PaKMas (Jurnal Pengabdian Kepada Masyarakat)
https://journal.yp3a.org/index.php/PaKMas
ISSN Media Elektronik 2808-0920
Vol. 6 No. 1 (Mei 2026) 96-104
DOI: 10.54259/pakmas.v6i1.4497

Peningkatan Prediksi Kelainan Tekanan Darah dengan *Logistic Regression* dan *Random Forest*: Pendekatan *Sequence Machine Learning*

Florentina Yuni Arini[1*], Rahmat Hidayat[2], Arzaki Zunior Putra[3], Muhammad Nur Furqon[4], Muhammad Zuniar Hilmi[5]
[1*,2,3,4,5]Program Studi Teknik Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Semarang, Kota Semarang, Indonesia
Email: [1*]floyuna@mail.unnes.ac.id, [2]rhmtsekolah@students.unnes.ac.id, [3]zeeks@students.unnes.ac.id, [4]muhammadnurfurqon3@students.unnes.ac.id, [5]zuniarhilmi10@students.unnes.ac.id

Abstract

Early detection of blood pressure abnormalities plays a critical role in preventing and managing cardiovascular diseases, which remain the leading cause of death globally. This study proposes a sequence machine learning approach that combines Random Forest (RF) and Logistic Regression (LR) to enhance the accuracy of abnormal blood pressure prediction. The dataset, obtained from Kaggle, includes various clinical and lifestyle-related features. Data preprocessing involved handling missing values, label encoding, and normalization of numerical features. Evaluation of individual models showed that Random Forest achieved an accuracy of 0.83, while Logistic Regression reached 0.75. The sequence model, which incorporates Random Forest-generated prediction probabilities as an additional feature in Logistic Regression, improved the prediction performance with an accuracy of 0.84. Feature importance analysis identified hemoglobin level, chronic kidney disease, and genetic pedigree coefficient as the most influential predictors in classifying abnormal blood pressure. These findings highlight the effectiveness of the sequence approach in addressing the complexity of medical data and improving the precision of clinical decision support systems for hypertension diagnosis and management. Recommendations include developing advanced ensemble models, collecting longitudinal data, and conducting external validation to enhance model generalizability across diverse clinical populations.

Keywords: *Abnormal Blood Pressure, Random Forest, Logistic Regression, Sequence Model, Medical Classification, Hypertension.*

Abstrak

Deteksi dini abnormalitas tekanan darah termasuk aspek krusial dalam upaya pencegahan dan penanganan penyakit kardiovaskular yang menjadi penyebab utama kematian global. Penelitian ini mengusulkan pendekatan *sequence machine learning* yang mengintegrasikan *Random Forest* (RF) dan *Logistic Regression* (LR) untuk meningkatkan akurasi prediksi tekanan darah abnormal. Dataset yang digunakan diperoleh dari Kaggle, mencakup berbagai parameter klinis dan gaya hidup. Proses *preprocessing* mencakup penanganan *missing values*, *label encoding*, serta normalisasi fitur numerik. Evaluasi model individual menunjukkan bahwa *Random Forest* memiliki akurasi sebesar 0,83, sedangkan *Logistic Regression* mencapai akurasi 0,75. Model *sequence* yang memanfaatkan probabilitas prediksi dari *Random Forest* sebagai fitur tambahan pada *Logistic Regression* menghasilkan akurasi 0,84, mengindikasikan peningkatan performa prediksi. Analisis *feature importance* menunjukkan bahwa kadar hemoglobin, penyakit ginjal kronis, dan koefisien keturunan genetik termasuk prediktor dominan dalam klasifikasi tekanan darah abnormal. Temuan ini menegaskan efektivitas pendekatan *sequence* dalam menangani kompleksitas data medis dan meningkatkan akurasi sistem pendukung keputusan klinis untuk diagnosis dan manajemen hipertensi. Rekomendasi diarahkan pada pengembangan model lanjutan, pengumpulan data longitudinal, serta validasi eksternal untuk meningkatkan generalisasi model dalam berbagai konteks populasi klinis.

Kata Kunci: Tekanan Darah Abnormal, *Random Forest*, *Logistic Regression*, Model *Sequence*, Klasifikasi Medis, Hipertensi.

Diterima Redaksi: 27-05-2025 | Selesai Revisi: 01-11-2025 | Diterbitkan Online: 15-05-2026

A. PENDAHULUAN

Deteksi dini abnormalitas tekanan darah sangat penting karena hipertensi termasuk faktor risiko utama namun dapat dikendalikan dalam pencegahan dan penanganan penyakit kardiovaskular, yang penyebab kematian utama secara global, dengan jumlah kematian mencapai sekitar 10,5 juta orang setiap tahun (Kuneinen dkk., 2024). Dengan melakukan deteksi dini dapat menghindari komplikasi lanjutan pada gagal jantung, memperbaiki prognosis pasien, serta mengatasi tantangan dalam proses diagnosis (Nugraha, Wahyu & Muhamad Syarif, 2024). Berbagai pendekatan dengan menggunakan *machine learning* telah diterapkan untuk meningkatkan hasil prediksi kondisi tekanan darah, namun masih terdapat kekurangan yang dapat disempurnakan, terutama dalam mengintegrasikan kekuatan beberapa algoritma. Berbagai jenis algoritma klasifikasi sudah diterapkan pada berbagai studi sebelumnya, tetapi masih belum mendapatkan hasil yang disetujui bersama mengenai algoritma terbaik (Nugraha, dkk, 2024). Sebagian besar penelitian yang ada juga terbatas pada data yang lebih kecil atau tidak mencakup penggabungan model untuk meningkatkan akurasi secara keseluruhan (Aziz, dkk, 2025). Penelitian ini mengusulkan bahwa pendekatan *sequence* yang menggabungkan *Random Forest* (RF) dan *Logistic Regression* (LR), dengan ini memanfaatkan probabilitas prediksi dari *Random Forest* sebagai fitur tambahan dalam model *Logistic Regression*, sehingga menghasilkan model prediksi yang lebih komprehensif.

