

Analisis Komparatif Model Pembelajaran Mesin Untuk Memprediksi Hipertensi Ke Dalam Empat Kelas Berdasarkan JNC 8

Ramadhan Maulana Yusup^{1*}, Estiko Rijanto²

¹Magister Sistem Informasi Universitas Komputer Indonesia Bandung, Indonesia; e-mail: ramadhan.75122002@mahasiswa.unikom.ac.id¹

²Pusat Riset Mekatronika Cerdas, Badan Riset dan Inovasi Nasional Bandung, Indonesia; e-mail: estiko.rijanto@brin.go.id

* Korespondensi: e-mail: ramadhan.75122002@mahasiswa.unikom.ac.id

Diterima:27/09/2024; Review:27/12/2024; Disetujui: 31/12/2024

ABSTRAK – Teknologi pembelajaran mesin telah muncul sebagai alat untuk memprediksi hipertensi dengan algoritma seperti Random Forest, Support Vector Machine (SVM), Gradient Boosting, dan XGBoost. Namun, sebagian besar penelitian sebelumnya mengkategorikan hipertensi menjadi dua kelompok berdasarkan ketentuan World Health Organization (WHO), yaitu normal dan hipertensi, yang kurang memberikan informasi rinci untuk penatalaksanaan optimal. Laporan kedelapan dari Joint National Committee 8 (JNC 8) memperkenalkan klasifikasi yang lebih rinci ke dalam empat kelas: normal, prehipertensi, hipertensi stadium 1, dan hipertensi stadium 2, yang memungkinkan penilaian risiko yang lebih spesifik dan personal. Digunakan dua dataset yang berbeda untuk mengevaluasi kinerja model, yaitu dataset dengan atribut faktor risiko yang mudah dikumpulkan (easy-to-collect), dan dataset dengan tambahan atribut biokimia berupa gula darah, kolesterol, dan trigliserida. Pendekatan langsung dan tidak langsung juga diterapkan untuk menganalisis pengaruhnya terhadap kinerja model. Validasi internal dan eksternal digunakan untuk mengevaluasi kinerja model, dimana pada validasi internal digunakan 10-fold cross validation. Hasil menunjukkan bahwa penambahan atribut biokimia secara signifikan meningkatkan kinerja model pembelajaran mesin dalam hal akurasi, precision, recall, specificity, F1-score, dan AUC. Pendekatan tidak langsung menunjukkan peningkatan nilai akurasi model, meski disertai dengan penurunan nilai AUC. Kinerja terbaik didapatkan oleh algoritma XGBoost, dimana validasi internal dan eksternal menunjukkan nilai akurasi tertinggi dengan menggunakan pendekatan tidak langsung sebesar 86,38% dan 80,22%, sedangkan AUC tertinggi didapatkan dengan pendekatan langsung dengan nilai sebesar 97,25% dan 94,54%.

Kata Kunci – faktor risiko, hipertensi, klasifikasi multi kelas, pembelajaran mesin.

Comparative Analysis of Machine Learning Models for Predicting Hypertension into Four Classes

ABSTRACT – Machine learning technology has emerged as a tool for predicting hypertension using algorithms such as Random Forest, Support Vector Machine (SVM), Gradient Boosting, and XGBoost. However, most previous studies categorized hypertension into two groups based on the World Health Organization (WHO) guidelines, namely normal and hypertension, which provides limited detailed information for optimal management. The Eighth Report of the Joint National Committee (JNC 8) introduced a more detailed classification into four classes: normal, prehypertension, stage 1 hypertension, and stage 2 hypertension, enabling more specific and personalized risk assessment. Two different datasets were used to evaluate the model's performance: a dataset with easy-to-collect risk factor attributes and a dataset with additional biochemical attributes such as blood glucose, cholesterol, and triglycerides. Both direct and indirect approaches were applied to analyze their impact on model performance. Internal and external validation were employed to assess the model's performance, with internal validation using 10-fold cross-validation. The results showed that the addition of biochemical attributes significantly improved the performance of machine learning models in terms of accuracy, precision, recall, specificity, F1-score, and AUC. The indirect approach demonstrated an increase in model accuracy, albeit accompanied by a decrease in AUC values. The best performance was achieved by the XGBoost algorithm, where internal and external validation showed the highest accuracy values using the indirect approach, at 86.38% and 80.22%, respectively, while the highest AUC was obtained with the direct approach, with values of 97.25% and 94.54%, respectively."

Keywords – risk factors, hypertension, multi-class classification, machine learning.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



1. PENDAHULUAN

Hipertensi merupakan salah satu penyakit kronis yang menjadi perhatian serius dalam bidang kesehatan global [1]. Hipertensi atau tekanan darah tinggi sering dikenal juga sebagai "the silent killer" karena biasanya terjadi tanpa menunjukkan gejala apapun. Hipertensi merupakan penyumbang tunggal utama morbiditas dan mortalitas akibat stroke, gagal ginjal, dan penyakit jantung [2]. Berdasarkan hasil survei nasional kementerian kesehatan tahun 2018, hipertensi menempati posisi teratas sebagai penyebab kematian di Indonesia akibat penyakit tidak menular [3]. Pada tahun 2018, angka prevalensi penderita hipertensi di Indonesia meningkat sangat tinggi. Dalam kurun waktu sekitar satu dekade, angka tersebut tidak mengalami penurunan dan angka prevalensi hipertensi mencapai 34,1% [3]. Data ini menjadi acuan yang menggambarkan tingginya angka penderita penyakit tidak menular yang disebabkan oleh hipertensi.

Mengingat dampak signifikan dari penyakit hipertensi, deteksi dini dan prediksi akurat individu yang berisiko menderita hipertensi menjadi sangat penting. Upaya untuk mendeteksi dan mengelola hipertensi secara efektif sangat penting dalam meningkatkan kesehatan masyarakat dan mengurangi beban penyakit tersebut. Sebagian besar penelitian tentang prediksi hipertensi mengkategorikan status hipertensi menjadi dua kelompok berdasarkan standar *World Health Organization* (WHO) yaitu, normal dan hipertensi [4]. Namun, pendekatan ini tidak memberikan informasi yang cukup rinci untuk penatalaksanaan hipertensi yang optimal. Laporan kedelapan dari *Joint National Committee 8* (JNC 8) memperkenalkan klasifikasi hipertensi yang lebih rinci ke dalam empat kelas, yaitu normal, prehipertensi, hipertensi stadium 1, dan hipertensi stadium 2 [5]. Pembagian menjadi empat kelas ini memungkinkan penilaian risiko yang lebih spesifik dan personal [6].

Dalam beberapa tahun terakhir, teknologi pembelajaran mesin telah muncul sebagai alat untuk memprediksi penyakit hipertensi. *Random Forest* (RF) telah digunakan untuk memprediksi hipertensi dan menunjukkan bahwa algoritma ini efektif untuk menangani dataset yang besar dan kompleks dalam melakukan prediksi penyakit hipertensi [8-11]. Beberapa penelitian lainnya juga menunjukkan performa yang baik saat menggunakan *Support Vector Machine* (SVM) dalam memprediksi hipertensi [12-14]. Adapun metode statistik sederhana seperti *Logistic*

Regression (LR) memberikan hasil yang baik juga dalam memprediksi hipertensi [15-18]. Begitu juga dengan algoritma *Boosting* seperti *Gradient Boosting* (GB) dan *Extreme Gradient Boosting* (XGBoost), dalam beberapa studi menunjukkan hasil prediksi yang cukup baik [18-20].

