

PREDIKSI PENYAKIT HIPERTENSI MENGGUNAKAN MACHINE LEARNING DENGAN ALGORITMA REGRESI LOGISTIK

Annisa Az'zahra Tarimana, Muhammad Ryan Septian Fajar,
Muhammad Azriel Saktiawan, Rizal Adi Saputra

Teknik Informatika, Universitas Halu Oleo
Jalan H.E.A. Mokodompit Kendari, Indonesia
annisatarimana555@gmail.com

ABSTRAK

Hipertensi adalah masalah kesehatan serius yang mempengaruhi sebagian besar populasi global. Deteksi dini dan pengelolaan yang tepat sangat penting untuk mencegah komplikasi serius, seperti penyakit kardiovaskular yang dapat berujung pada kematian. Penelitian ini menggunakan pendekatan machine learning dengan algoritma regresi logistik untuk mengidentifikasi hipertensi berdasarkan data klinis, termasuk informasi demografis, faktor gaya hidup, dan riwayat medis pasien. Melalui seleksi fitur dan pelatihan model, regresi logistik terbukti efektif dalam memprediksi hipertensi. Hasil penelitian menunjukkan bahwa regresi logistik mencapai akurasi 91% pada data uji dan F1-score sebesar 0.934, menandakan keseimbangan yang baik antara presisi dan recall. Model ini memberikan manfaat signifikan bagi klinisi dalam merancang intervensi dini dan strategi pengobatan yang lebih personal. Selain itu, penggunaan machine learning dalam konteks ini dapat mempercepat proses diagnosis dan meningkatkan akurasi penanganan. Penelitian lebih lanjut diharapkan memanfaatkan dataset yang lebih luas dan mengembangkan algoritma machine learning yang lebih canggih untuk meningkatkan akurasi prediksi dan memperluas aplikasi klinis, sehingga membantu lebih banyak pasien mendapatkan perawatan tepat waktu dan efektif.

Kata kunci : deteksi dini, hipertensi, machine learning, regresi logistik.

1. PENDAHULUAN

Hipertensi menjadi salah satu penyebab utama kematian tertinggi secara global dan sering disebut sebagai *silent killer* karena gejalanya yang tidak terlihat namun mematikan. Di Indonesia, tercatat 427.218 kasus kematian akibat hipertensi. Kasus hipertensi paling banyak ditemukan pada kelompok usia 31-44 tahun (31,6%), 45-54 tahun (45,3%), dan 55-64 tahun (55,2%). Dari total prevalensi hipertensi yang mencapai 34,1%, hanya 8,8% yang sudah terdiagnosis, dan dari jumlah tersebut, 13,3% penderita tidak rutin mengonsumsi obat. Selain itu, 32,3% penderita lainnya juga tidak patuh dalam minum obat secara teratur. Hal ini menunjukkan bahwa banyak penderita hipertensi tidak menyadari kondisinya dan tidak mendapatkan pengobatan yang tepat [1].

Gejala umum yang dialami penderita hipertensi meliputi sakit kepala, pusing, mual, muntah, nyeri di tengkuk, kelelahan, kelemahan, rasa cemas, sesak napas, mimisan, dan bahkan bisa menyebabkan penurunan kesadaran. Hipertensi dulu lebih sering dikaitkan dengan usia lanjut, namun kini semakin banyak anak muda yang juga mengalami kondisi ini. Faktor-faktor yang dapat memicu hipertensi mencakup jenis kelamin, usia, faktor genetik, tingkat stres, obesitas, kebiasaan konsumsi alkohol, merokok, asupan garam yang tinggi, kurangnya aktivitas fisik, serta riwayat diabetes dan masalah ginjal. Oleh karena itu, angka prevalensi hipertensi terus meningkat [2].

Berdasarkan permasalahan yang telah dipaparkan sebelumnya, diperlukan upaya yang tepat untuk mendeteksi hipertensi secara dini agar risiko

komplikasi serius dapat diminimalkan. Dengan adanya deteksi dini, para tenaga medis dapat memprediksi risiko hipertensi pada pasien dan merancang strategi penanganan yang efektif. Penelitian ini bertujuan untuk memberikan solusi melalui penerapan sistem prediksi hipertensi menggunakan algoritma regresi logistik. Sistem ini diharapkan mampu memprediksi kondisi hipertensi berdasarkan data klinis, sehingga dapat membantu klinisi dalam menentukan langkah intervensi dan pengobatan yang sesuai [3].

Algoritma regresi logistik adalah metode statistik yang digunakan untuk menentukan hubungan antara variabel dependen (dalam hal ini, status hipertensi) dan beberapa variabel independen seperti usia, riwayat kesehatan, dan gaya hidup [3]. Ini dapat membantu dalam memprediksi apakah seseorang memiliki risiko terkena penyakit hipertensi berdasarkan faktor-faktor tersebut [4].

