

# Pengaruh Pola Konsumsi Terhadap Hipertensi Menggunakan Regresi Linear Berganda

Nova Aulia Burhan<sup>1</sup>, Maimunah<sup>2</sup>, Emilyya Ully Artha<sup>3</sup>

<sup>1,2,3</sup>Departemen Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Magelang, Alamat Jl. Mayjen Bambang Soegeng, Glagak, Sumberrejo, Mertoyudan, Magelang, 56172, Indonesia  
E-mail: novaaulia206@gmail.com<sup>1</sup>, maimunah@unimma.ac.id<sup>2</sup>, ully@ummgl.ac.id<sup>3</sup>

**Abstract** — Hypertension is often referred to as the "silent killer" because it often develops without obvious symptoms, but can cause serious illnesses such as kidney failure, heart failure, and even stroke, which have the potential to cause permanent disability and even sudden death. The aim of this research is to study how consumption patterns impact the incidence of hypertension using the multiple linear regression method using machine learning using the Python programming language. Data was obtained from the Bandongan Community Health Center and the Magelang Regency BPS website, covering 70 data, 10 columns and 14 villages. The independent variables used are population, vegetable consumption, consumption of ready-made food and beverages and smoking, while the dependent variable is people with hypertension. Food and beverage consumption, as well as cigarette consumption, are used as independent variables in multiple linear regression modeling. The results of the analysis show that consumption of ready-made food and beverages and smoking have a significant influence, with a P-value of  $0.000 < 0.005$ . The R-square value is 0.741, indicating that the variables consumption of processed food and beverages and smoking have an influence of 74.1% on the incidence of hypertension, while other factors influence 25.9%. For future research, it is recommended to collect a wider and more diverse dataset to increase the validity of research findings.

**Key word** — Consumption, Hypertension, Multiple Linear Regression.

**Abstrak** — Hipertensi sering disebut sebagai "silent killer" karena sering kali berkembang tanpa gejala yang jelas, namun dapat menyebabkan penyakit serius seperti gagal ginjal, gagal jantung, bahkan stroke, yang berpotensi menyebabkan cacat permanen hingga kematian mendadak. Tujuan dari penelitian ini adalah untuk mempelajari bagaimana pola konsumsi berdampak pada kejadian hipertensi menggunakan metode regresi linear berganda dengan pemanfaatan *machine learning* menggunakan bahasa pemrograman *Python*. Data diperoleh dari Puskesmas Bandongan dan situs web BPS Kabupaten Magelang, mencakup 70 data, 10 kolom, dan 14 desa. Variabel bebas yang digunakan adalah jumlah penduduk, konsumsi sayur, konsumsi makanan minuman jadi serta merokok, sedangkan variabel terikatnya adalah penderita hipertensi. Konsumsi makanan minuman jadi, serta konsumsi rokok, digunakan untuk variabel bebas dalam pemodelan regresi linear berganda. Hasil analisis menunjukkan bahwa konsumsi makanan minuman jadi dan merokok memiliki pengaruh signifikan, dengan nilai *P-value*  $0,000 < 0,005$ . Nilai *R-square* ialah 0,741 menunjukkan bahwa variabel konsumsi makanan minuman jadi serta merokok memiliki pengaruh sebesar 74,1% terhadap kejadian hipertensi, sementara faktor lain mempengaruhi 25,9%. Untuk penelitian berikutnya, disarankan agar mengumpulkan dataset yang lebih luas dan beragam guna meningkatkan validitas temuan penelitian.

**Kata kunci**— Hipertensi, Konsumsi, Regresi Linear Berganda.

## I. PENDAHULUAN

Hipertensi tergolong sebagai penyakit tidak menular yang memiliki tingkat prevalensi sangat tinggi secara global[1]. Hipertensi kerap disebut sebagai "silent killer" karena gangguan ini cenderung menyerang secara diam-diam tanpa gejala atau keluhan[2]. Kondisi ini sangat berbahaya karena dapat menimbulkan gagal ginjal, gagal jantung bahkan stroke[3]. Penyakit-penyakit tersebut juga berpotensi menimbulkan cacat permanen atau kematian mendadak[4]. WHO memperkirakan bahwa seiring dengan pertumbuhan populasi global, prevalensi hipertensi akan terus meningkat. Diperkirakan pada tahun 2025, sekitar 29% penduduk dunia akan mengidap hipertensi. Tingkat prevalensi hipertensi juga cenderung lebih tinggi di negara-negara berkembang, mencapai sekitar 40%, dibandingkan dengan

negara maju yang sebesar 35%. Sebanyak 1,5 juta orang meninggal setiap tahun di Asia karena hipertensi, yang merupakan 25% dari semua orang yang terkena penyakit ini di dunia[5].