Studi-studi sebelumnya telah membuktikan bahwa algoritma *Random Forest* mampu menangani dengan baik dataset medis yang kompleks karena sifatnya yang *robust* terhadap *noise* dan *overfitting*, berkat penggabungan banyak *Decision Tree* yang meningkatkan akurasi, kestabilan prediksi, serta mengurangi risiko *overfitting* (Bimo, 2024). Sementara itu, *Logistic Regression* menawarkan interpretabilitas yang lebih baik dan kemampuan untuk mengestimasi probabilitas yang terkalibrasi dengan baik. Terdapat berbagai kekurangan dari metode Logistic Regression meliputi ketidakmampuannya untuk memprediksi hasil yang berkelanjutan dan membutuhkan sampel yang berukuran besar, agar hasil yang didapatkan stabil (Al Azhima, Silmi Ath Thahirah, dkk., 2022). Algoritma *Random Forest* menggunakan sejumlah pohon keputusan untuk meningkatkan ketepatan dan stabilitas prediksi (Lukman Rangga Aditya Tarigan, 2024). Dalam penelitian ini, kami menggabungkan kelebihan kedua metode tersebut, dimana *output* probabilistik dari model *Random Forest* digunakan sebagai fitur tambahan pada model *Logistic Regression*. Pendekatan integrasi ini menghasilkan model *sequence* yang menunjukkan performa prediksi lebih baik dibandingkan dengan penggunaan model individual mendemonstrasikan keefektifan metode *sequence* dalam konteks prediksi abnormalitas tekanan darah. Metode Logistic Regression dipilih berdasarkan hasil sejumlah penelitian yang menyimpulkan bahwa metode ini memberikan tingkat akurasi tertinggi (Al Azhima, Silmi Ath Thahirah, dkk., 2022).

Artikel ini menyajikan metodologi lengkap dari pendekatan *sequence* yang diusulkan, mulai dari *preprocessing* data, pemodelan individual, hingga integrasi model. Algoritma gabungan memiliki tingkat tingkat kesalahan yang lebih kecil serta tingkat akurasi yang lebih tinggi dibandingkan algoritma individual (Simamora, 2024). Analisis komprehensif juga dilakukan terhadap fitur-fitur yang berpengaruh signifikan dalam prediksi abnormalitas tekanan darah berdasarkan perhitungan *feature importance* dari model *Random Forest*. Kekurangan dari metode Random Forest meliputi proses pembelajaran yang lambat, risiko overfitting, kebutuhan komputasi yang cenderung tinggi, serta ukuran model yang dihasilkan cenderung besar (Al Azhima, Silmi Ath Thahirah, dkk., 2022). Diharapkan bahwa hasil temuan ini mampu memberikan peran serta yang penting dalam merancang sistem pendukung keputusan klinis yang lebih tepat dan akurat untuk diagnosis dan manajemen hipertensi, serta dapat diadaptasi untuk aplikasi prediksi penyakit lainnya dalam bidang kesehatan.

B. PELAKSANAAN METODE

Dataset

Dataset berupa himpunan data yang terorganisir dan dapat dianalisis untuk memperoleh jawaban atas pertanyaan, melakukan prediksi, serta membangun suatu model. Model Machine Learning sangat bergantung pada data yang digunakan untuk melatihnya (Andhika, 2025). Dataset yang kami teliti didapatkan dari Kaggle dengan judul *Blood Pressure Data for disease Prediction* (Bodanki, 2021). Data ini berisi kumpulan dari laporan medis peserta secara acak pada tahun 2019 yang berisi *Patient Number, Blood Pressure Abnormality, Level of Hemoglobin, Genetic Pedigree Coefficient, Age, BMI, Sex, Pregnancy, Smoking, Physical activity, salt content in the diet, alcohol consumption per day, Level of Stress, Chronic kidney disease*, dan *Adrenal and thyroid disorders*.

Machine Learning

Machine Learning termasuk sistem yang dirancang untuk melakukan prediksi berdasarkan pengalaman atau data historis (Pranandito & Hendry, 2023). Metode ini memanfaatkan data pelatihan untuk

membangun model yang dapat menghasilkan informasi baru sebagai masukan, yang kemudian digunakan dalam memprediksi penyakit jantung (Pranandito & Hendry, 2023). Machine learning sering digunakan untuk menyelesaikan kasus seperti klasifikasi dan *clustering*, terutama ketika menangani data dalam skala besar atau *big* data (Tarimana, Fajar, dkk, 2024). Dalam penelitian ini, *Machine Learning* diterapkan untuk memprediksi risiko hipertensi menggunakan indikator klinis seperti tekanan darah, Indeks Massa Tubuh (*IMT*), dan faktor risiko lainnya.

Metode *Logistic Regression*

Dalam ranah *machine learning*, *Logistic Regression* dikenal sebagai sebuah algoritma yang kerap diaplikasikan untuk menyelesaikan berbagai permasalahan klasifikasi (Tarimana, Fajar, dkk, 2024). Berdasarkan satu atau lebih variabel prediktor yang dapat berbentuk kategorik maupun kontinu, Logistic Regression digunakan sebagai teknik statistik untuk memodelkan variabel respons dalam skala nominal atau ordinal (Habibi, Hibatullah, dkk, 2023).. Dalam penelitian ini, *Logistic Regression* diterapkan untuk memprediksi abnormalitas tekanan darah berdasarkan berbagai fitur kesehatan dalam dataset. Tidak seperti beberapa teknik lain, *Logistic Regression* tidak membutuhkan asumsi distribusi normal multivariat atau kesamaan matriks kovarian, sehingga fleksibel untuk diaplikasikan pada berbagai jenis skala data (Tarimana, Fajar, dkk, 2024). Dengan mempertimbangkan sejumlah atribut klinis pasien, seperti usia, jenis kelamin, tekanan darah, kadar kolesterol, dan faktor risiko lainnya, metode regresi logistik dapat menghasilkan prediksi probabilitas terhadap kondisi kesehatan pasien yang menjadi fokus analisis (Saputro, Ajie, dkk, 2023).

