JUISI Jurnal Informatika dan Sistem Informasi

Volume 9 Nomor 2
e-ISSN: 2477-5452; p-ISSN: 2460-1306
DOI: *doi.org/10.37715/juisi.v9i2.4403*

Pengembangan Model Prediksi Risiko Hipertensi Menggunakan Algoritma Gradient Boosting Decision Tree Yang Dioptimalkan dengan Hyperparameter Tuning Tree Parzer Estimation

Development of a Hypertension Risk Prediction Model Using a Gradient Boosting Decision Tree Algorithm Optimized with Hyperparameter Tuning Tree Parzer Estimation

Franciscus Valentinus Ongkosianbhadra, Caecilia Citra Lestari*
Program Studi Informatika, Universitas Ciputra, Surabaya 60219, Indonesia
*(*Email Korespondensi: caecilia.citra@ciputra.ac.id)*

Abstrak: Penelitian ini dilatarbelakangi adanya kebutuhan akan sistem prediksi yang dapat membantu dalam mengidentifikasi pasien yang memiliki risiko terkena hipertensi sejak dini. Hipertensi adalah kondisi medis dimana tekanan darah dalam arteri menjadi terlalu tinggi secara berkepanjangan. Hal ini menyebabkan beban tambahan pada jantung dan pembuluh darah, yang pada gilirannya dapat menyebabkan kerusakan pada organ dan meningkatkan risiko serangan jantung, stroke, dan penyakit ginjal. Hipertensi sering tidak menimbulkan gejala dan hanya dapat di diagnosa melalui pemeriksaan darah. Pengobatan melibatkan perubahan gaya hidup seperti mengurangi asupan garam, berolahraga secara teratur, dan mengontrol berat badan. Obat-obatan juga dapat diberikan untuk membantu menurunkan tekanan darah. Algoritma *Gradient Boosting Decision Tree* merupakan salah satu teknik *machine learning* yang memiliki akurasi tinggi dalam mengatasi masalah pembelajaran biner. Model dibuat dengan melatih algoritma *Gradient Boosting Decision Tree* pada 70693 baris dataset dari Pusat Pengendalian dan Pencegahan Penyakit Amerika Serika. Dataset tersebut memiliki 17 perilaku dan riwayat kesehatan seseorang yang dapat mengindikasikan risiko hipertensi orang tersebut. Algoritma *Gradient Boosting Decision Tree* dioptimalkan menggunakan beberapa metode *hyperparameter tuning* yaitu *Random Search, Grid Search, Bayesian Optimization, Grading Search*, dan *Tree Parzer Estimation*. Validasi tertinggi diperoleh dari pengoptimalan model menggunakan *Tree Parzer Estimation*, yaitu mencapai akurasi 74,43%.

Kata Kunci: Tekanan Darah Tinggi, Machine Learning, Gradient Boosting Decision Tree, XGBoost, Tree Parzer Etimation.

Abstract: This research is motivated by the need for a prediction system that can help identify patients who are at risk of developing hypertension early on. Hypertension is a medical condition where the blood pressure in the arteries becomes too high for a long time. This causes additional stress on the heart and blood vessels, which in turn can cause damage to organs and increase the risk of heart attack, stroke, and kidney disease. Hypertension often causes no symptoms and can only be diagnosed through blood tests. Treatment involves lifestyle changes such as reducing salt intake, exercising regularly, and controlling body weight. Medicines may also be given to help lower blood pressure. The Gradient Boosting Decision Tree algorithm is a machine learning technique that has high accuracy in solving binary learning problems. The model was created by training the Gradient Boosting Decision Tree algorithm on a 70693 row dataset from the United States Centers for Disease Control and Prevention. This dataset has 17 behaviors and a person's health history that can indicate that person's risk of hypertension. The Gradient Boosting Decision Tree algorithm is optimized using several hyperparameter tuning methods, namely Random Search, Grid Search, Bayesian Optimization, Grading Seach, and Tree Parzer Estimation. The highest validation was obtained from optimizing the model using Tree Parzer Estimation, which achieved an accuracy of 74.43%

JUISI Jurnal Informatika dan Sistem Informasi

Volume 9 Nomor 2
e-ISSN: 2477-5452; p-ISSN: 2460-1306
DOI: *doi.org/10.37715/juisi.v9i2.4403*

Keywords: Hypertension, Machine Learning, Gradient Boosting Decision Tree, XGBoost. Tree Parzer Extiomation

Naskah diterima 13 Desember 2023; direvisi 23 Januari 2024; dipublikasi 30 Desember 2023.

1. PENDAHULUAN

Hipertensi merupakan salah satu penyakit kardiovaskular yang umum terjadi di masyarakat. Masalah hipertensi saat ini menjadi serius tidak hanya di Indonesia tetapi juga di seluruh dunia. Hipertensi dapat menyebabkan penyakit komplikasi yang berat, seperti gagal ginjal, diabetes, stroke, dan masalah jantung (Pratama, et.al., 2020).

Proyeksi menunjukkan bahwa jumlah kasus hipertensi di negara berkembang diperkirakan akan meningkat sekitar 80% pada tahun 2025 dari 639 juta pada tahun 2000. Pada tahun 2025, diperkirakan jumlah penderita hipertensi akan mencapai 1,15 miliar (Lisiswanti & Dananda, 2016). Hipertensi sering disebut sebagai "penyakit pembunuh diam-diam" dan menyebabkan biaya pengobatan yang tinggi, seperti yang terlihat dari data pelayanan BPJS yang menunjukkan biaya pengobatan sebesar Rp. 3 triliun per tahun untuk pasien hipertensi (Wantoro, et.al., 2021).