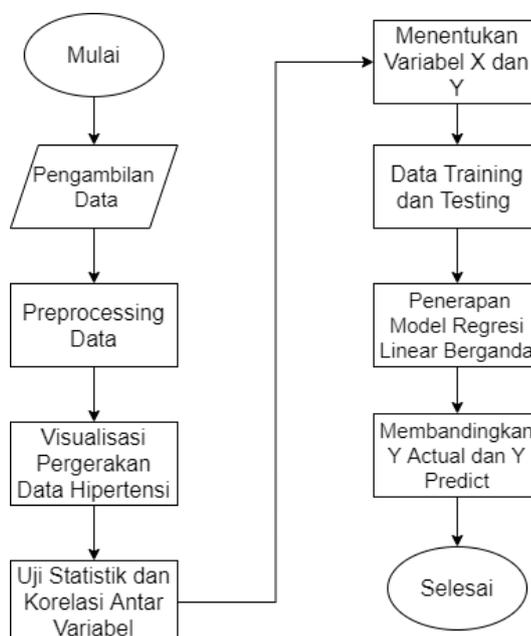
Terdapat beberapa faktor yang berpengaruh terhadap kejadian hipertensi. Faktor-faktor yang tidak dapat diubah termasuk umur, gender, dan etnis[6]. Selain itu, pola hidup sehari-hari juga berpotensi berperan, seperti kebiasaan merokok[7], asupan garam yang berlebihan, konsumsi makanan cepat saji[8], kebiasaan konsumsi sayur[9], mengonsumsi alkohol, mengalami stres, penggunaan estrogen dan kelebihan berat badan[10].

Saat ini, teknologi sangat berkembang pesat, terutama dalam hal penggunaan *machine learning*[11][12]. Aplikasi komputer dan algoritma matematika untuk *machine learning* menggunakan data untuk membuat prediksi[13]. Prediksi adalah dugaan tentang suatu kejadian atau peristiwa yang akan datang[14][15]. Penelitian Ghebyla (2019) menunjukkan bahwa metode regresi linear digunakan untuk memprediksi penjualan properti pada PT XYZ. Dengan menggunakan periode penjualan sebagai variabel bebas dan jumlah penjualan sebagai variabel terikat, penelitian ini berhasil memprediksi penjualan berbagai jenis properti berdasarkan data historis[16].

Data yang diperoleh dari Puskesmas Bandongan dan web Badan Pusat Statistik (BPS) Kabupaten Magelang, yang mencakup 70 data. Desa yang tercatat di Puskesmas Bandongan terdiri dari 14 desa. Di dalam data terdapat 10 kolom yang akan digunakan oleh peneliti sebagai acuan untuk melakukan prediksi. Data yang digunakan mencakup rentang waktu lima tahun, yaitu dari tahun 2019 hingga 2023. Jumlah penduduk, konsumsi sayur, konsumsi makanan minuman jadi, serta konsumsi rokok adalah variabel bebas yang digunakan dalam penelitian ini. Sementara itu, variabel terikatnya adalah penderita hipertensi. Untuk menggunakan variabel dalam pemodelan regresi linear berganda adalah konsumsi makanan minuman jadi ( $X_1$ ), konsumsi rokok ( $X_2$ ), dan penderita hipertensi ( $Y$ ). Metode regresi linear berganda digunakan dalam penelitian ini untuk menentukan apakah ada hubungan antara konsumsi makanan minuman jadi dan konsumsi rokok dengan hipertensi

## II. METODE PENELITIAN

Flowchart penelitian dapat dilihat pada Gambar 1, yang menjelaskan tahapan-tahapan penelitian yang dilakukan oleh penulis. Alur metode penelitian tersebut meliputi beberapa langkah, dimulai dari pengumpulan data, preprocessing data, dilanjutkan visualisasi data, melakukan uji analisis statistik dan korelasi antar variabel. Selanjutnya, dilakukan analisis data dan interpretasi hasil untuk mencapai kesimpulan yang relevan. Alur metode penelitian yang dilakukan, akan dijelaskan sebagai berikut :



Gambar 1. Flowchart penelitian

### A. Pengambilan Data

Dataset yang digunakan dalam penelitian ini diperoleh dari Puskesmas Bandongan dan situs web Badan Pusat Statistik (BPS) Kabupaten Magelang. Dataset tersebut bernama "hipertensi.csv" dan mencakup 70 data, terdiri dari 10 kolom yang mencakup periode tahun 2019 hingga 2023. Kolom-kolom dalam dataset ini meliputi nomor, tahun, desa, kode desa, jumlah penduduk, jumlah estimasi penderita, jumlah penderita hipertensi, konsumsi sayur, konsumsi makanan minuman jadi, dan konsumsi rokok. Dataset ini akan diolah dan dianalisis, sebagaimana dapat dilihat pada gambar 2.

	A	B	C	D	E	F	G	H	I	J
	no	tahun	desa	kode_desa	jumlah_penduduk	jumlah_estimasi_penderita	penderita_hipertensi	konsumsi_sayur	konsumsi_makanan_minuman_jadi	konsumsi_rokok
1										
2	1	2019	sukosari	01	1853	547	66	3,31	16,97	5,04
3	2	2019	kedungsar	02	2544	726	80	3,31	16,97	5,04
4	3	2019	salamkanc	03	4120	1194	175	3,31	16,97	5,04
5	4	2019	banyuwana	04	6665	1892	136	3,31	16,97	5,04
6	5	2019	trasan	05	7543	2245	205	3,31	16,97	5,04
7	6	2019	bandonga	06	7034	2085	271	3,31	16,97	5,04
8	7	2019	sukodadi	07	2644	774	108	3,31	16,97	5,04
9	8	2019	tonoboyo	08	3462	973	134	3,31	16,97	5,04
10	9	2019	kebonagui	09	2930	843	91	3,31	16,97	5,04
11	10	2019	kalegen	10	3042	858	87	3,31	16,97	5,04
12	11	2019	ngepanrej	11	4773	1325	60	3,31	16,97	5,04
13	12	2019	gandusari	12	4822	1385	152	3,31	16,97	5,04
14	13	2019	sidorejo	13	3600	978	115	3,31	16,97	5,04
15	14	2019	rejosari	14	7969	2208	132	3,31	16,97	5,04
16	15	2020	sukosari	01	1840	547	33	3,87	16,12	5,32
17	16	2020	kedungsar	02	2526	726	31	3,87	16,12	5,32