Metode Regresi Linier dan variannya telah digunakan untuk memberikan solusi pada beragam permasalahan yang memerlukan prediksi. Beberapa diantaranya adalah untuk prediksi penyakit jantung koroner [5], prediksi risiko diabetes prediksi [6], prediksi klasifikasi kecelakaan lalu lintas [7], akurasi prediksi penyakit diabetes mellitus [8], sedangkan kombinasinya dengan metode K-Nearest Neighbor telah digunakan untuk memprediksi penyakit diabetes [9].

Dari hasil ini, dapat disimpulkan bahwa metode regresi linier efektif dalam memprediksi beberapa penyakit dan permasalahan lainnya. Penelitian ini juga menggunakan metode regresi linier sederhana untuk memprediksi penyakit hipertensi, dengan hasil yang

menunjukkan kinerja baik berdasarkan pengujian menggunakan data train dan data test [10].

Dalam penelitian ini, algoritma Regresi Logistik digunakan untuk mengatasi permasalahan tersebut untuk meningkatkan pemahaman dan informasi dalam memprediksi risiko hipertensi, yang nantinya dapat menjadi acuan dalam upaya pencegahan. Hasil prediksi dari pengolahan data diharapkan dapat memberikan informasi yang akurat dan cepat, serta menghasilkan pengetahuan yang lebih baik dalam analisis pencegahan hipertensi dan penemuan potensi baru. Ini juga dapat digunakan untuk merumuskan strategi pencegahan yang efektif dan menjadi dasar untuk pengambilan keputusan yang tepat.

2. TINJAUAN PUSTAKA

2.1. Hipertensi

Tekanan darah tinggi, atau hipertensi, adalah kondisi di mana tekanan darah dalam arteri meningkat. Secara umum, hipertensi adalah keadaan tanpa gejala di mana tekanan darah yang tinggi dalam arteri meningkatkan risiko terkena stroke, aneurisma, gagal jantung, serangan jantung, dan kerusakan ginjal [11].

Saat mengukur tekanan darah, akan diperoleh dua angka, angka yang lebih tinggi disebut tekanan sistolik, diukur saat jantung berkontraksi, dan angka yang lebih rendah disebut tekanan diastolik, diukur saat jantung berelaksasi. Tekanan darah biasanya ditulis dalam format tekanan sistolik diikuti oleh tekanan diastolik, seperti 120/80 mmHg, yang dibaca seratus dua puluh per delapan puluh [12].

Seseorang dikatakan memiliki tekanan darah tinggi jika tekanan sistolik mencapai 140 mmHg atau lebih, atau tekanan diastolik mencapai 90 mmHg atau lebih, atau kedua-duanya. Dalam kasus tekanan darah tinggi, biasanya terjadi kenaikan pada tekanan sistolik dan diastolik. Pada hipertensi sistolik terisolasi, tekanan sistolik mencapai 140 mmHg atau lebih, sementara tekanan diastolik tetap di bawah 90 mmHg, yang masih dalam kisaran normal. Jenis hipertensi ini sering ditemukan pada orang yang lebih tua [12].

2.2. Machine learning

Machine learning adalah metode berbasis komputer yang bisa belajar sendiri dari data tanpa perlu diatur secara manual oleh manusia. Metode ini semakin pintar seiring dengan banyaknya data yang diolah, mirip dengan belajar dari pengalaman. *Machine learning* sering digunakan untuk menyelesaikan kasus seperti klasifikasi dan clustering, terutama ketika menangani data dalam skala besar atau big data [13].

Dalam *Machine learning*, pengolahan data dibagi menjadi dua bagian: data *training* dan data *testing*. Data *training* digunakan untuk melatih algoritma atau metode yang dipakai dalam *Machine learning*, sedangkan data *testing* digunakan untuk mengevaluasi kinerja algoritma yang sudah dilatih tersebut pada data baru yang belum pernah ditemui sebelumnya dalam proses data mining [14].

Teknik pembelajaran dalam *Machine learning* dibagi menjadi dua yaitu *supervised learning* dan *unsupervised learning*. *Supervised learning* adalah teknik di mana dataset yang digunakan sudah memiliki label, sehingga mesin bisa mempelajari dan mengidentifikasi label tersebut prediksi. Sebaliknya, *unsupervised learning* adalah teknik di mana mesin menarik kesimpulan sendiri dari informasi yang diperlukan tanpa adanya label pada data, sehingga cocok untuk data yang belum terlabeli [15].