Hipertensi adalah masalah serius dalam kesehatan masyarakat yang memerlukan perhatian khusus. Oleh karena itu, deteksi dini dan pencegahan hipertensi sangat penting. Tekanan darah manusia terdiri dari tekanan darah sistolik dan tekanan darah diastolik. Tekanan darah sistolik adalah tekanan dalam pembuluh darah saat jantung berdetak, sedangkan tekanan darah diastolik terjadi di antara ketukan denyut jantung (Roihan, et.al., 2020). Hipertensi terjadi ketika tekanan darah sistolik ≥ 140 mmHg dan tekanan darah diastolik ≥ 90 mmHg. Kondisi ini dapat menghambat suplai oksigen dan nutrisi ke jaringan tubuh yang membutuhkannya, menyebabkan kerusakan parah pada organ dan berpotensi berdampak fatal (Lisiswanti & Dananda, 2016).

Perkembangan teknologi dalam bidang medis terus meningkat. Salah satu contohnya adalah pemanfaatan *machine learning* dalam bidang medis (Roihan, et.al., 2020). *Machine learning* merupakan teknologi berbasis komputer dan matematika yang menggunakan data sebagai media pembelajaran oleh mesin komputer dan menghasilkan prediksi di masa depan (Mustafa & Rahimi Azghadi, 2021). Teknik *machine learning* memiliki potensi untuk membantu dalam prediksi risiko hipertensi dengan tingkat akurasi yang tinggi.

Beberapa penelitian telah dilakukan untuk mendiagnosis hipertensi menggunakan berbagai algoritma, termasuk *Artificial Neural Network* (Purwono, et.al., 2022), *Decision Tree, Random Forest, Gradient Boosting Machine, Gradient Boosting Decision Tree, Logistic Regression*, dan *Linear Discriminant Analysis* (Islam, et.al., 2022). Perbandingan performa dari enam teknik di atas, selain *Artificial Neural Network*, diketahui *Gradient Boosting Decision Tree* memiliki akurasi 90%, presisi 9-%, dan recall 100% (Islam, et.al., 2022). Performa dari *Gradient Boosting Decision Tree* yang baik juga ditunjukkan pada studi yang membandingkan tiga teknik *machine learning*, yaitu *Random Forest, Support Vector Machine*, dan *Gradient Boosting Decision Tree* pada model persetujuan pengajuan kredit (Givari, et.al., 2022). Dalam studi tersebut *Gradient Boosting Decision Tree* memiliki performa tertinggi dengan akurasi 82%, presisi 92%, dan recall 72%.

Kontribusi penelitian ini adalah pengembangan model prediksi risiko hipertensi menggunakan algoritma *Gradient Boosting Decision Tree*. Berbeda dengan model yang telah dilakukan sebelumnya (Purwono, et.al., 2022; Islam, et.al., 2022), model dalam penelitian ini diharapkan dapat memprediksi risiko hipertensi, atau dengan kata lain calon penderita hipertensi. Prediksi dilakukan dengan melihat pola perilaku, gaya hidup, dan status kesehatan seseorang. Kontribusi lain adalah algoritma *Gradient Boosting Decision Tree* mendapatkan perlakuan *hyperparameter tuning*. Perlakuan tersebut bertujuan untuk mengoptimalkan performa *Gradient Boosting Decision Tree* dalam membuat model prediksi risiko hipertensi ini. Terdapat lima metode *hyperparameter tuning* yang diterapkan, yaitu *Random Search, Grid Search, Bayesian Optimization, Gradient-based*, dan *Tree Parzer Estimation*.

JUISI Jurnal Informatika dan Sistem Informasi
Volume 9 Nomor 2
e-ISSN: 2477-5452; p-ISSN: 2460-1306
DOI: *doi.org/10.37715/juisi.v9i2.4403*

2. KAJIAN PUSTAKA

2.1 Studi Terdahulu

Penelitian sebelumnya yang dilakukan oleh Syahroni Damanik dan Lisa Novianti Sitompul pada tahun 2020 membahas tentang hubungan antara gaya hidup dan hipertensi pada lansia. Kesimpulan dari penelitian tersebut adalah bahwa gaya hidup yang tidak sehat, seperti merokok, kelebihan berat badan, kelebihan asupan natrium, lemak, dan kolesterol, dapat menjadi faktor penyebab terjadinya hipertensi (Islam, et.al., 2022).

Selanjutnya, pada tahun 2022, Sheikh Mohammed Shariful Islam dan rekannya melakukan penelitian yang bertujuan untuk memprediksi penyakit hipertensi dan faktor-faktor yang mempengaruhinya menggunakan enam model *machine learning*, termasuk *Decision Tree, Random Forest, Gradient Boosting Machine, XGBoost, Logistic Regression*, dan *Linear Discriminant Analysis*. Hasil penelitian menunjukkan bahwa model *XGBoost, Gradient Boosting Machine, Logistic Regression*, dan *Linear Discriminant Analysis* memiliki tingkat akurasi sebesar 90%. Sedangkan model *Random Forest* memperoleh tingkat akurasi sebesar 89%, dan model *Decision Tree* mencapai tingkat akurasi sebesar 83% (Givari, et.al., 2022).

