

Penerapan Algoritma Decision Tree Untuk Identifikasi Pola Kejadian Hipertensi Di Puskesmas Jayengan Surakarta

Bherlia Wahyu Pranitis¹⁾, Sri Wulandari, S.Si., M.M., M.Kom²⁾, Rizka Licia, S.KM., M.PH³⁾

^{1,2,3}Politeknik Indonusa Surakarta

^{1,2,3} JL. KH. Samanhudi No.31, Bumi Laweyan, Surakarta

¹f22092@poltekindonusa.ac.id, ²sriwulandari@poltekindonusa.ac.id, ³rizkalicia@poltekindonusa.ac.id,

Abstrak

Hipertensi merupakan salah satu masalah kesehatan utama yang terus mengalami peningkatan, termasuk di tingkat pelayanan kesehatan dasar seperti puskesmas. Penelitian ini bertujuan untuk mengidentifikasi pola kejadian hipertensi berdasarkan faktor risiko menggunakan *algoritma Decision Tree* C4.5. Metode yang digunakan adalah deskriptif kualitatif dengan pendekatan studi dokumentasi terhadap data rekam medis pasien hipertensi di Puskesmas Jayengan periode Januari-Desember 2024. Variabel yang dianalisis meliputi usia, jenis kelamin, indeks massa tubuh (IMT), dan riwayat diabetes melitus, dengan derajat hipertensi sebagai variabel dependen. Proses analisis dilakukan menggunakan metode *Knowledge Discovery in Database* (KDD) dan perangkat lunak WEKA. Hasil penelitian menunjukkan bahwa faktor yang paling berpengaruh terhadap kejadian hipertensi adalah IMT, diikuti oleh riwayat diabetes melitus dan usia. Model *Decision Tree* menghasilkan beberapa aturan (*rule*) yang menggambarkan pola hubungan antarvariabel. Namun, hasil evaluasi menunjukkan performa model masih rendah dengan akurasi sebesar 40,32%. Meskipun demikian, model mampu memberikan interpretasi yang mudah dipahami dalam mengidentifikasi faktor risiko hipertensi. Penelitian ini menunjukkan bahwa algoritma *Decision Tree* dapat digunakan sebagai alat bantu analisis pola, namun diperlukan peningkatan kualitas data dan metode untuk memperoleh hasil yang lebih optimal.

Kata kunci: Hipertensi, Data Mining, Decision Tree, C4.5, Rekam Medis, Faktor Risiko

Abstract

Hypertension is a major health problem that continues to increase, including at the primary healthcare level, such as community health centers. This study aims to identify patterns of hypertension incidence based on risk factors using the Decision Tree C4.5 algorithm. The method used was a descriptive qualitative approach with a documentation study approach on medical records of hypertension patients at the Jayengan Community Health Center from January to December 2024. The variables analyzed included age, gender, body mass index (BMI), and history of diabetes mellitus, with the degree of hypertension as the dependent variable. Process analysis was conducted using the Knowledge Discovery in Database (KDD) method and WEKA software. The results showed that the most influential factor in the incidence of hypertension was BMI, followed by history of diabetes mellitus and age. The Decision Tree model generated several rules describing the relationship patterns between variables. However, evaluation results showed that the model's performance was still low, with an accuracy of 40.32%. Nevertheless, the model was able to provide easy-to-understand interpretations in identifying hypertension risk factors. This study demonstrates that the Decision Tree algorithm can be used as a pattern analysis tool, but improvements in data quality and methods are needed to obtain more optimal results.

Keywords: Hypertension, Data Mining, Decision Tree, C4.5, Medical Records, Risk Factors

1. PENDAHULUAN

Hipertensi masih menjadi salah satu permasalahan kesehatan utama di masyarakat. Kondisi ini tergolong serius karena berpotensi menyebabkan berbagai komplikasi, seperti stroke, penyakit jantung koroner, dan gangguan fungsi ginjal (Sari, 2024).

Menurut *World Health Organization*, kasus hipertensi pada orang dewasa meningkat hampir dua kali lipat dari 650 juta pada tahun 1990 menjadi 1,3 miliar pada tahun 2019. Pada tahun 2024, jumlah tersebut diperkirakan mencapai 1,4 miliar pada usia 30-79 tahun, dengan 44% penderita tidak menyadari dirinya mengalami hipertensi (WHO, 2025).

Berdasarkan Survei Kesehatan Indonesia (SKI) tahun 2023, prevalensi hipertensi pada penduduk usia ≥ 18 tahun mencapai 34,5%. Angka ini menunjukkan bahwa kejadian hipertensi belum mengalami penurunan yang signifikan dibandingkan dengan hasil Riset Kesehatan Dasar (Riskesdas) tahun 2018 yang melaporkan prevalensi sebesar 34,1% (Lydia et al., 2023).