2.3. Regresi Logistik

Model regresi logistik adalah model statistik yang digunakan untuk mengetahui pengaruh variabel prediktor (X) terhadap variabel respon (Y) di mana variabel responnya berupa data dikotomi. Artinya, nilai dari variabel respon adalah 1 jika memenuhi kriteria tertentu dan 0 jika tidak memenuhi kriteria tersebut. Menurut Hosmer dan Lemeshow, model ini membantu kita memahami hubungan antara variabel prediktor dan variabel respon dalam bentuk yang sederhana dan jelas [13].

$$\pi(x_i) = \frac{e^{\beta_0 + \beta_1 x_{1i}}}{1 + e^{\beta_0 + \beta_1 x_{1i}}} \quad (1)$$

Dengan fungsi logit $g(x_i)$ yaitu:

$$g\pi(x_i) = \left[\ln \frac{\pi(x_i)}{1 - \pi(x_i)} \right] = \beta_0 + \beta_1 x_{1i} \quad (2)$$

Dari persamaan (1) dan (2) dapat disederhanakan menjadi:

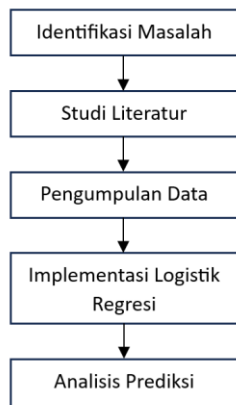
$$\pi(x_i) = \frac{e^{g(x_i)}}{1 + e^{g(x_i)}} \quad (3)$$

Regresi logistik adalah algoritma dalam *Machine learning* yang sering digunakan untuk tugas klasifikasi. Metode ini merupakan jenis regresi khusus yang bekerja dengan variabel respon biner, sementara variabel prediktornya bisa berupa data kontinu, kategori, atau kombinasi keduanya. Tidak seperti beberapa teknik lain, regresi logistik tidak membutuhkan asumsi distribusi normal multivariat atau kesamaan matriks kovarian, sehingga fleksibel untuk diaplikasikan pada berbagai jenis skala data [16].

3. METODE PENELITIAN

Jenis penelitian yang digunakan dalam penelitian ini adalah penelitian kuantitatif. Penelitian kuantitatif bersifat objektif dan dapat diukur dengan instrumen data atau statistik yang valid. Metode ini memungkinkan pengumpulan data numerik yang dapat dianalisis secara statistik untuk mendapatkan hasil yang jelas. Desain penelitian ini mengikuti pendekatan eksperimental, bertujuan untuk menguji hipotesis dan mengidentifikasi hubungan antar variabel. Analisis regresi logistik digunakan untuk

mendapatkan wawasan tentang faktor-faktor yang berkontribusi terhadap hipertensi.



Gambar 1. Desain penelitian

Desain penelitian yang ditampilkan dalam Gambar 1 memberikan gambaran mengenai alur dan metodologi yang diterapkan dalam penelitian ini.

3.1. Identifikasi Masalah

Pada tahap ini, kita mengidentifikasi dan merumuskan masalah dalam proses diagnosis hipertensi, yaitu bahwa prediksi penyakit masih dilakukan secara manual. Metode ini menghambat efisiensi tim medis dan meningkatkan risiko kesalahan manusia, yang dapat mengurangi akurasi informasi untuk pasien. Dengan meningkatnya jumlah penderita hipertensi, diperlukan sistem yang dapat memprediksi penyakit ini secara otomatis menggunakan teknologi *Machine learning*. Sistem ini diharapkan dapat menganalisis data klinis secara cepat dan akurat, sehingga tim medis dapat melakukan intervensi lebih awal dan meningkatkan kualitas pelayanan kesehatan bagi pasien.

3.2. Studi Literatur

Pada tahap ini, kami mencari metode yang tepat untuk menyelesaikan masalah dalam memprediksi penyakit hipertensi. Dari tinjauan jurnal ilmiah dan penelitian sebelumnya, algoritma Regresi Logistik ditemukan sangat cocok untuk tujuan ini. Metode ini dipilih karena kemampuannya memberikan hasil prediksi yang cepat dan akurat berdasarkan data yang tersedia. Regresi Logistik efektif dalam menangani data biner, sehingga relevan untuk memprediksi keberadaan hipertensi pada pasien dengan mempertimbangkan faktor-faktor seperti usia, tekanan darah sistolik dan diastolik, indeks massa tubuh (BMI), kadar kolesterol, dan riwayat hipertensi. Selain itu, algoritma ini sederhana dan mudah diinterpretasikan, sehingga hasilnya dapat dengan mudah dipahami oleh tim medis, meningkatkan akurasi diagnosis dan mempercepat proses pengambilan keputusan.

3.3. Pengumpulan Data

Data dikumpulkan melalui data pada Puskesmas Mokoau Kota Kendari yaitu data 1 tahun terakhir yakni dari April 2023 sampai dengan Desember 2023, dengan jumlah total data aktual sebanyak 1.018 data.