2.2 Teknologi

Dalam pelaksanaan penelitian ini, digunakan berbagai teknologi yang terkait dengan masalah atau sistem yang menjadi fokus penelitian. Teknologi tersebut digunakan sebagai sumber referensi dan panduan dalam penyusunan penelitian ini, berdasarkan penelitian terdahulu yang telah dilakukan.

2.2.1 XGBoost

XGBoost adalah sebuah perpustakaan (*library*) *machine learning* yang berbasis pada algoritma *Gradient Boosting Decision Tree* yang dikembangkan oleh Tianqi Chen. *XGBoost* memiliki beberapa fitur unik yang membuatnya menjadi salah satu algoritma yang populer dan efektif untuk masalah *machine learning* seperti klasifikasi, regresi, dan *scoring*. Fitur-fitur tersebut meliputi peningkatan akurasi melalui pemrosesan paralel yang lebih baik, optimasi yang lebih optimal melalui penggunaan teknik regulasi, dan peningkatan performa melalui penanganan yang lebih baik terhadap data yang memiliki *skewness* dan *outlier*. *XGBoost* memiliki implementasi yang sangat baik dalam bahasa pemrograman Python dan sering digunakan oleh para data *scientist* dan *engineer machine learning*. Algoritma ini mampu melakukan optimasi hingga 10 kali lebih cepat dibandingkan dengan algoritma *Gradient Boosting Machine* lainnya. Namun, nilai akurasi yang dihasilkan juga sangat bergantung pada parameter-parameter yang digunakan (Givari, et.al., 2022).

2.2.2 Randomized Search

Randomized search merupakan sebuah metode optimasi *hyperparameter* yang digunakan dalam *machine learning*. Pendekatan ini melibatkan pemilihan acak dari himpunan nilai yang mungkin untuk setiap *hyperparameter* yang akan dioptimalkan, dan kemudian melakukan evaluasi model untuk setiap kombinasi acak tersebut. Tujuan dari *randomized search* adalah untuk secara acak mengeksplorasi ruang *hyperparameter* dengan harapan menemukan konfigurasi *hyperparameter* yang menghasilkan performa model yang baik (Givari, et.al., 2022).

Berbeda dengan *grid search* yang mencoba semua kombinasi nilai *hyperparameter* yang mungkin, *randomized search* hanya memilih sejumlah kombinasi secara acak untuk dievaluasi. Keuntungan dari pendekatan ini adalah kemampuannya untuk mengeksplorasi ruang *hyperparameter* secara lebih efisien, terutama ketika ruang parameter sangat besar atau ketika hanya beberapa *hyperparameter* yang memiliki dampak signifikan terhadap performa model (Givari, et.al., 2022).

2.2.3 Tree Parzer Estimator

Tree Parzer Estimator (TPE) merupakan sebuah metode untuk mengoptimasi *hyperparameter* yang menggunakan model probabilistik untuk memperkirakan distribusi posterior *hyperparameter*. Pendekatan ini didasarkan pada teknik Parzen Window, di mana pendekatan probabilistik digunakan untuk menentukan probabilitas nilai-nilai *hyperparameter* yang optimal (Givari, et.al., 2022).

Volume 9 Nomor 2
e-ISSN: 2477-5452; p-ISSN: 2460-1306
DOI: *doi.org/10.37715/juisi.v9i2.4403*

2.2.4 Grid Search

Grid search adalah sebuah metode optimasi *hyperparameter* yang digunakan dalam *machine learning* untuk mencari kombinasi terbaik dari sejumlah *hyperparameter* yang telah ditentukan sebelumnya. Pendekatan ini melibatkan pencarian sistematis melalui ruang parameter dengan menguji setiap kombinasi *hyperparameter* yang mungkin, dan memilih kombinasi yang menghasilkan performa model yang optimal (Alfat, et.al., 2022).

Dalam *grid search*, setiap *hyperparameter* ditentukan dengan sejumlah nilai yang telah ditentukan sebelumnya. Selanjutnya, semua kombinasi nilai-nilai tersebut diuji secara sistematis dengan melatih dan menguji model untuk setiap kombinasi yang ada. Dengan demikian, *grid search* memberikan pemahaman mengenai kombinasi *hyperparameter* mana yang menghasilkan performa terbaik untuk model yang sedang dilatih (Alfat, et.al., 2022).

2.2.5 Gradient Based Optimization

Gradient-based optimization adalah suatu metode optimasi yang memanfaatkan gradien atau turunan parsial dari fungsi objektif untuk melakukan update pada nilai-nilai parameter model secara iteratif dengan tujuan mencapai nilai minimum atau maksimum dari fungsi tersebut. Pendekatan ini banyak digunakan dalam *machine learning* untuk melatih model dengan melakukan pembaruan parameter berdasarkan gradien fungsi kerugian terhadap parameter (Damanik & Sitompul, 2020).

2.2.6 Bayesian Optimization

Bayesian optimization adalah suatu metode optimasi *hyperparameter* yang menggabungkan pendekatan probabilistik dengan model bayesian untuk mencari kombinasi optimal dari *hyperparameter* dalam ruang pencarian yang kompleks. Pendekatan ini memanfaatkan informasi yang dikumpulkan dari iterasi sebelumnya untuk memperbaiki estimasi model dan mengarahkan pencarian ke area yang lebih menjanjikan. Dalam *Bayesian optimization*, digunakan model bayesian untuk memodelkan fungsi objektif yang ingin di optimasi sebagai fungsi probabilistik. Model ini menghasilkan distribusi posterior yang diperbarui setelah setiap iterasi. Pada setiap langkah, model tersebut memberikan estimasi untuk kombinasi *hyperparameter* yang memiliki probabilitas tinggi dalam memberikan performa yang baik (Alfat, et.al., 2022).