Di Provinsi Jawa Tengah, kasus Penyakit Tidak Menular (PTM), khususnya hipertensi, masih tergolong tinggi dan menjadi perhatian utama di sektor kesehatan. Data Profil Dinas Kesehatan Jawa Tengah tahun 2024 menunjukkan bahwa hipertensi merupakan kasus terbanyak di antara PTM dengan proporsi sebesar 73,1%. Diperkirakan sebanyak 7.985.534 penduduk usia >15 tahun mengalami hipertensi, atau sekitar 26,9% dari total populasi pada kelompok usia tersebut. Dari jumlah tersebut, sekitar 84,6% telah mendapatkan pelayanan kesehatan. Meskipun cakupan pelayanan cukup luas, tingginya angka kejadian menunjukkan bahwa upaya pencegahan belum berjalan secara optimal. Oleh karena itu, diperlukan strategi yang lebih komprehensif dan berbasis data, terutama pada tingkat pelayanan dasar seperti puskesmas (Dinkes Jawa Tengah, 2024).

Kasus hipertensi di Kota Surakarta mengalami peningkatan tajam dan menjadi penyakit terbanyak di puskesmas wilayah Surakarta. Pada tahun 2023 kasus hipertensi tercatat sebanyak 67.355 kasus (Dinkes Surakarta, 2023). Sedangkan pada tahun 2024, kasus hipertensi meningkat menjadi 134.157 kasus, atau hampir dua kali lipat dibandingkan tahun sebelumnya (Dinkes Surakarta, 2024). Kondisi tersebut menunjukkan bahwa hipertensi mendominasi kasus penyakit di

wilayah tersebut serta mengindikasikan bahwa upaya pengendalian penyakit tidak menular di Kota Surakarta belum optimal, khususnya dalam efektivitas intervensi dan perubahan perilaku masyarakat.

Sejalan dengan kondisi tersebut, Puskesmas Jayengan sebagai salah satu fasilitas pelayanan kesehatan di Kota Surakarta, khususnya Kecamatan Serengan, juga menghadapi permasalahan yang sama. Kasus hipertensi terus mengalami peningkatan dari tahun ke tahun. Berdasarkan studi pendahuluan pada Oktober 2025, jumlah kunjungan pasien hipertensi tercatat sebanyak 9.702 kasus pada tahun 2024 dan meningkat menjadi 10.862 kasus pada tahun 2025. Berbagai upaya promotif dan preventif telah dilakukan, seperti sosialisasi, penyuluhan kesehatan, senam rutin, serta pelayanan posyandu lansia. Namun demikian, meskipun program tersebut telah dilaksanakan secara berkelanjutan, angka kejadian hipertensi masih terus meningkat.

Kejadian hipertensi dipengaruhi oleh berbagai faktor risiko, seperti riwayat keluarga, kebiasaan merokok, asupan garam tinggi, pola makan tidak sehat, kurangnya aktivitas fisik, serta kondisi metabolik seperti diabetes dan dislipidemia (Kemenkes No. HK.01.07/Menkes/4634, 2021). Beragamnya faktor tersebut menyebabkan pola kejadian hipertensi menjadi kompleks, sehingga diperlukan analisis data yang tepat untuk mengidentifikasinya. Namun, pemanfaatan data rekam medis di tingkat pelayanan dasar untuk mengkaji pola tersebut masih terbatas.

Penelitian yang menggunakan metode Random Forest dalam klasifikasi hipertensi menunjukkan kinerja yang sangat baik, dengan tingkat akurasi sebesar 98% pada data training dan 95% pada data testing. Namun, penelitian tersebut lebih menitikberatkan pada performa model klasifikasi, sehingga belum mengkaji secara mendalam hubungan antarvariabel yang memengaruhi kejadian hipertensi (Novianti et al., 2024). Penerapan algoritma Naïve Bayes dalam klasifikasi hipertensi di Rumah Sakit Undata Palu juga menghasilkan akurasi yang tinggi, yaitu sebesar 98,38%. Meskipun demikian, penelitian ini masih terbatas pada lingkungan rumah sakit, sehingga belum mampu menggambarkan variasi faktor risiko yang mungkin terjadi pada tingkat pelayanan kesehatan dasar seperti puskesmas (Dwiyanas Silolongan et al., 2025). Sementara itu,

metode Support Vector Machine dengan pendekatan Grid Search dan Genetic Algorithm (GA) difokuskan pada optimasi parameter model untuk memperoleh akurasi terbaik sebesar 89,22%. Namun, pendekatan tersebut belum mencakup analisis pola kejadian hipertensi berdasarkan karakteristik populasi (Awalullaili et al., 2023).

Berdasarkan beberapa penelitian sebelumnya, sebagian besar studi masih berfokus pada kinerja algoritma klasifikasi dan belum mengkaji pola faktor risiko hipertensi di tingkat pelayanan dasar. Hal ini penting diteliti mengingat meningkatnya kasus hipertensi di Puskesmas Jayengan serta belum optimalnya pemanfaatan data rekam medis untuk analisis pola kejadian. Oleh karena itu, penelitian ini bertujuan menerapkan algoritma Decision Tree C4.5 untuk mengidentifikasi pola hipertensi secara sistematis sebagai dasar perumusan strategi pencegahan yang lebih efektif.