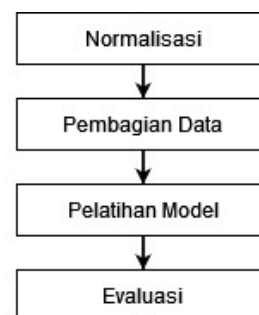
Tabel 1. Data Hipertensi

Age	SBP	DBP	CHOL	BMI	HTN
69	96	116	256	24.77854221	1
32	97	84	168	21.19008305	0
89	154	61	181	31.47239326	1
78	146	94	286	30.61832487	1
38	156	66	161	24.84185813	1
41	177	80	288	18.91569133	1
20,	176	77	195	32.34345590	1
...	...	...	...	...	...
67	166	63	230	25.06179283	1

Pada Tabel 1 dapat dilihat bahwa data yang dikumpulkan mencakup informasi seperti usia, tekanan darah sistolik dan diastolik, indeks massa tubuh (BMI), kadar kolesterol, dan riwayat hipertensi.

3.4. Tahapan Logistik Regresi

Pada tahap ini, data yang telah dikumpulkan diolah menggunakan algoritma Regresi Logistik.



Gambar 2. Metode Regresi Logistik

Pada gambar 2 terdapat langkah - langkah dalam mengolah data menggunakan algoritma regresi logistik yaitu sebagai berikut

3.5. Normalisasi Data

Pada tahap ini data telah dinormalisasikan Data yang dimasukkan ke dalam Regresi Logistik akan diskalakan menggunakan metode Standard Scaling. Proses ini melibatkan pengurangan setiap nilai pada titik data dengan rata-rata variabel tersebut, lalu membaginya dengan standar deviasi. Tujuan dari standardisasi ini adalah untuk memastikan semua variabel memiliki skala yang sama, sehingga dapat meningkatkan kinerja model. Dengan mengubah data menjadi distribusi dengan rata-rata nol dan deviasi standar satu, model akan berfungsi lebih efisien dan menghindari bias terhadap variabel dengan skala lebih besar. Selain itu, standard scaling juga mempercepat proses pelatihan model, membantu algoritma untuk konvergen lebih cepat, dan meningkatkan akurasi prediksi hipertensi.

3.6. Pembagian data *training* dan *testing*

Pembagian data untuk pelatihan (*training*) dan pengujian (*testing*) adalah langkah penting dalam pengembangan model. Dalam penelitian ini, 80% dari total dataset akan dipilih secara acak sebagai data *training*, sementara 20% sisanya akan digunakan sebagai data *testing*. Data *training* digunakan untuk melatih model Regresi Logistik agar dapat mengenali pola dalam data. Sementara itu, data *testing* berfungsi untuk mengevaluasi kinerja model pada data yang belum pernah dilihat sebelumnya. Untuk meningkatkan akurasi, setiap titik data akan distandarisasi menggunakan rumus *z-score*, dimana:

$$Z = \frac{x_i - \mu}{\sigma} \quad (4)$$

x_i = Titik data point
 μ = Rata – rata
 σ = standar deviasi
 z = nilai *Z* – score

Dengan proses ini, model diharapkan dapat memberikan prediksi yang lebih baik terhadap penyakit hipertensi.

3.7. Pelatihan Model

Dalam proses pelatihan, model akan mempelajari hubungan antara variabel-variabel independen seperti tekanan darah sistolik, tekanan darah diastolik, dan BMI, dengan variabel dependen atau target, yaitu status hipertensi. Dengan memeriksa multikolinearitas antara fitur-fitur dalam set pelatihan. Regresi Logistik menghitung probabilitas seorang individu mengalami hipertensi berdasarkan nilai input, dan hasil pelatihannya menghasilkan bobot atau koefisien yang menunjukkan kontribusi masing-masing variabel terhadap kemungkinan terjadinya hipertensi. Proses scaling dilakukan menggunakan *StandardScaler* untuk menstandarkan fitur-fitur dalam set pelatihan dan pengujian.

3.8. Evaluasi

Evaluasi model Regresi Logistik dilakukan dengan menggunakan *confusion matrix* untuk menilai kinerjanya dalam memprediksi hipertensi. *Confusion matrix* menghasilkan beberapa metrik utama, seperti akurasi, presisi, recall, dan F1-score. Recall atau sensitivitas menilai kemampuan model dalam menangkap semua kasus positif yang sebenarnya ada, dan F1-score memberikan keseimbangan antara presisi dan recall, terutama jika data tidak seimbang. Dengan metrik-metrik ini, kita dapat mengukur seberapa baik model bekerja dalam mendeteksi hipertensi berdasarkan data yang telah diuji.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Data yang dikumpulkan melalui data pada Puskesmas Mokoau Kota Kendari, terdiri dari 1018

observasi tanpa nilai yang hilang dengan variabel usia, tekanan darah sistolik, tekanan darah diastolik, kolesterol, BMI, dan hipertensi.

