebut berada pada usia rentan dan memiliki kebiasaan mengonsumsi gula dan lemak secara berlebihan, peluangnya meningkat drastis menjadi sekitar 98,5%, yang menunjukkan adanya peningkatan risiko yang sangat tinggi. Sementara itu, untuk laki-laki yang juga berada pada usia yang tidak rentan dan memiliki pola konsumsi yang sehat, peluang mengalami hipertensi jauh lebih rendah, yaitu sekitar 2,3%. Namun, ketika laki-laki tersebut berada dalam kondisi yang lebih berisiko, yakni pada usia lanjut dan mengonsumsi gula serta lemak berlebihan, maka peluang terkena hipertensi meningkat tajam menjadi 93,0%.

2. Tingkat akurasi kesalahan model regresi logistik pada reduksi *imbalanced data* diagnosa hipertensi dengan *Tomek Links* diperoleh berdasarkan nilai APER sebesar 10,5%. Sebaliknya untuk persentase tingkat akurasi ketepatan model diperoleh 89,5%. Dengan demikian dapat disimpulkan bahwa, model regresi logistik mampu mengklasifikasikan dengan baik pada reduksi *imbalanced data* diagnosa hipertensi dengan *Tomek Links*.

5.2 Saran

Saran penulis untuk penelitian selanjutnya dan instansi terkait dalam upaya pengembangan dan penyempurnaan yang berkelanjutan sebagai berikut:

1. Penelitian selanjutnya disarankan untuk menggunakan kombinasi baru pada metode *imbalanced data*, metode klasifikasi yang berbeda, serta faktor-faktor penyebab hipertensi yang bervariasi guna memperoleh tingkat akurasi yang lebih optimal.
2. Dinas Kesehatan terkait disarankan untuk memanfaatkan model regresi logistik yang telah dikembangkan sebagai alat bantu dalam proses pengambilan keputusan berbasis data. Selanjutnya, variabel-variabel signifikan tersebut juga dapat dijadikan dasar dalam penyusunan strategi pencegahan yang lebih efektif dan tepat sasaran. Dinas Kesehatan diharapkan dapat mengintegrasikan temuan ini ke dalam program edukasi masyarakat, khususnya dalam meningkatkan kesadaran akan faktor risiko hipertensi.
3. Masyarakat umum disarankan untuk menggunakan informasi ini untuk menjaga kesehatan dengan lebih baik. Faktor yang berpengaruh signifikan seperti usia dan pola makan sangat berpengaruh terhadap tekanan darah.

DAFTAR PUSTAKA

- Afifah, D. N. (2020). Penerapan Metode Regresi Logistik Biner pada Kesejahteraan Rumah Tangga Di Kabupaten Mojokerto. [Skripsi, Universitas Islam Negeri Maulana Malik Ibrahim]. Malang.
- Agresti, A. (1990). *Categorical Data Analysis*. New York: John Wiley & Sons, Inc.
- Agresti, A. (2002). *Categorical Data Analysis* (3rd ed). New York: John Wiley & Sons, Inc.
- Agresti, A. (2007). *An Introduction to Categorical Data Analysis*. Wiley-Interscience.
- Agresti, A. (2013). *Categorical Data Analysis* (3rd ed., Vol. 742). New Jersey: John Wiley & Sonc, Inc.
- Al-Hakim, A. A. (1990). *Al-Mustadrak 'ala al-Sahihain*. Dar al-Kutub al-Ilmiyah.
- American Heart Association (2023). *High Blood Pressure and Women*. Diakses pada 24 Mei 2025, dari <https://www.heart.org>.
- Antipov, E., & Pokryshevskaya, E. (2009). *Applying CHAID for logistic regression diagnostics and classification accuracy improvement*. The State University Higher School of Economics. Munich Personal RePEc Archive.
- An-Nawawi, M. Y. B. S. (2010). *Hadits Arba'in Nawawiyah (Indonesia)*. Abdullah Haidhir (Penerj.), Dr. Muh. Mu'inudinillah Basri & Maerwandi Tarmizi (Murajaah). Maktab Dakwah dan Bimbingan Jaliyat Rabwah.
- Batista, G. E., Bazzan, A. L., & Monard, M. C. (2003). Balancing Training Data for Automated Annotation of Keywords: a Case Study. [https://doi.org/10.1016/S0031-3203\(02\)00257-1](https://doi.org/10.1016/S0031-3203(02)00257-1). Diakses pada 15 November 2024.
- Batista, G., Prati, R. C., & Monard, M. C. (2004). A Study of The Behavior of Several Methods for Balancing Machine Learning Training Data. *SIGKDD Explorations Newsletter*, 6(1), 20-29. <https://doi.org/10.1145/1007730.1007735>. Diakses pada 15 November 2024.
- Choirunnisa, S. (2019). *Metode hibrida oversampling dan undersampling untuk menangani ketidakseimbangan data kegagalan akademik Universitas XYZ* [Thesis, Institut Teknologi Sepuluh Nopember]. Surabaya.
- Dewi, I. A. M. C., Dharmendra, I. K., & Setiasih, N. W. (2023). Analisis sentimen review aplikasi Satu Sehat Mobile menggunakan model sampling Tomek Links.

Jurnal Informatika dan Komputer, 12(3), 45-55.
<https://doi.org/10.36002/jutik.v9i5.2644>. Diakses pada 15 November 2024.

Draper, N. R., & Smith, H. (1998). *Applied Regression Analysis* (3rd ed.). Canada: John Wiley & Sons.

Fatah, K. S., & Mahmood, R. F. (2016). Parameter estimation for binary logistic regression using different iterative methods. *Journal of Zankoy Sulaimani*, 19(2), 177–178. <https://doi.org/10.17656/jzs.10621>. Diakses pada 15 Maret 2025.

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Technique* (3rd ed.). United States of America: Elsevier Inc.

Hocking, R. R., (2003). *Methods and Applications of Linear Models* (2nd ed.). New Jersey: John Wiley & Sons.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression* (2nd ed.). New York, United State of America: John Wiley & Sons.

Johnson, R. A., & Wichern, D. (2007). *Applied Multivariate Statistical Analysis* (6th ed.). New Jersey: Pearson Education, Inc.