Tujuan dari penelitian ini adalah untuk melakukan analisis komparatif terhadap berbagai model pembelajaran mesin, termasuk *Random Forest*, *Support Vector Machine* (SVM), *Gradient Boosting*, dan *XGBoost*, dalam memprediksi status hipertensi ke dalam empat kelas berdasarkan klasifikasi yang diperkenalkan oleh *Joint National Committee 8* (JNC 8), yaitu normal, prehipertensi, hipertensi stadium 1, dan hipertensi stadium 2. Penelitian ini bertujuan untuk mengevaluasi performa model-model tersebut dengan menggunakan dua dataset berbeda, yaitu dataset yang terdiri dari atribut faktor risiko yang mudah dikumpulkan dan dataset yang dilengkapi dengan atribut biokimia seperti gula darah, kolesterol, dan trigliserida. Melalui validasi internal dan eksternal, penelitian ini bertujuan untuk mengidentifikasi model dengan kinerja terbaik yang dapat digunakan untuk meningkatkan akurasi, precision, recall, specificity, F1-score, dan AUC dalam prediksi hipertensi yang lebih rinci dan personal.

2. METODE PENELITIAN

2.1. Sumber Data

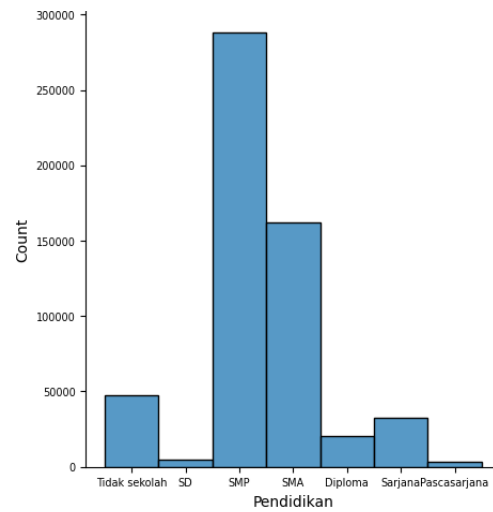
Penelitian ini telah memperoleh persetujuan klirens etik dari Badan Riset dan Inovasi Nasional (BRIN) dengan nomor persetujuan 061/KE.03/SK/03/2024. Proses pengajuan klirens etik dilakukan dengan mengikuti prosedur yang telah ditetapkan oleh BRIN, termasuk penyampaian proposal penelitian, dokumen-dokumen pendukung, serta penjelasan mengenai metode penelitian yang akan digunakan. Semua aspek penelitian, termasuk pengumpulan data, analisis, dan publikasi hasil, telah ditinjau dan disetujui oleh komite etik BRIN untuk memastikan bahwa penelitian ini mematuhi prinsip-prinsip etika yang ketat dan bertanggung jawab.

Tabel 1. Daftar Atribut

Nama	Indikator
Status Hipertensi (JNC-8)	<ul style="list-style-type: none"> Normal: Tekanan sistolik < 120 mmHg dan tekanan diastolik < 80 mmHg. Prehipertensi: Tekanan sistolik 120-139 mmHg dan tekanan diastolik ≥ 80-89 mmHg. Hipertensi Stadium 1: Tekanan sistolik 140-159 mmHg dan tekanan diastolik ≥ 90-99 mmHg. Hipertensi Stadium 2: Tekanan sistolik ≥ 160 mmHg dan tekanan diastolik ≥ 100 mmHg.

Status Hipertensi (WHO)	<ul style="list-style-type: none"> Normal: Tekanan sistolik < 120 mmHg dan tekanan diastolic < 80 mmHg. Hipertensi Stadium 1: Tekanan sistolik 140-159 mmHg dan tekanan diastolik \geq 90-99 mmHg.
Riwayat Hipertensi Keluarga	Ya atau Tidak
Konsumsi Alkohol	Ya atau Tidak
Kurang Konsumsi Sayur dan Buah	Ya atau Tidak
Kurang Aktivitas Fisik	Ya atau Tidak
Merokok	Ya atau Tidak
Indeks Massa Tubuh	Numerik
Lingkar perut	Numerik
Gula Darah Sewaktu	Numerik
Kolesterol Total	Numerik
Trigliserida	Numerik
Jenis Kelamin	<ul style="list-style-type: none"> L: Laki-laki P: Perempuan
Usia	Numerik
Provinsi	Terdiri dari 38 provinsi di Indonesia
Aktivitas Pekerjaan	<ul style="list-style-type: none"> H: Aktivitas Tinggi I: Aktivitas Sedang L: Aktivitas Rendah
Pendidikan	<ul style="list-style-type: none"> Didik0: Tidak Sekolah Didik1: Sekolah Dasar dan Sekolah Menengah Pertama Didik2: Sekolah Menengah Atas Didik3: Diploma, Sarjana, dan Pascasarjana

intensitas fisiknya menjadi tiga kategori: aktivitas tinggi, aktivitas sedang, dan aktivitas rendah. Nilai-nilai yang tidak dapat dipahami dibiarkan kosong dan akan diproses lebih lanjut saat menangani nilai yang hilang.



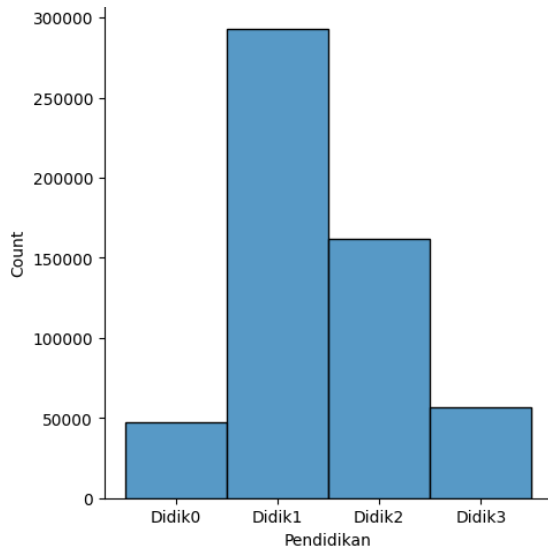
Gambar 1. Histogram Kategori Pendidikan

Setelah pengelompokan nilai pada atribut pendidikan, diilustrasikan pada gambar 1 bahwa ditemukan distribusi kategori pendidikan yang tidak seimbang, dengan mayoritas responden berpendidikan SMP dan SMA, sementara jumlah responden dengan pendidikan lebih rendah atau lebih tinggi jauh lebih sedikit. Untuk mengatasi distribusi yang tidak seimbang ini, maka kategori pendidikan SD digabungkan dengan SMP. Penggabungan ini dilakukan karena responden dengan pendidikan SD dan SMP memiliki dasar pendidikan yang tidak berbeda jauh, terutama dalam hal pengetahuan dasar dan kemampuan penalaran. Demikian Diploma, Sarjana, dan Pascasarjana, memiliki latar belakang pendidikan yang lebih tinggi dan memiliki kemampuan analisis dan penalaran yang lebih kompleks.