Metode regresi logistik termasuk salah satu pendekatan analisis kuantitatif yang dapat diaplikasikan untuk mengkaji hubungan antara satu atau sejumlah variabel independen dengan variabel dependen bertipe biner (dua kategori) (Ibnas, dkk, 2023). Apabila variabel dependen memiliki dua kategori, analisis ini disebut regresi logistik biner, di mana nilai 1 merepresentasikan kejadian sukses dan nilai 0 menunjukkan kejadian gagal (Ibnas, dkk, 2023).

Secara matematis, *Logistic Regression* memodelkan probabilitas keluaran seperti pada persamaan 1 (Sitanggang, dkk, 2022).

$$Z_l = \beta_0 + \beta_1 y_1 + \beta_2 y_2 + \ldots + \beta_n y_n \qquad (1)$$

Persamaan 1 menyatakan skor nilai linear yang dinyatakan pada Z_l. Nilai skor ini berupa konstanta β_0, koefisien regresi $\beta_1, \beta_2, \ldots, \beta_n$ dikalikan dengan variabel input $y_1, y_2, \ldots, y_n$. Kemudian semua hasil perkalian dari koefisien regresi dengan variabel input dijumlahkan dengan konstanta β_0 sehingga menghasilkan nilai Z_l.

Metode *Random Forest*

Random forest termasuk salah satu algoritma *Machine Learning* yang paling banyak digunakan (Sun, dkk, 2024). Metode *machine learning* yang disebut *Random Forest* mengurangi korelasi antara atribut data dengan menggabungkan banyak *Decision Tree* (Salman, dkk, 2024). Setiap pohon dalam *Random Forest* dibentuk dari subset data yang dipilih secara acak dengan penggantian (bootstrap sampling) dan subset fitur yang juga dipilih secara acak. *Random Forest* berupa kombinasi dari teknik pohon keputusan yang ada, kemudian digabungkan dan dikombinasikan pada suatu model (Adrian, dkk, 2021). Pendekatan ini membuat *Random Forest* mampu menangani data yang kompleks, tidak linear, dan memiliki banyak dimensi dengan baik. Selain itu, *Random Forest* menawarkan metrik *feature importance* yang dapat digunakan untuk menentukan faktor mana yang memiliki pengaruh terbesar pada model prediksi. Dalam penelitian ini, model *Random Forest* dikonfigurasi dengan parameter *n_estimators*=50, *max_depth*=5, dan *max_features*=1.

Secara konseptual, *Random Fores*t bekerja dengan menggabungkan prediksi dari sejumlah pohon keputusan seperti pada persamaan 2 (Sari & Suryono, 2024).

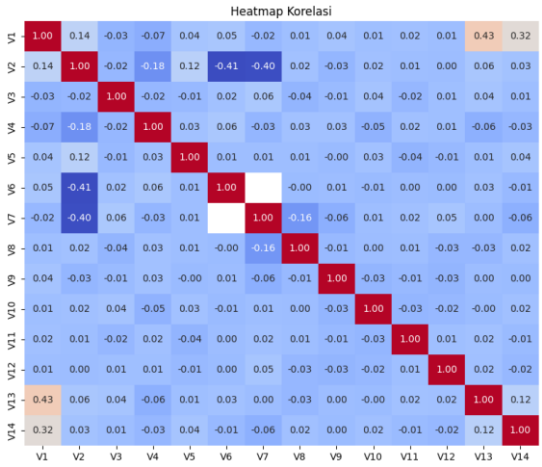
$$F(x) = \frac{1}{J} \sum_{j=1}^{J} h_j(x) \qquad (2)$$

Pada persamaan 2 tersebut, *F(x)* merepresentasikan *output* dari model *Random Forest*. Nilai $\frac{1}{J}$ berfungsi untuk menghitung rata-rata dari seluruh *output* pohon keputusan. Faktor ini memastikan bahwa setiap *output* pohon mendapatkan bobot yang sama dalam proses agregasi, sehingga hasil akhir $F(x)$ menyatakan nilai rata-rata dari seluruh prediksi pohon yang ada (Sari & Suryono, 2024). Parameter *J* menunjukkan banyaknya pohon keputusan (*decision tree*) yang membentuk *ensemble*. Sementara itu, $h_j(x)$ menyatakan *output* dari pohon keputusan ke-*j* terhadap input *x*. *Output* akhir *F(x)* diperoleh dengan menghitung rata-rata dari seluruh *output* pohon dalam *ensemble*, sehingga meningkatkan akurasi dan mengurangi risiko *overfitting* dibandingkan dengan penggunaan satu pohon tunggal.