Kementerian Agama Republik Indonesia. (2023). *Hukum Makan dan Minum Sambil Berdiri*. Diakses pada 6 November 2024, dari <https://kemenag.go.id>.

Kementerian Kesehatan Republik Indonesia. (2023). *Hipertensi – The Silent Killer*. Diakses pada 6 November 2024, dari <https://www.kemkes.go.id>.

Kementerian Kesehatan Republik Indonesia. (2024). *Bahaya Hipertensi, Upaya Pencegahan dan Pengendalian Hipertensi*. Diakses pada 6 November 2024, dari <https://www.kemkes.go.id>.

Kumalasanti, R., & Aprilianti, N. M. D. (2024). Sentiment analysis of Bali calendar application reviews using K-nearest neighbour. *International Journal of Engineering Technology and Natural Sciences*, 6(1), 70-73. <https://doi.org/10.46923/ijets.v6i1.339>. Diakses pada 15 November 2024.

Manembu, M., Rumampuk, J., & Danes, V. R. (2015). Pengaruh posisi duduk dan berdiri terhadap tekanan darah sistolik dan diastolik pada pegawai negeri sipil Kabupaten Minahasa Utara. *Jurnal e-Biomedik (eBm)*, 3(3), 817-819. <https://doi.org/10.35790/ebm.v3i3.10150>. Diakses pada 15 November 2024.

Misna, Rais, & Utami, I. T. (2018). Analisis regresi logistik biner untuk mengklasifikasi penderita hipertensi berdasarkan kebiasaan merokok di RSU Mokopido Toli-Toli. *Natural Science: Journal of Science and Technology*, 7(3), 341–348. <https://api.core.ac.uk/oai/oai:ojs.pkp.sfu.ca:article/11464>. Diakses pada 15 November 2024.

- Nilamsari, N. (2014). Memahami Studi Dokumen dalam Penelitian Kualitatif. *Wacana Jurnal Ilmiah Ilmu Komunikasi*, 13(2), 177–181. <https://doi.org/10.32509/wacana.v13i2.143>. Diakses pada 15 Oktober 2024.
- Prasetyorini, H., & Prawesti, D. (2012). Stress With the Incidence of Hipertension Complications to Patients with Hypertension. *Nursing Journal*, 5, 61-70. <https://www.neliti.com/id/publications/210184/stress-with-the-incidence-of-hipertension-complications-to-patients-with-hyperte>. Diakses pada 15 November 2024.
- Pratama, R. N. (2018). Regresi Logistik Biner Untuk Mengetahui Faktor-Faktor yang Mempengaruhi Penerimaan Mahasiswa Melalui Jalur Masuk Perguruan Tinggi SNMPTN FMIPA Universitas Brawijaya. [Skripsi, Universitas Brawijaya]. Malang.
- Pratiwi, D. L. (2018). Penerapan metode combine sampling pada klasifikasi imbalanced data biner status ketertinggalan desa di Jawa Timur [Undergraduate thesis, Institut Teknologi Sepuluh Nopember]. Surabaya.
- Qiong, G., Wang, X.-M., Wu, Z., Ning, B., & Xin, C.-S. (2016). An improved SMOTE algorithm based on genetic algorithm for imbalanced data classification. *Journal of Digital Information Management*, 14 (2), 92-103. <https://digitalcommons.odu.edu/ecefapubs/462/>. Diakses pada 15 November 2024.
- Qur'an Kemenag. (2024). Diakses pada 6 November 2024, dari <https://quran.kemenag.go.id>.
- Ramandhani, R., Sudarno., & Safitri, D. (2017). Metode Bootstrap Aggregating Regresi Logistik Biner untuk Ketepatan Klasifikasi Kesejahteraan Rumah Tangga di Kota Pati. *Jurnal Aussen*. 6, 121-130. <https://doi.org/10.14710/j.gauss.6.1.121-130>. Diakses pada 15 November 2024.
- Reckelhoff, J.F. (2001). *Gender differences in the regulation of blood pressure*. *Hypertension*, 37(5), 1199-1208. <https://doi.org/10.1161/01.HYP.37.5.1199>. Diakses pada 24 Mei 2025.
- Ryan, T. P. (1997). *Modern Regression Methods*. New York: John Wiley & Sons.
- Sain, H., & Purnami, S. W. (2015). Combine sampling Support Vector Machine for Imbalanced data Classification. *International Journal of Applied Mathematics and Statistics*, 72(59-66). <https://doi.org/10.1016/j.procs.2015.12.105>. Diakses pada 15 November 2024.
- Sahila, R., Widiarini, T., & Utami, I. T. (2024). Analisis Klasifikasi Menggunakan Regresi Logistik Biner Dan Algoritma Naïve Bayes Classifier pada Penyakit Hipertensi. *Jurnal Gaussian*, 13(2), 319-327.

<https://doi.org/10.14710/j.gauss.13.2.319-327>. Diakses pada 15 November 2024.

Solberg, A., & Solberg, R. (1996). A Large-Scale Evaluation of Features for Automatic Detection of Oil Spills in ERS SAR Image. *In International Geoscience and Remote Sensing Symposium*, (pp. 1484-1486). <https://doi.org/10.1109/IGARSS.1996.516705>. Diakses pada 15 Oktober 2024.

Sungkono, J., & Nugrahaningsih, T. (2017). Simulasi Dampak Multikolinearitas pada Kondisi Penyimpangan Asumsi Normalitas. *Magistra*, 29(101), 45-50.

Supandi, H., Istikhori, I., & Suryana, T. (2025). *Hadis Ibnu Abbas tentang menjaga lima perkara sebelum datang lima perkara yang lain (Al-Hakim)*. *Al-Tarbiyah: Jurnal Ilmu Pendidikan Islam*, 3(2), 01–09. <https://doi.org/10.59059/al-tarbiyah.v3i2.2143>. Diakses pada 15 November 2024.

Tafsir Web. (2024, November). Tafsir Web. Diakses dari <https://tafsirweb.com>.

Taufiq, L. M., Diliyanti, S., Taswin, & Muriman, Y. (2020). Hubungan Gaya Hidup Dengan Kejadian Hipertensi pada Lansia di Wilayah Kerja Puskesmas Meo-Meo Kota Bau Bau. *Jurnal Industri Kreatif (JIK)*, 4(1), 45–56. <https://doi.org/10.36352/jik.v4i01.55>. Diakses pada 15 November 2024.