3. METODOLOGI PENELITIAN

3.1 Analisis Kebutuhan Model

3.1.1 Analisis Kebutuhan Dataset

Analisis kebutuhan dataset untuk penelitian yang berkaitan dengan pengembangan model prediksi risiko hipertensi menggunakan algoritma *gradient boosting decision tree* yang telah dioptimalkan dilakukan dengan menggunakan metode wawancara dengan seorang ahli sebagai metode seleksi atribut. Ahli tersebut adalah dokter Widyarta, seorang dokter umum yang memiliki pengalaman selama 46 tahun dan telah menjabat sebagai kepala puskesmas, direktur rumah sakit umum Wajo, rumah sakit umum Tana Toraja, serta kepala UPF rumah sakit jiwa.

Selama wawancara, beberapa pertanyaan diajukan kepada ahli tersebut. Pertanyaan pertama adalah apakah risiko hipertensi dapat diprediksi di masa depan melalui gaya hidup, seperti indeks massa tubuh, merokok, aktivitas fisik, kebiasaan mengonsumsi buah, kebiasaan mengonsumsi sayur, dan kebiasaan mengonsumsi alkohol. Pertanyaan selanjutnya adalah apakah risiko hipertensi dapat diprediksi di masa depan melalui faktor usia, jenis kelamin, tingginya kadar kolesterol, riwayat pemeriksaan kolesterol dalam 5 tahun terakhir, riwayat serangan jantung, kondisi keseluruhan tubuh, kondisi mental, kondisi fisik dalam 30 hari terakhir, kesulitan berjalan atau menaiki tangga, riwayat penyakit diabetes, dan riwayat penyakit stroke. Pertanyaan terakhir adalah apakah terdapat atribut lain yang dapat memprediksi risiko hipertensi selain atribut-atribut tersebut.

Berdasarkan hasil wawancara, ditemukan bahwa atribut usia, jenis kelamin, tingginya kadar kolesterol, riwayat pemeriksaan kolesterol dalam 5 tahun terakhir, indeks massa tubuh, merokok, riwayat serangan jantung, aktivitas fisik, kebiasaan mengonsumsi buah, kebiasaan mengonsumsi sayur, kebiasaan mengonsumsi alkohol, kondisi keseluruhan tubuh, kondisi mental, kondisi fisik dalam 30 hari terakhir, kesulitan berjalan atau menaiki tangga, riwayat penyakit diabetes, dan riwayat penyakit stroke diperlukan untuk memprediksi risiko hipertensi. Selain itu, terdapat atribut lain yang juga dapat memprediksi risiko hipertensi, yaitu riwayat hipertensi pada orang tua dan kebiasaan mengonsumsi daging. Namun, karena keterbatasan dataset yang tersedia, atribut riwayat hipertensi pada

Volume 9 Nomor 2
e-ISSN: 2477-5452; p-ISSN: 2460-1306
DOI: *doi.org/10.37715/juisi.v9i2.4403*

orang tua dan kebiasaan mengonsumsi daging tidak akan digunakan dalam pembuatan model prediksi risiko hipertensi.

Dengan demikian, berdasarkan wawancara dengan ahli, dapat disimpulkan bahwa semua atribut dalam dataset akan digunakan untuk membuat model prediksi risiko hipertensi. Namun, atribut riwayat hipertensi pada orang tua dan kebiasaan mengonsumsi daging tidak akan digunakan dalam pembuatan model prediksi risiko hipertensi karena keterbatasan dataset.

3.1.2 Analisis Kebutuhan Hardware

Analisis kebutuhan perangkat keras untuk penelitian yang berkaitan dengan pengembangan model prediksi risiko hipertensi menggunakan algoritma *gradient boosting decision tree* yang telah dioptimalkan bertujuan untuk memastikan pelatihan model *machine learning* dapat berjalan dengan baik. Menurut (K, 2021), berikut adalah spesifikasi minimum komputer yang diperlukan untuk melakukan pelatihan model *machine learning*:

- Spesifikasi CPU yang direkomendasikan setidaknya adalah Intel i3-10100F atau prosesor Intel i3 generasi lainnya dengan seri F. Dalam konteks *machine learning*, CPU (*Central Processing Unit*) berperan dalam melakukan pemodelan dan pembelajaran model. CPU bertugas mengeksekusi instruksi-instruksi yang diberikan oleh kode program, seperti pengolahan data, perhitungan matematis, dan penentuan hasil dari model. Pada proses pelatihan model *machine learning*, CPU harus bekerja secara cepat dan efisien untuk memproses data serta melakukan pembaruan pada model agar dapat belajar dan meningkatkan performa. Oleh karena itu, memiliki spesifikasi CPU yang baik dan kuat dapat mempengaruhi kecepatan dan efisiensi dalam proses pelatihan *machine learning*.
- RAM (*Random Access Memory*) merupakan jenis memori sementara yang digunakan oleh komputer untuk menyimpan data dan instruksi yang sedang diproses. Dalam konteks *machine learning*, RAM berperan dalam menyimpan data *training* dan data *testing* selama proses pemodelan dan pembelajaran. Model *machine learning* sering membutuhkan kapasitas memori yang besar untuk menangani volume data *training* yang besar, sehingga memiliki jumlah RAM yang memadai menjadi faktor penting untuk memastikan kelancaran proses pelatihan. Ketika RAM tidak mencukupi, komputer akan menggunakan *harddisk* sebagai memori virtual, yang dapat melambatkan proses pemodelan dan pembelajaran. Oleh karena itu, memiliki spesifikasi RAM yang baik dan dengan kapasitas yang mencukupi menjadi sangat penting dalam konteks *machine learning* untuk memastikan kelancaran dan kecepatan proses pelatihan. Rekomendasi minimal adalah RAM sebesar 8 GB.
- Media penyimpanan, atau *storage*, digunakan sebagai tempat untuk menyimpan data oleh komputer. Dalam konteks *machine learning*, *storage* digunakan untuk menyimpan data *training*, data *testing*, dan hasil dari proses pemodelan dan pembelajaran. Banyak model *machine learning* membutuhkan ruang penyimpanan yang besar untuk menampung data *training* yang volumin, sehingga memiliki kapasitas *storage* yang memadai menjadi faktor penting dalam melaksanakan proses pelatihan. Jenis penyimpanan yang umum digunakan dalam *machine learning* adalah *harddisk* atau *solid state drive* (SSD), yang memiliki perbedaan dalam kecepatan akses dan kapasitas penyimpanan. Oleh karena itu, spesifikasi penyimpanan yang baik dengan kapasitas yang mencukupi sangat penting dalam konteks *machine learning* untuk memastikan data *training* dan hasil pemodelan dapat disimpan dengan aman. Rekomendasi minimal adalah 128 GB SSD dan 1 TB SSD.
- Kartu grafis, atau *Graphics Cards*, merupakan perangkat yang digunakan untuk memproses dan menampilkan gambar, animasi, dan video pada komputer. Dalam konteks *machine learning*, kartu grafis juga dapat digunakan untuk melakukan pemodelan dan pembelajaran, karena memiliki kemampuan untuk melakukan perhitungan paralel dengan cepat. Beberapa algoritma *machine learning*, seperti *neural network*, memerlukan perhitungan matematis yang kompleks, dan kartu grafis dapat menjalankan perhitungan tersebut lebih efisien daripada CPU. Selain itu, kartu grafis juga memiliki kapasitas memori yang besar, yang memungkinkan penyimpanan data *training* dan data *testing* selama proses pemodelan dan pembelajaran. Oleh karena itu, spesifikasi kartu grafis yang baik dan handal dapat mempengaruhi kecepatan dan efisiensi dalam proses pelatihan model *machine learning*. Rekomendasi minimal adalah NVIDIA GTX 1050.

Volume 9 Nomor 2
e-ISSN: 2477-5452; p-ISSN: 2460-1306
DOI: *doi.org/10.37715/juisi.v9i2.4403*

3.2 Desain

3.2.1 Desain Dataset

Penelitian ini menggunakan dataset tahun 2015 dari Pusat Pengendalian dan Pencegahan Penyakit Amerika Serikat yang didapatkan dari situs Kaggle. Dataset ini terdiri dari beberapa fitur yang dapat digunakan untuk memprediksi hipertensi, seperti usia, jenis kelamin, keberadaan kolesterol tinggi, riwayat cek kolesterol dalam 5 tahun terakhir, indeks massa tubuh, kebiasaan merokok, riwayat serangan jantung, aktivitas fisik, kebiasaan mengonsumsi buah, sayur, dan alkohol, kondisi keseluruhan tubuh, kondisi mental, kondisi fisik dalam 30 hari terakhir, kesulitan berjalan atau menaiki tangga, riwayat diabetes, dan riwayat stroke.

Seperti yang terlihat pada Tabel 1, dataset ini terdiri dari 18 fitur dan 70.693 baris data, di mana 39.832 baris merupakan data penderita hipertensi. Dari 18 fitur tersebut, satu fitur akan menjadi kelas (target) dan 17 fitur lainnya akan menjadi atribut (*input*). Fitur yang akan menjadi kelas adalah hipertensi, sehingga terdapat 17 fitur yang akan digunakan dalam penelitian ini, termasuk usia, jenis kelamin, keberadaan kolesterol tinggi, riwayat cek kolesterol dalam 5 tahun terakhir, indeks massa tubuh, kebiasaan merokok, riwayat serangan jantung, aktivitas fisik, kebiasaan mengonsumsi buah, sayur, dan alkohol, kondisi keseluruhan tubuh, kondisi mental, kondisi fisik dalam 30 hari terakhir, kesulitan berjalan atau menaiki tangga, riwayat diabetes, dan riwayat stroke.

Sebelum membangun model *machine learning*, dataset perlu melalui tahap *preprocessing*. Namun, dataset yang telah diperoleh sudah melewati proses pembersihan, manipulasi, dan memiliki kelas yang seimbang, sehingga dataset tersebut siap digunakan.