	Age	Systolic_BP	Diastolic_BP	Cholesterol	BMI	Hypertension
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	52.881000	134.110000	89.317000	224.862000	26.612228	0.695000
std	20.958915	26.038662	17.357067	43.200176	4.737398	0.460638
min	18.000000	90.000000	60.000000	150.000000	18.510781	0.000000
25%	34.750000	111.000000	75.000000	188.000000	22.491363	0.000000
50%	52.500000	133.000000	89.000000	225.000000	26.530516	1.000000
75%	71.000000	157.000000	104.000000	261.250000	30.675236	1.000000
max	89.000000	179.000000	119.000000	299.000000	34.992702	1.000000

Gambar 3. Analisis Statistik

Dapat dilihat pada gambar 3, rata-rata usia adalah 52,88 tahun dengan standar deviasi 20,96, menunjukkan variasi usia yang cukup besar. Tekanan darah sistolik rata-rata adalah 134,11 mmHg dan diastolik 89,32 mmHg, dengan distribusi yang mencakup rentang dari nilai normal hingga hipertensi. Rata-rata kadar kolesterol adalah 224,86 mg/dL, dan BMI rata-rata 26,61, yang termasuk dalam kategori berat badan normal hingga overweight. Sebanyak 69,5% dari sampel mengalami hipertensi. Data ini memberikan gambaran yang luas mengenai faktor-faktor risiko kardiovaskular dalam populasi yang diteliti.

4.2. Normalisasi Data

Pada analisis ini, terdapat lima variabel input yang digunakan untuk memprediksi kemungkinan seseorang mengalami hipertensi. Variabel-variabel tersebut meliputi *Age*, *Systolic_BP*, *Diastolic_BP*, *Cholesterol*, dan *BMI*. Data yang ditampilkan merupakan hasil normalisasi, di mana setiap nilai telah disesuaikan untuk berada dalam rentang antara 0 dan 1. Ini penting untuk memastikan bahwa setiap variabel memiliki pengaruh yang seimbang dalam model prediksi. Kolom target, yang menunjukkan apakah seseorang menderita hipertensi atau tidak, tidak termasuk dalam daftar variabel input ini. Dengan pendekatan ini, analisis bertujuan untuk meningkatkan akurasi dan efektivitas model dalam memprediksi hipertensi berdasarkan data klinis yang relevan. Selanjutnya, dilakukan normalisasi agar data memiliki rentang antara 0 dan 1.

Age	Systolic_BP	Diastolic_BP	Cholesterol	BMI
0.181495	0.277580	0.348754	0.857651	0.024123
0.049822	0.281139	0.234875	0.533808	0.011353
0.252669	0.483986	0.153025	0.580071	0.047944
0.213523	0.455516	0.270463	0.953737	0.044905
0.071174	0.491103	0.170819	0.508897	0.024348

Gambar 4. Normalisasi Data

Dari data pada Gambar 4, semua data telah melalui proses normalisasi, sehingga nilai setiap fitur berada dalam rentang antara 0 dan 1, yang bertujuan

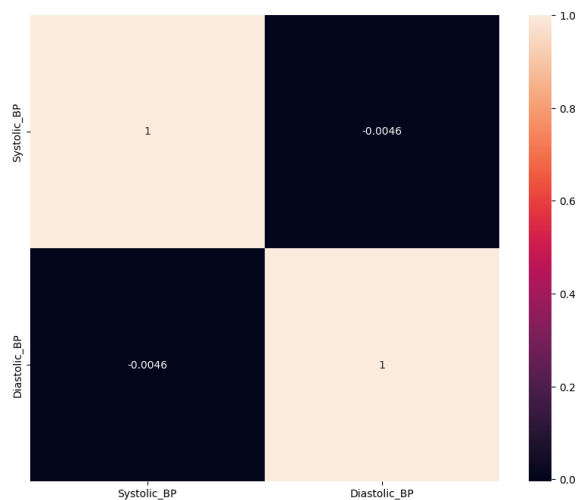
untuk memudahkan model *Machine learning* dalam memproses dan menganalisis data.

4.3. Pembagian Data

Setelah kolom yang tidak diperlukan di-drop, data kemudian dibagi menjadi dua subset yaitu data pelatihan (*training set*) dan data pengujian (*testing set*). Pembagian ini dilakukan menggunakan fungsi *train_test_split* dari *scikit-learn*. Set data dibagi dengan proporsi 80% untuk pelatihan dan 20% untuk pengujian, dengan parameter *random_state=42* untuk memastikan pembagian yang konsisten, serta *stratify=y* untuk menjaga proporsi target variabel tetap sama di kedua set.

4.4. Pelatihan Model

Langkah selanjutnya adalah memeriksa multikolinearitas antara fitur-fitur dalam set pelatihan. Ini dilakukan dengan visualisasi *heatmap* menggunakan *sns.heatmap*.