Turana, Y., Widyantoro, B., Situmorang, T. D., Delliana, J., Roesli, R. M. A., Danny, S. S., Suhardjono, S., Sofiatin, Y., Hermiawaty, E., Kuncoro, A. S., Barack, R., Beaney, T., Chis Ster, A., Poulter, N. R., & Santoso, A. (2020). May Measurement Month 2018: An analysis of blood pressure screening results from Indonesia. *European Heart Journal Supplements*, 22(Suppl H), H66–H69. <https://doi.org/10.1093/eurheartj/suaa031>. Diakses pada 15 November 2024.

Tomek, I. (1997). Two Modifications of CNN. *IEEE Transactions of Systems, Man, and Communications*, 6, 769-772. <http://dx.doi.org/10.1109/TSMC.1976.4309452>. Diakses pada 15 November 2024.

Umaroh, A. K. (2020). Perbandingan Metode Regresi Logistik Biner dan Classification and Regression Tree Pada Klasifikasi Status Kesejahteraan Rumah Tangga Di Kota Batu. [Skripsi, Universitas Islam Negeri Maulana Malik Ibrahim]. Malang.

Wan, C. M., Nosedal-Sanchez, A., Nosedal-Sanchez, J., Asgary, A., & Pantin, B. (2019). Modeling provision of disaster mutual assistance by electricity utilities using logistic regression. *International Journal of Disaster Risk Reduction*, 8-9. <https://doi.org/10.1016/j.ijdr.2019.101110>. Diakses pada 15 November 2024.

World Health Organization. (2023). *Global Report On Hypertension: The Race Against A Silent Killer*. Diakses pada 6 November 2024, dari <https://www.who.int>.

World Health Organization. (2023). *Hypertension*. Diakses pada 6 November 2024, dari <https://www.who.int>.

World Health Organization. (2024). *World Hypertension Day 2024: Measure Your Blood Pressure Accurately, Control It, Live Longer*. Diakses pada 6 November 2024, dari <https://www.who.int>.

Zhang, Y., Wu, L., & Wang, S. (2011). Magnetic Resonance Brain Image Classification by An Improved Artificial Bee Colony Algorithm. *Progress In Electromagnetics Research*, 116, 67-79. <http://dx.doi.org/10.2528/PIER11031709>. Diakses pada 15 November 2024.

LAMPIRAN

Lampiran 1. Data Diagnosa Hipertensi

No	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	y
1	Laki-Laki	Tidak	Ya	Tidak	Tidak	Tidak	Tidak	Tidak	39	Tidak Hipertensi
2	Perempuan	Tidak	Ya	Tidak	Ya	Ya	Ya	Tidak	40	Tidak Hipertensi
3	Laki-Laki	Ya	Ya	Tidak	Tidak	Ya	Tidak	Tidak	61	Tidak Hipertensi
4	Perempuan	Tidak	Ya	Tidak	Tidak	Tidak	Ya	Tidak	58	Hipertensi
5	Laki-Laki	Ya	Ya	Tidak	Tidak	Tidak	Tidak	Tidak	63	Tidak Hipertensi
6	Perempuan	Tidak	Ya	Tidak	Ya	Ya	Ya	Tidak	44	Tidak Hipertensi
7	Laki-Laki	Tidak	Ya	Tidak	Tidak	Tidak	Tidak	Tidak	23	Tidak Hipertensi
8	Perempuan	Tidak	Ya	Ya	Tidak	Ya	Tidak	Tidak	60	Tidak Hipertensi
9	Laki-Laki	Ya	Ya	Tidak	Tidak	Tidak	Ya	Tidak	57	Tidak Hipertensi
10	Perempuan	Tidak	Ya	Ya	Tidak	Tidak	Tidak	Tidak	18	Tidak Hipertensi
11	Perempuan	Tidak	Ya	Tidak	Tidak	Tidak	Ya	Tidak	63	Tidak Hipertensi
12	Perempuan	Tidak	Tidak	Ya	Ya	Ya	Ya	Tidak	43	Tidak Hipertensi
13	Perempuan	Tidak	Ya	Ya	Tidak	Tidak	Tidak	Tidak	56	Hipertensi
14	Laki-Laki	Ya	Ya	Tidak	Ya	Tidak	Tidak	Tidak	12	Tidak Hipertensi
15	Perempuan	Tidak	Ya	Tidak	Tidak	Tidak	Ya	Tidak	66	Tidak Hipertensi
16	Perempuan	Tidak	Ya	Tidak	Tidak	Ya	Tidak	Tidak	76	Tidak Hipertensi
17	Perempuan	Tidak	Tidak	Tidak	Ya	Ya	Ya	Tidak	22	Tidak Hipertensi
18	Perempuan	Tidak	Ya	Tidak	Tidak	Ya	Tidak	Tidak	43	Hipertensi
19	Laki-Laki	Ya	Ya	Tidak	Tidak	Tidak	Tidak	Tidak	59	Tidak Hipertensi
20	Laki-Laki	Ya	Ya	Tidak	Tidak	Tidak	Tidak	Tidak	64	Tidak Hipertensi
21	Laki-Laki	Ya	Ya	Tidak	Tidak	Tidak	Tidak	Tidak	4	Tidak Hipertensi
22	Laki-Laki	Ya	Ya	Tidak	Tidak	Tidak	Ya	Tidak	59	Tidak Hipertensi
23	Perempuan	Tidak	Ya	Tidak	Tidak	Tidak	Ya	Tidak	37	Tidak Hipertensi
24	Perempuan	Tidak	Tidak	Tidak	Ya	Tidak	Ya	Tidak	55	Hipertensi
25	Perempuan	Tidak	Tidak	Ya	Tidak	Tidak	Ya	Tidak	53	Tidak Hipertensi
26	Perempuan	Tidak	Tidak	Ya	Tidak	Ya	Ya	Tidak	61	Tidak Hipertensi
27	Perempuan	Tidak	Ya	Tidak	Ya	Ya	Tidak	Tidak	43	Tidak Hipertensi
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2387	Perempuan	Tidak	Tidak	Ya	Tidak	Tidak	Tidak	Tidak	44	Hipertensi