Setelah dilakukan penggabungan, maka didapatkan distribusi nilai seperti pada gambar 2. Adapun keterangan dari setiap kategori nilainya seperti yang telah dicantumkan pada tabel 1.

2.2. Menangani Data yang Tidak Konsisten

Terdapat data-data yang tidak konsisten pada dataset yang digunakan. Oleh karena itu perlu dilakukan penanganan terhadap data-data tersebut. Langkah-langkah penanganan data yang dilakukan meliputi: (1) Mengganti koma dengan titik pada semua atribut numerik untuk memastikan keseragaman format data. (2) Menghapus karakter selain angka dan titik pada atribut numerik guna menjaga integritas data. (3) Menghapus karakter selain angka pada atribut nomor KTP untuk memastikan konsistensi dan validitas data demografi. (4) Mengosongkan atribut jenis kelamin jika nilai yang ada tidak konsisten, agar dapat diproses lebih lanjut saat menangani nilai yang hilang. (5) Mengelompokkan nilai pada atribut pendidikan menjadi tujuh jenjang: Tidak Sekolah, SD, SMP, SMA, Diploma, Sarjana, dan Pascasarjana, serta mengosongkan nilai yang tidak dapat dipahami untuk penanganan lebih lanjut. (6) Mengelompokkan pekerjaan berdasarkan



Gambar 2. Histogram Pada Atribut Pendidikan

2.3. Menangani Nilai yang Hilang

Seperti yang terlihat pada tabel 2, dataset ini memiliki berbagai tingkat kehilangan data, dengan atribut seperti trigliserida, kolesterol, dan gula darah mengalami kehilangan data yang sangat tinggi. Nilai yang hilang dalam dataset dapat menyebabkan masalah dalam analisis data dan menghambat pelatihan model pembelajaran mesin. Oleh karena itu, penanganan nilai yang hilang menjadi langkah krusial. Berikut langkah-langkah yang dilakukan: (1) Menghapus observasi dengan nilai hilang pada atribut Status Hipertensi (JNC-8) dan Status Hipertensi (WHO) karena keduanya merupakan target prediksi. (2) Mengisi nilai hilang pada atribut provinsi, jenis kelamin, tanggal lahir, dan usia berdasarkan validasi nomor KTP sesuai Peraturan Pemerintah No. 37 Tahun 2007 dan Peraturan Menteri Dalam Negeri No. 58 Tahun 2021 [20], [21]. (3) Mengisi nilai kosong pada atribut numerik menggunakan algoritma *Multiple Imputation by Chained Equations* (MICE) dan *Bayesian Ridge Regression*. (4) Mengisi nilai kosong pada atribut kategorikal menggunakan MICE dan *Random Forest*, di mana inisialisasi awal menggunakan nilai mode yang paling sering muncul.

2.4. Menangani Outlier

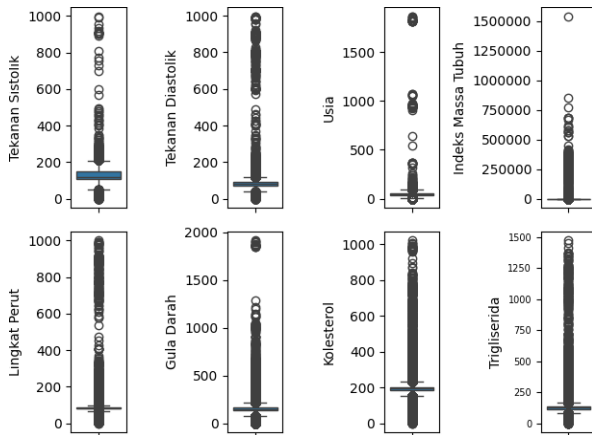
Outlier merupakan sebuah perturbansi acak di dalam nilai atribut numerik [22]. Outlier biasanya terlihat sebagai nilai ekstrem yang jauh dari rentang nilai umum data lainnya, baik di ujung tinggi maupun rendah. Adapun pada setiap atribut numerik yang ada yaitu tekanan sistolik, tekanan diastolik, usia, indeks massa tubuh, lingkar perut, gula darah, kolesterol dan trigliserida, nilai negatif tidak termasuk ke dalam rentang nilai dari semua atribut tersebut. Sehingga semua nilai negatif yang ada pada

atribut tersebut diidentifikasi sebagai outlier. Selain itu, outlier diidentifikasi juga menggunakan box plot. Seperti pada gambar 3, dapat dilihat bahwa semua atribut numerik memiliki nilai yang ekstrim. Nilai-nilai yang ada diluar dari batas bawah dan batas atas dari setiap *whisker* diidentifikasi sebagai outlier.

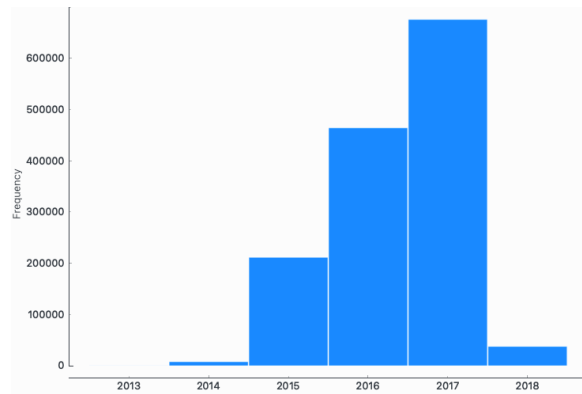
Tabel 2. Persentase Data Hilang

Atribut	Persentase data hilang
Tanggal Pemeriksaan	0%
Nomor KTP	0.31%
Jenis Kelamin	0.58%
Provinsi	0.92%
Tanggal Lahir	5.23%
Usia	5.29%
Status Hipertensi (JNC-8)	14.71%
Status Hipertensi (WHO)	14.71%
Tekanan Sistolik	14.71%
Tekanan Diastolik	14.85%
Kurang Aktivitas Fisik	27.09%
Kurang Konsumsi Sayur dan Buah	27.76%
Konsumsi Alkohol	27.99%
Merokok	25.84%
Indeks Massa Tubuh	31.47%
Lingkar Perut	46.64%
Riwayat Hipertensi Keluarga	63.19%
Aktivitas Pekerjaan	64.97%
Pendidikan	68.68%
Gula Darah Sewaktu	84.84%
Kolesterol Total	93.76%
Trigliserida	98.63%

Setelah semua outlier diidentifikasi, langkah selanjutnya adalah penanganannya. Untuk outlier pada atribut usia, nilainya diperbaiki dengan mengekstraksi tanggal lahir dari nomor KTP dan menghitung ulang usia jika nomor KTP valid. Jika usia masih berada di luar batas *whisker* setelah perhitungan ulang, atau jika nomor KTP tidak valid atau tidak ada, observasi tersebut dihapus. Sementara itu, pada atribut numerik lainnya, semua observasi dengan nilai outlier dihapus tanpa proses tambahan.



