JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 4, Agustus 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i4.8744
Hal 462-471
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Implementasi Algoritma Support Vector Machine (SVM) dan Random Forest Untuk Klasifikasi Penyakit Hipertensi Berdasarkan Data Kesehatan

Siti Alia Azhaar*, Tohirin Al Mudzakir, Hilda Yulia Novita, Sutan Faisal

Fakultas Ilmu Komputer, Teknik Informatika, Universitas Buana Perjuangan Karawang, Karawang, Indonesia
Email: [1,*]if21.sitiazhaar@mhs.ubpkarawang.ac.id, [2]tohirin@ubpkarawang.ac.id, [3]hilda.yulia@ubpkarawang.ac.id, [4]sutan.faisal@ubpkarawang.ac.id.
Email Penulis Korespondensi: if21.sitiazhaar@mhs.ubpkarawang.ac.id
Submitted **14-06-2025;** Accepted **28-07-2025;** Published **14-08-2025**

Abstrak

Salah satu penyakit tidak menular yang paling banyak menyebabkan kematian di Indonesia adalah hipertensi. Pada salah satu puskesmas, tingkat prevalensi penderita hipertensi cukup tinggi. Berdasarkan hasil pemeriksaan, lebih dari 1000 pasien terdiagnosis hipertensi setiap tahunnya. Permasalahan yang dihadapi di puskesmas tersebut adalah belum adanya klasifikasi data yang terstruktur mengenai pasien hipertensi dan normal. Tujuan dari penelitian ini adalah untuk membandingkan kinerja algoritma Support Vector Machine (SVM) dan Random Forest (RF) dalam menciptakan model klasifikasi hipertensi berdasarkan data pemeriksaan kesehatan dari Puskesmas Anggadita. Data sebanyak 2.500 pasien dikumpulkan dan dilakukan preprocessing, termasuk penanganan missing value, penghapusan data duplikat, transformasi data menggunakan label encoding, serta pembagian data menjadi training dan testing. Metode SVM menerapkan kernel Radial Basis Function (RBF), sedangkan RF terdiri dari 100 pohon keputusan. Evaluasi dilakukan dengan confusion matrix untuk menghitung akurasi, presisi, recall, dan f1-score. Hasil menunjukkan bahwa metode SVM memperoleh akurasi 93%, presisi 0.96 (Normal) dan 0.90 (Hipertensi), serta f1-score 0.94 dan 0.92. Sementara itu, model RF menunjukkan kinerja lebih unggul dengan akurasi 96%, presisi 0.97 (Normal) dan 0.95 (Hipertensi), serta masing-masing f1-score bernilai 0.97 dan 0.95. Dengan demikian, algoritma Random Forest memiliki performa lebih baik dalam klasifikasi data hipertensi, dapat diimplementasikan sebagai alat bantu bagi institusi kesehatan dalam mengelola data pasien.

Kata Kunci: Hipertensi; Support Vector Machine; Random Forest; Klasifikasi;

Abstract

One of the most common non-communicable diseases causing death in Indonesia is hypertension. At one community health center, the prevalence of hypertension is quite high. Based on examination results, more than 1,000 patients are diagnosed with hypertension each year. The issue faced at this health center is the lack of structured data classification for hypertensive and normal patients. The objective of this study is to compare the performance of the Support Vector Machine (SVM) and Random Forest (RF) algorithms in creating a hypertension classification model based on health examination data from the Anggadita Health Center. Data from 2,500 patients was collected and preprocessed, including handling missing values, removing duplicate data, transforming data using label encoding, and dividing the data into training and testing sets. The SVM method applied a Radial Basis Function (RBF) kernel, while the RF consisted of 100 decision trees. Evaluation was conducted using a confusion matrix to calculate accuracy, precision, recall, and F1-score. The results showed that the SVM method achieved an accuracy of 93%, precision of 0.96 (Normal) and 0.90 (Hypertension), and F1-scores of 0.94 and 0.92. Meanwhile, the RF model showed superior performance with an accuracy of 96%, precision of 0.97 (Normal) and 0.95 (Hypertension), and F1-scores of 0.97 and 0.95, respectively. Thus, the Random Forest algorithm performs better in classifying hypertension data and can be implemented as a tool to assist healthcare institutions in managing patient data.

Keywords: Hypertension; Support Vector Machine; Random Forest; Classification;

1. PENDAHULUAN

Penyakit tidak menular umumnya dikategorikan sebagai penyakit *degeneratif* yang muncul seiring dengan proses penuaan. Sebagian besar penyakit ini merupakan penyebab kematian terbanyak di Indonesia [1]. Berdasarkan informasi dari *World Health Organization (WHO)*, hipertensi merupakan penyebab utama kematian di Indonesia, yakni mencapai 66% dari total kematian, setelah penyakit jantung dan diabetes [2]. Hipertensi, yang juga dikenal sebagai tekanan darah tinggi, adalah penyakit kronis yang tidak menular dan umum terjadi di masyarakat. Hipertensi didefinisikan sebagai suatu kondisi yang ditandai dengan kenaikan tekanan darah yang tidak normal, baik pada tekanan sistolik maupun diastolik [3]. Tekanan darah tinggi umumnya diartikan sebagai kondisi di mana angka tekanan sirkulasi darah melebihi 140/90 mmHg, adapun tekanan sirkulasi darah yang normal berkisar di angka 120/80 mmHg [4].

Pada salah satu puskesmas tingkat prevalensi yang terkena hipertensi cukup tinggi, berdasarkan hasil pemeriksaan menunjukkan bahwa lebih dari 1000 pasien terdiagnosis hipertensi setiap tahunnya. Sebagian besar kasus ini dipengaruhi oleh berbagai faktor risiko, termasuk jenis kelamin, usia, konsumsi garam, kadar kolesterol, kebiasaan merokok, obesitas, kurangnya aktivitas fisik, stres, dan riwayat keluarga [5]. Di sisi lain, kesadaran masyarakat untuk memeriksa tekanan darah sejak dini masih rendah, mengakibatkan banyak kasus hipertensi tidak terdiagnosis.

Permasalahan yang dihadapi di salah satu puskesmas yaitu belum ada klasifikasi data yang terstruktur mengenai pasien hipertensi dan normal, sehingga menyulitkan pemantauan dan tindakan yang tepat. Akibatnya, puskesmas kesulitan mengidentifikasi pasien yang memerlukan perhatian khusus dan menilai hasil pengobatan. Hal ini menghambat upaya meningkatkan kesadaran masyarakat tentang pentingnya pemeriksaan tekanan darah rutin dan mengurangi risiko

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 4, Agustus 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i4.8744
Hal 462-471
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

komplikasi akibat hipertensi yang tidak tertangani dengan cepat. Oleh karena itu, untuk mengelompokkan data pasien yang berkaitan dengan hipertensi, diperlukan prosedur klasifikasi yang sistematis.