C. HASIL DAN PEMBAHASAN

Preprocessing Data

Setelah pengumpulan data, *preprocessing* melibatkan proses pengolahan data melalui beberapa langkah, termasuk membersihkan, mengubah, dan mengintegrasikan data agar dapat digunakan oleh algoritma *machine learning* (Putra, dkk, 2024). Dalam tahap *preprocessing*, ditemukan sejumlah *missing values* pada beberapa variabel seperti *Level of Hemoglobin* (2,3%), *Physical activity* (1,5%), dan *Level of Stress* (0,8%). Beberapa proses preprocessing data dilakukan guna meningkatkan akurasi dari model prediksi (Gori, dkk, 2024). *Missing values* pada variabel numerik diimputasi menggunakan nilai median, sementara pada variabel kategorikal menggunakan nilai modus. Variabel kategorikal seperti *Sex*, *Smoking*, *Chronic kidney disease*, *Pregnancy*, *Adrenal and thyroid disorders*+ dikonversi menjadi representasi numerik menggunakan teknik *Label Encoding*. Seluruh fitur numerik kemudian dinormalisasi menggunakan *StandardScaler* untuk menyeragamkan skala data.



Gambar 1. Heatmap korelasi antar variabel

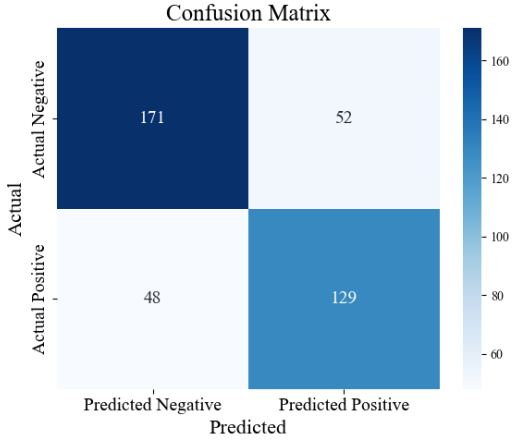
Gambar 1 menampilkan *heatmap* korelasi untuk variabel-variabel yang digunakan dalam penelitian ini. *Heatmap* ini menggambarkan hubungan linier antara pasangan variabel, dengan nilai korelasi Pearson ditunjukkan dalam bentuk matriks simetris. Korelasi positif yang tinggi ditunjukkan oleh warna merah, sedangkan korelasi negatif ditunjukkan oleh warna biru. Koefisien korelasi yang mendekati 1 menandai hubungan positif yang erat, sedangkan nilai yang mendekati -1 mengindikasikan hubungan negatif yang kuat. Sementara itu, nilai korelasi sekitar 0 mencerminkan tidak terdapat hubungan linier signifikan antara variabel.

Variabel-variabel yang digunakan pada heatmap meliputi:

- V1 = *Blood_Pressure_Abnormality*
- V2 = *Level_of_Hemoglobin*
- V3 = *Genetic_Pedigree_Coefficient*
- V4 = *Age*
- V5 = *BMI*
- V6 = *Sex*
- V7 = *Pregnancy*
- V8 = *Smoking*
- V9 = *Physical_activity*
- V10 = *salt_content_in_the_diet*
- V11 = *alcohol_consumption_per_day*
- V12 = *Level_of_Stress*
- V13 = *Chronic_kidney_disease*
- V14 = *Adrenal_and_thyroid_disorders*

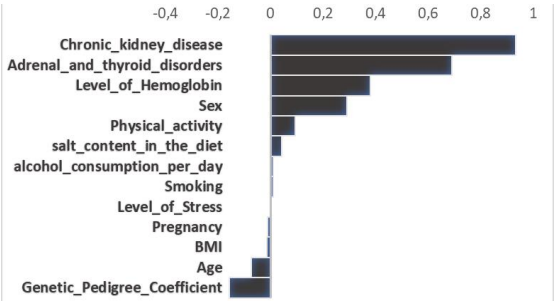
Evaluasi Model *Logistic Regression*

Untuk menilai efektivitas model *Logistic Regression* dalam memprediksi kelainan tekanan darah, dilakukan evaluasi pada data testing menggunakan *confusion matrix*. *Confusion matrix* membagi hasil prediksi ke dalam empat kategori *True Negative*, *False Positive*, *False Negative*, dan *True Positive* sehingga memungkinkan perhitungan metrik utama seperti akurasi, presisi, *recall*, dan *F1-score*.



Gambar 2. *Confusion matrix* model *Logistic Regression*

Confusion matrix model *Logistic Regression* pada **gambar 2** menunjukkan bahwa dari 223 kasus tekanan darah normal, model berhasil mengidentifikasi sebanyak 171 kasus dengan benar (*true negative*) dan 52 kasus lain teridentifikasi sebagai abnormal (*false positive*). Dari 177 kasus abnormal, model mengidentifikasi 129 kasus dengan benar (*true positive*) dan 48 kasus teridentifikasi sebagai normal (*false negative*).



Gambar 3. Pengaruh Variabel terhadap *Blood Pressure* Abnormality model *Logistic Regression*

Analisis koefisien dari model *Logistic Regression* pada gambar 3 mengungkapkan variabel-variabel yang memiliki pengaruh paling signifikan terhadap prediksi abnormalitas tekanan darah:

1. *Chronic kidney disease* (0.9325): Penyakit ginjal kronis memiliki koefisien positif tertinggi, mengindikasikan pengaruh kuat terhadap peningkatan probabilitas abnormalitas tekanan darah. Target tekanan darah sistolik kurang dari 120 mm Hg menggunakan pembacaan kantor standar bagi kebanyakan orang dengan penyakit ginjal kronis (CKD) yang tidak menjalani dialisis, pengecualiannya ialah anak-anak dan penerima transplantasi ginjal (Cheung et al., 2021). Patogenesis hipertensi pada pasien CKD bersifat kompleks dan multifaktorial, dan seringkali resisten terhadap pengobatan karena kompleksitasnya (Hamrahian, 2022).
2. *Adrenal and thyroid disorders* (0.6917): Gangguan adrenal dan tiroid menunjukkan pengaruh positif yang substansial, sesuai dengan patofisiologi gangguan endokrin terhadap regulasi tekanan darah. Tekanan darah tinggi pada hipertiroidisme terkait dengan keadaan hiperaktif dan hipermetabolisme. Orang yang terkena dampaknya gelisah dan gelisah dan mungkin mengalami takikardia, tekanan darah tinggi, penurunan berat badan, insomnia, kelelahan, diare, dan intoleransi panas (Saeed, S.M., 2023).
3. *Level of Hemoglobin* (0.3807): Kadar hemoglobin memiliki koefisien positif moderat, menegaskan peran penting parameter hematologi dalam tekanan darah.