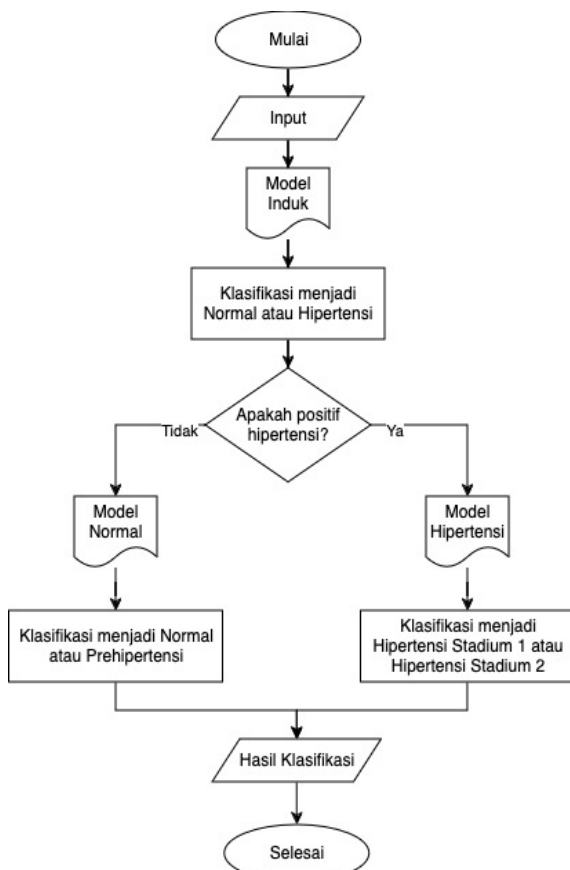
Gambar 3. Box Plot Atribut Numerik



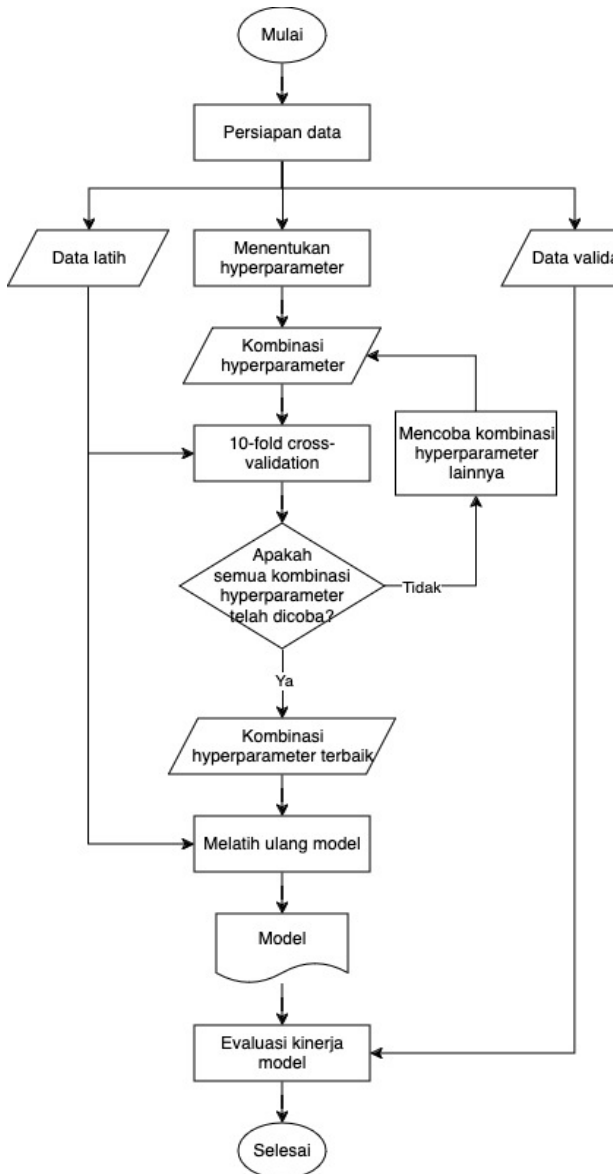
Gambar 5. Histogram Data Dalam Periode Tahun

2.5. Teknik Pengujian Penelitian

Untuk mengklasifikasi status hipertensi ke dalam 4 kelas, 2 pendekatan digunakan dalam membangun model pembelajaran mesin yaitu langsung dan tidak langsung. Pendekatan langsung merupakan model klasifikasi multi kelas seperti pada umumnya, dimana algoritma langsung mengklasifikasi ke salah satu dari empat kelas berdasarkan atribut penjelasnya. Adapun pendekatan tidak langsung merupakan model klasifikasi yang terdiri dari 3 model, yaitu model induk, model normal, dan model hipertensi.



Gambar 4. Alur Kerja Klasifikasi Menggunakan Pendekatan Tidak Langsung

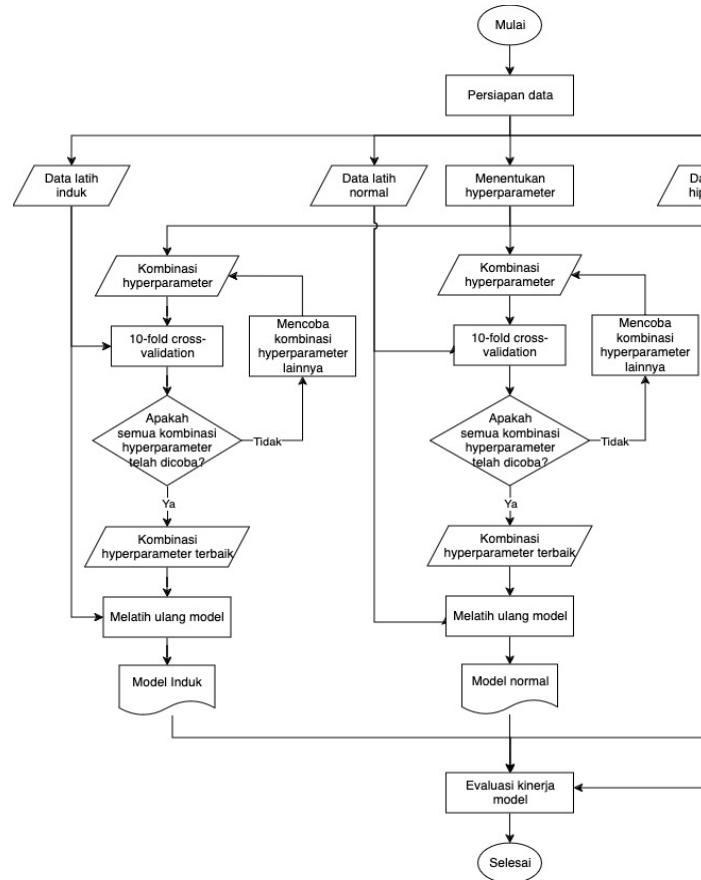


Gambar 6. Alur Pengembangan Model Langsung

Seperti yang diilustrasikan pada gambar 4, pendekatan tidak langsung bekerja secara bertahap. Mula-mula data dimasukkan diklasifikasi oleh model induk untuk diketahui status hipertensinya apakah normal atau hipertensi berdasarkan ketentuan WHO. Setelah itu jika hasilnya adalah normal, maka dilakukan pengklasifikasian lagi menggunakan model normal untuk menentukan apakah responden tersebut termasuk ke dalam kelas normal atau prehipertensi. Begitu pula sebaliknya, jika hasil klasifikasi dari model induk adalah hipertensi, maka dilakukan pengklasifikasi kembali menggunakan model hipertensi untuk menentukan apakah responden tersebut termasuk ke dalam kelas hipertensi stadium 1 atau hipertensi stadium 2.