Sebelumnya telah dilakukan beberapa penelitian terkait. Penelitian oleh [2] membahas klasifikasi tekanan darah tinggi menggunakan algoritma SVM dan *Naïve Bayes* pada pasien Klinik Polresta Samarinda tahun 2022. Hasil penelitian menunjukkan bahwa akurasi metode SVM adalah 96,67%, sedangkan akurasi algoritma *Naïve Bayes* adalah 93,33%. Temuan ini menunjukkan bahwa bahwa metode SVM memiliki keunggulan dibandingkan metode *Naive Bayes* dalam aspek akurasi. Penelitian lain oleh [6] memperkirakan kemungkinan munculnya diabetes di fase awal dengan memanfaatkan metode klasifikasi *Random Forest.* Temuan menunjukkan efektivitas pendekatan ini yang tinggi, dengan nilai ROC 0,998 dan akurasi 97,88%. Kemudian dalam penelitian [7], menggunakan *Naïve Bayes Classifier* dan *Regresi Logistik Biner* untuk klasifikasi hipertensi. Hasilnya menunjukkan *sensitivitas Regresi Logistik Biner* mencapai 93,33%, sementara *Naïve Bayes Classifier* hanya mencapai 63,64%, menunjukkan bahwa *Regresi Logistik Biner* memiliki kinerja yang lebih baik.

Sementara itu penelitian oleh [8] membandingkan pendekatan *Random Forest* dan *Naïve Bayes* untuk memprediksi hipertensi. Tingkat akurasi untuk *Naïve Bayes* mencapai 85,9%, dengan presisi di angka 85,5%, dan tingkat *recall* sebesar 82,6%. Sebaliknya, *Random Forest* mencapai *accuracy*, *precision*, dan *recall* sebesar 100%, yang menunjukkan keunggulan *Random Forest* dalam prediksi hipertensi. Selanjutnya pada penelitian [9] mengkaji klasifikasi *multiclass* pada data pasien penyakit tiroid menggunakan algoritma SVM. Pendekatan *One Against One* menghasilkan akurasi 99,53%, lebih baik dibandingkan pendekatan *One Against All*. Pada penelitian [10], menggunakan metode *Random Forest* dan SVM untuk mengimplementasikan *data mining* dalam klasifikasi diabetes. Metode terbaik adalah *Random Forest* dengan pembagian data 80%:20%, yang memiliki *f1-score* 0,98, akurasi 98%, presisi 0.96, *recall* 1, dan *spesifisitas* 0,95. Faktor yang paling berpengaruh adalah *polyuria*, *polydipsia*, dan jenis kelamin. Selanjutnya pada penelitian [11] menggunakan metode *Multilayer Perceptron, Support Vector Machine,* dan *Decision Tree* untuk kajian model jaringan syaraf tiruan untuk memprediksi secara dini tingkat kelulusan mahasiswa. Hasil menunjukkan model *Decision Tree* menghasilkan tingkat *error rate* 0 dan akurasi 100%. Diikuti oleh SVM dengan *error rate* 0,011 dan akurasi 98,9%, sedangkan model *Multilayer Perceptron* memiliki *error rate* 0,029 dengan akurasi 97,1%. Berdasarkan pengujian *Confusion Matrix*, *Decision Tree* direkomendasikan untuk sistem prediksi kelulusan mahasiswa tepat waktu. Kemudian pada penelitian [12] melakukan perbandingan performa algoritma *Support Vector Machine* dan algoritma *Regresi Logistik*. Hasil evaluasi menunjukkan model SVM dengan SMOTE unggul dengan akurasi, presisi, dan *recall* masing-masing 1,0. Sementara itu, model LR tanpa SMOTE memiliki akurasi 0,97, presisi 1,0, dan *recall* 0,90. Hal ini menunjukkan bahwa model SVM dengan SMOTE cenderung lebih unggul dibandingkan yang tanpa SMOTE. Sementara itu, penelitian yang dilakukan oleh [13] untuk analisis sentimen pindah ibu kota negara (ikn) baru pada twitter menggunakan algoritma *Naive Bayes* dan *Support Vector Machine (SVM).* Hasil menunjukkan *Naive Bayes* memiliki nilai akurasi sebesar 86.94%, presisi 96.24% dan *recall* 86.66%. Sedangkan metode SVM menghasilkan nilai akurasi 90.81%, presisi 90.12% dan nilai *recall* 99.12% yang menunjukkan bahwa metode SVM lebih unggul dari *Naive Bayes*. Penelitian selanjutnya oleh [14] melakukan klasifikasi peringkat pengguna aplikasi seluler berdasarkan data dari google play store menggunakan *Logistic Regression, K-Nearest Neighbor*, dan *Support Vector Machine.* Hasil diperoleh bahwa algoritma *Logistic Regression* memiliki akurasi 84,78%, presisi 87,24%, dan *recall* 62,16%. Algoritma SVM mencatat akurasi 86,67% dan *recall* 76,88%, sedangkan KNN menghasilkan tingkat akurasi 88,59%, presisi 91,07%, dan *recall* 71,88% yang menunjukkan performa lebih unggul dalam mengklasifikasikan data.

Beberapa penelitian sebelumnya telah berhasil menerapkan algoritma klasifikasi seperti SVM, *Random Forest*, dan lainnya pada berbagai kasus kesehatan, sebagian besar studi tersebut tidak secara spesifik membahas klasifikasi hipertensi pada tingkat pelayanan kesehatan seperti puskesmas. Selain itu, belum banyak studi yang secara langsung membandingkan kinerja algoritma *Random Forest* dan SVM dalam konteks klasifikasi hipertensi. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan model klasifikasi yang mampu membedakan antara pasien dengan tekanan darah normal dan hipertensi, serta mengevaluasi performa dari model *Random Forest* dan *Support Vector Machine (SVM)* guna mendukung pengambilan keputusan medis yang cepat dan terstruktur di puskesmas.

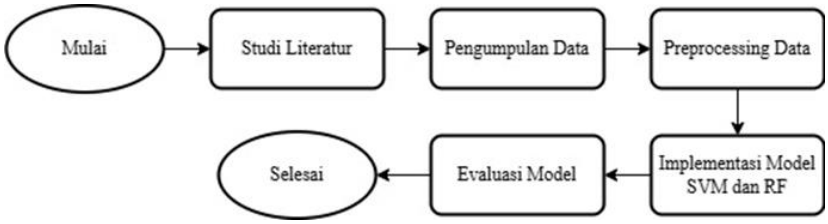
2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Perancangan studi ini dilakukan melalui pendekatan terencana dan sistematis guna memastikan setiap tahapannya dapat terlaksana secara maksimal. Langkah-langkah penelitian dirancang secara berurutan agar saling melengkapi dalam mencapai tujuan yang telah ditentukan.