Pemilihan algoritma Decision Tree C4.5 dalam penelitian ini didasarkan pada kemampuannya dalam mengidentifikasi pola serta hubungan antarvariabel secara sistematis dan terstruktur. Algoritma ini mampu mengolah data yang kompleks menjadi model pohon keputusan yang sederhana dan mudah dipahami, dengan kinerja klasifikasi yang baik serta cenderung lebih unggul dibandingkan beberapa metode lainnya. Oleh karena itu, aturan (*rule*) yang dihasilkan dapat dimanfaatkan secara optimal dalam proses analisis dan pengambilan keputusan (Sholikhah et al., 2023). *Decision Tree* C4.5 juga mampu menangani data dengan berbagai jenis atribut, baik numerik maupun kategorik, tanpa memerlukan asumsi statistik yang rumit (Ramadhon et al., 2024). Dengan demikian, penerapan algoritma *Decision Tree* C4.5 dalam penelitian ini diharapkan mampu mengidentifikasi pola kejadian hipertensi secara lebih menyeluruh serta menghasilkan informasi yang mudah dipahami oleh tenaga kesehatan di Puskesmas Jayengan.

2. TINJAUAN PUSTAKA

a. Puskesmas

Puskesmas adalah fasilitas pelayanan kesehatan tingkat pertama yang berperan penting dalam meningkatkan kualitas kesehatan masyarakat melalui berbagai upaya, seperti promosi kesehatan, pencegahan penyakit, pengobatan, serta pemulihan

kesehatan. Selain itu, puskesmas juga dapat memberikan pelayanan paliatif yang bertujuan untuk meningkatkan kualitas hidup pasien dengan penyakit kronis maupun terminal. Dalam pelaksanaannya, puskesmas bertugas menyelenggarakan pelayanan kesehatan secara menyeluruh, terintegrasi, dan berkesinambungan, serta berfungsi dalam pemenuhan kebutuhan kesehatan sepanjang siklus kehidupan, perbaikan determinan kesehatan, hingga pemberdayaan masyarakat melalui berbagai program kesehatan (Permenkes RI Nomor 19 Tahun 2024, 2024).

b. Rekam Medis

Rekam medis merupakan dokumen yang berisi data identitas pasien, hasil pemeriksaan, diagnosis, tindakan, serta pelayanan kesehatan yang telah diberikan kepada pasien. Dokumen ini mencakup aspek administratif dan klinis yang berfungsi sebagai sumber informasi dalam proses pelayanan kesehatan serta sebagai dasar pengambilan keputusan medis (Permenkes RI Nomor 24, 2022). Selain itu, rekam medis juga memiliki peran penting sebagai sumber data penelitian, bahan pendidikan, serta bukti hukum dan dokumentasi pelayanan kesehatan yang lengkap dan berkesinambungan (Siregar, 2024). Dengan demikian, rekam medis menjadi komponen penting dalam mendukung peningkatan mutu pelayanan kesehatan (Prasetyo et al., 2023).

c. Hipertensi

Hipertensi merupakan kondisi peningkatan tekanan darah di atas batas normal, yaitu $\geq 140/90$ mmHg, yang dapat menimbulkan berbagai komplikasi serius seperti penyakit jantung, stroke, dan gangguan ginjal (Kristina et al., 2024). Berdasarkan penyebabnya, hipertensi dibagi menjadi dua jenis, yaitu hipertensi primer yang penyebabnya tidak dapat dipastikan, serta hipertensi sekunder yang terjadi akibat adanya kondisi medis tertentu. (Direktorat P2PTM Kemenkes RI, 2024). Tekanan darah juga diklasifikasikan ke dalam beberapa kategori, mulai dari optimal hingga

hipertensi derajat 3, yang digunakan untuk menentukan tingkat keparahan dan penatalaksanaan yang tepat (Kemenkes No. HK.01.07/Menkes/4634, 2021). Kejadian hipertensi dipengaruhi oleh berbagai faktor risiko, seperti riwayat keluarga, usia, pola makan tidak sehat, obesitas, kebiasaan merokok, konsumsi alkohol, stres, serta kondisi metabolik seperti diabetes dan dislipidemia (Ekasari et al., 2021). Kompleksitas faktor-faktor tersebut menyebabkan pola kejadian hipertensi menjadi beragam dan memerlukan analisis yang tepat untuk mengidentifikasinya.

d. Data Mining

Data mining adalah suatu proses mengolah dan menganalisis data dalam skala besar untuk mengidentifikasi pola, keterkaitan, serta informasi tersembunyi yang dapat dimanfaatkan sebagai pengetahuan baru (Maulani et al., 2023). Dalam penerapannya, data mining memiliki berbagai teknik, seperti klasifikasi, clustering, asosiasi, serta prediksi, yang digunakan sesuai dengan tujuan analisis data (Atika & Priatna, 2020). Dalam penerapannya, data mining memiliki berbagai teknik, seperti klasifikasi, klustering, asosiasi, serta prediksi, yang digunakan sesuai dengan tujuan analisis data.

e. Decision Tree

Decision Tree merupakan salah satu metode dalam data mining yang digunakan untuk membantu proses pengambilan keputusan secara sistematis melalui struktur berbentuk pohon yang terdiri atas akar, cabang, dan daun. Metode ini bekerja dengan membagi data berdasarkan atribut tertentu hingga diperoleh keputusan akhir yang paling optimal (Muhyadi, 2022). Terdapat beberapa jenis algoritma dalam *Decision Tree*, di antaranya CART, ID3, dan C4.5, yang masing-masing memiliki karakteristik dan keunggulan dalam proses klasifikasi data (Arhami & Nasir, 2020).

f. Algoritma C4.5

Algoritma C4.5 merupakan pengembangan dari ID3 yang digunakan untuk membentuk model pohon keputusan dalam proses klasifikasi data.