Gambar 2. Dataset yang akan diolah

### B. Preprocessing Data

Preprocessing data adalah tahap awal dalam analisis data yang melibatkan serangkaian langkah untuk mempersiapkan data mentah menjadi bentuk yang bersih dan siap untuk diolah lebih lanjut. Tahap preprocessing dilakukan untuk memastikan kualitas data agar lebih optimal [17]. Pada penelitian ini, dilakukan pengecekan terhadap data yang tidak memiliki isi dan memilih kolom yang bernilai tipe data numerik saja.

### C. Visualisasi Pergerakan Data Hipertensi

Visualisasi ini digunakan untuk menampilkan grafik pergerakan data hipertensi dan jumlah penduduk dari tahun 2019 hingga 2023. Visualisasi pergerakan data hipertensi memberikan gambaran yang jelas mengenai perubahan prevalensi hipertensi dan jumlah penduduk seiring waktu, serta membantu mengidentifikasi tren dan pola yang mungkin tidak terlihat dalam data mentah.

### D. Uji Statistik dan Korelasi Antar Variabel

Uji statistik dan analisis korelasi digunakan untuk memeriksa integritas data serta memahami hubungan antar variabel dalam dataset. Uji statistik memungkinkan identifikasi adanya data yang tidak wajar, membantu memastikan bahwa data yang digunakan dalam penelitian valid dan dapat diandalkan. Selain itu, analisis korelasi antar variabel bertujuan untuk mengidentifikasi sejauh mana variabel-variabel berhubungan satu sama lain. Dengan memahami hubungan ini, peneliti dapat mengungkap interaksi kompleks antara berbagai faktor yang mempengaruhi hipertensi, seperti konsumsi makanan minuman jadi dan kebiasaan merokok, dalam konteks jumlah penduduk dan prevalensi hipertensi dari tahun 2019 hingga 2023.

### E. Menentukan Variabel X dan Y

Penelitian ini menentukan variabel-variabel yang digunakan untuk menganalisis hubungan antara konsumsi makanan minuman jadi, konsumsi rokok, dan prevalensi hipertensi. Variabel independen (X)

yang digunakan adalah konsumsi makanan minuman jadi (X1) dan konsumsi rokok (X2). Variabel dependen (Y) dalam penelitian ini adalah jumlah penderita hipertensi. Penentuan variabel-variabel ini didasarkan pada hipotesis bahwa konsumsi makanan minuman jadi dan kebiasaan merokok memiliki pengaruh signifikan terhadap prevalensi hipertensi. Dengan mengidentifikasi dan mengukur variabel-variabel ini, penelitian bertujuan untuk mengungkap hubungan kuantitatif antara faktor pola konsumsi dan kejadian hipertensi, yang dianalisis melalui regresi linear berganda.

#### F. Data Training dan Testing

Sebelum proses pemodelan dimulai, dataset akan dibagi menjadi dua subset yaitu data training dan data testing. Data training digunakan untuk melatih model, dan data testing digunakan untuk menguji kinerja model[9][18]. Pengujian dengan data testing bertujuan untuk mengevaluasi kemampuan model dalam membuat prediksi pada data yang belum pernah dilihat sebelumnya, sehingga dapat memberikan gambaran yang akurat mengenai keandalan dan generalisasi model dalam konteks yang lebih luas. Pembagian data ini esensial untuk memastikan validitas dan akurasi model prediktif yang dikembangkan dalam penelitian.

#### G. Model Regresi Linear Berganda

Metode regresi linier berganda digunakan untuk memperkirakan pengaruh beberapa variabel independen terhadap satu variabel dependen, serta untuk menentukan apakah ada hubungan antara variabel bebas (X) dan variabel terikat (Y)[19][20]. Teknik ini memungkinkan analisis mendalam mengenai bagaimana setiap variabel independen berkontribusi terhadap variabel dependen, baik secara individu maupun bersama-sama. Penggunaan regresi linier berganda membantu dalam memahami kompleksitas hubungan antara variabel-variabel yang terlibat, memberikan pandangan yang lebih jelas tentang dinamika yang ada di antara mereka. Adapun rumus perhitungan regresi linier berganda dapat dilihat pada persamaan 1, sebagai berikut :

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n \quad (1)$$

Penjelasan :

Y : Variabel terikat

a : Bilangan Konstan

b1 : Koefisien Regresi 1

b2 : Koefisien Regresi 2

X1 : Variabel bebas 1

X2 : Variabel bebas 2

#### H. Membandingkan Y Actual dan Y Predict

Evaluasi model prediksi dilakukan dengan membandingkan hasil prediksi (Y Predict) dengan nilai aktual (Y Actual). Langkah ini penting untuk menilai seberapa dekat hasil prediksi yang dihasilkan oleh model dengan data sebenarnya. Hasil evaluasi ini akan memberikan wawasan tentang keandalan model dalam membuat prediksi dan kemampuannya untuk diterapkan dalam konteks yang lebih luas.