Evaluasi kinerja model, baik langsung maupun tidak langsung, dilakukan melalui dua jenis validasi, yaitu validasi internal dan validasi eksternal. Kedua jenis validasi ini penting untuk

memastikan bahwa model yang dikembangkan memiliki kinerja yang baik dan dapat digeneralisasi ke dataset yang belum pernah



Gambar 7. Alur Pengembangan Model Tidak Langsung

Tabel 3. Rekapitulasi Kinerja Validasi

Algoritma	Atribut yang Digunakan	Pendekatan	Akurasi	Precision	Recall
Logistic Regression	Easy-to-collect	Langsung	42.62%	43.01%	38.77%
		Tidak Langsung	42.76%	42.85%	39.74%
	Easy-to-collect · Biokimia	Langsung	47.17%	47.02%	43.94%
		Tidak Langsung	48.16%	47.81%	45.82%
SVM	Easy-to-collect	Langsung	44.65%	46.07%	40.32%
		Tidak Langsung	44.82%	44.84%	41.31%
	Easy-to-collect · Biokimia	Langsung	50.99%	51.97%	47.13%
		Tidak Langsung	51.73%	51.85%	48.59%
Random Forest	Easy-to-collect	Langsung	64.59%	65.12%	64.17%
		Tidak Langsung	65.28%	66.07%	64.67%
	Easy-to-collect · Biokimia	Langsung	85.31%	85.80%	85.15%
		Tidak Langsung	85.32%	85.83%	85.19%
Gradient Boosting	Easy-to-collect	Langsung	64.04%	65.10%	63.12%
		Tidak Langsung	64.59%	65.73%	63.78%
	Easy-to-collect · Biokimia	Langsung	85.59%	86.04%	85.45%
		Tidak Langsung	85.68%	86.26%	85.51%
XGBoost	Easy-to-collect	Langsung	65.88%	66.57%	65.33%
		Tidak Langsung	66.09%	66.79%	65.53%
	Easy-to-collect · Biokimia	Langsung	86.32%	86.90%	86.14%
		Tidak Langsung	86.38%	86.90%	86.24%

dilihat sebelumnya. Untuk melakukan validasi ini, dataset dibagi berdasarkan periode waktu. Data untuk validasi internal mencakup periode tahun 2013 hingga semester awal tahun 2017, sementara validasi eksternal menggunakan data dari semester akhir tahun 2017 hingga tahun 2018. Pemilihan periode ini didasarkan pada gambar 5, yang menunjukkan bahwa jumlah observasi pada tahun 2018 lebih sedikit dibandingkan tahun-tahun sebelumnya, sehingga observasi dari semester akhir tahun 2017 ditambahkan untuk memperbesar ukuran dataset validasi eksternal.

Validasi internal dilakukan dengan metode *10-fold cross-validation*, di mana dataset dibagi menjadi 10 subset, dan model dilatih pada 9 subset serta diuji pada 1 subset [22]. Proses ini diulangi 10 kali, dan hasilnya dirata-rata untuk mengestimasi kinerja model. Validasi eksternal digunakan untuk menguji model pada dataset baru yang tidak digunakan selama pelatihan, penting untuk menilai generalisasi model terhadap data baru yang mungkin berbeda dari data pelatihan. Hasil validasi eksternal memberikan wawasan tentang kinerja model dalam situasi dunia nyata. Kedua validasi ini dievaluasi menggunakan berbagai indikator kinerja, diantaranya akurasi, *precision*, *recall*, *specificity*, *F1-score*, kurva ROC, dan nilai AUC.

2.6. Perancangan dan Analisis Data

Pada gambar 6 digambarkan proses pengembangan dan evaluasi model pembelajaran mesin yang dilakukan. Dimulai dengan persiapan data, dimana data dibagi menjadi dataset pelatihan dan validasi. Selanjutnya, berbagai kombinasi *hyperparameter* ditentukan dan diuji menggunakan *10-fold cross-validation* pada data pelatihan. Proses ini berulang hingga semua kombinasi *hyperparameter* dicoba dan kinerjanya dievaluasi. Kombinasi *hyperparameter* yang memberikan kinerja terbaik kemudian dipilih. Kemudian model dilatih ulang dengan kombinasi *hyperparameter* terbaik menggunakan seluruh data pelatihan untuk memaksimalkan pembelajaran. Model akhir kemudian dievaluasi menggunakan data validasi untuk menilai kinerjanya.

Adapun gambar 7 proses pengembangan model prediksi hipertensi menggunakan pendekatan tidak langsung. Proses dimulai dengan persiapan data, di mana data dibagi menjadi data latih induk, data latih normal, data latih hipertensi, dan data validasi. Setiap set data latih kemudian menjalani proses optimisasi *hyperparameter* secara terpisah. Untuk setiap set data, berbagai kombinasi *hyperparameter* dicoba dan dievaluasi menggunakan *10-fold cross-validation*. Proses ini berulang hingga semua kombinasi *hyperparameter* dicoba, dan kombinasi

terbaik dipilih untuk melatih ulang model masing-masing. Setelah itu, model induk, model normal, dan model hipertensi dibangun berdasarkan kombinasi *hyperparameter* terbaik yang dipilih

Jika dibandingkan dengan alur pengembangan model langsung menunjukkan bahwa pendekatan langsung hanya menggunakan satu model dengan satu proses optimisasi *hyperparameter* yang diuji dengan *10-fold cross-validation*. Sedangkan, dalam pendekatan tidak langsung bertingkat, proses ini dipecah menjadi tiga tahap terpisah: model induk untuk klasifikasi awal, model normal untuk sub-kategori normal, dan model hipertensi untuk sub-kategori hipertensi. Setiap model dilatih dan dioptimalkan secara independen sebelum dilakukan evaluasi kinerja model secara keseluruhan.