Algoritma ini bekerja dengan pendekatan top-down melalui proses pemilihan atribut terbaik berdasarkan nilai gain ratio untuk membagi data secara optimal (Utama & Nugroho, 2023). Dalam prosesnya, C4.5 melibatkan perhitungan entropy, information gain, dan gain ratio untuk menentukan atribut yang paling berpengaruh dalam pembentukan pohon keputusan (Buululo, 2020). Keunggulan algoritma C4.5 terletak pada kemampuannya dalam menangani data numerik dan kategorik, mengatasi missing value, serta melakukan pruning untuk meningkatkan akurasi model. Dengan karakteristik tersebut, C4.5 mampu menghasilkan model yang lebih sederhana, mudah dipahami, dan efektif digunakan dalam analisis serta pengambilan keputusan (Ramadhon et al., 2024).

3. METODE PENELITIAN

Penelitian ini merupakan penelitian kualitatif deskriptif dengan pendekatan studi dokumentasi yang bertujuan untuk menggambarkan pola kejadian hipertensi berdasarkan data rekam medis pasien. Data yang digunakan merupakan data sekunder yang diperoleh dari rekam medis pasien hipertensi di Puskesmas Jayengan periode Januari hingga Desember 2024. Variabel yang digunakan meliputi variabel independen, yaitu usia, jenis kelamin, Indeks Massa Tubuh (IMT), dan riwayat diabetes melitus, sedangkan variabel dependen adalah derajat hipertensi. Penentuan sampel dilakukan teknik *purposive sampling*, dan proses serta analisis data dilakukan menggunakan perangkat lunak WEKA (*Waikato Environment for Knowledge Analysis*). WEKA digunakan untuk melakukan tahap klasifikasi dengan algoritma *Decision Tree* C4.5 (J48), mulai dari pembagian data, pembangunan model, hingga evaluasi kinerja model. Dalam penerapannya, proses analisis data tersebut mengikuti tahapan dalam metode *Knowledge Discovery in Database* (KDD), yang meliputi beberapa tahap sebagai berikut:

a. Pemilihan Data

Tahap awal dilakukan dengan mengumpulkan data rekam medis pasien hipertensi di Puskesmas Jayengan selama periode Januari-Desember 2024. Data yang digunakan berupa data sekunder

yang diperoleh melalui penelusuran dokumen pada sistem informasi puskesmas (SIMPUS).

b. Seleksi Data

Menyeleksi data sesuai dengan tujuan penelitian. Variabel yang digunakan meliputi usia, jenis kelamin, indeks massa tubuh (IMT), riwayat diabetes, dan derajat hipertensi.

c. Prapemrosesan Data

Tahap ini dilakukan untuk memastikan bahwa data yang digunakan telah bersih dan siap dianalisis. Proses yang dilakukan meliputi penghapusan data yang tidak sesuai dengan tujuan penelitian, pengecekan serta penghilangan data duplikat, penanganan data yang tidak lengkap (*missing value*) baik dengan pengisian maupun penghapusan, serta identifikasi *outlier* untuk mendeteksi data yang menyimpang dari pola umum.

d. Transformasi Data

Data yang sudah melalui proses pembersihan selanjutnya dikonversi kedalam format yang sesuai dengan kebutuhan analisis.

e. Visualisasi Data

Tahap visualisasi data dilakukan untuk mengetahui gambaran umum dari dataset sebelum dianalisis lebih lanjut. Melalui proses ini, peneliti dapat memahami distribusi setiap variabel, mengidentifikasi pola yang terbentuk, serta melihat hubungan antar faktor yang diduga memengaruhi kejadian hipertensi.

f. Penerapan Algoritma *Decision Tree* (*Data Mining*)

Penelitian ini menerapkan algoritma *Decision Tree* C4.5 untuk proses klasifikasi. Adapun langkah-langkah yang dilakukan meliputi pembagian data menjadi data latih dan data uji, penentuan variabel target serta fitur sebagai *input* dan *output*, penetapan kriteria pemisahan (*splitting*), pembangunan model pohon keputusan, serta proses pemangkasan (*pruning*) untuk menyederhanakan struktur pohon dan meningkatkan kinerja model.

g. Pola (*Knowledge*)

Hasil dari proses klasifikasi berupa pola atau aturan (*rule*) yang menggambarkan faktor-faktor utama yang memengaruhi kejadian hipertensi. Pola tersebut dapat digunakan sebagai dasar untuk