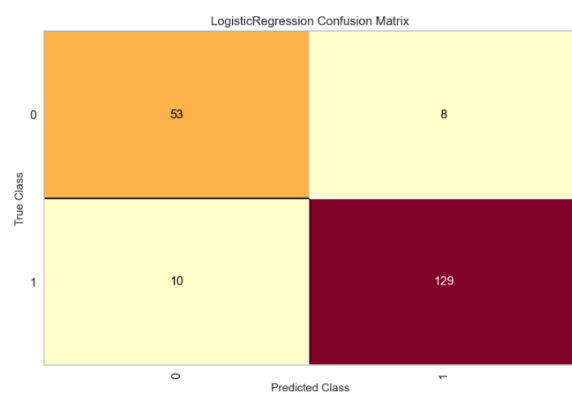


Gambar 5. Visualisasi *Heatmap*

Dapat dilihat pada Gambar 5, *heatmap* menunjukkan bahwa tidak ada hubungan korelasi yang signifikan antara dua fitur utama, yaitu *Systolic_BP* dan *Diastolic_BP*. Nilai korelasi antara kedua fitur ini sangat kecil (-0.0046), yang menunjukkan bahwa tidak ada multikolinearitas, sehingga kedua fitur tersebut dapat digunakan secara bersamaan dalam model. Proses *scaling* dilakukan menggunakan *StandardScaler* untuk menstandarkan fitur-fitur dalam set pelatihan dan pengujian. Standarisasi ini penting untuk memastikan bahwa fitur-fitur yang berbeda memiliki skala yang sama, terutama karena algoritma seperti regresi logistik sensitif terhadap perbedaan skala antar fitur. Setelah *scaling*, data yang dihasilkan adalah *X_train_scaled* untuk pelatihan dan *X_test_scaled* untuk pengujian. Tujuan dari model ini adalah untuk memprediksi status hipertensi berdasarkan fitur-fitur seperti *Systolic_BP* dan *Diastolic_BP*.

4.5. Evaluasi Model

Setelah model dilatih, langkah berikutnya adalah menguji kinerjanya dengan memasukkan data *testing*. Data ini berisi contoh yang belum pernah dilihat oleh model, sehingga memungkinkan kita menilai seberapa baik model dapat memprediksi hasil yang benar. Hasil prediksi kemudian dibandingkan dengan label sebenarnya dari data *testing*. Untuk mengevaluasi keakuratan prediksi, kita menggunakan *confusion matrix*, yang menunjukkan jumlah prediksi yang benar dan salah, terbagi menjadi *True Positives* (TP), *True Negatives* (TN), *False Positives* (FP), dan *False Negatives* (FN). Melalui analisis ini, kita dapat menghitung metrik kinerja seperti akurasi, presisi, *recall*, dan *F1-score*, serta mengidentifikasi area yang perlu diperbaiki untuk meningkatkan performa model dalam memprediksi hipertensi.



Gambar 6. *Confusion matrix*

Pada Gambar 6 *confusion matrix* menunjukkan hasil evaluasi model regresi logistik dalam memprediksi hipertensi. Terdapat empat bagian pada matriks ini:

- Kelas *true* negatif (tidak hipertensi), di mana model berhasil mengidentifikasi 53 pasien yang tidak mengalami hipertensi.
- Kelas *false* positif (prediksi hipertensi, tetapi sebenarnya tidak). Ini menunjukkan bahwa ada 8 pasien yang diidentifikasi oleh model sebagai hipertensi padahal mereka tidak.
- Kelas *false* negatif (prediksi tidak hipertensi, tetapi sebenarnya hipertensi). Ini berarti 10 pasien yang seharusnya terdeteksi sebagai hipertensi tidak teridentifikasi oleh model.
- Kelas *true* positif, di mana model dengan akurat mengidentifikasi 129 pasien yang benar-benar mengalami hipertensi.

Dari informasi ini, kita dapat menghitung beberapa metrik evaluasi, seperti akurasi, presisi, *recall*, dan *F1-score*, untuk mendapatkan gambaran menyeluruh tentang performa model. Hasil test Regresi Logistik ke data test menunjukkan bahwa Regresi Logistik berhasil memprediksi *Class 0* dan *1* secara baik.

4.6. Analisa Prediksi

Tahap ini melibatkan evaluasi model menggunakan matriks evaluasi seperti akurasi, presisi, *recall*, dan *F1-score* untuk menilai kinerja model dalam memprediksi hipertensi dengan hasil yang dapat dilihat pada tabel 2.

Tabel 2. Hasil Prediksi

<i>Acc Train</i>	<i>Prec Train</i>	<i>Rec Train</i>	<i>F1 Train</i>	<i>Acc Test</i>	<i>F1 Test</i>
0.892	0.941	0.928	0.923	0.91	0.934

Berdasarkan Tabel 2 hasil model menunjukkan akurasi sebesar 0,892 untuk data *training* dan 0,923 untuk data *testing*, mencerminkan keakuratan yang baik. Presisi mencapai 0,941 pada data *training* dan 0,91 pada data *testing*, yang menunjukkan bahwa sebagian besar prediksi positif adalah benar. Selain itu, nilai *recall* sebesar 0,928 untuk data *training* dan 0,934 untuk data *testing* menandakan kemampuan model dalam mendeteksi kasus hipertensi. Dengan nilai *F1-score* yang mencapai 0,923 pada data *training* dan 0,934 pada data *testing*, model regresi logistik terbukti efektif dalam mengidentifikasi hipertensi.