Tabel 1. Tampilan Dataset

Atribut	Data ke-1	Data ke-2	Data ke-3
Age	4.0	12.0	13.0
Sex	1.0	1.0	1.0
HighChol	0.0	1.0	0.0
CholCheck	1.0	1.0	1.0
BMI	26.0	26.0	26.0
Smoker	0.0	1.0	0.0
HeartDiseaseorAttack	0.0	0.0	0.0
PhysActivity	1.0	0.0	1.0
Fruits	0.0	1.0	1.0

Untuk mengembangkan model prediksi risiko hipertensi menggunakan algoritma *gradient boosting decision tree* yang dioptimalkan, berikut adalah beberapa langkah yang dapat diikuti:

- Seleksi fitur: Melakukan seleksi fitur yang paling relevan dan penting untuk model tersebut. Hal ini dapat dilakukan dengan menggunakan metode wawancara dengan ahli, di mana ahli akan membantu dalam memilih fitur-fitur yang memiliki kontribusi terbesar dalam memprediksi risiko hipertensi.
- Pembagian data: Membagi data menjadi dua bagian, yaitu data latih dan data uji. Data latih akan digunakan untuk melatih model, sedangkan data uji akan digunakan untuk menguji performa model dan melakukan validasi. Pembagian ini penting untuk memastikan bahwa model dapat menggeneralisasi dengan baik pada data yang belum pernah dilihat sebelumnya.

Dengan mengikuti langkah-langkah ini, akan memungkinkan pengembangan model prediksi risiko hipertensi yang lebih efektif dan akurat.

3.2.2 Desain Model Machine Learning

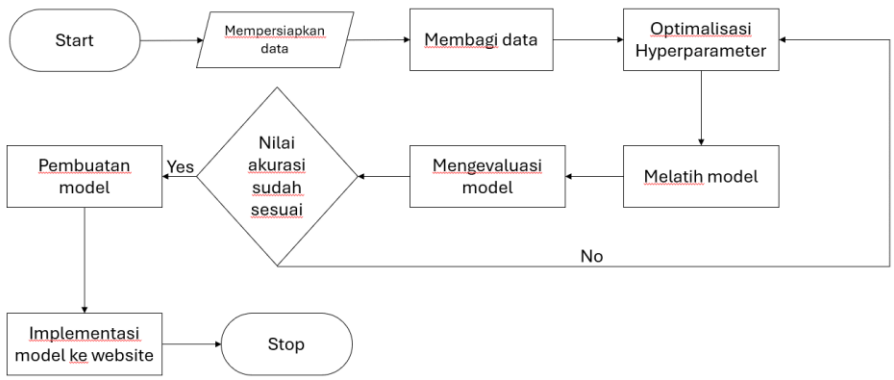
Berikut adalah langkah-langkah yang dapat dilakukan untuk membuat model *machine learning* dalam mengembangkan prediksi risiko hipertensi menggunakan algoritma *gradient boosting decision tree* yang telah dioptimalkan seperti pada Gambar 1:

- Persiapan Data: Mengumpulkan data yang relevan untuk analisis prediksi risiko hipertensi. Dalam tahap ini, data yang diperlukan untuk melatih dan menguji model dikumpulkan dan dipersiapkan agar siap digunakan dalam proses pengembangan model.
- Pembagian Data: Data yang telah disiapkan kemudian dibagi menjadi dua kelompok, yaitu data pelatihan

JUISI Jurnal Informatika dan Sistem Informasi
Volume 9 Nomor 2
e-ISSN: 2477-5452; p-ISSN: 2460-1306
DOI: *doi.org/10.37715/juisi.v9i2.4403*

(*training* data) dan data pengujian (testing data). Data pelatihan digunakan untuk melatih model, sedangkan data pengujian digunakan untuk menguji kinerja dan evaluasi model yang telah dikembangkan.

- Optimalisasi *Hyperparameter*: Algoritma *gradient boosting decision tree* memiliki beberapa *hyperparameter* yang dapat diatur untuk meningkatkan kinerja model. Pada tahap ini, teknik seperti *randomized search* dapat digunakan untuk mencari kombinasi *hyperparameter* yang optimal untuk model yang sedang dikembangkan.
- Pelatihan Model: Menggunakan algoritma *gradient boosting decision tree* yang telah dioptimalkan untuk melatih model prediksi risiko hipertensi. Dalam tahap ini, model akan mempelajari pola-pola yang terdapat dalam data pelatihan dan berusaha untuk memprediksi risiko hipertensi berdasarkan fitur-fitur yang relevan.
- Evaluasi Model: Melakukan evaluasi model dengan menggunakan metrik-metrik seperti akurasi, presisi, dan *recall*. Evaluasi ini membantu untuk memahami sejauh mana model mampu melakukan prediksi dengan benar dan memberikan *insight* tentang kinerja model yang telah dikembangkan.
- Implementasi Model: Setelah model dianggap memadai, model tersebut dapat diimplementasikan dalam website atau platform lainnya agar dapat digunakan dengan mudah oleh pengguna. Dengan demikian, hasil prediksi risiko hipertensi dapat diakses dan dimanfaatkan dengan lebih praktis dan efisien.



Gambar 1. Diagram Alir Pembuatan Model

4. IMPLEMENTASI MODEL PREDIKSI RISIKO HIPERTENSI

Sebelum digunakan untuk membuat model prediksi risiko hipertensi, dilakukan *hyperparameter* tuning pada algoritma *Gradient Boosting Decision Tree* untuk menentukan nilai-nilai parameter yang dapat menghasilkan model yang optimal. Tabel 2 adalah nilai-nilai parameter algoritma *Gradient Boosting Decision Tree* yang dihasilkan oleh lima metode *hyperparameter tuning* yang telah disebutkan pada pendahuluan. Dari tabel tersebut terlihat bahwa nilai-nilai parameter hasil dari metode *Tree Parzer Estimation* membuat model dengan akurasi yang lebih tinggi, yaitu 74,43%.