Beberapa variabel dengan koefisien negatif mencakup *Genetic Pedigree Coefficient* (-0.1582), *Age* (-0.0741), BMI (-0.0127), *Pregnancy* (-0.0123) mengindikasikan bahwa peningkatan pada variabel-variabel ini dalam model *Logistic Regression* ini justru menurunkan probabilitas abnormalitas tekanan darah. Hal ini menarik karena kontras dengan temuan *feature importance* pada model *Random Forest*, di mana *Genetic Pedigree Coefficient* justru menjadi salah satu prediktor positif terkuat. Sedangkan variabel *Pregnancy* justru yang pada keadaan nyata seharusnya meningkatkan probabilitas abnormalitas tekanan darah, tetapi pada model ini justru didapatkan hasil yang sebaliknya. Hal ini dapat diakibatkan oleh dataset yang digunakan memiliki banyak missing value pada variabel *Pregnancy* yang mencapai 1558 data.

Model *Logistic Regression* dibangun setelah data mengalami *preprocessing* dan normalisasi menggunakan *StandardScaler*. Hasil evaluasi model *Logistic Regression* pada data testing menunjukkan performa sebagai berikut:

Tabel 1. Evaluasi Model Logistic Regression

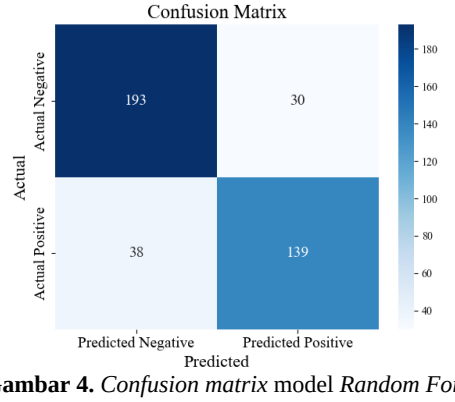
Metrix	Kelas Normal (0,0)	Kelas Abnormal(0,1)	Rata-Rata
Precision	0,78	0.71	0,75
Recall	0,77	0.73	0,75
F1-score	0,77	0.72	0,75
Accuracy	-	-	0,75

Tabel 1 menunjukkan hasil evaluasi model *Logistic Regression* setelah data melalui proses preprocessing dan normalisasi menggunakan *StandardScaler*. Evaluasi dilakukan terhadap dua kelas, yaitu Normal (0,0) dan Abnormal (0,1), menggunakan metrik precision, recall, F1-score, dan *accuracy*. Secara keseluruhan, model memperlihatkan kinerja yang cukup optimal dengan tingkat akurasi utama sebesar 0,75.

Evaluasi Model *Random Forest*

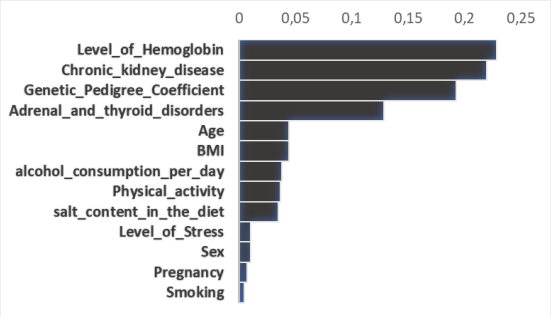
Untuk menilai efektivitas model *Random Forest* dalam memprediksi kelainan tekanan darah, dilakukan evaluasi pada data uji menggunakan *confusion matrix*. Matriks ini mengelompokkan hasil prediksi menjadi *True Negative*, *False Positive*, *False Negative*, dan *True Positive*, sehingga memungkinkan perhitungan metrik utama seperti akurasi, presisi, *recall*, dan *F1-score*.

Confusion matrix model *Random Forest* pada gambar 4 menunjukkan bahwa dari 223 kasus tekanan darah normal, model berhasil mengidentifikasi 193 kasus dengan benar (*true negative*) dan 30 kasus teridentifikasi sebagai abnormal (*false positive*).



Gambar 4. *Confusion matrix* model *Random Forest*

Sementara itu, dari 177 kasus tekanan darah abnormal, model mengidentifikasi 139 kasus dengan benar (*true positive*) dan 38 kasus teridentifikasi sebagai normal (*false negative*).



Gambar 5. Pengaruh Variabel terhadap *Blood Pressure* Abnormality model *Random Forest*

Analisis *feature importance* dari model *Random Forest* pada gambar 5 mengungkapkan bahwa tiga variabel teratas yang paling berpengaruh dalam prediksi abnormalitas tekanan darah yaitu (1) *Level of Hemoglobin* (0.2285): Kadar hemoglobin memiliki pengaruh terbesar dalam model, mengindikasikan hubungan kuat antara faktor hematologi dan tekanan darah, (2) *Chronic kidney disease* (0.2195): Penyakit ginjal kronis muncul sebagai prediktor kuat kedua, menegaskan hubungan yang diketahui antara fungsi ginjal dan regulasi tekanan darah, dan (3) *Genetic Pedigree Coefficient* (0.1928): Koefisien keturunan genetik memiliki pengaruh signifikan, menunjukkan peran penting faktor genetik dalam menentukan risiko abnormalitas tekanan darah.