4.7. Feature Importance

Dalam menganalisis hasil prediksi, penting untuk mengevaluasi *Feature importance* yang menunjukkan pengaruh setiap variabel terhadap kinerja model regresi logistik. *Feature importance* ini diukur berdasarkan koefisien dari model regresi logistik, yang menggambarkan kontribusi relatif masing-masing fitur dalam memprediksi hipertensi. Seperti yang terlihat pada Tabel 3, terdapat dua fitur utama yang dievaluasi yaitu *Systolic_BP* dan *Diastolic_BP*.

Tabel 3. Feature Importance

<i>Feature</i>	<i>Importance</i>
<i>Systolic_BP</i>	0.114182
<i>Diastolic_BP</i>	0.195557

Dari Tabel 3 dapat dilihat hasil fitur *Systolic_BP* memiliki bobot sebesar 0.114 (dibulatkan), yang menunjukkan bahwa setiap peningkatan satu unit pada *Systolic_BP* akan meningkatkan *log likelihood* dari risiko hipertensi sebesar 0.11%. Ini menunjukkan bahwa tekanan darah sistolik memiliki pengaruh signifikan tetapi relatif lebih rendah dibandingkan dengan fitur lainnya.

Sementara itu, *Diastolic_BP* menunjukkan bobot sebesar 0.195, yang berarti setiap penambahan satu unit pada *Diastolic_BP* berkontribusi terhadap peningkatan *log likelihood* hipertensi sebesar 0.195%. Hal ini menegaskan bahwa tekanan darah diastolik memiliki dampak yang lebih besar terhadap peningkatan risiko hipertensi dibandingkan tekanan darah sistolik. Dengan demikian, hasil ini memberikan wawasan penting bagi tim medis untuk fokus pada kedua variabel tersebut dalam upaya pencegahan dan pengelolaan hipertensi.

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa metode regresi logistik efektif digunakan untuk memprediksi risiko hipertensi pada individu. Penelitian ini memanfaatkan algoritma *machine learning* untuk menganalisis variabel kesehatan seperti usia, jenis kelamin, tekanan darah sistolik, tekanan darah diastolik, BMI, dan kadar kolesterol dalam upaya mendeteksi potensi hipertensi. Hasil penelitian menunjukkan bahwa algoritma regresi logistik mampu menghasilkan prediksi dengan tingkat akurasi yang tinggi, yaitu sebesar 91% pada data uji, serta *F1-score* sebesar 0.934, yang menunjukkan keseimbangan yang sangat baik antara presisi (93.2%) dan *recall* (93.6%). Hal ini membuktikan bahwa regresi logistik dapat menjadi alat yang efektif dalam mendukung diagnosis dini hipertensi dan membantu tenaga kesehatan memantau pasien berisiko. Selain itu, variabel yang paling berpengaruh dalam prediksi adalah usia, tekanan darah sistolik, tekanan darah diastolik, BMI, dan kadar kolesterol, di mana peningkatan pada variabel-variabel ini berhubungan dengan risiko hipertensi yang lebih tinggi.

Penelitian ini menyoroti potensi regresi logistik sebagai alat klinis yang dapat membantu pengambilan keputusan awal dan merancang intervensi yang lebih personal bagi pasien hipertensi. Namun, penelitian ini juga memiliki keterbatasan, seperti jumlah data yang terbatas dan variabel yang belum mencakup semua faktor yang mungkin mempengaruhi hipertensi, seperti riwayat keluarga dan kebiasaan hidup. Untuk meningkatkan akurasi dan generalisasi model, penelitian lanjutan disarankan menggunakan dataset yang lebih besar dan lebih beragam. Selain itu, disarankan juga untuk menguji algoritma *machine learning* yang lebih kompleks, seperti Random Forest atau Support Vector Machines (SVM), untuk membandingkan hasil prediksi. Secara keseluruhan, penelitian ini membuktikan bahwa regresi logistik memiliki potensi yang besar dalam prediksi hipertensi dan memberikan manfaat signifikan dalam aplikasi klinis dan manajemen kesehatan.

DAFTAR PUSTAKA

- [1] K. Abdul Khalim, U. Hayati, and A. Bahtiar, "Perbandingan Prediksi Penyakit Hipertensi Menggunakan Metode Random Forest Dan Naïve Bayes," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 498–504, 2023, doi: 10.36040/jati.v7i1.6376.
- [2] S. Choerunnisa Nurzanah, S. Alam, and T. Iman Hermanto, "Analisis Association Rule Untuk Identifikasi Pola Gejala Penyakit Hipertensi Menggunakan Algoritma Apriori (Studi Kasus: Klinik Rafina Medical Center)," *JIKO (Jurnal Inform. dan Komputer)*, vol. 5, no. 2, pp. 132–141, 2022, doi: 10.33387/jiko.v5i2.4792.
- [3] A. Datumaya, W. Sumari, A. K. Febrianto, and Y. Pramitarini, "Sistem Prediksi Permintaan Darah Menggunakan Metode Regresi Linier