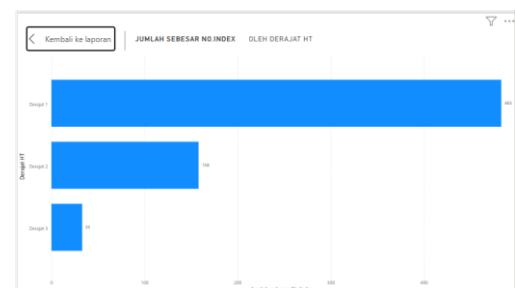
mengidentifikasi dan memahami karakteristik pasien yang memiliki risiko hipertensi.

h. Evaluasi

Tahap akhir penelitian adalah mengevaluasi model *Decision Tree* yang telah dibangun. Evaluasi dilakukan dengan mengukur nilai *akurasi*, *precision*, *recall*, serta menggunakan *confusion matrix* untuk menilai kemampuan model dalam mengidentifikasi pola kejadian hipertensi.

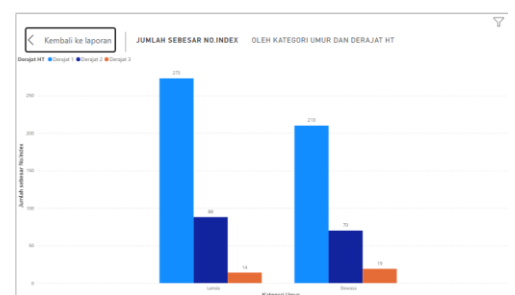
4. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data yang berasal dari rekam medis pasien hipertensi di Puskesmas Jayengan pada periode Januari hingga Desember 2024 dengan total 9.702 data. Setelah melalui proses seleksi, prapemrosesan, dan transformasi, data yang layak dianalisis berjumlah 674. Selanjutnya, dilakukan visualisasi data menggunakan Power BI untuk memperoleh gambaran umum terkait distribusi variabel, pola data, serta hubungan antarvariabel.



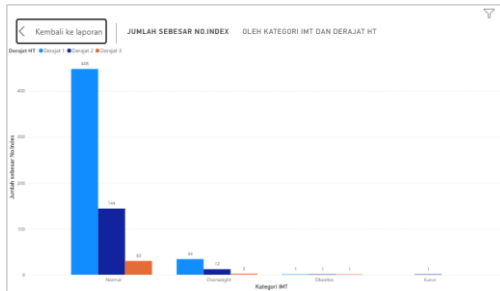
Gambar 1. Tingkat Kejadian Hipertensi

Visualisasi menunjukkan bahwa sebagian besar pasien berada pada hipertensi derajat 1 sebanyak 483 orang, diikuti derajat 2 sebanyak 158 orang, dan paling sedikit pada derajat 3 sebanyak 33 orang.



Gambar 2. Pengaruh Usia terhadap Tingkat Hipertensi

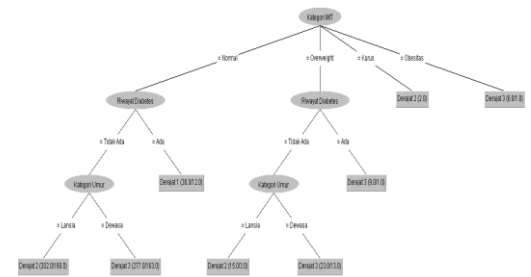
Visualisasi ini menunjukkan distribusi tingkat hipertensi pada kelompok usia lansia dan dewasa. Pada lansia, hipertensi derajat 1 mendominasi, diikuti derajat 2 dan derajat 3 dengan jumlah yang lebih sedikit. Pola yang sama juga terlihat pada kelompok dewasa, namun dengan jumlah kasus yang lebih rendah dibandingkan lansia.



Gambar 3. Pengaruh IMT terhadap Tingkat Hipertensi

Sebagian besar pasien hipertensi berada pada kategori IMT normal, dengan 448 pasien pada derajat 1, 144 pasien pada derajat 2, dan 30 pasien pada derajat 3. Pada kategori *overweight*, jumlahnya lebih sedikit, yaitu 34 pasien pada derajat 1, 12 pasien pada derajat 2, dan 2 pasien pada derajat 3. Sementara itu, kategori IMT obesitas dan kurus memiliki jumlah kasus yang sangat rendah, masing-masing hanya sekitar 1 pasien pada beberapa derajat hipertensi.

Selanjutnya dilakukan proses analisis menggunakan teknik data mining dengan menerapkan algoritma *Decision Tree* C4.5. Pengujian dilakukan menggunakan dataset yang telah melalui proses penyeimbangan kelas (*resampling*). Setelah proses tersebut, jumlah data berubah dari 674 menjadi 672 karena metode *resample* melakukan pengambilan sampel secara acak sehingga tidak seluruh data asli terikut dalam proses tersebut. Dari analisis tersebut terbentuk pohon keputusan (*Decision Tree*) sebagai berikut:



Gambar 4. Pohon Keputusan

Berdasarkan hasil *Decision Tree*, diperoleh beberapa aturan sebagai berikut:

- IF IMT = Kurus THEN Derajat Hipertensi = Derajat 2
- IF IMT = Obesitas THEN Derajat Hipertensi = Derajat 3
- IF IMT = Normal AND Riwayat Diabetes = Ya THEN Derajat Hipertensi = Derajat 1
- IF IMT = Normal AND Riwayat Diabetes = Tidak AND Umur = Lansia THEN Derajat Hipertensi = Derajat 2
- IF IMT = Normal AND Riwayat Diabetes = Tidak AND Umur = Dewasa THEN Derajat Hipertensi = Derajat 3
- IF IMT = *Overweight* AND Riwayat Diabetes = Ya THEN Derajat Hipertensi = Derajat 3
- IF IMT = *Overweight* AND Riwayat Diabetes = Tidak AND Umur = Lansia THEN Derajat Hipertensi = Derajat 2
- IF IMT = *Overweight* AND Riwayat Diabetes = Tidak AND Umur = Dewasa THEN Derajat Hipertensi = Derajat 3

Tabel 1. Faktor Penentu Derajat Hipertensi

Derajat HT	Faktor Penentu
------------	----------------

Derajat 1	a. IMT kategori normal b. Memiliki riwayat diabetes
Derajat 2	a. IMT kategori kurus b. IMT normal, tidak memiliki riwayat diabetes, dan termasuk usia lansia (≥ 60 tahun) c. IMT <i>overweight</i> , tidak memiliki riwayat diabetes, dan termasuk usia lansia (≥ 60 tahun)
Derajat 3	a. IMT dalam kategori obesitas b. IMT normal, tidak memiliki riwayat diabetes, dan termasuk usia dewasa (18–59 tahun) c. IMT <i>overweight</i> dan memiliki riwayat diabetes d. IMT <i>overweight</i> , tidak memiliki riwayat diabetes, dan termasuk usia dewasa (18–59 tahun)

Setelah diperoleh aturan (*rule*) dari model *Decision Tree*, dilakukan evaluasi terhadap model untuk mengetahui kinerja model yang dihasilkan. Berdasarkan hasil pengujian menggunakan metode *10-fold cross validation* dengan algoritma J48 pada WEKA, diperoleh nilai akurasi sebesar 40,32% dari total 672 data. Nilai *kappa* sebesar 0,1049 mengindikasikan bahwa tingkat kesepakatan model masih rendah. Selain itu, nilai *weighted average precision* sebesar 0,413, *recall* sebesar 0,403, dan *F1-score* sebesar 0,376 juga menunjukkan bahwa performa model dalam melakukan klasifikasi belum optimal dan belum merata pada setiap kelas.

Secara teoritis performa algoritma klasifikasi sangat dipengaruhi oleh karakteristik data yang digunakan. Ketidakseimbangan distribusi kelas, jumlah dan kualitas atribut, serta adanya

tumpang tindih pola antar data dapat menyulitkan model dalam membedakan setiap kelas secara akurat, sehingga berdampak pada rendahnya nilai akurasi dan *kappa* (Putra, 2020).

Kondisi tersebut sejalan dengan hasil penelitian ini. Berdasarkan visualisasi distribusi data, jumlah kasus hipertensi tidak merata, di mana derajat 1 mendominasi dengan 483 kasus, sedangkan derajat 2 dan derajat 3 memiliki jumlah yang jauh lebih sedikit. Ketidakseimbangan ini menyebabkan model cenderung memprediksi kelas mayoritas, sehingga kinerja klasifikasi pada kelas minoritas menjadi kurang optimal. Meskipun telah dilakukan proses *resampling* hingga diperoleh 672 data, distribusi kelas masih belum sepenuhnya seimbang, sehingga model tetap menunjukkan kecenderungan bias terhadap kelas mayoritas. Hal ini mengindikasikan bahwa jumlah data yang besar pada suatu kelas tidak selalu menjamin tingginya performa model dalam mengklasifikasikan kelas tersebut.

Hasil penelitian ini menunjukkan bahwa nilai akurasi yang diperoleh masih lebih rendah dibandingkan dengan beberapa penelitian sebelumnya. Metode *Random Forest* dilaporkan mampu mencapai akurasi sebesar 95% pada data testing (Novianti et al., 2024). Selain itu, *Support Vector Machine* juga memiliki performa yang cukup baik dengan akurasi sebesar 89,22% (Awalullaili et al., 2023). Sementara itu, metode *K-Nearest Neighbor* menghasilkan akurasi sebesar 62,60%, yang masih lebih tinggi dibandingkan dengan hasil pada penelitian ini (Soraya et al., 2023). Perbedaan tersebut mengindikasikan bahwa pemilihan algoritma memiliki pengaruh yang signifikan terhadap performa model, di mana metode seperti *Random Forest* dan *Support Vector Machine* cenderung lebih mampu menangani kompleksitas data dibandingkan dengan *Decision Tree*.

Meskipun akurasi yang diperoleh masih rendah, *Decision Tree* tetap mampu menghasilkan aturan IF-THEN yang mudah dipahami. Hal ini menjadi keunggulan dalam interpretasi, sehingga dapat membantu tenaga kesehatan memahami faktor risiko hipertensi.