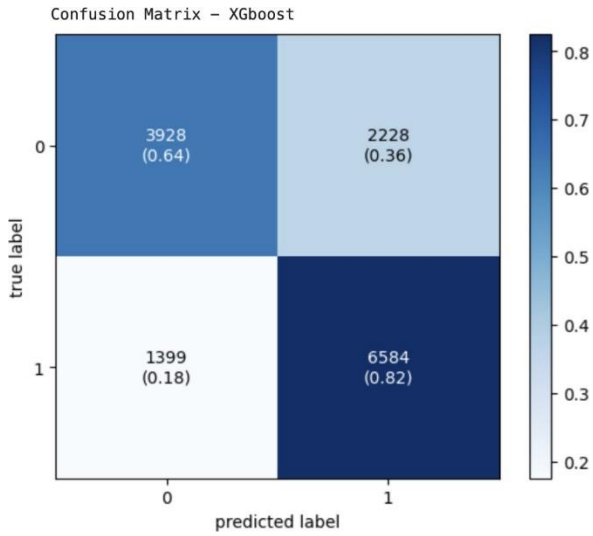
Tabel 2. Hasil Perbandingan Nilai Akurasi

Metode Hyperparameter Tuning	Parameter Terbaik	Nilai Akurasi
Randomized Search	‘n_estimators’: 50, ‘max_depth’: 3, ‘learning_rate’: 0.001, ‘colsample_bytree’: 1	74,10%
Tree Parzer Estimator	'n_estimators':150,'max_depth': 3, 'learning_rate' : 0.10190909090909092, 'colsample_bytree' : 0.8	74,43%
Grid Search	'colsample_bytr ee': 0.5, 'learning_rate': 0.1, 'max_depth': 5, 'n_estimators': 150	74,24%
Gradient Based Optimization	'n_estimators':75, 'max_depth':1, 'learning_rate': 0.9264029810861985, 'colsample_bytree': 0.5457054495536402	74,17%
Bayesian Optimization	'colsample_bytree':0.7962072844310213,	74,20%

Jurnal Informatika dan Sistem Informasi
Volume 9 Nomor 2
e-ISSN: 2477-5452; p-ISSN: 2460-1306
DOI: *doi.org/10.37715/juisi.v9i2.4403*

'learning_rate': 0.02393512381599932,
'max_depth': 3, 'n_estimators': 746

Seperti yang terlihat pada Gambar 2, hasil dari *confusion matrix* menunjukkan *true positive* sebesar 64%, *true negative* sebesar 82%, *false positive* sebesar 36%, dan *false negative* sebesar 18%. Dari nilai tersebut dapat diketahui presisi model prediksi risiko hipertensi memiliki nilai sebesar 74,71% dan *recall* memiliki nilai sebesar 82,47%.



Gambar 2. Confusion Matrix

5. PENGUJIAN DAN PEMBAHASAN

5.1 Pengujian Model

5.1.1 Hasil Pengujian

Pengujian ini dilakukan untuk mengevaluasi akurasi model yang telah diimplementasikan pada website. Pengujian dilakukan dengan menginputkan secara acak 10 data dari dataset melalui website. Kemudian, hasil prediksi dari model dibandingkan dengan label yang ada pada dataset untuk melihat kesesuaian. Pada pengujian pertama, prediksi menyatakan bahwa seseorang tidak terkena hipertensi, sesuai dengan label data pada dataset. Hasil pengujian selanjutnya dapat dilihat pada Tabel 3, yang menampilkan nilai dari fungsi dataset dan prediksi, serta hasil yang diperoleh. Dari hasil pengujian ini, didapatkan nilai akurasi sebesar 40%.

Tabel 3. Pengujian Dataset

Percobaan	1	2	3	4	5	6	7	8	9	10
Age	4.0	13.0	12.0	2.0	8.0	13.0	4.0	2.0	6.0	9.0
Sex	1.0	1.0	0.0	0.0	1.0	1.0	0.0	1.0	0.0	0.0
HighChol	0.0	0.0	1.0	0.0	1.0	1.0	0.0	0.0	1.0	1.0
CholCheck	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
BMI	26.0	26.0	37.0	26.0	39.0	42.0	26.0	21.0	37.0	25.0
Smoker	0.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0
HeartDiseaseorAttack	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
PhysActivity	1.0	1.0	1.0	1.0	1.0	0.0	1.0	1.0	0.0	1.0

Volume 9 Nomor 2
e-ISSN: 2477-5452; p-ISSN: 2460-1306
DOI: *doi.org/10.37715/juisi.v9i2.4403*

Fruits	0.0	1.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	1.0
Veggies	1.0	1.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0
HvyAlcoholConsump	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
GenHlth	3.0	1.0	2.0	1.0	3.0	4.0	1.0	1.0	4.0	2.0
MentHlth	5.0	0.0	0.0	4.0	0.0	0.0	0.0	0.0	0.0	0.0
PhysHlth	30.0	10.0	0.0	0.0	10.0	0.0	0.0	2.0	0.0	0.0
DiffWalk	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0
Stroke	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Diabetes	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0
Hasil	0.0	0.0	1.0	0.0	1.0	1.0	0.0	0.0	1.0	1.0
Prediksi	1.0	0.0	0.0	1.0	0.0	1.0	1.0	0.0	0.0	1.0

5.1.2 Pembahasan

Dari hasil pengujian ini, diperoleh akurasi model sebesar 40%. Fungsi pada dataset telah sesuai dengan variabel pemicu hipertensi yang diperoleh dari wawancara dengan dokter Widyarta. Dataset telah melalui proses pembersihan, augmentasi, memiliki kelas yang seimbang, dan siap digunakan. Selain itu, model juga memiliki nilai akurasi, presisi, dan *recall* yang baik. Namun, hasil ini bisa dipengaruhi oleh variasi yang mungkin ada dalam subset data yang digunakan. Dengan menggunakan lebih banyak data, dapat mengurangi efek variasi dan mendapatkan hasil yang lebih stabil.