Variabel lain yang memiliki pengaruh cukup signifikan yaitu *Adrenal and thyroid disorders* (0.1287), sedangkan variabel seperti *Smoking* (0.0048) dan *Pregnancy* (0.0070) menunjukkan pengaruh yang relatif kecil dalam model ini.

Model *Random Forest* dikonfigurasi dengan parameter *n_estimators*=50, *max_depth*=5, dan *max_features*=1 untuk mencegah *overfitting* dan mengoptimalkan performa. Teknik *Random Forest* memiliki performa yang baik pada data testing berdasarkan hasil evaluasi dengan menggunakan metrik berikut:

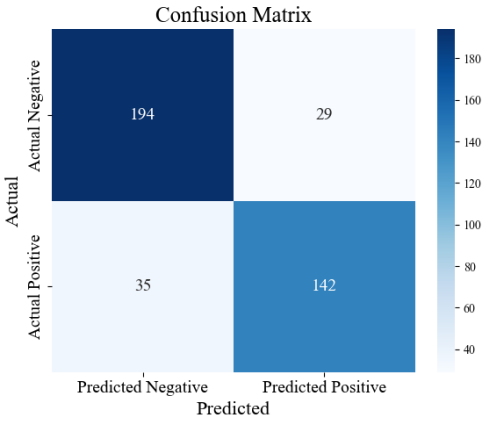
Tabel 2. Evaluasi Model *Random Forest*

Metrix	Kelas Normal (0,0)	Kelas Abnormal (1,0)	Rata - rata
Precision	0,84	0,82	0,83
Recall	0,87	0,79	0,83
F1-score	0,85	0,80	0,83
Accuracy	-	-	0,83

Dapat dilihat pada Tabel 2 bahwa model *Random Forest* menunjukkan kemampuan yang baik dalam mengidentifikasi kedua kelas, dengan sedikit performa lebih baik untuk kelas normal (*F1-score* 0.85) dibandingkan kelas abnormal (*F1-score* 0.80). Hal ini mengindikasikan bahwa model relatif seimbang dalam prediksinya, meskipun sedikit lebih sensitif terhadap kelas normal. Nilai *accuracy* keseluruhan sebesar 0,83 juga mengindikasikan kinerja model yang solid.

Evaluasi Model *sequence Logistic Regression* dan *Random Forest*

Pendekatan *sequence* yang kami usulkan memanfaatkan probabilitas prediksi dari model *Random Forest* sebagai fitur tambahan untuk model *Logistic Regression*. Probabilitas ini memberikan informasi tentang "keyakinan" model *Random Forest* dalam prediksinya, yang dapat dimanfaatkan oleh model *Logistic Regression* untuk meningkatkan akurasinya.



Gambar 6. *Confusion matrix* Model *sequence Logistic Regression* dan *Random Forest*

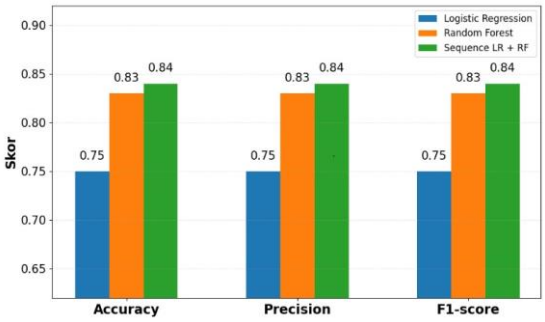
Confusion matrix model *sequence* pada **gambar 6** menunjukkan bahwa dari 223 kasus tekanan darah normal, model berhasil mengidentifikasi 194 kasus dengan benar (*true negative*) dan 29 kasus teridentifikasi sebagai abnormal (*false positive*). Dari 177 kasus abnormal, model mengidentifikasi 142 kasus dengan benar (*true positive*) dan 35 kasus teridentifikasi sebagai normal (*false negative*).

Hasil evaluasi model *sequence* pada data testing menunjukkan performa yang meningkat dibandingkan model *Random Forest* individual, dengan *matrix* sebagai berikut:

Tabel 3. Evaluasi Model *sequence Logistic Regression* dan *Random Forest*

Metrik	Kelas Normal (0,0)	Kelas Abnormal (1,0)	Rata - rata
Precision	0,85	0,83	0,84
Recall	0.87	0,80	0,84
F1-score	0.86	0,82	0,84
Accuracy	-	-	0,84

Dapat dilihat dari tabel 3 terdapat peningkatan performa dalam model *sequence* dibandingkan dengan model *Random Forest* individual (*accuracy* 0.84 *vs* 0.83) mungkin tampak kecil, namun signifikan mengingat konteks klinis di mana setiap peningkatan akurasi dapat berarti diagnosis yang lebih tepat dan manajemen yang lebih efektif untuk pasien dengan risiko abnormalitas tekanan darah.



Gambar 7. *Perbandingan skor evaluasi antara model Logistic Regression, Random Forest, dan Sequence (LR + RF)*

Gambar 7 menyajikan perbandingan performa tiga model klasifikasi dalam memprediksi abnormalitas tekanan darah berdasarkan 3 metrik evaluasi utama yaitu *accuracy*, *precision*, dan *F1-score.* Dari hasil yang ditampilkan, model sequence menunjukkan performa paling unggul secara konsisten dibandingkan kedua model individual. Model *sequence* berhasil mencapai nilai *accuracy* sebesar 0,84, mengungguli *Random Forest* (0,83) dan *Logistic Regression* (0,75). Selain itu, nilai *precision* dan *F1-score* model *sequence* masing-masing mencapai 0,84 dan 0,84. Konsistensi skor yang tinggi pada semua metrik mencerminkan stabilitas dan kekuatan generalisasi model sequence dalam klasifikasi dua kelas (normal dan abnormal) secara seimbang. Keunggulan ini menjadikan model *sequence* sebagai pilihan yang lebih efektif dalam konteks prediksi klinis, khususnya untuk deteksi dini hipertensi, di mana peningkatan akurasi sekecil apa pun dapat berdampak signifikan terhadap kualitas diagnosis dan intervensi medis.