Untuk menjelaskan alur proses yang diterapkan dalam penelitian ini secara menyeluruh, mulai dari studi literatur, pengumpulan data, *preprocessing data*, implementasi model SVM dan RF, dan evaluasi model. Gambar 1 menyajikan diagram tahapan penelitian dari awal hingga akhir.

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 4, Agustus 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i4.8744
Hal 462-471
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom



Gambar 1. Tahapan Penelitian

2.2 Studi Literatur

Studi literatur dilaksanakan untuk meneliti dan menganalisis penerapan model SVM dan *Random Forest* untuk pengklasifikasian penyakit hipertensi. Studi ini mengacu pada beragam studi sebelumnya untuk menilai efektivitas, tingkat akurasi, dan kelebihan dari masing-masing algoritma dalam mendeteksi pasien hipertensi berdasarkan data kesehatan dan faktor risiko yang berkaitan.

2.3 Pengumpulan Data

Pengumpulan data dilakukan dari hasil pemeriksaan di Puskesmas Anggadita, dengan total 2.500 data pasien dari tahun 2023, 2024 dan 2025 pada bulan Januari sampai Februari. Data ini terbagi menjadi dua kelompok: "Hipertensi" bagi pasien yang telah didiagnosis menderita tekanan darah tinggi dan "Normal" untuk pasien yang tidak memiliki diagnosis tekanan darah tinggi. Dataset tersebut mencakup berbagai informasi, seperti nama, usia, jenis kelamin, riwayat penyakit pada keluarga atau diri sendiri, kebiasaan merokok, kurangnya aktivitas fisik, pola makan, konsumsi alkohol, tekanan darah (sistolik dan diastolik), serta indeks massa tubuh (IMT).

2.4 *Preprocessing* Data

Tahap *preprocessing* pada studi ini dilakukan guna mengoptimalkan kondisi dataset sebelum digunakan dalam proses klasifikasi hipertensi. Berikut ini adalah beberapa langkah *preprocessing* data yang dijelaskan:

a. Mengatasi *Missing Values*

Mengatasi *missing values* dilakukan untuk mengidentifikasi kolom yang memiliki nilai kosong *(missing values)*. Jika kolom penting, isi nilai kosong dengan metode imputasi, contohnya mean/median untuk kolom numerik. Modus (nilai yang paling sering muncul) untuk kolom kategorikal. Kemudian jika kolom tidak relevan, maka dapat dihapus.

b. Menghapus Data Duplikat

Menghapus data duplikat dilakukan untuk meningkatkan akurasi, dan kualitas data dengan memastikan bahwa setiap entri yang disimpan bersifat unik dan tidak berulang. Data duplikat ini dilakukan dengan mencari data yang tercatat lebih dari satu kali, kemudian hapus data duplikat atau gabungkan data jika informasi tambahan tersedia.

c. *Transformasi* Data

Transformasi data dilakukan untuk mengubah format, skala, atau distribusi data agar lebih sesuai untuk analisis atau pemodelan. Dalam mengonversi variabel kategori menjadi format angka, penelitian ini menerapkan *label encoding*. Contoh penerapan *label encoding* untuk mengubah data kategorikal menjadi numerik dalam proses *transformasi* data ditampilkan pada Tabel 1.

Tabel 1. *Label Encoding*

Kategori	Nilai Asli	*Label Encoding*
Jenis Kelamin	Laki-laki	0
	Perempuan	1
Diagnosis	Normal	0
	Hipertensi	1
Merokok	Tidak	0
	Ya	1
Kurang Aktivitas Fisik	Tidak	0
	Ya	1
Gula Berlebihan	Tidak	0
	Ya	1
Garam Berlebihan	Tidak	0
	Ya	1
Lemak Berlebihan	Tidak	0
	Ya	1
Kurang Makan Buah dan Sayur	Tidak	0
	Ya	1
Konsumsi Alkohol	Tidak	0

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 4, Agustus 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i4.8744
Hal 462-471
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

	Ya	1

d. Pembagian Dataset

Setelah data dipersiapkan, data dibagi menjadi dua bagian: 20% digunakan untuk pengujian dan 80% digunakan untuk pelatihan. Satu bagian dari data dipakai untuk melatih model, sedangkan bagian lainnya digunakan untuk menilai seberapa baik model yang sudah dilatih.

2.5 Implementasi Model SVM dan RF

Salah satu tahap penting dalam pengelompokan hipertensi adalah penggunaan model SVM dan RF. Kedua pendekatan ini memproses data dengan cara yang berbeda.

a. *Support Vector Machine (SVM)*

Support Vector Machine (SVM) adalah metode klasifikasi yang berfungsi untuk memisahkan berbagai kategori data. Tujuannya adalah menemukan *hyperplane* atau batas pemisah paling efektif yang mampu membedakan satu kelompok data dengan kelompok data lainnya secara optimal [6]. Dalam studi ini, digunakan algoritma *Support Vector Machine (SVM)* dengan kernel *Radial Basis Function (RBF)* sebagai metode klasifikasi untuk mengidentifikasi pasien yang menderita hipertensi. Adapun rumus untuk memisahkan dua klasifikasi, digunakan *hyperplane* yang optimal dengan menerapkan persamaan sebagai berikut:

$$f(x) = \omega.x + b \quad (1)$$

Dalam hal ini, ω merepresentasikan vektor bobot yang tegak lurus terhadap *hyperplane*, x merupakan vektor fitur dari data input, dan b berperan sebagai nilai bias. Pada proses klasifikasi, algoritma SVM berupaya untuk meningkatkan margin, yang merupakan jarak antara *hyperplane* dan titik data yang paling dekat di setiap kelas. Jarak ini dihitung dengan rumus $\frac{2}{||\omega||}$ di mana $||\omega||$ menyatakan norma *Euclidean* dari vektor bobot tersebut. Untuk kasus di mana data tidak dapat dipisahkan secara linear, SVM dapat menerapkan fungsi kernel agar pemisahan antar kelas tetap memungkinkan dalam ruang berdimensi lebih tinggi [15].

b. *Random Forest (RF)*

Random Forest dalah algoritma yang bekerja dengan mengombinasikan sejumlah pohon keputusan *(decision trees)* guna menghasilkan prediksi yang lebih andal dan akurat [16]. Pada setiap simpul pohon, pemilihan atribut dilakukan dengan mengevaluasi tingkat ketidakmurnian menggunakan nilai *entropy* serta menghitung *information gain* untuk menentukan atribut yang paling relevan dalam pengambilan keputusan [15]. Untuk menghitung *entropy* dan nilai *information gain* menggunakan rumus:

$$Entropy\ (Y) = -\sum_i \mathrm{p}(c|Y)\ log_2\ \mathrm{p}(c|Y) \quad (2)$$

Di mana Y adalah himpunan data kasus, $p(c|Y)$ menggambarkan *persentase* munculnya Y yang tergolong sebagai kelas c.

$$Information\ Gain\ (Y, a) = Entropy\ (Y) - \sum_{ve}\ values\ \frac{Y_v}{Y_a}\ Entropy\ (Y_v) \quad (3)$$

Values(a) merujuk pada seluruh nilai yang mungkin ada dalam kumpulan contoh *a*. *Yv* adalah subset dari *Y* yang sesuai berdasarkan nilai *v* yang berhubungan pada atribut *a* [17].