Lampiran 2. Data Diagnosa Hipertensi Setelah Pengkodean

No	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	y
1	1	0	1	0	0	0	0	0	1	0
2	0	0	1	0	1	1	1	0	1	0
3	1	1	1	0	0	1	0	0	1	0
4	0	0	1	0	0	0	1	0	1	1
5	1	1	1	0	0	0	0	0	1	0
6	0	0	1	0	1	1	1	0	1	0
7	1	0	1	0	0	0	0	0	0	0
8	0	0	1	1	0	1	0	0	1	0
9	1	1	1	0	0	0	1	0	1	0
10	0	0	1	1	0	0	0	0	0	0
11	0	0	1	0	0	0	1	0	1	0
12	0	0	0	1	1	1	1	0	1	0
13	0	0	1	1	0	0	0	0	1	1
14	1	1	1	0	1	0	0	0	0	0
15	0	0	1	0	0	0	1	0	1	0
16	0	0	1	0	0	1	0	0	1	0
17	0	0	0	0	1	1	1	0	0	0
18	0	0	1	0	0	1	0	0	1	1
19	1	1	1	0	0	0	0	0	1	0
20	1	1	1	0	0	0	0	0	1	0
21	1	1	1	0	0	0	0	0	0	0
22	1	1	1	0	0	0	1	0	1	0
23	0	0	1	0	0	0	1	0	1	0
24	0	0	0	0	1	0	1	0	1	1
25	0	0	0	1	0	0	1	0	1	0
26	0	0	0	1	0	1	1	0	1	0
27	0	0	1	0	1	1	0	0	1	0
28	0	0	0	0	1	1	1	0	1	0
29	0	0	1	1	0	0	0	0	1	0
30	0	0	1	0	0	0	0	0	1	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2387	0	0	0	1	0	0	0	0	1	1

Lampiran 3. Syntax Tomek Links dan Regresi Logistik

```

# Install dan load package
install.packages("psych")
install.packages("summarytools")
install.packages("dplyr")
install.packages("UBL")
install.packages("themis")
install.packages("car")

library(readxl)
library(psych)
library(dplyr)
library(themis)
library(UBL)
library(car)

DataHipertensiKategorik <- read_excel("D:/Document/Tugas
Kuliah/Bismillah Skripsi/SKRIPSI/HIPERTENSI/DATA/BARU/Rasio ke
Nominal/DataHipertensiKategorik.xlsx")
View(DataHipertensiKategorik)

# Statistik deskriptif umum untuk semua kolom numerik
summary(DataHipertensiKategorik)
# Melihat struktur data
str(DataHipertensiKategorik)
# Melihat dimensi data
dim(DataHipertensiKategorik)
# Melihat 6 baris pertama
head(DataHipertensiKategorik)
# Statistik deskriptif lengkap (mean, sd, min, max, skewness,
kurtosis, dll.)
describe(DataHipertensiKategorik)

DataHipertensiKategorik$Y=as.factor(DataHipertensiKategorik$Y)
str(DataHipertensiKategorik)

#-----
# Set seed untuk reproducibility
set.seed(303)

# Membuat indeks acak untuk 70% data
index1 <- sample(nrow(DataHipertensiKategorik), 0.7 *
nrow(DataHipertensiKategorik))

# Split data
train_data1 <- DataHipertensiKategorik[index1, ]
test_data1 <- DataHipertensiKategorik[-index1, ]
table(train_data1$Y)
table(test_data1$Y)

# Cek jumlah baris
nrow(train_data1)
nrow(test_data1)

#-----
#mengubah variabel 'class'
colnames(train_data1)[10] <- 'class'
# Tomek Links

```

```

train_tomek_sample1 <- recipe(~., train_data1) %>%
  step_tomek(class) %>%
  prep() %>%
  bake(new_data = NULL)

train_tomek_sample1
table(train_tomek_sample1$class)

#-----
# Melakukan Estimasi Parameter (MLE dengan Newton Raphson)
# Ubah target class menjadi 0 dan 1 jika masih faktor
train_tomek_sample1$class <-
as.numeric(as.character(train_tomek_sample1$class))
str(train_tomek_sample1)

# Matriks X: prediktor, tambahkan kolom intercept
X1 <- as.matrix(cbind(Intercept = 1, train_tomek_sample1[, -
which(names(train_tomek_sample1) == "class")]))
y1 <- as.matrix(train_tomek_sample1$class)

#-----
XtX <- t(X1) %*% X1
XtX_inv <- solve(t(X1) %*% X1)
Xty <- t(X1) %*% y1
beta_ols <- solve(t(X1) %*% X1) %*% t(X1) %*% y1

#-----
# Tampilkan beta awal (iterasi ke-0)
cat("Estimasi beta pada iterasi ke-0 (inisialisasi):\n")
print(beta_ols)

# Fungsi sigmoid
sigmoid1 <- function(z1) {
  1 / (1 + exp(-z1))
}

#-----
# Hitung p1 pada iterasi ke-0
p1_0 <- sigmoid1(X1 %*% beta_ols)
W <- y1 - p1_0
V <- p1_0 * (1 - p1_0)
V_matrix <- diag(as.vector(V))
XtV <- t(X1) %*% V_matrix
XtVX <- t(X1) %*% V_matrix %*% X1
XtVX_inv <- solve(XtVX)
XtW <- t(X1) %*% W
XtVX_inv_XtW <- XtVX_inv %*% XtW
beta_new <- beta_ols + XtVX_inv_XtW

#-----
# Newton-Raphson
max_iter1 <- 100
epsilon1 <- 1e-6
converged1 <- FALSE

for (i in 1:max_iter1) {
  p1 <- sigmoid1(X1 %*% beta_ols)

  # Cegah elemen nol pada diagonal matriks W