### III. HASIL DAN PEMBAHASAN

#### A. Preprocessing Data

Pada Gambar 3, dapat dilihat bahwa semua data yang digunakan lengkap dan tidak terdapat nilai yang hilang. Selanjutnya, ditampilkan bahwa terdapat 10 kolom data yang akan diolah dalam penelitian ini.

```
data = data.dropna()

#Cek missing value/nilai kosong
print("Jumlah missing value tiap kolom:")
print(data.isnull().sum())

Jumlah missing value tiap kolom:
no                0
tahun             0
desa              0
kode_desa        0
jumlah_penduduk  0
jumlah_estimasi_penderita  0
penderita_hipertensi  0
konsumsi_sayur   0
konsumsi_makanan_minuman_jadi  0
konsumsi_rokok   0
dtype: int64
```

Gambar 3. Memeriksa data kosong

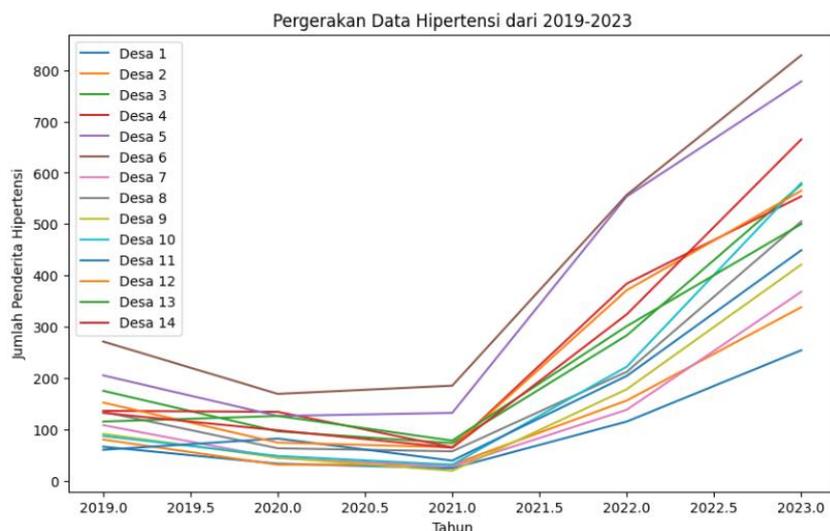
Proses pengolahan data dilanjutkan dengan menampilkan perintah untuk memproses hanya data bertipe numerik, sebagaimana dilihat pada Gambar 4.

```
[ ] #memilih kolom yang bernilai tipe data numerik saja
data = data.select_dtypes(include='number')
```

Gambar 4. Memilih kolom numerik

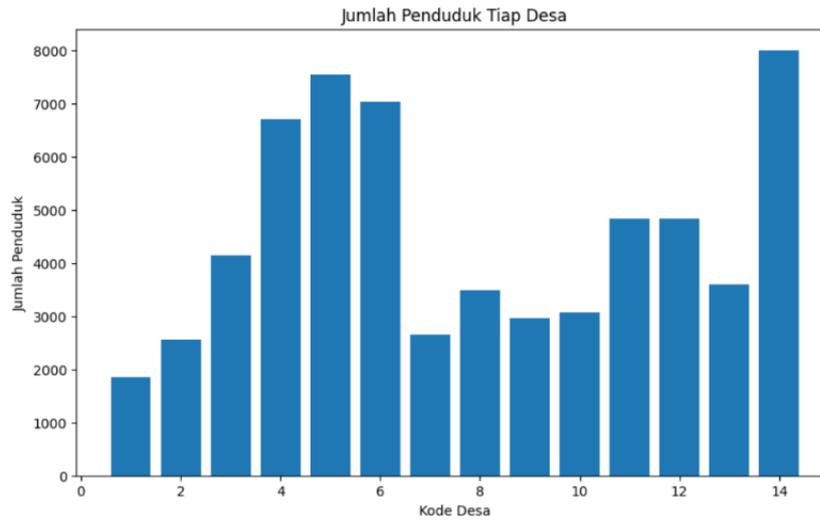
### B. Visualisasi Pergerakan Data Hipertensi

Pergerakan data hipertensi dari tahun 2019 hingga 2023 menunjukkan kenaikan data yang akan diolah. Seperti dilihat pada Gambar 5, terdapat penurunan kasus hipertensi dari tahun 2019 ke tahun 2020. Namun, dari tahun 2020 hingga 2023, jumlah kasus hipertensi meningkat secara signifikan.



Gambar 5. Visualisasi data hipertensi

Pada gambar 6 menampilkan data jumlah penduduk berdasarkan kode desa dari tahun 2019 hingga 2023. Desa dengan populasi penduduk tertinggi adalah desa dengan kode 14, sedangkan desa dengan populasi penduduk terendah adalah desa dengan kode 1.