2.6 Evaluasi Model

Evaluasi terhadap model dilakukan menggunakan *confusion matrix*, yang menyajikan hasil klasifikasi dalam bentuk tabel matriks. Terdapat empat komponen utama yang menunjukkan hasil klasifikasi dalam *confusion matrix*, yaitu *True Positive (TP), True Negative (TN), False Positive (FP),* dan *False Negative (FN)* [18]. *True Negative (TN)* menggambarkan jumlah kasus negatif yang berhasil diklasifikasikan dengan benar, sementara *False Positive (FP)* menunjukkan jumlah data negatif yang keliru teridentifikasi sebagai positif. *True Positive (TP)* merupakan jumlah data positif yang berhasil dikenali secara tepat, sedangkan *False Negative (FN)* mencerminkan kasus di mana data positif salah diklasifikasikan sebagai negatif [19]. Untuk menilai kinerja suatu model, terdapat empat metode yang dapat digunakan [20].

a. *Accuracy*:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \times 100\% \quad (4)$$

b. *Precision*:

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (5)$$

c. *Recall*:

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (6)$$

d. *F1-Score*:

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 4, Agustus 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i4.8744
Hal 462-471
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

$$F1 - score = 2 \times \frac{Presisi \times Recall}{Presisi + Recall} \quad (7)$$

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pengumpulan Data

Studi ini memanfaatkan hasil pemeriksaan hipertensi dari Puskesmas Anggadita, yang terdiri dari 2.500 data pasien. Data diklasifikasikan ke dalam dua kelompok: "Hipertensi" untuk pasien yang terdiagnosis, dan "Normal" untuk yang tidak terdiagnosis. Dataset ini berisi data pasien yang mencakup nama, tanggal lahir, usia, jenis kelamin, riwayat penyakit baik dari keluarga/diri sendiri, kebiasaan merokok, tingkat aktivitas fisik yang rendah, pola makan, konsumsi alkohol, tekanan darah (sistolik dan diastolik), serta indeks massa tubuh (IMT) dan diagnosis.
Untuk memberikan gambaran umum mengenai isi data yang digunakan dalam penelitian ini, berikut disajikan data hipertensi dari sebagian pasien dalam Tabel 2.

Tabel 2. Hasil yang Diperoleh dari Pengumpulan Data

No.	Nama Pasien	Tanggal Lahir	Usia	Jenis Kelamin	Riwayat Penyakit Pada Keluarga/Diri Sendiri	…	Sistol	Diastol	Tinggi Badan (Cm)	Berat Badan (Kg)	Diagnosis
1.	Kaspi	25-10-1970	54	Laki-Laki	Penyakit Diabetes	…	137	85	155	67	Normal
2.	Amah Maryamah	23-12-1975	49	Perempuan	Penyakit Diabetes	…	129	88	155	54	Normal
3.	Dadang	12-04-1968	57	Laki-Laki	Penyakit Diabetes	…	118	73	158	40	Normal
4.	Emis	07-01-1954	71	Perempuan	Penyakit Hipertensi	…	142	95	155	48	Hipertensi
…	…		…	…	…	…	…	…	…	…	…
2500.	Heny Sukmawati	07-05-1977	47	Perempuan	Penyakit Hiperetnsi	…	184	106	158	70	Hipertensi

3.2 Hasil *Preprocessing* Data

Dalam prosedur ini, dilakukan serangkaian *preprocessing* data untuk mempersiapkan dataset agar siap untuk analisis dan pemodelan. Prosedur ini terdiri dari beberapa tahap utama, seperti penanganan nilai kosong *(missing value)*, menghapus data duplikat, *transformasi data*, dan pembagian dataset. Tahap awal dalam proses ini dimulai dengan mengatasi nilai kosong *(missing value)* pada data.

Untuk melihat atribut apa saja yang mengandung nilai kosong dalam dataset, sebelum dilakukan penanganan ditampilkan informasi pada Tabel 3 berikut:

Tabel 3. Mengecek *Missing Value*

Nama Pasien	0
Tanggal Lahir	24
…	…
Riwayat Penyakit Pada Keluarga/Diri Sendiri	1663
Merokok	0
Kurang Aktivitas Fisik	0
Gula Berlebihan	0
Garam Berlebihan	0
Lemak Berlebihan	0
Kurang Makan Buah dan Sayur	0
Konsumsi Alkohol	0
Sistol	0
Diastol	0
Tinggi Badan (Cm)	0
Berat Badan (Kg)	0
Diagnosis	0

Berdasarkan Tabel 3, ditemukan beberapa atribut dalam dataset yang memiliki nilai kosong *(missing value)*. Atribut dengan jumlah nilai kosong tertinggi adalah riwayat penyakit pada keluarga/diri sendiri dengan 1.663 entri yang kosong, kemudian pada tanggal lahir sebanyak 24 entri.

Setelah dilakukan proses pengecekan terhadap atribut yang memiliki nilai kosong seperti dijelaskan pada tahap sebelumnya, langkah selanjutnya yaitu mengatasi *missing value* yang ditampilkan pada Tabel 4.

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 4, Agustus 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i4.8744
Hal 462-471
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Tabel 4. Mengatasi *Missing Value*

Nama Pasien	0
Tanggal Lahir	0
…	…
Riwayat Penyakit Pada Keluarga/Diri Sendiri	0
Merokok	0
Kurang Aktivitas Fisik	0
Gula Berlebihan	0
Garam Berlebihan	0
Lemak Berlebihan	0
Kurang Makan Buah dan Sayur	0
Konsumsi Alkohol	0
Sistol	0
Diastol	0
Tinggi Badan (Cm)	0
Berat Badan (Kg)	0
Diagnosis	0

Berdasarkan dari Tabel 4, untuk mengatasi nilai yang kosong pada kolom kategorikal seperti riwayat penyakit pada keluarga/ diri sendiri diisi dengan modus (nilang yang sering muncul) atau fungsi *mode()*. Kemudian pada kolom bertipe datetime seperti tanggal lahir diisi dengan perkiraan tanggal lahir yang dihitung dari usia menggunakan fungsi *today - (usia * 365 hari)*. Hasil akhir dari proses ini menunjukkan bahwa dataset sudah bersih dan tidak terdapat lagi nilai kosong.