```

```

w_diag1 <- as.vector(p1 * (1 - p1))
w_diag1[w_diag1 < 1e-6] <- 1e-6 # stabilisasi numerik
W1 <- diag(w_diag1)

# Tangani kemungkinan error pada solve()
tryCatch({
  z1 <- X1 %*% beta_ols + solve(W1) %*% (y1 - p1)
  beta_new1 <- solve(t(X1) %*% W1 %*% X1) %*% t(X1) %*% W1 %*%
z1

# Tampilkan hasil estimasi beta pada iterasi ke-1 dan ke-2
if (i == 1 || i == 2 || i == 10) {
  cat("Estimasi beta pada iterasi ke-", i, ":\n")
  print(beta_new1)
}

# Tampilkan estimasi beta pada iterasi ke-11
if (i == 11) {
  cat("Estimasi beta pada iterasi ke-11:\n")
  print(beta_new1)
}

cat("Iterasi", i, "- Perubahan total:", sum(abs(beta_new1 -
beta_ols)), "\n")

if (sum(abs(beta_new1 - beta_ols)) < epsilon1) {
  cat("Konvergen pada iterasi ke-", i, "\n")
  converged1 <- TRUE
  beta_ols <- beta_new1
  break
}

beta_ols <- beta_new1
}, error = function(e) {
  cat("Error pada iterasi", i, ":", e$message, "\n")
  break
})
}

# Tampilkan hasil akhir estimasi parameter
colnames(beta_ols) <- "Estimate"
rownames(beta_ols) <- colnames(X1)
cat("\n Estimasi Parameter Akhir:\n")
print(beta_ols)

#-----
# Pastikan data hanya berisi prediktor dan target sebagai faktor
train_tomek_sample1$class <- as.factor(train_tomek_sample1$class)
str(train_tomek_sample1)

# Buat model regresi logistik untuk perhitungan VIF
# Gunakan formula semua prediktor terhadap target 'class'
model_vif1 <- glm(class ~ ., data = train_tomek_sample1, family =
binomial(link="logit"))
model_vif1
summary(model_vif1)

#-----
# Ambil hanya variabel prediktor (tanpa kolom class)

```

```

predictors <- train_tomek_sample1[, -
which(names(train_tomek_sample1) == "class")]

# Data frame untuk menyimpan hasil
vif_manual_df <- data.frame(
  Variable = character(),
  SS_total = numeric(),
  SS_error = numeric(),
  SS_regresi = numeric(),
  R2 = numeric(),
  Tolerance = numeric(),
  VIF = numeric(),
  stringsAsFactors = FALSE
)

# Loop untuk setiap variabel prediktor
for (var in names(predictors)) {
  # Tentukan target dan prediktor lainnya
  y <- predictors[[var]]
  X_others <- predictors[, names(predictors) != var, drop = FALSE]

  # Gabungkan target y ke dalam data
  df <- data.frame(y = y, X_others)

  # Model regresi linier
  model <- lm(y ~ ., data = df)

  # Prediksi dan residual
  y_pred <- predict(model)
  residuals <- y - y_pred

  # Hitung SS_total, SS_error, dan SS_regresi
  SS_total <- sum((y - mean(y))^2)
  SS_error <- sum((residuals)^2)
  SS_regresi <- SS_total - SS_error

  # Hitung R^2, Tolerance, dan VIF
  R2 <- SS_regresi / SS_total
  Tolerance <- 1 - R2
  VIF <- 1 / Tolerance

  # Simpan hasil
  vif_manual_df <- rbind(vif_manual_df, data.frame(
    Variable = var,
    SS_total = SS_total,
    SS_error = SS_error,
    SS_regresi = SS_regresi,
    R2 = R2,
    Tolerance = Tolerance,
    VIF = VIF
  ))
}

# Tampilkan hasil
print(vif_manual_df)

#-----
# Uji Signifikansi Parameter Secara Simultan dengan Uji G

```

```

# Mengubah semua variabel prediktor menjadi faktor (kecuali
variabel target 'class')
train_tomek_sample1$X1 <- as.factor(train_tomek_sample1$X1)
train_tomek_sample1$X2 <- as.factor(train_tomek_sample1$X2)
train_tomek_sample1$X3 <- as.factor(train_tomek_sample1$X3)
train_tomek_sample1$X4 <- as.factor(train_tomek_sample1$X4)
train_tomek_sample1$X5 <- as.factor(train_tomek_sample1$X5)
train_tomek_sample1$X6 <- as.factor(train_tomek_sample1$X6)
train_tomek_sample1$X7 <- as.factor(train_tomek_sample1$X7)
train_tomek_sample1$X8 <- as.factor(train_tomek_sample1$X8)
train_tomek_sample1$X9 <- as.factor(train_tomek_sample1$X9)

str(train_tomek_sample1)
# Model penuh: semua prediktor
model_penuh <- glm(class ~ ., data = train_tomek_sample1, family =
binomial(link = "logit"))
summary(model_penuh)

# Model kosong (hanya intercept)
model_kosong <- glm(class ~ 1, data = train_tomek_sample1, family
= binomial(link = "logit"))

# Hitung statistik G
G_stat <- 2 * (logLik(model_penuh)[1] - logLik(model_kosong)[1])
df_g <- attr(logLik(model_penuh), "df") -
attr(logLik(model_kosong), "df")
p_value_g <- 1 - pchisq(G_stat, df = df_g)

#-----
# Uji Signifikansi Parameter Secara Parsial dengan Uji Wald

# Ekstrak koefisien dan kovarians dari model penuh
coef_model <- coef(model_penuh)
cov_matrix <- vcov(model_penuh)

# Hitung statistik Wald (z^2)
wald_stat <- (coef_model / sqrt(diag(cov_matrix)))^2

# Hitung p-value dari distribusi Chi-square dengan df = 1 untuk
masing-masing parameter
p_values_wald <- 1 - pchisq(wald_stat, df = 1)

# Gabungkan hasil ke dalam satu data frame
wald_results <- data.frame(
  Estimate = coef_model,
  Std_Error = sqrt(diag(cov_matrix)),
  Wald_Stat = round(wald_stat, 4),
  p_value = round(p_values_wald, 4)
)

# Tampilkan hasil
cat("\nUji Wald - Signifikansi Parsial Koefisien:\n")
print(wald_results)

#-----
# Uji Signifikansi Parsial Kedua: Uji Wald pada Variabel
Signifikan
# Ambil nama variabel signifikan (kecuali intercept)