Gambar 6. Visualisasi data jumlah penduduk

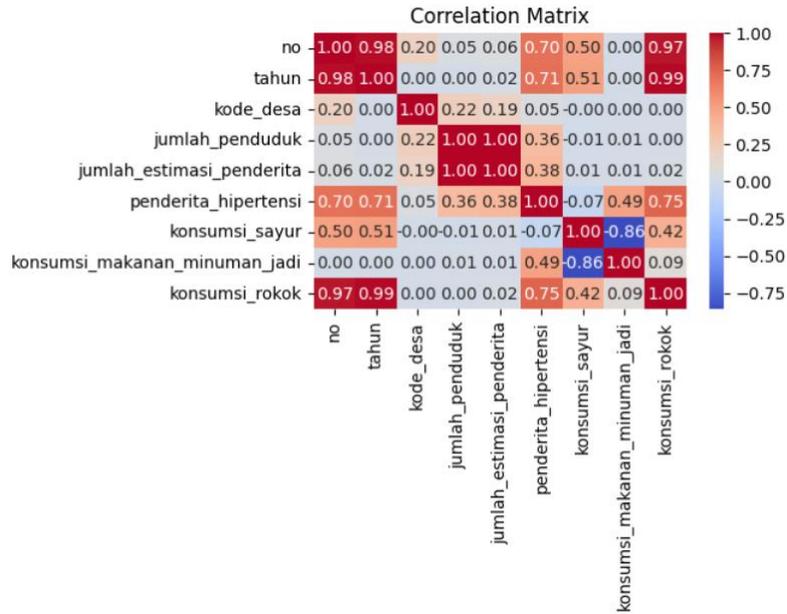
### C. Uji Statistik dan Korelasi Antar Variabel

Hasil analisis uji statistik deskriptif dilihat pada gambar 7. Data tersebut digunakan untuk mendeskripsikan karakteristik sampel dan menentukan apakah data dapat digunakan lebih lanjut dalam penelitian. Dari data tersebut, misal penderita hipertensi memiliki nilai tertinggi sebesar 829.000000, konsumsi sayur memiliki nilai tertinggi sebesar 4.620000, konsumsi makanan dan minuman jadi memiliki nilai tertinggi sebesar 17.270000, dan konsumsi rokok (X2) memiliki nilai tertinggi sebesar 7.380000. Dapat disimpulkan data tersebut masih wajar karena menggambarkan batas atas dari distribusi data dalam masing-masing sampel.

	no	tahun	kode_desa	jumlah_penduduk	jumlah_estimasi_penderita	penderita_hipertensi	konsumsi_sayur	konsumsi_makanan_minuman_jadi	konsumsi_rokok
count	70.000000	70.000000	70.000000	70.000000	70.000000	70.000000	70.000000	70.000000	70.000000
mean	35.500000	2021.000000	7.500000	4484.485714	1300.671429	217.914286	4.010000	16.142000	6.076000
std	20.351085	1.424425	4.060235	1961.228628	574.327199	201.085009	0.458965	0.911529	0.857158
min	1.000000	2019.000000	1.000000	1840.000000	547.000000	19.000000	3.310000	14.810000	5.040000
25%	18.250000	2020.000000	4.000000	2920.250000	843.250000	67.750000	3.870000	15.540000	5.320000
50%	35.500000	2021.000000	7.500000	3846.500000	1116.000000	134.000000	3.870000	16.120000	6.040000
75%	52.750000	2022.000000	11.000000	6649.000000	1892.000000	318.250000	4.380000	16.970000	6.600000
max	70.000000	2023.000000	14.000000	8006.000000	2286.000000	829.000000	4.620000	17.270000	7.380000

Gambar 7. Uji statistik deskriptif

Analisis korelasi antar variabel dalam konteks data hipertensi dilihat pada gambar 8. Berdasarkan rencana penelitian, terdapat keterkaitan antara penderita hipertensi dengan pola makan seseorang. Hasil analisis menunjukkan bahwa terdapat korelasi positif rendah antara jumlah penduduk dengan penderita hipertensi bernilai 0,36, korelasi negatif sangat rendah antara konsumsi sayur dengan penderita hipertensi sebesar -0,07, korelasi positif sedang antara konsumsi makanan dan minuman jadi dengan penderita hipertensi sebesar 0,49, serta korelasi positif yang lebih kuat antara konsumsi rokok dengan penderita hipertensi bernilai 0,75.



Gambar 8. Korelasi antar variabel

#### D. Memilih Variabel X dan Y

Dilanjutkan dengan pembagian data menjadi variabel independen dan variabel dependen dapat dilihat pada gambar 9. Variabel independen (X) adalah konsumsi makanan minuman jadi dan konsumsi rokok sedangkan variabel dependen (Y) adalah penderita hipertensi.

```
[ ] # Pilih variabel independen dan dependen
X = data[['konsumsi_makanan_minuman_jadi', 'konsumsi_rokok']]
y = data['penderita_hipertensi']
```

Gambar 9. Variabel X dan Y

#### E. Data Training dan Testing

Tahap selanjutnya adalah membagi dataset tersebut menjadi dua bagian, yaitu 80% sebagai data latih dan 20% sebagai data uji. Pembagian ini dilakukan untuk memastikan model dapat dilatih dan diuji dengan data yang berbeda. Proses pembagian data ini dapat dilihat secara rinci pada Gambar 10.

```
[ ] # Pisahkan data menjadi set pelatihan dan pengujian
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Gambar 10. Pembagian data training dan testing

F. Perhitungan Regresi Linear Berganda dengan Library Statsmodels

Penelitian ini memanfaatkan library statsmodels. Gambar 11 menunjukkan hasil dari metode analisis regresi linear berganda.