- (Studi Kasus Pada UTD PMI Kabupaten Bojonegoro),” *JIP (Jurnal Inform. Polinema)*, vol. 7, pp. 85–90, 2021.
- [4] N. S. Niko, A. Rahman, D. Marini Umi Atmaja, and A. Basri, “Prediksi Penyakit Diabetes Untuk Pencegahan Dini Dengan Metode Regresi Linear,” *Bull. Inf. Technol.*, vol. 4, no. 3, pp. 313–219, 2023, doi: 10.47065/bit.v4i3.739.
- [5] K. Chandra, D. Juan, and S. Prasetyo, “Prosiding SENAM 2024: Prediksi Penyakit Jantung Koroner Menggunakan Metode K-NN dan Regresi Logistik Berdasarkan Kerangka Kerja CRISP-DM,” *Sist. Inf. Inform.*, vol. 4, no. 2020, pp. 241–248, 2024.
- [6] E. Oktavia, A. Id Hadiana, and F. Rahmat Umbara, “Penerapan Metode Regresi Logistik Dalam Prediksi Risiko Diabetes Melitus Gestasional,” *J. Informatics Commun. Technol.*, vol. 5, no. 2, pp. 177–185, 2023.
- [7] B. L. Fauzan, T. Agustin, and A. M. H. Mahmudah, “Prediksi Klasifikasi Kecelakaan Lalu Lintas di Kota Surakarta dengan Menggunakan Metode Regresi Logistik Multinomial,” *Sustain. Civ. Build. Manag. Eng.*, vol. 1, no. 4, p. 9, 2024, doi: 10.47134/scbmej.v1i4.3159.
- [8] M. I. Gunawan, D. Sugiarto, and I. Mardianto, “Peningkatan Kinerja Akurasi Prediksi Penyakit Diabetes Mellitus Menggunakan Metode Grid Search pada Algoritma Logistic Regression,” *J. Edukasi dan Penelit. Inform.*, vol. 6, no. 3, p. 280, 2020, doi: 10.26418/jp.v6i3.40718.
- [9] A. Aprilia, P. Lo, V. Jennifer, and E. Tjioe, “Penerapan Model CRISP-DM untuk Prediksi Penyakit Diabetes Menggunakan Metode K-Nearest Neighbor dan Logistic regression,” vol. 4, pp. 48–57, 2024.
- [10] A. Anggrawan, H. Hairani, and N. Azmi, “Prediksi Penjualan Produk Unilever Menggunakan Metode Regresi Linear,” *J. Bumigora Inf. Technol.*, vol. 4, no. 2, pp. 123–132, 2022, doi: 10.30812/bite.v4i2.2416.
- [11] P. Purwono, P. Dewi, S. K. Wibisono, and B. P. Dewa, “Model Prediksi Otomatis Jenis Penyakit Hipertensi dengan Pemanfaatan Algoritma Machine Learning Artificial Neural Network,” *Insect (Informatics Secur. J. Tek. Inform.)*, vol. 7, no. 2, pp. 82–90, 2022, doi: 10.33506/insect.v7i2.1828.
- [12] U. Hasanah, “Tekanan Darah Tinggi (Hipertensi),” *J. Keperawatan Jiwa*, vol. 7, no. 1, p. 87, 2019, [Online]. Available: <https://p2ptm.kemkes.go.id/uploads/2016/10/Tekanan-Darah-Tinggi-Hipertensi.pdf>
- [13] J. Junifer Pangaribuan, H. Tanjaya, and K. Kenichi, “Mendeteksi Penyakit Jantung Menggunakan Machine Learning Dengan Algoritma Logistic Regression,” *J. Inf. Syst. Dev.*, vol. 06, no. 02, pp. 1–10, 2021.
- [14] Suherman, F. Muammar, and I. Afriantoro, “Analisa Data Mining untuk Prediksi Penyakit Kanker Paru dengan Algoritma Regresi Linear,” *J. Teknol. Pelita Bangsa*, vol. 13, no. 3, pp. 167–172, 2022.
- [15] A. Kurniadi Hermawan, A. Nugroho, and Edora, “Analisa Data Mining Untuk Prediksi Penyakit Ginjal Kronik Dengan Algoritma Regresi Linier,” *Bull. Inf. Technol.*, vol. 4, no. 1, pp. 37–48, 2023, doi: 10.47065/bit.v4i1.475.
- [16] Q. R. Cahyani *et al.*, “Prediksi Risiko Penyakit Diabetes menggunakan Algoritma Regresi Logistik Diabetes Risk Prediction using Logistic Regression Algorithm Article Info ABSTRAK,” *JOMLAI J. Mach. Learn. Artif. Intell.*, vol. 1, no. 2, pp. 2828–9099, 2022, doi: 10.55123/jomlai.v1i2.598.