D. PENUTUP

Simpulan

Pendekatan *sequence* yang menggabungkan *Random Forest* dan *Logistic Regression* terbukti meningkatkan performa prediksi abnormalitas tekanan darah, ditandai dengan peningkatan akurasi dari 0,75 pada *Logistic Regression* 0,83 pada *Random Forest* menjadi 0,84 pada model *sequence*. Integrasi *output* probabilistik *Random Forest* sebagai fitur tambahan dalam model *Logistic Regression* memanfaatkan kekuatan komplementer kedua algoritma—kemampuan *Random Forest* dalam menangkap hubungan *non-linear* dan ketangguhan terhadap *noise*, serta interpretabilitas dan kejelasan kerangka probabilistik dari *Logistic Regression*. Analisis *feature importance* menunjukkan dominasi faktor klinis, fisiologis, dan genetik seperti kadar hemoglobin, penyakit ginjal kronis, dan koefisien keturunan genetik dalam memprediksi tekanan darah abnormal, yang memiliki implikasi penting bagi praktik skrining dan manajemen hipertensi. Implementasi model *sequence* ini dapat mendukung pengambilan keputusan klinis yang lebih akurat, memungkinkan intervensi dini terhadap pasien berisiko tinggi, serta berpotensi digunakan dalam riset epidemiologi dan uji klinis terkait hipertensi.

Saran

Berdasarkan analisis terhadap model *sequence* yang dikembangkan, direkomendasikan untuk meningkatkan kualitas dataset dengan mengumpulkan data yang lebih komprehensif dan meminimalkan *missing values*, terutama pada variabel *Pregnancy* yang memerlukan imputasi ekstensif. Pengembangan model *ensemble* lanjutan seperti *stacking* dengan beberapa model dasar atau *gradient boosting* perlu dipertimbangkan untuk meningkatkan performa prediksi abnormalitas tekanan darah. Penting juga untuk melakukan validasi eksternal pada populasi dengan karakteristik demografis dan klinis yang berbeda guna menilai generalisabilitas model dalam konteks klinis yang beragam. Untuk meningkatkan presisi, pengembangan model spesifik bagi subpopulasi tertentu berdasarkan usia, jenis kelamin, atau komorbiditas dapat memberikan hasil prediksi yang lebih akurat untuk kelompok dengan karakteristik unik. Selain itu, integrasi data *longitudinal* dengan memasukkan pengukuran tekanan darah berulang dari waktu ke waktu akan memungkinkan prediksi tren dan risiko jangka panjang yang lebih komprehensif, memperkuat aplikasi klinis dari model yang dikembangkan.

E. DAFTAR PUSTAKA

Adrian, M. R., Putra, M. P., Rafialdy, M. H., & Rakhmawati, N. A. (2021). Perbandingan Metode Klasifikasi Random Forest dan SVM Pada Analisis Sentimen PSBB. *Jurnal Informatika Upgris*, *7*(1). https://doi.org/10.26877/jiu.v7i1.7099

Al Azhima, S. A. T., Darmawan, D., Arief Hakim, N. F., Kustiawan, I., Al Qibtiya, M., & Syafei, N. S. (2022). Hybrid Machine Learning Model untuk memprediksi Penyakit Jantung dengan Metode Logistic Regression dan Random Forest. *Jurnal Teknologi Terpadu*, *8*(1), 40–46. https://doi.org/10.54914/jtt.v8i1.539

Andhika, R. (n.d.). *MACHINE LEARNING DALAM PENGEMBANGAN PERANGKAT LUNAK | Integrative Perspectives of Social and Science Journal*. Retrieved September 19, 2025, from https://ipssj.com/index.php/ojs/article/view/163

Aziz, F., & Abasa, S. (2025). PENGEMBANGAN DAN VALIDASI MODEL HYBRID MACHINE LEARNING UNTUK DIAGNOSIS AWAL DEPRESI. *Journal Pharmacy and Application of Computer Sciences*, *3*(1), 8–15. https://doi.org/10.59823/jopacs.v3i1.69

Az'zahra Tarimana, A., Ryan Septian Fajar, M., Azriel Saktiawan, M., & Adi Saputra, R. (2024). PREDIKSI PENYAKIT HIPERTENSI MENGGUNAKAN MACHINE LEARNING DENGAN ALGORITMA REGRESI LOGISTIK. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(6), 12062–12068. https://doi.org/10.36040/jati.v8i6.11793

Bimo, A. A. (n.d.). *Pemanfaatan Decision Tree pada Algoritma Random Forest untuk Klasifikasi Kanker Payudara*. https://informatika.stei.itb.ac.id/~rinaldi.munir/Matdis/2024-2025/Makalah/Makalah-IF1220-Matdis-2024%20(121).pdf

Bodanki, P. (2021). *Blood pressure data for disease prediction* [Dataset]. Kaggle. https://www.kaggle.com/datasets/pavanbodanki/blood-press

Cheung, A. K., Chang, T. I., Cushman, W. C., Furth, S. L., Hou, F. F., Ix, J. H., Knoll, G. A., Muntner, P., Pecoits-Filho, R., Sarnak, M. J., Tobe, S. W., Tomson, C. R. V., Lytvyn, L., Craig, J. C., Tunnicliffe, D. J., Howell, M., Tonelli, M., Cheung, M., Earley, A., & Mann, J. F. E. (2021). Executive summary of the KDIGO 2021 Clinical Practice Guideline for the Management of Blood Pressure in Chronic Kidney Disease. *Kidney International*, *99*(3), 559–569. https://doi.org/10.1016/j.kint.2020.10.026