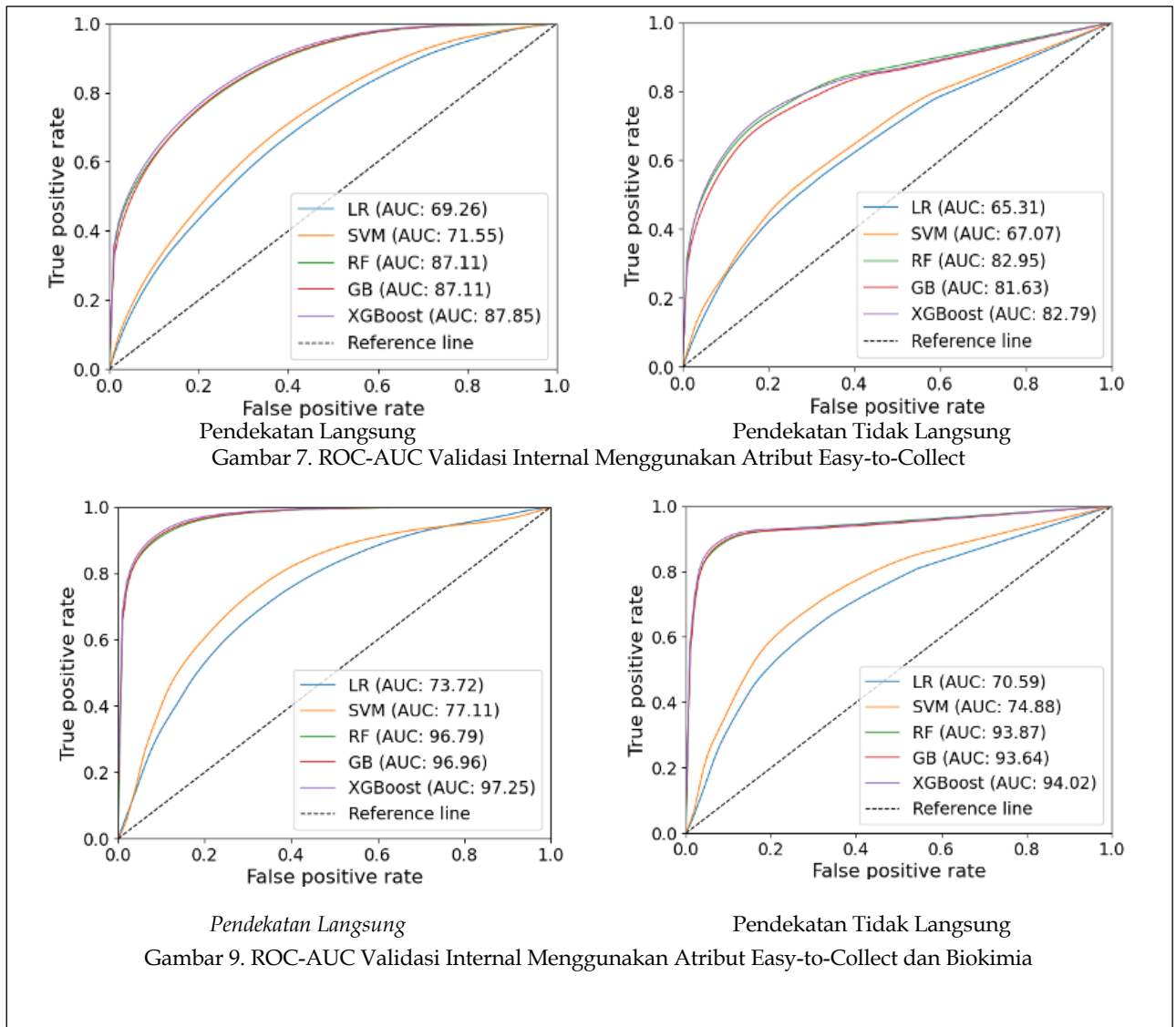
3. HASIL DAN PEMBAHASAN

Kinerja model pembelajaran mesin yang telah dilatih terbagi atas dataset, pendekatan, dan algoritma yang digunakan. Lalu, kurva ROC dan nilai AUC ikut disertakan juga untuk menunjukkan kemampuan model dalam melakukan klasifikasi.

3.1. Validasi Internal

Berdasarkan rekapitulasi kinerja validasi internal pada tabel 3, dapat dilihat bahwa pada validasi internal penambahan atribut biokimia secara signifikan meningkatkan kinerja model pembelajaran mesin pada hampir semua metrik evaluasi. Lalu dari kelima algoritma yang digunakan, RF, GB dan XGBoost menunjukkan peningkatan kinerja yang paling signifikan dengan adanya tambahan atribut biokimia. Jika dibandingkan nilai akurasi dan nilai AUC dari semua algoritma, dapat dilihat bahwa XGBoost memiliki kinerja yang paling baik.

Berdasarkan pendekatan yang dilakukan, dapat dilihat bahwa semua akurasi mendapatkan peningkatan kinerja dengan menggunakan pendekatan tidak langsung. Kinerja lainnya seperti *precision*, *recall*, *specificity* dan *f1-score* juga terlihat ada beberapa peningkatan. Namun hal ini disertai juga dengan adanya penurunan pada nilai AUC. Dari segi waktu latih, menggunakan atribut tambahan biokimia terlihat menambah waktu latih pada setiap algoritma. Jika dibandingkan semua algoritmanya, dapat dilihat bahwa logistic regression memiliki waktu latih yang tercepat dan SVM memiliki waktu yang latih yang terlama. Adapun jika dilihat dari pendekatan yang dilakukan, pendekatan langsung terlihat lebih cepat dibandingkan dengan pendekatan tidak langsung.



Adapun berdasarkan grafik ROC-AUC pada gambar 8 dan 9 menunjukkan bahwa RF, GB dan XGBoost secara konsisten memberikan performa terbaik dalam memprediksi hipertensi, baik pada dataset *easy-to-collect* maupun dataset biokimia. RF, GB dan XGBoost menunjukkan kemampuannya dalam menangani data yang kompleks dan menghasilkan prediksi yang akurat. Sebaliknya, model linear seperti LR dan SVM memiliki performa yang rendah, terutama pada data biokimia yang lebih kompleks.

3.2. Validasi Eksternal

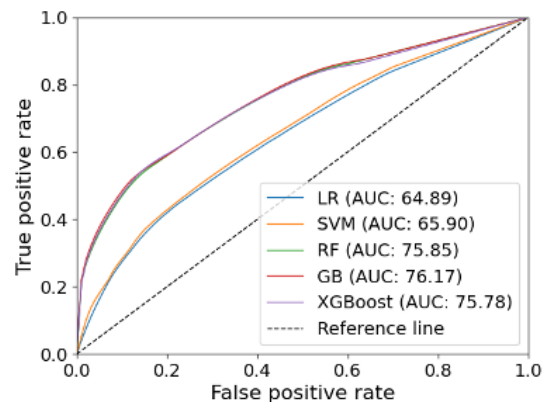
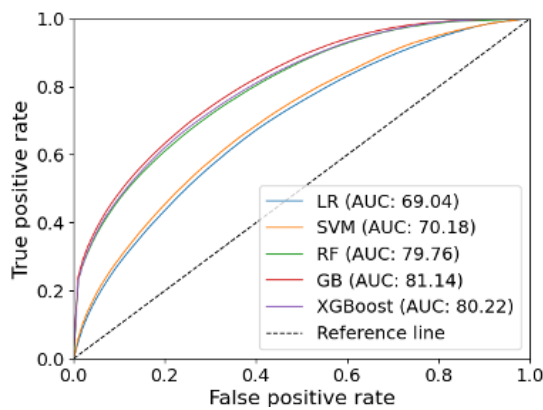
Sama seperti validasi internal, penambahan atribut biokimia juga secara signifikan meningkatkan kinerja model pembelajaran mesin pada hampir semua metrik evaluasi pada saat validasi eksternal. Lalu dari kelima algoritma yang digunakan, RF, GB dan XGBoost juga menunjukkan peningkatan kinerja yang paling signifikan dengan adanya tambahan atribut biokimia. Jika dibandingkan nilai akurasi dan nilai AUC dari semua algoritma, dapat dilihat

bahwa XGBoost juga memiliki kinerja yang paling baik. berdasarkan pendekatan yang dilakukan pun, semua akurasi mendapatkan peningkatan kinerja dan AUC mendapatkan penurunan nilai dengan menggunakan pendekatan tidak langsung. Dari segi waktu latih pun terlihat pola yang sama, penggunaan atribut tambahan biokimia serta pendekatan tidak langsung menunjukkan penambahan waktu latih pada setiap algoritma.