Tahap berikutnya dalam proses *preprocessing* adalah mengidentifikasi dan menghapus data duplikat. Proses ini dilakukan dengan memeriksa seluruh entri pada dataset untuk mendeteksi baris-baris yang tercatat lebih dari satu kali. Jumlah data duplikat yang terdeteksi ditunjukkan pada Gambar 2.

```
Jumlah data duplikat: 38
```

Gambar 2. Mengecek Data Duplikat

Berdasarkan Gambar 2, dapat disimpulkan masih terdapat 38 data duplikat pada dataset. Untuk menjaga keunikan data dan meningkatkan keakuratan hasil analisis maupun pemodelan, data yang sama atau berulang pada baris tersebut dapat dihapus.

Setelah proses identifikasi dan penghapusan data duplikat dilakukan, jumlah baris dalam dataset mengalami pengurangan yaitu berjumlah 2462 baris setelah 38 data duplikat dihapus. Hal ini bertujuan untuk memastikan bahwa data yang digunakan dalam proses analisis bebas dari pengulangan entri yang dapat memengaruhi akurasi model. Data yang telah dibersihkan dari duplikat ditunjukkan pada Tabel 5.

Tabel 5. Setelah Hapus Data Duplikat

No.	Nama Pasien	Tanggal Lahir	Usia	Jenis Kelamin	Riwayat Penyakit Pada Keluarga/Diri Sendiri	…	Sistol	Diastol	Tinggi Badan (Cm)	Berat Badan (Kg)	Diagnosis
1.	Kaspi	25-10-1970	54	Laki-Laki	Penyakit Diabetes	…	137	85	155	67	Normal
2.	Amah Maryamah	23-12-1975	49	Perempuan	Penyakit Diabetes	…	129	88	155	54	Normal
3.	Dadang	12-04-1968	57	Laki-Laki	Penyakit Diabetes	…	118	73	158	40	Normal
4.	Emis	07-01-1954	71	Perempuan	Penyakit Hipertensi	…	142	95	155	48	Hipertensi
…	…		…	…	…	…	…	…	…	…	…
2462.	Heny Sukmawati	07-05-1977	47	Perempuan	Penyakit Hiperetnsi	…	184	106	158	70	Hipertensi

Pada tahap selanjutnya adalah *transformasi* data. Dalam tahap *transformasi* data, dilakukan proses konversi nilai kategorikal ke dalam bentuk numerik menggunakan metode *label encoding*. *Transformasi* ini bertujuan untuk memudahkan algoritma dalam membaca dan memproses data non-numerik. Tabel 6 berikut menyajikan data sebelum dilakukan proses *label encoding*.

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 4, Agustus 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i4.8744
Hal 462-471
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Tabel 6. Sebelum *Label Encoding*

No.	Nama Pasien	Usia	Jenis Kelamin	…	Merokok	Kurang Aktivitas Fisik	Gula Berlebihan	Garam Berlebihan	Lemak Berlebihan	Kurang Makan Buah dan Sayur	Konsumsi Alkohol	Diagnosis
1.	Kaspi	54	Laki-Laki	…	Tidak	Ya	Tidak	Ya	Tidak	Ya	Tidak	Normal
2.	Amah Maryamah	49	Perempuan	…	Tidak	Ya	Tidak	Tidak	Ya	Tidak	Tidak	Normal
3.	Dadang	57	Laki-Laki	…	Tidak	Ya	Ya	Tidak	Tidak	Tidak	Tidak	Normal
4.	Emis	71	Perempuan	…	Tidak	Ya	Ya	Tidak	Tidak	Tidak	Tidak	Hipertensi
…	…	…	…	…	…	...	...	…	…	…	…	…
2462.	Heny Sukmawati	47	Perempuan	…	Tidak	Ya	Ya	Tidak	Tidak	Tidak	Tidak	Hipertensi

Berdasarkan Tabel 6, dapat dilihat bahwa atribut seperti jenis kelamin, merokok, aktivitas fisik, gula berlebihan, garam berlebeihan, lemak berlebihan, kurang makan buah dan sayur, konsumsi alkohol serta diagnosis masih menggunakan nilai kategorikal seperti "Laki-Laki", "Perempuan", "Ya", dan "Tidak". Nilai-nilai tersebut nantinya akan dikonversi ke dalam bentuk numerik agar dapat digunakan dalam proses pelatihan model klasifikasi.

Setelah dilakukan proses *transformasi* dengan metode *label encoding*, nilai-nilai kategorikal pada data dikonversi ke dalam bentuk numerik. Tabel 7 menyssajikan hasil dari proses tersebut, di mana setiap nilai kategorikal telah digantikan dengan representasi numerik yang sesuai.

Tabel 7. Setelah *Label Encoding*

No.	Nama Pasien	Usia	Jenis Kelamin	…	Merokok	Kurang Aktivitas Fisik	Gula Berlebihan	Garam Berlebihan	Lemak Berlebihan	Kurang Makan Buah dan Sayur	Konsumsi Alkohol	Diagnosis
1.	Kaspi	54	0	…	0	1	0	1	0	1	0	0
2.	Amah Maryamah	49	1	…	0	1	0	0	1	0	0	0
3.	Dadang	57	0	…	0	1	1	0	0	0	0	0
4.	Emis	71	1	…	0	1	1	0	0	0	0	1
…	…	…	…	…	…	...	...	…	…	…	…	…
2462.	Heny Sukmawati	47	1	…	0	1	1	0	0	0	0	1

Berdasarkan Tabel 7, dapat dilihat bahwa seluruh atribut kategorikal seperti jenis kelamin, kebiasaan merokok, aktivitas fisik, gula berlebihan, garam berlebeihan, lemak berlebihan, kurang makan buah dan sayur, konsumsi alkohol dan diagnosis kini telah direpresentasikan dengan angka 0 dan 1. *Transformasi* ini bertujuan untuk menyesuaikan format data agar dapat digunakan dalam proses pelatihan algoritma klasifikasi *machine learning*.