Ermawati, Ibnas, R., & Kurniawan, B. A. (2024). Klasifikasi Penderita Anemia Menggunakan Metode Regresi Logistik. *Jurnal MSA (Matematika Dan Statistika Serta Aplikasinya)*, *11*(2), 93–101. https://doi.org/10.24252/msa.v11i2.45083

Gori, T., Sunyoto, A., & Al Fatta, H. (2024). Preprocessing Data dan Klasifikasi untuk Prediksi Kinerja Akademik Siswa. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *11*(1), 215–224. https://doi.org/10.25126/jtiik.20241118074

Habibi, M. R., Hibatullah, F., Kusrini, D. E., Putri, D. A. P., Pratiwi, N. Y., Putri, F. E., Aisha, N., Putra, F. P., Diana, A. P. A., & Ramadlana, A. R. (2023). ANALISIS FAKTOR YANG MEMENGARUHI PRESTASI IPK MAHASISWA DENGAN MENGGUNAKAN REGRESI LOGISTIK. *Journal of Innovation Research and Knowledge*, *3*(7), 1387–1394.

Hamrahian, S. M. (2022). Hypertension and Cardiovascular Disease in Patients with Chronic Kidney Disease. In J. McCauley, S. M. Hamrahian, & O. H. Maarouf (Eds.), *Approaches to Chronic Kidney Disease* (pp. 281–295). Springer International Publishing. https://doi.org/10.1007/978-3-030-83082-3_15

Kuneinen, S. M., Kautiainen, H., Ekblad, M. O., & Korhonen, P. E. (2024). Multifactorial prevention program for cardiovascular disease in primary care: Hypertension status and effect on mortality. *Journal of Human Hypertension*, *38*(4), 322–328. https://doi.org/10.1038/s41371-024-00900-x

Nugraha, W., & Syarif, M. (2024). Evaluasi Performa Algoritma Klasifikasi dalam Prediksi Gagal Jantung: Studi Kasus Dataset Heart Failure Prediction. *Techno.Com*, *23*(4), 897–908. https://doi.org/10.62411/tc.v23i4.11685

Pranandito, R., & Hendry, H. (2023). PERBANDINGAN PREDIKSI PENYAKIT SERANGAN JANTUNG MENGGUNAKAN MODEL MACHINE LEARNING. *JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, *8*(4), 1228–1237. https://doi.org/10.29100/jipi.v8i4.4165

Putra, F., Tahiyat, H. F., Ihsan, R. M., Rahmaddeni, R., & Efrizoni, L. (2024). Penerapan Algoritma K-Nearest Neighbor Menggunakan Wrapper Sebagai Preprocessing untuk Penentuan Keterangan Berat Badan Manusia: Application of K-Nearest Neighbor Algorithm Using Wrapper as Preprocessing for Determination of Human Weight Information. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *4*(1), 273–281. https://doi.org/10.57152/malcom.v4i1.1085

Rangga Aditya Tarigan, L., & Dahlan, D. (2024). OPTIMALISASI FITUR DENGAN FORWARD SELECTION PADA ESTIMASI TINGKAT PENYAKIT PARU-PARU MENGGUNAKAN ALGORITMA KLASIFIKASI RANDOM FOREST. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(5), 10341–10348. https://doi.org/10.36040/jati.v8i5.11064

Saeed, S. M. A. (2023). The Effect of the Thyroid Gland on High Blood Pressure. *Journal of Prevention, Diagnosis and Management of Human Diseases*, *3*(02), 13–17. https://doi.org/10.55529/jpdmhd.32.13.17

Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random Forest Algorithm Overview. *Babylonian Journal of Machine Learning*, *2024*, 69–79. https://doi.org/10.58496/BJML/2024/007

Saputro, D. K., Ajie, M. F. R., Azizah, S., & Hartanti, D. (2023). Penerapan Logistic Regression untuk Mendeteksi Penyakit Jantung pada Pasien. *Prosiding Seminar Nasional Teknologi Informasi Dan Bisnis*, 666–671.

Sari, P. K., & Suryono, R. R. (2024). KOMPARASI ALGORITMA SUPPORT VECTOR MACHINE DAN RANDOM FOREST UNTUK ANALISIS SENTIMEN METAVERSE. *Jurnal Mnemonic*, *7*(1), 31–39. https://doi.org/10.36040/mnemonic.v7i1.8977

Simamora, P., Pasaribu, S. A., & Vera Wijaya. (2025). Peningkatan dan Optimalisasi Prediksi Harga Emas Menggunakan Metode Combine Machine Learning Random Forest dan Gradient Boosting. *Jurnal Mahkota Informatika*, *1*(1), 42–52.

Sitanggang, D., Nicholas, N., Wilson, V., Sinaga, A. R. A., & Simanjuntak, A. D. (2022). IMPLEMENTASI DATA MINING UNTUK MEMPREDIKSI PENYAKIT JANTUNG MENGGUNAKAN METODE K-NEAREST NEIGHBOR DAN LOGISTIC REGRESSION. *Jurnal Teknik Informasi Dan Komputer (Tekinkom)*, *5*(2), 493. https://doi.org/10.37600/tekinkom.v5i2.698

Sun, Z., Wang, G., Li, P., Wang, H., Zhang, M., & Liang, X. (2024). An improved random forest based on the classification accuracy and correlation measurement of decision trees. *Expert Systems with Applications*, *237*, 121549. https://doi.org/10.1016/j.eswa.2023.121549