6. KESIMPULAN DAN SARAN

6.1 Kesimpulan

Model prediksi risiko hipertensi menggunakan algoritma *gradient boosting decision tree* yang dioptimalkan dengan metode *tree parzer estimator* mencapai akurasi sebesar 74,43%. Namun, berdasarkan pengujian dengan 10 data acak dari dataset yang dimasukkan melalui website, model hanya memperoleh akurasi sebesar 40%.

6.2 Saran

Pengembangan model prediksi risiko hipertensi menggunakan algoritma *gradient boosting decision tree* yang dioptimalkan memiliki potensi untuk mendeteksi dini dan mencegah hipertensi, serta mengurangi risiko komplikasi kesehatan yang terkait. Namun, perlu dilakukan perbaikan karena pengujian dengan data acak hanya menghasilkan akurasi sebesar 40%. Untuk pengujian di masa depan, disarankan untuk menggunakan lebih banyak data acak. Dengan meningkatnya jumlah data yang digunakan dalam pengujian, variabilitas dalam hasil pengujian dapat dikurangi. Penggunaan sedikit data dalam pengujian dapat menghasilkan hasil yang sangat dipengaruhi oleh variasi yang mungkin ada dalam subset tersebut. Dengan memperluas jumlah data, efek variasi tersebut dapat diminimalkan dan hasil yang lebih stabil dapat diperoleh. Selain itu, model prediksi risiko hipertensi dengan akurasi 74,43% juga perlu ditingkatkan. Berdasarkan wawancara dengan dokter Widyarta, beberapa fitur dalam dataset seperti riwayat hipertensi pada orang tua dan kebiasaan konsumsi daging dirasa kurang. Oleh karena itu, saran untuk penelitian masa depan adalah menggunakan dataset dengan lebih banyak data dan fitur yang lebih lengkap, sehingga akurasi model prediksi risiko hipertensi dapat ditingkatkan.

Daftar Pustaka

Alfat, L., Hermawan, H., Rustandiputri, A., Inzhagi, R., & Tandjilal, R. (2022). Prediksi Saham PT. Aneka Tambang Tbk. dengan K-Nearest Neighbors. JSAI (Journal Scientific and Applied Informatics), 5(3), 236-243.

Volume 9 Nomor 2
e-ISSN: 2477-5452; p-ISSN: 2460-1306
DOI: *doi.org/10.37715/juisi.v9i2.4403*

Damanik, S., & Sitompul, L. N. (2020). Hubungan Gaya Hidup dengan Hipertensi Pada Lansia di Klinik Tutun Sehati Tahun 2019. Nursing Arts, 14(1), 30-36.

Givari, M. R., Sulaeman, M. R., & Umaidah, Y. (2022). Perbandingan Algoritma SVM, Random Forest Dan XGBoost Untuk Penentuan Persetujuan Pengajuan Kredit. Nuansa Informatika, 16(1), 141-149.

Islam, S. M. S., Talukder, A., Awal, M. A., Siddiqui, M. M. U., Ahamad, M. M., Ahammed, B., ... & Maddison, R. (2022). Machine learning approaches for predicting hypertension and its associated factors using population-level data from three south asian countries. Frontiers in Cardiovascular Medicine, 9, 839379.

K, B. (2021, May 17). Best PC Builds For Deep Learning In Every Budget Ranges. Diambil kembali dari Medium: https://towardsdatascience.com/best-pc-builds-for-deep-learning-in-every-budget-ranges-3e83d1351a8

Lisiswanti, R., & Dananda, D. N. A. (2016). Upaya pencegahan hipertensi. Jurnal Majority, 5(3), 50-54.

Mustafa, A., & Rahimi Azghadi, M. (2021). Automated machine learning for healthcare and clinical notes analysis. Computers, 10(2), 24.

Pratama, I. B. A., Fathnin, F. H., & Budiono, I. (2020). Analisis Faktor yang Mempengaruhi Hipertensi di Wilayah Kerja Puskesmas Kedungmundu. In Prosiding Seminar Nasional Pascasarjana (PROSNAMPAS) (Vol. 3, No. 1, pp. 408-413).

Purwono, P., Dewi, P., Wibisono, S. K., & Dewa, B. P. (2022). Model Prediksi Otomatis Jenis Penyakit Hipertensi dengan Pemanfaatan Algoritma Machine Learning Artificial Neural Network. Insect (Informatics and Security): Jurnal Teknik Informatika, 7(2), 82-90.

Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang. Jurnal Khatulistiwa Informatika, 5(1), 490845.

Wantoro, A., Syarif, A., Berawi, K. N., Muludi, K., Sulistiyanti, S. R., & Sutyarso, S. (2021). Implementasi Metode Pembobotan Berbasis Aturan Dan Metode Profile Matching Pada Sistem Pakar Medis Untuk Prediksi Risiko Hipertensi. Jurnal Teknoinfo, 15(2), 134-145.