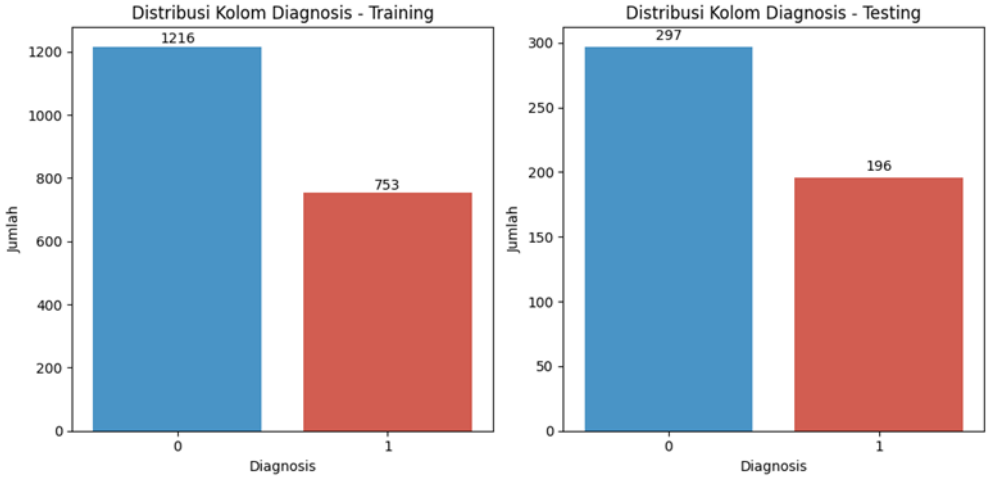
Setelah proses *transformasi* dan *label encoding* selesai dilakukan, langkah selanjutnya adalah membagi dataset menjadi dua bagian, yaitu *training* (pelatihan) dan *testing* (pengujian) dengan fungsi *train_test_split* dari pustaka *sklearn.model_selection*. Pembagian ini bertujuan untuk melatih model menggunakan sebagian data dan kemudian menguji performa model tersebut pada data yang belum pernah dilihat sebelumnya. Rasio pembagian yang digunakan dalam penelitian ini adalah 80% untuk data pelatihan dan 20% untuk data pengujian. Jumlah data hasil pembagian ditunjukkan pada Gambar 3.

```
Jumlah data training:    1969
Jumlah data testing:   493
```

Gambar 3. Jumlah *Training* dan *Testing*

Untuk mengetahui distribusi proporsi antara label klasifikasi Normal dan Hipertensi pada data *training* dan *testing*, dilakukan visualisasi data terhadap kolom Diagnosis. Visualisasi ini bertujuan untuk memastikan bahwa proses pembagian data tidak menyebabkan ketidakseimbangan yang signifikan antara kelas. Gambar 4 berikut menyajikan hasil distribusi label diagnosis pada masing-masing subset data.

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 4, Agustus 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i4.8744
Hal 462-471
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom



Gambar 4. Visualisasi Label Diagnosis Data *Training* dan *Testing*

Berdasarkan Gambar 4, pada visualisasi tersebut, label 0 merepresentasikan kondisi Normal, sedangkan label 1 merepresentasikan kondisi Hipertensi. Terlihat bahwa pada data *training* terdapat 1.216 data dengan label 0 (Normal) dan 753 data dengan label 1 (Hipertensi). Sementara itu, pada data *testing* ditemukan 297 data dengan label 0 dan 196 data dengan label 1. Berdasarkan distribusi ini, dapat disimpulkan bahwa proporsi kedua label relatif seimbang pada masing-masing subset data, yaitu sekitar 60–62% untuk kelas Normal dan 38–40% untuk kelas Hipertensi.

3.3 Implementasi Model SVM dan RF

Pada tahap implementasi, model SVM dengan kernel *Radial Basis Function (RBF)* diatur menggunakan parameter *`C=1`*, *`gamma='scale'`*, dan *`random_state=42`* untuk memastikan bahwa proses pelatihan menghasilkan output yang konsisten. Parameter *`C`* mengatur keseimbangan antara kesalahan pada data *training* dan kompleksitas model, sedangkan *`gamma='scale'`* menyesuaikan nilai gamma secara otomatis berdasarkan jumlah fitur. Selanjutnya, model dilatih dengan menggunakan data pelatihan *(`X_train`* dan *`y_train`)* guna mempelajari pola-pola yang diperlukan untuk melakukan klasifikasi. Sesudah proses *training* selesai, model digunakan untuk memprediksi kelas pada data *testing (`X_test`)* dengan metode *`predict()`*. Hasil prediksi *(y_pred_svm)* dimanfaatkan untuk menilai seberapa baik model dalam mengategorikan data baru dengan akurat.

Kemudian untuk memastikan konsistensi dalam hasil pelatihan, model *Random Forest (RF)* diinisialisasi dengan 100 pohon keputusan *('n_estimators=100')* dan *'random_state=42'*. Model dilatih pada data *training ('X_train' dan 'y_train')* sehingga dapat mempelajari pola klasifikasi dari beberapa pohon keputusan. Setelah *training*, model digunakan untuk memprediksi kelas pada data *testing ('X_test')* menggunakan metode *'predict()'*. Hasil prediksi *('y_pred_rf')* kemudian digunakan untuk mengevaluasi sejauh mana model mampu mengklasifikasikan data secara akurat dan efektif.

3.4 Hasil Evaluasi Model

Pada tahap evaluasi, *confusion matrix* digunakan untuk menilai tingkat akurasi model dalam mengklasifikasikan kasus hipertensi. *Confusion matrix* memberikan gambaran lengkap mengenai jumlah prediksi yang benar maupun salah pada masing-masing kelas, sehingga dapat dimanfaatkan untuk menghitung berbagai metrik evaluasi lainnya seperti akurasi, presisi, *recall*, dan *f1-score*.

Untuk mengetahui performa model klasifikasi dalam mendeteksi kondisi hipertensi, dilakukan evaluasi menggunakan metrik seperti akurasi, presisi, *recall,* dan *f1-score*. Evaluasi ini bertujuan untuk menilai sejauh mana model mampu mengklasifikasikan data dengan benar, baik untuk kelas "Normal" maupun "Hipertensi". Hasil evaluasi model *Support Vector Machine (SVM)* tersebut ditampilkan pada Tabel 8.

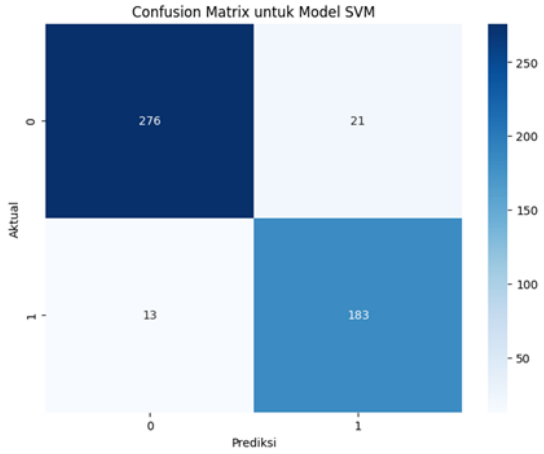
Tabel 8. Evaluasi Performa Model SVM

Akurasi	Presisi	*Recall*	*F1-Score*	Kelas
93%	0.96	0.93	0.94	0
	0.90	0.93	0.92	1

Berdasarkan hasil evaluasi pada Tabel 8, model SVM mencapai akurasi sebesar 93%, yang menunjukkan performa klasifikasi yang sangat baik. Untuk kelas "Normal", presisi mencapai 0.96, *recall* 0.93, dan *f1-score* 0.94 yang membuktikan bahwa model memiliki tingkat akurasi yang baik dalam mengidentifikasi data normal. Sementara itu, untuk kelas "Hipertensi", presisi mencapai 0.90, *recall* 0.93, dan *f1-score* 0.92 menunjukkan model sangat baik dalam mengidentifikasi data hipertensi. Secara keseluruhan, model SVM menunjukkan kinerja yang seimbang dan efektif dalam membedakan kedua kelas.