```

=====
                        OLS Regression Results
=====
Dep. Variable:          y          R-squared:                0.741
Model:                 OLS       Adj. R-squared:           0.732
Method:                Least Squares   F-statistic:              75.93
Date:                  Wed, 05 Jun 2024   Prob (F-statistic):       2.75e-16
Time:                  04:22:17       Log-Likelihood:           -341.28
No. Observations:     56           AIC:                     688.6
Df Residuals:         53           BIC:                     694.6
Df Model:              2
Covariance Type:      nonrobust
=====
                        coef      std err          t      P>|t|      [0.025      0.975]
-----
const      -2341.5790    263.851     -8.875     0.000    -2870.798    -1812.360
x1          95.7975     16.453      5.822     0.000      62.797     128.798
x2         165.8573     17.851     9.291     0.000     130.054     201.661
=====
Omnibus:                15.049   Durbin-Watson:           2.287
Prob(Omnibus):          0.001   Jarque-Bera (JB):        21.705
Skew:                   0.923   Prob(JB):                 1.94e-05
Kurtosis:               5.428   Cond. No.                 310.
=====

```

Gambar 11. Hasil perhitungan regresi linear berganda menggunakan statsmodels

Nilai R-square atau koefisien determinasi yang didapat adalah 0,741 dan sisanya 0,259 dipengaruhi oleh faktor lain. Uji F (Uji Hipotesis Secara Simultan) menghasilkan sebesar 75,93 dengan p-value  $2.75e-16 > 0,05$ . Untuk Uji t (Uji Hipotesis Secara Parsial) menunjukkan bahwa p-value untuk variabel X1 adalah  $0,000 < 0,05$  dan p-value untuk variabel X2 juga sebesar  $0,000 < 0,05$ . Disimpulkan keduanya bahwa H1 diterima, yang menunjukkan bahwa adanya pengaruh antara variabel X1 dan X2 terhadap variabel Y. Berdasarkan hasil, persamaan model regresi yang dihasilkan adalah:  $Y = -2341,58 + 95,80X1 + 165,86X2$ . Hasil dari analisis regresi linear berganda sebelumnya dapat disimpulkan sebagai berikut :

1. Nilai konstan adalah -2341,58 yang menunjukkan bahwa nilai penderita hipertensi (Y) = -2341,58 jika nilai konsumsi makanan minuman jadi (X1) = 0 dan nilai konsumsi rokok (X2) = 0. Nilai ini digunakan sebagai nilai titik awal dalam prediksi variabel dependen Y.
2. Koefisien regresi konsumsi makanan minuman jadi (X1) sebesar 95,80 yang berarti konsumsi makanan minuman jadi berpengaruh searah terhadap penderita hipertensi (Y). Koefisien ini menunjukkan perubahan rata-rata dalam Y untuk setiap peningkatan 1% pada X. Dengan asumsi bahwa X2 tidak berubah, koefisien ini menunjukkan bahwa Y tetap sama. Artinya setiap peningkatan 1% dalam X1 diasosiasikan dengan peningkatan 95.80 dalam penderita hipertensi (Y).
3. Koefisien regresi konsumsi rokok (X2) sebesar 165,86 menunjukkan bahwa konsumsi rokok berdampak terhadap penderita hipertensi (Y). Untuk setiap kenaikan 1% pada X2, dengan asumsi X1 tidak berubah, koefisien ini menunjukkan bahwa setiap kenaikan 1% pada X2 dikaitkan dengan peningkatan 165,86 dalam penderita hipertensi (Y).

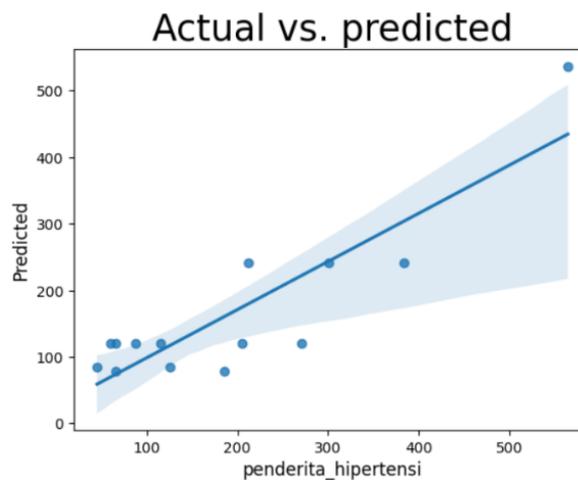
G. Membandingkan Y Actual dan Y Predict

Hasil perbandingan antara y aktual dan y prediksi, yang dapat dilihat pada Gambar 12, menunjukkan selisih angka yang tidak begitu besar. Hal ini mengindikasikan bahwa model memiliki tingkat akurasi yang baik.

	Actual	Predicted
22	45	85.037720
0	66	120.025549
49	212	241.772534
4	205	120.025549
54	301	241.772534
18	126	85.037720
10	60	120.025549
33	185	78.960258
45	384	241.772534
12	115	120.025549

Gambar 12. Hasil perhitungan data asli dan data prediksi

Selanjutnya, pada Gambar 13, hasil perbandingan y aktual dan y prediksi ditampilkan dalam bentuk grafik, yang memudahkan visualisasi dan analisis lebih lanjut terhadap kinerja model.



Gambar 13. Grafik y actual dan y predict