Adapun berdasarkan grafik ROC-AUC pada gambar 10 dan 11, hasil yang sama menunjukkan bahwa RF, GB dan XGBoost menghasilkan performa terbaik dalam memprediksi hipertensi, baik menggunakan dataset yang menggunakan atribut *easy-to-collect* saja maupun dengan dataset yang telah ditambahkan atribut biokimia.

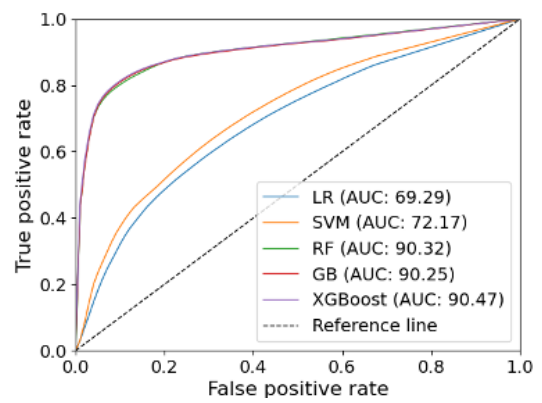
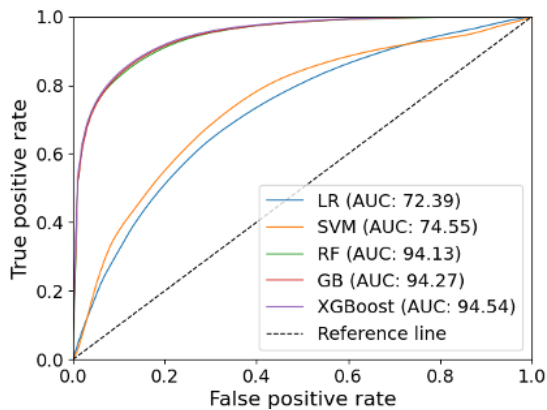
Tabel 3. Rekapitulasi Kinerja Validasi Eksternal

Algoritma	Atribut yang Digunakan	Pendekatan	Akurasi	Precision	Recall	Specificity	F1-Score	AUC	Waktu Latih
Logistic Regression	Easy-to-collect	Langsung	43.91%	42.76%	38.88%	79,01%	36.53%	69.04%	10 detik
		Tidak Langsung	44.22%	42.24%	39.89%	79,23%	38.36%	64.89%	14 detik
	Easy-to-collect + Biokimia	Langsung	47.38%	45.64%	43.24%	80,41%	41.59%	72.39%	15 detik
		Tidak Langsung	47.82%	46.41%	45.12%	80,82%	44.05%	69.29%	25 detik
SVM	Easy-to-collect	Langsung	45.17%	45.27%	39.66%	79,19%	36.59%	70.18%	250 menit 49 detik
		Tidak Langsung	45.26%	44.34%	40.71%	79,46%	38.22%	65.90%	397 menit 52 detik
	Easy-to-collect + Biokimia	Langsung	49.48%	50.06%	45.31%	80,80%	43.38%	74.58%	254 menit 32 detik
		Tidak Langsung	49.70%	49.85%	46.81%	81,24%	44.66%	72.17%	399 menit 13 detik
Random Forest	Easy-to-collect	Langsung	56.28%	55.22%	54.31%	83,88%	54.57%	79.76%	4 menit 40 detik
		Tidak Langsung	55.32%	56.61%	54.72%	84,10%	55.31%	75.85%	7 menit 53 detik
	Easy-to-collect + Biokimia	Langsung	79.23%	80.24%	78.09%	92,43%	79.01%	94.13%	6 menit 34 detik
		Tidak Langsung	79.37%	80.40%	78.34%	92,49%	79.20%	90.32%	10 menit 57 detik
Gradient Boosting	Easy-to-collect	Langsung	57.05%	57.94%	55.22%	84,26%	55.97%	81.14%	4 menit 7 detik
		Tidak Langsung	57.31%	58.33%	55.48%	84,33%	56.20%	76.17%	9 menit 45 detik
	Easy-to-collect + Biokimia	Langsung	79.74%	80.95%	78.69%	92,59%	79.65%	94.27%	5 menit 11 detik
		Tidak Langsung	79.81%	80.84%	78.84%	92,65%	79.70%	90.25%	10 menit 42 detik
XGBoost	Easy-to-collect	Langsung	56.34%	56.51%	55.14%	84,18%	55.58%	80.22%	12 menit 37 detik
		Tidak Langsung	56.72%	56.99%	55.47%	84,30%	55.96%	75.78%	22 menit 42 detik
	Easy-to-collect + Biokimia	Langsung	80.11%	81.37%	79.08%	92,72%	80.04%	94.54%	15 menit 24 detik
		Tidak Langsung	80.22%	81.39%	79.26%	92,78%	80.15%	90.47%	32 menit 18 detik



Pendekatan Langsung Pendekatan Tidak Langsung

Gambar 8. ROC-AUC Validasi Eksternal Menggunakan Atribut Easy-to-Collect



Pendekatan Langsung Pendekatan Tidak Langsung

Gambar 11. ROC-AUC Validasi Eksternal Menggunakan Atribut Easy-to-Collect dan Biokimia

4. KESIMPULAN

Model pembelajaran mesin untuk memprediksi status hipertensi ke dalam empat kelas berdasarkan klasifikasi yang diperkenalkan oleh Joint National Committee 8 (JNC 8) telah berhasil dibuat. Lima algoritma pembelajaran mesin telah digunakan, yang diantaranya adalah LR, SVM, RF, GB dan XGBoost. Evaluasi kinerja dari setiap model juga telah dilakukan menggunakan validasi internal dan eksternal. Dari evaluasi yang dilakukan, XGBoost memiliki kinerja terbaik dibandingkan dengan algoritma lainnya, dimana validasi internal dan eksternal menunjukkan nilai akurasi tertinggi dengan menggunakan pendekatan tidak langsung sebesar 86,38% dan 80,22%, sedangkan AUC tertinggi didapatkan dengan pendekatan langsung dengan nilai sebesar 97,25% dan 94,54%.

Penelitian lebih lanjut dapat difokuskan pada optimalisasi waktu latihan tanpa mengorbankan akurasi atau AUC. Hal ini mungkin dapat dilakukan dengan menggabungkan metode-metode reduksi data atau menggunakan metode komputasi yang lebih efisien. Selain itu, pengembangan model hibrid yang menggabungkan keunggulan dari beberapa algoritma pembelajaran mesin juga dapat dieksplorasi untuk menghasilkan prediksi yang lebih akurat. Dalam penelitian lanjutan juga baik untuk mempertimbangkan integrasi model pembelajaran mesin dengan sistem kesehatan yang ada, sehingga memungkinkan aplikasi praktis yang lebih luas dalam mendeteksi dan mengelola hipertensi secara lebih efektif dan efisien.