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 4, Agustus 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i4.8744
Hal 462-471
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Untuk memperjelas performa klasifikasi model SVM secara visual, dilakukan visualisasi *confusion matrix*. Visualisasi ini membantu menunjukkan jumlah prediksi benar dan salah dari masing-masing kelas (Normal dan Hipertensi), sehingga dapat digunakan untuk mengevaluasi akurasi model secara lebih rinci. Gambar 5 berikut ini menggambarkan hasil evaluasi model berdasarkan *confusion matrix*.



Gambar 5. *Confusion Matrix* Model SVM

Pada Gambar 5, hasil dari *confusion matrix* model SVM menggambarkan bahwa label 0 merujuk pada kategori "Normal," sementara label 1 menunjukkan "Hipertensi". Model berhasil mengklasifikasikan 183 data penderita hipertensi secara akurat, namun terdapat 13 data yang salah prediksi sebagai "Normal". Sementara itu, dari kategori "Normal," sebanyak 276 data telah diklasifikasikan dengan tepat, dan 21 data lainnya salah diprediksi sebagai "Hipertensi". Hasil ini mengindikasikan bahwa model SVM berfungsi dengan baik untuk membedakan antara individu yang memiliki hipertensi dan yang dalam kondisi normal, dengan tingkat akurasi prediksi yang lebih tinggi daripada kesalahan dalam prediksinya. Adapun cara menghitung akurasi klasifikasi penyakit hipertensi pada *confuxion matrix* model svm sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \times 100\%$$
$$= \frac{183+276}{183+276+21+13} \times 100\%$$
$$= \frac{459}{493} \times 100\%$$
$$= 0{,}93 \times 100\%$$
$$= 93\%$$

Setelah mengevaluasi model SVM, langkah berikutnya adalah menilai performa model *Random Forest (RF)* menggunakan metrik evaluasi yang sama, yaitu akurasi, presisi, *recall*, dan *f1-score*. Evaluasi ini dilakukan untuk mengetahui sejauh mana model RF mampu membedakan antara kelas Normal dan Hipertensi dengan tepat. Hasil evaluasi performa model RF disajikan pada Tabel 9.

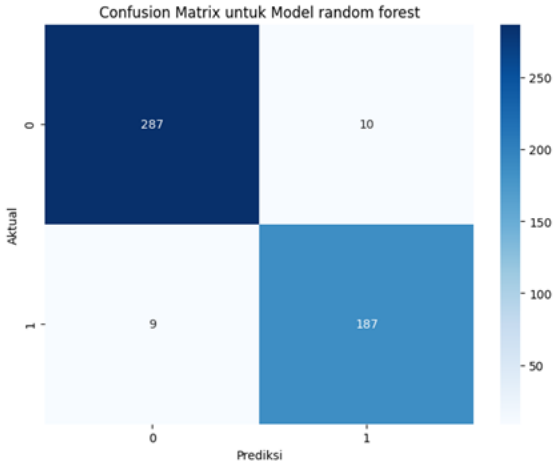
Tabel 9. Evaluasi Performa Model RF

Akurasi	Presisi	*Recall*	*F1-Score*	Kelas
96%	0.97	0.97	0.97	0
	0.95	0.95	0.95	1

Merujuk pada hasil evaluasi yang disajikan dalam Tabel 9, model *Random Forest (RF)* mencapai akurasi 96%, yang menunjukkan kinerja klasifikasi yang sangat baik. Model tersebut memiliki nilai presisi 0,97, *recall* 0,97, dan *f1-score* 0,97 untuk kelas "Normal", yang menunjukkan bahwa model tersebut dapat secara konsisten mengklasifikasikan data normal. Sementara itu, kelas "Hipertensi" memiliki nilai presisi 0,95, *recall* 0,95, dan *f1-score* 0,95, yang menampilkan bahwa model tersebut juga cukup dapat dipercaya untuk mengidentifikasi data hipertensi. Secara keseluruhan, model *Random Forest* berkinerja baik dan konsisten dalam membedakan antara kedua kelas tersebut.

Untuk memperjelas hasil evaluasi model *Random Forest (RF),* dilakukan visualisasi dalam bentuk *confusion matrix* yang menggambarkan perbandingan antara nilai aktual dan nilai prediksi pada masing-masing kelas (Normal dan Hipertensi). Hasil evaluasi model Random Forest dalam bentuk *confusion matrix* dapat dilihat pada Gambar 6.

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 4, Agustus 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i4.8744
Hal 462-471
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom



Gambar 6. *Confusion Matrix* Model RF

Merujuk pada hasil *confusion matrix* model *Random Forest (RF)* yang ditampilkan pada Gambar 6, diketahui bahwa label 0 mewakili kategori "Normal", sedangkan label 1 menunjukkan "Hipertensi." Model berhasil mengklasifikasikan 187 data penderita hipertensi secara akurat, meskipun terdapat 9 data yang salah diprediksi sebagai "Normal." Sementara itu, dari kategori "Normal," sebanyak 287 data berhasil diklasifikasikan dengan benar, dan 10 data lainnya keliru diprediksi sebagai "Hipertensi". Temuan ini memperlihatkan bahwa model RF memiliki kinerja klasifikasi yang sangat baik dalam membedakan individu yang mengalami hipertensi dan yang berada dalam kondisi normal, dengan jumlah prediksi yang akurat jauh lebih tinggi dibandingkan jumlah kesalahan. Adapun cara menghitung akurasi klasifikasi penyakit hipertensi pada *confusion matrix* model RF adalah sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \times 100\%$$
$$= \frac{187+287}{187+287+9+10} \times 100\%$$
$$= \frac{474}{493} \times 100\%$$
$$= 0{,}96 \times 100\%$$
$$= 96\%$$