#### H. Perbandingan dengan Penelitian Sebelumnya

Penelitian ini menunjukkan bahwa terdapat pengaruh signifikan antara konsumsi makanan dan minuman jadi terhadap kejadian hipertensi, dengan nilai  $P = 0,000$  yang lebih kecil dari nilai signifikan  $0,05$ . Hasil ini sejalan dengan penelitian sebelumnya oleh Annisa Destiani (2021), yang menemukan hubungan antara frekuensi konsumsi makanan cepat saji dan kejadian hipertensi pada masyarakat migran. Dalam penelitiannya, ditemukan bahwa migran yang sering mengonsumsi makanan cepat saji memiliki risiko lebih tinggi untuk mengalami hipertensi, dengan nilai  $P = 0,022$ . Oleh karena itu, dapat disimpulkan bahwa semakin sering seseorang mengonsumsi makanan cepat saji, semakin besar risiko mereka untuk mengalami hipertensi.

Penelitian ini juga menunjukkan bahwa terdapat pengaruh signifikan antara merokok dan kejadian hipertensi, dengan nilai  $P = 0,000$  yang lebih kecil dari nilai signifikan  $0,05$ . Hasil ini konsisten dengan penelitian sebelumnya oleh Irene (2019) yang membuktikan adanya hubungan antara merokok dan hipertensi di Puskesmas Kawangkoan. Studi tersebut, yang menggunakan 74 data, menemukan bahwa nilai p-value sebesar  $0,016$  lebih kecil dari  $0,05$ , menunjukkan hubungan yang signifikan antara merokok dan kejadian hipertensi. Oleh karena itu, penelitian ini mengonfirmasi bahwa merokok merupakan faktor risiko signifikan untuk hipertensi.

#### IV. SIMPULAN

Hasil dari korelasi antar variabel konsumsi sayur dengan penderita hipertensi mempunyai korelasi negatif yang lemah dibandingkan korelasi konsumsi makanan minuman jadi dan konsumsi rokok dengan penderita hipertensi yang memiliki korelasi positif yang lebih kuat. Maka variabel yang digunakan untuk diteliti ke tahap model adalah konsumsi makanan minuman jadi (X1), konsumsi rokok (X2), dan penderita hipertensi (Y).

Pada perhitungan regresi linear berganda menggunakan library statsmodels menunjukkan bahwa nilai R-square sebesar 0,741. Disimpulkan ada pengaruh variabel konsumsi makanan minuman jadi dan konsumsi rokok terhadap penderita hipertensi ialah 74,1% dan 25,9% adalah faktor yang tidak diteliti dalam penelitian ini, contohnya etnis, usia, atau faktor lain. Uji F menunjukkan bahwa model regresi secara keseluruhan signifikan, dengan *p-value* 2,75e-16 dan nilai 75,93. Uji t menunjukkan bahwa variabel konsumsi makanan minuman jadi dan konsumsi rokok memiliki dampak yang signifikan terhadap penderita hipertensi dengan *p-value* masing-masing sejumlah 0,000. Untuk nilai konstan memiliki nilai -1241,58 sedangkan nilai koefisien tiap-tiap variabel didapatkan untuk variabel konsumsi makanan minuman jadi 95,80, untuk variabel konsumsi merokok sebesar 165,86. Berdasarkan hasil pada perhitungan diatas, persamaan regresi linear berganda adalah :  $Y = -2341,58 + 95,80X1 + 165,86X2$ . Penelitian ini menunjukkan bahwa pemanfaatan *google collab* dengan bahasa pemrograman python sangat berguna dalam penggunaan analisis metode regresi linear berganda, karena memudahkan saat perhitungan model. Penelitian selanjutnya direkomendasikan untuk mengumpulkan dataset yang lebih luas dan beragam, sehingga dapat meningkatkan validitas hasil penelitian.

#### UCAPAN TERIMA KASIH

Penulis berterima kasih kepada Universitas Muhammadiyah Magelang dan Dosen Pembimbing yang telah memberikan bantuan selama penelitian.

#### DAFTAR PUSTAKA

- [1] M. Falah, "Hubungan Jenis Kelamin Dengan Angka Kejadian Hipertensi Pada Masyarakat Di Kelurahan Tamansari Kota Tasikmalaya," *J. Keperawatan Kebidanan STIKes Mitra Kencana Tasikmalaya*, vol. 3, no. 1, p. 88, 2019.
- [2] S. C. Nurzanah, S. Alam, and T. I. Hermanto, "Analisis Association Rule Untuk Identifikasi Pola Gejala Penyakit Hipertensi Menggunakan Algoritma Apriori ( Studi Kasus : Klinik Rafina Medical Center ) Association Rule Analysis for Identification of Hypertension Symptoms Patterns Using Apriori Algorithm," vol. 5, no. 2, pp. 132–141, 2022, doi: 10.33387/jiko.
- [3] M. M. Santoni, N. Chamidah, and N. Matondang, "Prediksi Hipertensi menggunakan Decision Tree, Naïve Bayes dan Artificial Neural Network pada software KNIME," *Techno.Com*, vol. 19, no. 4, pp. 353–363, 2020, doi: 10.33633/tc.v19i4.3872.
- [4] Y. Tedju, "Pengelompokan Penyakit Hipertensi Di Kota Kupang Menggunakan Metode K-Means," *HOAQ (High Educ. Organ. Arch. Qual. J. Teknol. Inf.)*, vol. 13, no. 2, pp. 98–108, 2022, doi: 10.52972/hoaq.vol13no2.p98-108.