DAFTAR PUSTAKA

- [1] B. Zhou, P. Perel, G. A. Mensah, and M. Ezzati, "Global epidemiology, health burden and effective interventions for elevated blood pressure and hypertension," *Nat Rev Cardiol*, vol. 18, no. 11, pp. 785–802, Nov. 2021, doi: 10.1038/s41569-021-00559-8.
- [2] X. Ji *et al.*, "Prediction Model of Hypertension Complications Based on GBDT and LightGBM," *J. Phys.: Conf. Ser.*, vol. 1813, no. 1, p. 012008, Feb. 2021, doi: 10.1088/1742-6596/1813/1/012008.
- [3] Kementerian Kesehatan RI, *Hasil Riset Kesehatan Dasar Tahun 2018*. 2018.
- [4] O. M. N. Khatib and M. S. El-Guindy, *Clinical guidelines for the management of hypertension*. Cairo: World Health Organization, Regional Office for the Eastern Mediterranean, 2005.
- [5] K. Bell, J. Twigg, B. R. Olin, and I. R. Date, "Hypertension: the silent killer: updated JNC-8 guideline recommendations," *Alabama pharmacy association*, vol. 334, p. 4222, 2015.
- [6] B. A. Sabol *et al.*, "Defining the risk profile of women with stage 1 hypertension: a time to event analysis," *American journal of obstetrics & gynecology MFM*, vol. 3, no. 4, p. 100376, 2021.
- [7] M. Ijaz, G. Alfian, M. Syafrudin, and J. Rhee, "Hybrid Prediction Model for Type 2 Diabetes and Hypertension Using DBSCAN-Based Outlier Detection, Synthetic Minority Over Sampling Technique (SMOTE), and Random Forest," *Applied Sciences*, vol. 8, no. 8, p. 1325, Aug. 2018, doi: 10.3390/app8081325.
- [8] S. Sakr *et al.*, "Using machine learning on cardiorespiratory fitness data for predicting hypertension: The Henry Ford Exercise Testing (FIT) Project," *PLoS ONE*, vol. 13, no. 4, p. e0195344, Apr. 2018, doi: 10.1371/journal.pone.0195344.
- [9] L. A. AlKaabi, L. S. Ahmed, M. F. Al Attiyah, and M. E. Abdel-Rahman, "Predicting hypertension using machine learning: Findings from Qatar Biobank Study," *PLoS ONE*, vol. 15, no. 10, p. e0240370, Oct. 2020, doi: 10.1371/journal.pone.0240370.
- [10] J. J. Bouillier, T. C. Y. Chan, M. Ranjan, and S. Deo, "Risk Stratification for Early Detection of Diabetes and Hypertension in Resource-Limited Settings: Machine Learning Analysis," *J Med Internet Res*, vol. 23, no. 1, p. e20123, Jan. 2021, doi: 10.2196/20123.
- [11] Z. Pei *et al.*, "Risk-Predicting Model for Incident of Essential Hypertension Based on Environmental and Genetic Factors with Support Vector Machine," *Interdiscip Sci Comput Life Sci*, vol. 10, no. 1, pp. 126–130, Mar. 2018, doi: 10.1007/s12539-017-0271-2.
- [12] R. Patnaik, M. Chandran, S.-C. Lee, A. Gupta, C. Kim, and C. Kim, "Predicting the occurrence of essential hypertension using annual health records," in *2018 Second International Conference on Advances in Electronics, Computers and Communications (ICAEECC)*, IEEE, 2018, pp. 1–5.
- [13] J.-H. Wu *et al.*, "Risk Assessment of Hypertension in Steel Workers Based on LVQ and Fisher-SVM Deep Excavation," *IEEE Access*, vol. 7, pp. 23109–23119, 2019, doi: 10.1109/ACCESS.2019.2899625.
- [14] B. M. Heo and K. H. Ryu, "Prediction of Prehypertension and Hypertension Based on Anthropometry, Blood Parameters, and Spirometry," *IJERPH*, vol. 15, no. 11, p. 2571, Nov. 2018, doi: 10.3390/ijerph15112571.
- [15] P. Andriani and N. Chamidah, "Modelling of Hypertension Risk Factors Using Logistic Regression to Prevent Hypertension in Indonesia," *J. Phys.: Conf. Ser.*, vol. 1306, no. 1, p. 012027, Aug. 2019, doi: 10.1088/1742-6596/1306/1/012027.
- [16] D. Nuryunarsih, L. Herawati, A. Badi'ah, J. D. T. Donsu, and Okatiranti, "Predicting Changes in Systolic and Diastolic Blood Pressure of Hypertensive Patients in Indonesia Using Machine Learning," *Curr Hypertens Rep*, vol. 25, no. 11, pp. 377–383, Nov. 2023, doi: 10.1007/s11906-023-01261-5.
- [17] R. Kurniawan *et al.*, "Hypertension prediction using machine learning algorithm among Indonesian adults," *IJ-AI*, vol. 12, no. 2, p. 776, Jun. 2023, doi: 10.11591/ijai.v12.i2.pp776-784.
- [18] C. Ye *et al.*, "Prediction of incident hypertension within the next year: prospective study using statewide electronic health records and machine learning," *Journal of medical Internet research*, vol. 20, no. 1, p. e22, 2018.
- [19] W. Chang *et al.*, "Prediction of Hypertension Outcomes Based on Gain Sequence Forward Tabu Search Feature Selection and XGBoost," *Diagnostics*, vol. 11, no. 5, p. 792, Apr. 2021, doi: 10.3390/diagnostics11050792.
- [20] Pemerintah Pusat Republik Indonesia, "Peraturan Pemerintah Nomor 37 Tahun 2007 tentang Pelaksanaan

- Undang Undang Nomor 23 Tahun 2006 Tentang Administrasi Kependudukan." 2007.
- [21] Kementerian Dalam Negeri Republik Indonesia, "Peraturan Menteri Dalam Negeri Nomor 58 Tahun 2021 tentang Kode, Data Wilayah Administrasi Pemerintahan, dan Pulau." 2021.
- [22] C. Vercellis, *Business intelligence: data mining and optimization for decision making*. John Wiley & Sons, 2011.
- [23] H. Zhao *et al.*, "Predicting the Risk of Hypertension Based on Several Easy-to-Collect Risk Factors: A Machine Learning Method," *Front. Public Health*, vol. 9, p. 619429, Sep. 2021, doi: 10.3389/fpubh.2021.619429.
- [24] S. M. S. Islam *et al.*, "Machine Learning Approaches for Predicting Hypertension and Its Associated Factors Using Population-Level Data From Three South Asian Countries," *Front. Cardiovasc. Med.*, vol. 9, p. 839379, Mar. 2022, doi: 10.3389/fcvm.2022.839379.
- [25] Md. M. Islam *et al.*, "Predicting the risk of hypertension using machine learning algorithms: A cross sectional study in Ethiopia," *PLoS ONE*, vol. 18, no. 8, p. e0289613, Aug. 2023, doi: 10.1371/journal.pone.0289613.
- [26] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A Comparative Analysis of XGBoost," *Artif Intell Rev*, vol. 54, no. 3, pp. 1937–1967, Mar. 2021, doi: 10.1007/s10462-020-09896-5.
- [27] J. Kim and I. C. Hwang, "Drawing Guidelines for Receiver Operating Characteristic Curve in Preparation of Manuscripts," *J Korean Med Sci*, vol. 35, no. 24, p. e171, 2020, doi: 10.3346/jkms.2020.35.e171.