```

```

variabel_signifikan <-
rownames(wald_results)[which(wald_results$p_value < 0.05)]
variabel_signifikan <- variabel_signifikan[variabel_signifikan !=
"(Intercept)"]
cat("\nVariabel Signifikan Parsial (p < 0.05):\n")
print(variabel_signifikan)

# Buat formula model hanya dengan variabel signifikan
formula_signifikan <- as.formula(paste("class ~",
paste(variabel_signifikan, collapse = " + ")))

# Bangun ulang model hanya dengan variabel signifikan
model_wald2 <- glm(formula_signifikan, data = train_tomek_sample1,
family = binomial(link = "logit"))

# Hitung ulang koefisien dan varians kovarian
coef_model2 <- coef(model_wald2)
cov_matrix2 <- vcov(model_wald2)

# Hitung ulang statistik Wald
wald_stat2 <- (coef_model2 / sqrt(diag(cov_matrix2)))^2
p_values_wald2 <- 1 - pchisq(wald_stat2, df = 1)

# Gabungkan hasilnya
wald_results2 <- data.frame(
  Estimate = coef_model2,
  Std_Error = sqrt(diag(cov_matrix2)),
  Wald_Stat = round(wald_stat2, 4),
  p_value = round(p_values_wald2, 4)
)

# Tampilkan hasil uji Wald kedua
cat("\nUji Wald Kedua - Signifikansi Parsial pada Variabel
Signifikan:\n")
print(wald_results2)

#-----
# Uji Signifikansi Parameter Secara Simultan Kedua dengan Uji G

# Model kosong (hanya intercept)
model_kosong2 <- glm(class ~ 1, data = train_tomek_sample1, family
= binomial(link = "logit"))

# Hitung statistik G
G_stat2 <- 2 * (logLik(model_wald2)[1] - logLik(model_kosong2)[1])

# Derajat kebebasan = jumlah parameter bebas dalam model_wald2
df_g2 <- attr(logLik(model_wald2), "df") -
attr(logLik(model_kosong2), "df")

# Hitung p-value
p_value_g2 <- 1 - pchisq(G_stat2, df = df_g2)

# Tampilkan hasil
cat("\nUji G Kedua - Likelihood Ratio Test pada Variabel
Signifikan:\n")
cat("LogLik Model Kosong:", logLik(model_kosong2), "\n")
cat("LogLik Model Parsial (Signifikan):", logLik(model_wald2),
"\n")

```

```

cat("Statistik G:", round(G_stat2, 4), "\n")
cat("Derajat kebebasan:", df_g2, "\n")
cat("p-value:", round(p_value_g2, 4), "\n")

#-----
# Uji Kesesuaian Model (Goodness-of-Fit) dengan Uji Deviance

# Dapatkan deviance residual dari model parsial
deviance_model <- deviance(model_wald2)

# Dapatkan derajat kebebasan residual
df_residual <- df.residual(model_wald2)

# Hitung p-value dari distribusi Chi-square
p_value_deviance <- 1 - pchisq(deviance_model, df_residual)

# Tampilkan hasil uji deviance
cat("\nUji Kesesuaian Model - Goodness of Fit Test (Deviance):\n")
cat("Deviance:", round(deviance_model, 4), "\n")
cat("Derajat Kebebasan:", df_residual, "\n")
cat("p-value:", round(p_value_deviance, 4), "\n")

#-----
# Interpretasi Koefisien dengan Odds Ratio

# Ambil koefisien dan standar error dari model parsial
(signifikan)
estimates <- coef(model_wald2)
std_errors <- sqrt(diag(vcov(model_wald2)))

# Hitung Odds Ratio
odds_ratios <- exp(estimates)

# Hitung 95% Confidence Interval untuk Odds Ratio
lower_ci <- exp(estimates - 1.96 * std_errors)
upper_ci <- exp(estimates + 1.96 * std_errors)

# Gabungkan dalam satu data frame
interpretasi_or <- data.frame(
  Estimate = round(estimates, 4),
  Odds_Ratio = round(odds_ratios, 4),
  CI_Lower = round(lower_ci, 4),
  CI_Upper = round(upper_ci, 4)
)

# Tampilkan hasil
cat("\nInterpretasi Koefisien dengan Odds Ratio:\n")
print(interpretasi_or)

#-----
#Ketepatan Klasifikasi
# Pastikan kolom target pada test data berupa faktor dengan label
0 dan 1
test_data1$class <- as.factor(as.character(test_data1$Y))

# Sesuaikan nama prediktor pada test data jika diperlukan
test_data_prediksi <- test_data1[, colnames(train_tomek_sample1)[-
which(colnames(train_tomek_sample1) == "class")]]

```

```
# Prediksi probabilitas dari model parsial
prob_pred <- predict(model_wald2, newdata = test_data_prediksi,
type = "response")

# Konversi probabilitas ke kelas (threshold default 0.5)
class_pred <- ifelse(prob_pred >= 0.5, 1, 0)
class_pred <- as.factor(class_pred)

# Confusion Matrix
conf_matrix <- table(Predicted = class_pred, Actual =
test_data1$class)
cat("\nConfusion Matrix:\n")
print(conf_matrix)

# Hitung nilai APER (Apparent Error Rate)
# APER = (jumlah prediksi salah) / total prediksi
jumlah_salah <- sum(conf_matrix) - sum(diag(conf_matrix))
total_prediksi <- sum(conf_matrix)
APER <- jumlah_salah / total_prediksi
cat("\nNilai APER (Apparent Error Rate):", round(APER, 4), "\n")

# Hitung akurasi (nilai ketepatan klasifikasi)
akurasi <- sum(diag(conf_matrix)) / total_prediksi
cat("Akurasi (Ketepatan Klasifikasi):", round(akurasi, 4), "\n")
```

Lampiran 4. Hasil Perhitungan Manual Estimasi Parameter

Mencari nilai estimasi parameter β^0

1. Membuat Matriks X

1	0	0	1	0	0	0	1	0	0
1	0	0	1	0	0	1	0	0	0
1	0	0	0	0	0	1	0	0	1
1	0	0	0	1	1	1	1	0	1
1	0	0	0	1	0	0	1	0	1
1	0	0	0	0	0	0	0	0	0
1	0	0	0	1	1	0	0	0	0
1	0	0	0	0	0	0	0	0	0
1	0	0	1	0	0	0	0	0	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1	1	0	1	0	0	0	0	0	0