- [5] M. Yunus, "Hubungan Usia dan Jenis Kelamin dengan Kejadian Hipertensi di Puskesmas Haji Pemanggilan Kecamatan Anak Tuha Kab. Lampung Tengah," *Anal. Pengetah. Keuangan, Kepribadian Dan Sikap Keuang. Terhadap Perilaku Manaj. Keuang.*, vol. 11, no. 1, pp. 192–201, 2021, [Online]. Available: [https://journals.ekb.eg/article\\_243701\\_6d52e3f13ad637c3028353d08aac9c57.pdf](https://journals.ekb.eg/article_243701_6d52e3f13ad637c3028353d08aac9c57.pdf).
- [6] Fajariyah, "Risiko Pola Konsumsi Dan Status Gizi Pada Kejadian Hipertensi Masyarakat Migran Di Indonesia," *Media Gizi Indones.*, vol. 16, no. 2, p. 194, 2021, [Online]. Available: <https://e-journal.unair.ac.id/MGI/article/view/21034>.

- 
- [7] H. N. Putri, E. Budianita, F. Syafria, and F. Insani, "Penerapan Algoritma Apriori Dalam Menentukan Pola Perilaku Dan Gaya Hidup Terhadap Penderita Hipertensi," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 5, no. 3, pp. 450–458, 2022, doi: 10.32672/jnkkti.v5i3.4402.
- [8] A. C. Stevenson *et al.*, "Neighborhood Fast-Food Environments and Hypertension in Canadian Adults," *Am. J. Prev. Med.*, vol. 65, no. 4, pp. 696–703, 2023, doi: 10.1016/j.amepre.2023.04.005.
- [9] A. A. Rihi Leo, S. MCh Willa, and Anita Bilaut, "Hubungan Konsumsi Laru, Gara, Sayur, dan Buah Terhadap Resiko Hipertensi Pria Dewasa Kupang," *J. Ilm. Gizi Kesehat.*, vol. 1, no. 02, pp. 1–9, 2020.
- [10] E. Garwahasada and B. Wirjatmadi, "Hubungan Jenis Kelamin, Perilaku Merokok, Aktivitas Fisik dengan Hipertensi Pada Pegawai Kantor," *Media Gizi Indones.*, vol. 15, no. 1, pp. 60–65, 2020, [Online]. Available: <https://e-journal.unair.ac.id/MGI/article/view/12314/9068>.
- [11] A. Mustafa and M. Rahimi Azghadi, "Automated machine learning for healthcare and clinical notes analysis," *Computers*, vol. 10, no. 2, pp. 1–31, 2021, doi: 10.3390/computers10020024.
- [12] H. Hartono, K. Khotimah, and A. Wibowo, "Deteksi Serangan Remote Code Execution Dan Cross Site Scripting Menggunakan Machine Learning," *J. Inform.*, vol. 23, no. 2, pp. 229–242, 2023, doi: 10.30873/ji.v23i2.3931.
- [13] A. Roihan, P. A. Sunarya, and A. S. Rafika, "Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper," *IJCIT (Indonesian J. Comput. Inf. Technol.)*, vol. 5, no. 1, pp. 75–82, 2020, doi: 10.31294/ijcit.v5i1.7951.
- [14] D. Novianty, N. D. Palasara, and M. Qomaruddin, "Algoritma Regresi Linear pada Prediksi Permohonan Paten yang Terdaftar di Indonesia," *J. Sist. dan Teknol. Inf.*, vol. 9, no. 2, p. 81, 2021, doi: 10.26418/justin.v9i2.43664.
- [15] N. Purwati, R. Nurlistiani, N. Purwati, R. Nurlistiani, and O. Devinsen, "Data Mining Dengan Algoritma Neural Network Dan Visualisasi Data," *J. Inform.*, vol. 20, no. 2, pp. 156–163, 2020, [Online]. Available: <https://jurnal.darmajaya.ac.id/index.php/JurnalInformatika/article/view/2273>.
- [16] G. N. Ayuni and D. Fitriannah, "Penerapan metode Regresi Linear untuk prediksi penjualan properti pada PT XYZ," *J. Telemat.*, vol. 14, no. 2, pp. 79–86, 2019, [Online]. Available: <https://journal.ithb.ac.id/telematika/article/view/321>.
- [17] R. Dahlia, S. Hadiani, R. Dahlia, N. Wuryani, W. Gata, and A. Selawati, "Penerapan Data Mining Terhadap Data Covid-19 Menggunakan Algoritma Klasifikasi," *Sekol. Tinggi Ilmu Komput. Nusa Mandiri Jakarta*, vol. 21, no. 1, pp. 44–52, 2021.
- [18] I. M. Latief, A. Subekti, and W. Gata, "Prediksi Tingkat Pelanggan Churn Pada Perusahaan Telekomunikasi Dengan Algoritma Adaboost," *J. Inform.*, vol. 21, no. 1, pp. 34–43, 2021, doi: 10.30873/ji.v21i1.2867.
- [19] A. Rismanitanti, R. Mawarni, S. Rahmatullah, D. M. Efendi, and S. Nurbaiti, "Perbandingan Pengolahan Data Prediksi Persediaan Gas Lpg 3Kg Menggunakan Regresi Linier Berganda Dan K-Means," *J. Inform.*, vol. 22, no. 2, pp. 183–195, 2022, doi: 10.30873/ji.v22i2.3376.
- [20] Y. Tabuni and C. Suruan, "Analisis Upaya Penggunaan Sistem Komputer di Kalangan Karyawan dengan Model Unified Theory of Acceptance and Use of Technology," vol. 24, no. 1, pp. 52–64, 2024.