4. KESIMPULAN

Penelitian ini menyimpulkan bahwa baik model *Support Vector Machine (SVM)* maupun *Random Forest (RF)* mampu mengklasifikasikan pasien hipertensi dengan baik. Akan tetapi, RF sementara SVM hanya mencapai 93%. Selain itu, dalam aspek presisi, *recall*, dan *f1-score* untuk kategori "Normal" dan "Hipertensi", RF memberikan hasil yang lebih optimal dengan tingkat akurasi mencapai 96%. Temuan ini mengindikasikan bahwa *Random Forest* lebih efisien dalam mendeteksi individu yang berisiko mengalami hipertensi berdasarkan informasi kesehatan yang ada. Diharapkan penerapan model klasifikasi ini dapat memberikan dukungan kepada tenaga medis dalam melakukan diagnosis lebih awal dan manajemen pasien yang lebih tepat serta terencana di fasilitas kesehatan seperti puskesmas.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada Puskesmas Anggadita atas izin serta dukungan yang telah diberikan selama proses pengumpulan data dalam penelitian ini. Ucapan terima kasih juga disampaikan kepada dosen pembimbing dan seluruh sivitas akademika Fakultas Ilmu Komputer Universitas Buana Perjuangan Karawang atas bimbingan, masukan, serta motivasi yang sangat berharga sepanjang pelaksanaan penelitian. Penghargaan yang tulus penulis berikan kepada orang tua dan keluarga tercinta atas doa, dukungan moral, dan semangat yang tak pernah surut. Tak lupa, penulis menyampaikan terima kasih kepada rekan-rekan atas kerja sama, kontribusi, dan dedikasi luar biasa yang turut berperan dalam penyelesaian penelitian ini.

REFERENCES

[1] S. A. Fani, S. Noviana, G. S. Tunabenany, and S. Aryanti, "Edukasi Pemeriksaan Molekuler pada Penyakit Hipertensi : Inovasi Watermelon Lemonade sebagai Nutrisi Preventif," vol. 1, no. 3, pp. 202–211, 2024.

[2] R. P. Pridiptama, W. Wasono, and F. D. . Amijaya, "Perbandingan Algoritma Support Vector Machine dan Na¨ıveNa¨ıve Bayes pada Klasifikasi Penyakit Tekanan Darah Tinggi (Studi Kasus: Klinik Polresta Samarinda)," *Basis*, vol. 3, no. 1, pp. 1–16, 2024.

[3] 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.

[4] R. Muhardina, J. Rekayasa, and S. Komputer, "Prediksi Risiko Penyakit Hipertensi Menggunakan Metode Extreme Learning Machine (ELM) [1]," vol. 12, no. 02, 2024.
[5] F. O. Awalullaili, D. Ispriyanti, and T. Widiharih, "Klasifikasi Penyakit Hipertensi Menggunakan Metode Svm Grid Search Dan Svm Genetic Algorithm (Ga)," *J. Gaussian*, vol. 11, no. 4, pp. 488–498, 2023.
[6] W. Apriliah, I. Kurniawan, M. Baydhowi, and T. Haryati, "Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest," *Sistemasi*, vol. 10, no. 1, p. 163, 2021.
[7] A. Naïve, B. Classifier, and P. Penyakit, "Analisis klasifikasi menggunakan regresi logistik biner dan algoritma naïve bayes classifier pada penyakit hipertensi 1,2,3," vol. 13, no. 2007, pp. 319–327, 2024.
[8] 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.
[9] R. S. Tantika and A. Kudus, "Penggunaan Metode Support Vector Machine Klasifikasi Multiclass pada Data Pasien Penyakit Tiroid," *Bandung Conf. Ser. Stat.*, vol. 2, no. 2, pp. 159–166, 2022.
[10] A. Ghozali, H. Pratiwi, and S. S. Handajani, "Implementasi Data Mining Menggunakan Metode Random Forest Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes," *Delta J. Ilm. Pendidik. Mat.*, vol. 11, no. 2, p. 147, 2023.
[11] T. Rohana, E. Nurlaelasari, E. E. Awal, and H. Y. Novita, "Kajian Model Jaringan Syaraf Tiruan Untuk Memprediksi Secara Dini Tingkat Kelulusan Mahasiswa," *Technol. J. Ilm.*, vol. 15, no. 4, p. 629, 2024.
[12] H. Hikmayanti, A. F. Nurmasruriyah, A. Fauzi, N. Nurjanah, and A. Nur Rani, "Performance Comparison of Support Vector Machine Algorithm and Logistic Regression Algorithm," *Int. J. Artif. Intelegence Res.*, vol. 7, no. 1, p. 1, 2023.
[13] A. M. Siregar, "Analisis Sentimen Pindah Ibu Kota Negara (IKN) Baru pada Twitter Menggunakan Algoritma Naive Bayes dan Support Vector Machine (SVM)," *Fakt. Exacta*, vol. 16, no. 3, pp. 170–181, 2023.
[14] K. A. Baihaqi, E. Sediyono, C. Dewi, I. R. Widiasari, and A. Fauzi, "Classification of Mobile Application User Ratings Based on Data from Google Play Store," *E3S Web Conf.*, vol. 500, pp. 1–8, 2024.
[15] N. F. Sahamony, T. Terttiaavini, and H. Rianto, "Analisis Perbandingan Kinerja Model Machine Learning untuk Memprediksi Risiko Stunting pada Pertumbuhan Anak," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 413–422, 2024.
[16] D. Sudrajat, A. I. Purnamasari, A. R. Dikananda, D. A. Kurnia, and A. Bahtiar, "Klasifikasi Mutu Pembelajaran Hybrid berdasarkan Algoritma C.45, Random Forest dan Naïve Bayes dengan Optimasi Bootsrap Areggating (Bagging) pada masa COVID-19," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 6, p. 2227, 2022.
[17] H. Nalatissifa, W. Gata, S. Diantika, and K. Nisa, "Perbandingan Kinerja Algoritma Klasifikasi Naive Bayes, Support Vector Machine (SVM), dan Random Forest untuk Prediksi Ketidakhadiran di Tempat Kerja," *J. Inform. Univ. Pamulang*, vol. 5, no. 4, p. 578, 2021.
[18] M. M. S. Jogo, M. K. Biddinika, and A. Fadlil, "Klasifikasi Penyakit Diabetes dengan Algoritma Decision Tree dan Naïve Bayes," *Resist. (Elektronika Kendali Telekomun. Tenaga List. Komputer) Vol.*, vol. 6, no. 2, pp. 113–118, 2023.
[19] B. W. Kurniadi, H. Prasetyo, G. L. Ahmad, B. Aditya Wibisono, and D. Sandya Prasvita, "Analisis Perbandingan Algoritma SVM dan CNN untuk Klasifikasi Buah," *Semin. Nas. Mhs. Ilmu Komput. dan Apl. Jakarta-Indonesia*, no. September, pp. 1–11, 2021.
[20] Y. N. Paramitha, A. Nuryaman, A. Faisol, E. Setiawan, and D. E. Nurvazly, "Klasifikasi Penyakit Stroke Menggunakan Metode Naïve Bayes," *J. Siger Mat.*, vol. 04, no. 01, pp. 11–16, 2023.