2. Membuat Matriks Y

0
0
1
1
1
0
0
0
1
⋮
0

3. Membuat Matriks X^T

1	1	1	1	1	1	1	1	1	...	1
0	0	0	0	0	0	0	0	0	...	1
0	0	0	0	0	0	0	0	0	...	0
1	1	0	0	0	0	0	0	1	...	1
0	0	0	1	1	0	1	0	0	...	0
0	0	0	1	0	0	1	0	0	...	0
0	1	1	1	0	0	0	0	0	...	0
1	0	0	1	1	0	0	0	0	...	0
0	0	0	0	0	0	0	0	0	...	0
0	0	1	1	1	0	0	0	1	...	0

4. Mengalikan Matriks X^T dengan X

1606	249	210	652	401	274	408	558	...	324
249	249	98	142	56	42	42	68	...	44
210	98	210	94	45	42	47	70	...	44
652	142	94	652	167	135	168	260	...	141
401	56	45	167	401	109	134	172	...	102
274	42	42	135	109	274	115	145	...	72
408	42	47	168	134	115	408	169	...	81
558	68	70	260	172	145	169	558	...	111
7	4	5	4	3	2	2	4	...	2
324	44	44	141	102	72	81	111	...	324

5. Menginverskan $X^T X$

0,00191	-0,00059	-0,00052	-0,00087	-0,00061	...	-0,00071
-0,00059	0,00554	-0,00192	-0,00057	0,00000	...	0,00017
-0,00052	-0,00192	0,00624	0,00008	0,00016	...	-0,00008
-0,00087	-0,00057	0,00008	0,00268	0,00002	...	-0,00010
-0,00061	0,00000	0,00016	0,00002	0,00347	...	-0,00025
-0,00017	-0,00002	-0,00019	-0,00022	-0,00050	...	-0,00027
-0,00072	0,00033	0,00001	-0,00001	-0,00027	...	0,00009
-0,00078	0,00030	-0,00003	-0,00024	-0,00022	...	0,00009
0,00046	-0,00119	-0,00286	-0,00017	-0,00058	...	-0,00028
-0,00071	0,00017	-0,00008	-0,00010	-0,00025	...	0,00391

6. Mengalikan Matriks X^T dengan Y

467
43
53
190
154
107
139
171
3
307

7. Mengalikan $(X^T X)^{-1} X^T Y$

0,11284
-0,10299
-0,01362
-0,01206
0,05273
0,03991
0,05063
0,00885
0,10510
0,81396

Mencari nilai estimasi parameter β^1

1. Mengalikan β^0 dengan variabel independen

β_0	$\beta_1 x_1$	$\beta_2 x_2$	$\beta_3 x_3$	$\beta_4 x_4$...	$\beta_9 x_9$
0,11284	0,00000	0,00000	-0,01206	0,00000	...	0,00000
0,11284	0,00000	0,00000	-0,01206	0,00000	...	0,00000
0,11284	0,00000	0,00000	0,00000	0,00000	...	0,81396
0,11284	0,00000	0,00000	0,00000	0,05273	...	0,81396
0,11284	0,00000	0,00000	0,00000	0,05273	...	0,81396
0,11284	0,00000	0,00000	0,00000	0,00000	...	0,00000
0,11284	0,00000	0,00000	0,00000	0,05273	...	0,00000
0,11284	0,00000	0,00000	0,00000	0,00000	...	0,00000
0,11284	0,00000	0,00000	-0,01206	0,00000	...	0,81396
⋮	⋮	⋮	⋮	⋮	⋮	⋮
0,11284	-0,10299	0,00000	-0,01206	0,00000	...	0,00000

2. Menghitung Matriks $\pi(x)$

0,52738
0,53778
0,72660
0,74629
0,72877
0,52818
0,55119
0,52818
0,71397
⋮
0,49945

6. Mengalikan Matriks $X^T V$

0,24925	0,24857	0,19865	0,18934	0,19766	...	0,25000
0,0000	0,0000	0,0000	0,0000	0,0000	...	0,25000
0,0000	0,0000	0,0000	0,0000	0,0000	...	0,25000
0,24925	0,24857	0,0000	0,0000	0,0000	...	0,25000
0,0000	0,0000	0,0000	0,18934	0,19766	...	0,25000
0,0000	0,0000	0,0000	0,18934	0,0000	...	0,0000
0,0000	0,24857	0,19865	0,18934	0,0000	...	0,0000
0,24925	0,0000	0,0000	0,18934	0,19766	...	0,0000
0,0000	0,0000	0,0000	0,0000	0,0000	...	0,0000
0,0000	0,0000	0,0000	0,18934	0,19766	...	0,0000
0,0000	0,0000	0,19865	0,18934	0,19766	...	0,0000
0,24925	0,24857	0,19865	0,18934	0,19766	...	0,25000

7. Mengalikan Matriks $X^T V X$

384,193	60,484	50,432	155,688	94,184	...	65,204
60,484	60,484	23,605	34,449	13,652	...	9,271
50,432	23,605	50,432	22,589	10,605	...	9,047
155,688	34,449	22,589	155,688	38,984	...	28,468
94,184	13,652	10,605	38,984	94,184	...	20,041
64,281	10,046	9,895	31,608	24,996	...	14,208
96,989	10,211	11,245	39,741	31,334	...	15,92
133,217	16,615	16,735	61,997	40,211	...	22,121
1,627	0,943	1,191	0,943	0,694	...	0,389
65,204	9,271	9,047	28,468	20,041	...	65,204

8. Menginverskan $X^T V X$

0,00788	-0,00242	-0,00220	-0,00364	-0,00258	...	-0,00285
-0,00242	0,02282	-0,00788	-0,00237	-0,00007	...	0,00062
-0,00220	-0,00788	0,02596	0,00034	0,00069	...	-0,00037
-0,00364	-0,00237	0,00034	0,01120	0,00010	...	-0,00043
-0,00258	-0,00007	0,00069	0,00010	0,01463	...	-0,00100
-0,00074	-0,00007	-0,00078	-0,00094	-0,00205	...	-0,00109
-0,00302	0,00136	0,00004	-0,00001	-0,00113	...	0,00041
-0,00326	0,00123	-0,00011	-0,00103	-0,00091	...	0,00038
0,00196	-0,00485	-0,01234	-0,00078	-0,00249	...	-0,00103
-0,00285	0,00062	-0,00037	-0,00043	-0,00100	...	0,01865