Namun, keterbatasan performa menunjukkan perlunya perbaikan pada kualitas dan distribusi data, terutama dalam mengatasi ketidakseimbangan kelas, serta penggunaan metode lain agar hasil lebih optimal.

Hasil penelitian menunjukkan bahwa Indeks Massa Tubuh (IMT), riwayat diabetes melitus, dan usia merupakan faktor yang berpengaruh terhadap kejadian hipertensi. IMT menjadi variabel utama dalam model klasifikasi, yang terlihat dari posisinya sebagai akar pada pohon keputusan. Hal ini menunjukkan bahwa IMT memiliki peran dominan dalam menentukan tingkat hipertensi. Hasil ini sejalan dengan penelitian sebelumnya yang menunjukkan adanya hubungan antara indeks massa tubuh (IMT) dengan kejadian hipertensi (Faiha et al., 2025)

Selain itu, riwayat diabetes melitus dan usia juga berkontribusi dalam proses klasifikasi, yang terlihat dari percabangan pohon keputusan. Secara teoritis, kedua faktor tersebut memang memiliki hubungan dengan peningkatan risiko hipertensi. Diabetes melitus dapat merusak pembuluh darah akibat tingginya kadar glukosa dalam darah. Kondisi hiperglikemia memicu pembentukan AGEs yang mempercepat aterosklerosis dan meningkatkan kekakuan pembuluh darah, sehingga dapat menyebabkan peningkatan tekanan darah (Hidayat et al., 2025). Faktor usia juga berperan dalam peningkatan risiko hipertensi. Bertambahnya usia menyebabkan pembuluh arteri menjadi lebih kaku dan kurang elastis, sehingga meningkatkan resistensi pembuluh darah dan memicu kenaikan tekanan darah, terutama tekanan sistolik (Rumaf et al., 2025).

Temuan tersebut diperkuat melalui interpretasi aturan (*rule*) yang dihasilkan oleh model *Decision Tree* dalam mengklasifikasikan setiap derajat hipertensi. Pada hipertensi derajat 1, model mengelompokkan pasien dengan IMT normal yang memiliki riwayat diabetes, yang menunjukkan bahwa diabetes tetap dapat meningkatkan risiko hipertensi meskipun IMT berada pada kategori normal. Hal ini didukung oleh penelitian yang menyatakan bahwa IMT

normal tidak selalu mencerminkan kondisi metabolik yang optimal, karena masih terdapat individu dengan kadar gula darah yang meningkat (Paransi et al., 2025). Pada hipertensi derajat 2, model mengidentifikasi pola pada pasien dengan IMT kurus atau kombinasi IMT normal dan *overweight* tanpa riwayat diabetes, terutama pada kelompok lansia. Hal ini menunjukkan bahwa usia menjadi faktor penting dalam peningkatan risiko hipertensi, meskipun tidak disertai diabetes. Temuan ini sejalan dengan penelitian yang menyatakan bahwa lansia lebih rentan mengalami hipertensi akibat penurunan fungsi jantung dan meningkatnya kekakuan pembuluh darah, yang berdampak pada peningkatan tekanan darah. Selain itu, proses penuaan juga berkontribusi terhadap peningkatan resistensi perifer dan aktivitas saraf simpatik yang memicu terjadinya hipertensi (Riyada et al., 2023). Sementara itu, pada hipertensi derajat 3, model menunjukkan pola yang lebih kompleks, seperti pasien dengan IMT obesitas serta kombinasi IMT *overweight* dengan riwayat diabetes maupun tanpa riwayat diabetes pada usia dewasa. Hal ini mengindikasikan bahwa tingkat keparahan hipertensi tidak hanya dipengaruhi oleh satu faktor, tetapi merupakan hasil interaksi berbagai faktor risiko secara bersamaan. Temuan ini sejalan dengan penelitian yang menyebutkan bahwa hipertensi dipengaruhi oleh kombinasi beberapa faktor, sehingga satu faktor saja tidak cukup untuk menyebabkan terjadinya hipertensi (Fadillah et al., 2023).

Dalam penelitian ini, variabel jenis kelamin tidak berperan dalam klasifikasi hipertensi, yang terlihat dari tidak adanya variabel tersebut dalam model pohon keputusan. Hal ini menunjukkan bahwa jenis kelamin tidak memiliki pengaruh signifikan dibandingkan IMT, riwayat diabetes melitus, dan usia. Temuan ini sejalan dengan penelitian sebelumnya yang juga menyatakan bahwa jenis kelamin bukan merupakan faktor dominan dalam kejadian hipertensi (Mohi et al., 2024)moh.

Hasil penelitian ini menunjukkan bahwa algoritma *Decision Tree* dapat dimanfaatkan sebagai alat bantu dalam

mengidentifikasi faktor risiko hipertensi secara lebih mudah dipahami. Namun, rendahnya performa model mengindikasikan perlunya perbaikan pada kualitas data, terutama dalam mengatasi ketidakseimbangan kelas. Oleh karena itu, penelitian selanjutnya disarankan untuk menerapkan teknik penyeimbangan data yang lebih efektif serta menggabungkan metode klasifikasi lain guna meningkatkan akurasi model.