9. Mengalikan $X^T W$

-447,344
-91,741
-64,622
-181,202
-83,014
-55,380
-98,161
-148,768
-1,208
73,573

10. Mengalikan $(X^T V X)^{-1} X^T W$

-1,67284
-0,32549
-0,04605
-0,03721
0,17425
0,12892
0,16034
0,02887
0,34669
2,73744

11. Menambahkan β^0 dengan $(X^T V X)^{-1} X^T W$

-1,56000
-0,42848
-0,05967
-0,04926
0,22699
0,16883
0,21097
0,03772
0,45179
3,55140

Lampiran 5. Output Iterasi *Newton-Raphson* dengan R Studio

Iterasi	β_0	β_1	β_2	β_3	β_4
1	-1,5600002	-0,4284753	-0,0596666	-0,0492638	0,2269881
2	-2,0170586	-0,9799221	-0,1219180	-0,1006449	0,4410180
3	-2,1322013	-1,4330928	-0,1489528	-0,1272928	0,5332481
4	-2,1430535	-1,5679671	-0,1497661	-0,1309877	0,5428678
5	-2,1432823	-1,5760427	-0,1496495	-0,1310744	0,5430353
6	-2,1432827	-1,5760695	-0,1496491	-0,1310746	0,5430355
7	-2,1432827	-1,5760695	-0,1496491	-0,1310746	0,5430355

Lampiran 6. Output Iterasi *Newton-Raphson* dengan R Studio Lanjutan

Iterasi	β_5	β_6	β_7	β_8	β_9
1	0,1688303	0,2109709	0,0377209	0,4517920	3,5514022
2	0,3231098	0,4015101	0,0695969	0,9513093	4,6181126
3	0,3869358	0,4772015	0,0798409	1,2643196	5,0864934
4	0,3943827	0,4848024	0,0802473	1,3307506	5,2027174
5	0,3945714	0,4849480	0,0802336	1,3337118	5,2097092
6	0,3945718	0,4849482	0,0802335	1,3337200	5,2097333
7	0,3945718	0,4849482	0,0802335	1,3337200	5,2097333

RIWAYAT HIDUP



Putri Aulia Fachreza, lahir di Tulungagung pada 23 Juni 2003, biasa dipanggil Reza atau Putri. Penulis merupakan anak pertama dari Ayah Zaenal Abidin dan Mama Yuntiani. Adik penulis bernama Muhammad Bachtiar. Penulis telah menempuh pendidikan mulai dari PAUD Barokah yang lulus pada tahun 2008, dilanjutkan menempuh pendidikan TK Islam Al-Khoiriyah dan lulus pada tahun 2010. Penulis melanjutkan Pendidikan sekolah dasar di SD Islam Al-Khoiriyah yang lulus pada tahun 2016 dan kemudian melanjutkan pendidikan sekolah menengah pertama di MTSN 1 Tulungagung yang lulus pada tahun 2019. Penulis kemudian menempuh sekolah menengah atas selama 2 tahun di MAN 1 Tulungagung dan lulus pada tahun 2021. Pada tahun yang sama, penulis melanjutkan pendidikan perguruan tinggi di Universitas Islam Negeri Maulana Malik Ibrahim Malang pada Program Studi Matematika di Fakultas Sains dan Teknologi.



KEMENTERIAN AGAMA RI
UNIVERSITAS ISLAM NEGERI
MAULANA MALIK IBRAHIM MALANG
FAKULTAS SAINS DAN TEKNOLOGI

Jl. Gajayana No.50 Dinoyo Malang Telp. / Fax. (0341)558933

BUKTI KONSULTASI SKRIPSI

Nama : Putri Aulia Fachreza
NIM : 210601110043
Fakultas / Jurusan : Sains dan Teknologi / Matematika
Judul Skripsi : Reduksi *Imbalanced Data* Diagnosa Hipertensi dengan
Tomek Links pada Regresi Logistik
Pembimbing I : Ria Dhea Layla Nur Karisma, M.Si.
Pembimbing II : Erna Herawati, M.Pd.

No	Tanggal	Hal	Tanda Tangan
1.	14 Oktober 2024	Konsultasi Topik dan Data	1.
2.	6 November 2024	Konsultasi Bab I, II, dan III	2.
3.	15 November 2024	Konsultasi Bab I, II, dan III	3.
4.	19 November 2024	Konsultasi Kajian Agama Bab I dan II	4.
5.	26 November 2024	Konsultasi Kajian Agama Bab I dan II	5.
6.	3 Desember 2024	Konsultasi Kajian Agama Bab I dan II	6.
7.	6 Desember 2024	ACC Bab I, II, dan III	7.
8.	6 Desember 2024	ACC Kajian Agama Bab I dan II	8.
9.	9 Desember 2024	ACC Seminar Proposal	9.
10.	25 Februari 2025	Konsultasi Revisi Seminar Proposal	10.
11.	15 April 2025	Konsultasi Bab IV	11.
12.	22 April 2025	Konsultasi Bab IV	12.
13.	29 April 2025	Konsultasi V	13.
14.	30 April 2025	Konsultasi Kajian Agama Bab IV	14.
15.	5 Mei 2025	Konsultasi Kajian Agama Bab IV	15.



**KEMENTERIAN AGAMA RI
UNIVERSITAS ISLAM NEGERI
MAULANA MALIK IBRAHIM MALANG
FAKULTAS SAINS DAN TEKNOLOGI**

Jl. Gajayana No.50 Dinoyo Malang Telp. / Fax. (0341)558933

16.	6 Mei 2025	ACC Bab IV dan V	16.
17.	6 Mei 2025	ACC Kajian Agama Bab IV	17.
18.	15 Mei 2025	ACC Seminar Hasil	18.
19.	23 Mei 2025	Konsultasi Revisi Seminar Hasil	19.
20.	13 Juni 2025	ACC Sidang Skripsi	20.
21.	16 Juni 2025	ACC Keseluruhan	21.

Malang, 16 Juni 2025

Mengetahui,

Ketua Program Studi Matematika



Dr. Lily Susanti, M.Sc.

NIP. 19741129 200012 2 005