5. PENUTUP

Berdasarkan hasil penelitian, algoritma *Decision Tree* C4.5 (J48) dapat digunakan untuk mengklasifikasikan kejadian hipertensi, namun performa model masih tergolong rendah. Meskipun demikian, model mampu mengidentifikasi faktor-faktor yang berpengaruh, yaitu Indeks Massa Tubuh (IMT), riwayat diabetes melitus, dan usia, dengan IMT sebagai faktor paling dominan, sedangkan jenis kelamin tidak memiliki pengaruh yang signifikan. Temuan ini menunjukkan bahwa hipertensi dipengaruhi oleh interaksi beberapa faktor risiko utama. Oleh karena itu, penelitian selanjutnya disarankan untuk meningkatkan kualitas dan distribusi data serta menggunakan metode klasifikasi yang lebih optimal.

6. REFERENSI

- Arhami, M., & Nasir, M. (2020). *Data Mining: Algoritma Dan Implementasi* (R. I. Utami (Ed.); 1st Ed.). Andi.
- Atika, Prima Dina, & Priatna, W. (2020). Modul Perkuliahan Data Mining. In Modul Perkuliahan.
- Awalullaili, F. O., Ispriyanti, D., & Widiarihi, T. (2023). Klasifikasi Penyakit Hipertensi Menggunakan Metode Svm Grid Search Dan Svm Genetic Algorithm (GA). *Jurnal Gaussian*, 11(4), 488–498.
<https://doi.org/10.14710/J.Gauss.11.4.488-498>
- Buululo, E. (2020). *Data Mining Untuk Perguruan Tinggi*. CV Budi Utama.
- Dinkes Jawa Tengah. (2024). *Profil Kesehatan Provinsi Jawa Tengah Tahun 2024*. Dinas Kesehatan Jawa Tengah, 1–260.
- Dinkes Surakarta. (2024). *Profil Kesehatan Kota Surakarta 2024*. Dinas Kesehatan Kota Surakarta, 1–207.
www.dinkes.surakarta.go.id
- Dinkes Surakarta. (2023). *Profil Kesehatan Kota Surakarta 2023*. Dinas Kesehatan Kota Surakarta.
- Direktorat P2PTM Kemenkes RI. (2024). *Buku Pedoman Hipertensi 2024*. Buku Pedoman Pengendalian Hipertensi Di Fasilitas Kesehatan Tingkat Pertama, 1–71.
- Dwiyana Silolongan, G., Masse, F. A., & Kusumawati, D. (2025). Klasifikasi Penyakit Hipertensi Menggunakan Algoritma Naive Di Rumah Sakit Undata Palu. *Jurnal Locus Penelitian Dan Pengabdian*, 4(8), 7467–7476.
<https://doi.org/10.58344/Locus.V4i8.4536>
- Ekasari, M. F., Suryati, E. S., Badriah, S., Narendra, S. R., & Amini, F. I. (2021). Hipertensi. In *Hipertensi*.
- Fadillah, N. A., Pujiyanti, N., Sari, A. R., & Hildawati, N. (2023). Pengaruh Perilaku Merokok, Konsumsi Buah Dan Sayur Terhadap Kejadian Hipertensi (Studi Cross Sectional Pada Masyarakat Di Wilayah Kerja Puskesmas Aluh-Aluh Kabupaten Banjar). *Jurnal Kesehatan Masyarakat*, 10(2).
- Faiha, A. T., Wulansari, R., Ernawati, K., Arsyad, M., & Kunci, K. (2025). Hubungan Antara Indeks Massa Tubuh (Imt) Dengan Kejadian Hipertensi Pada Lansia Di Rw 07 Kelurahan Sumur Batu, Jakarta Pusat, Serta Tinjauannya Dalam Pandangan Islam. *Journal Medical*, 3(4).
- Hidayat, R., Triswanti, N., Hutahut, A. F., & Sjahrian, T. (2025). Hubungan Kadar Gula Darah Dengan Tekanan Darah Pada Pasien Diabetes Melitus Tipe 2 Di Rumah Sakit Bintang Amin. *Jurnal Ilmu Pendidikan Nonformal*, 12(5), 1102–1108.
- Kemenkes No. Hk.01.07/Menkes/4634, 1 (2021).
- Kristina, Y., B.B, F., Suweni, K., & Sinaga, E. (2024). Peningkatan Pengetahuan Masyarakat Melalui Pendidikan Kesehatan Dalam Rangka Pencegahan Hipertensi Di Gmakh Jemaat Kolayinuk. *Jurnal Kreativitas Pengabdian Kepada Masyarakat*, 7, 1415–1428.
- Lydia, A., Tiksnadi, A., Indriani, A., Mumpuni, A. S. S., Lukito, A. A., Soenarta, A. A., Aryatama, Tiksnad, B.

- B., Pramono, B. A., Septiar, D., Pramukarso, D. T., Erwinanto, Nuraliyah, F., Fauzie, K., Hane, L. O., Lusiana, L., Hustrini, N. M., Dewi, P. P., Martosuwignyo, P., ... Turana, Y. (2023). Panduan Promotif Dan Preventif Hipertensi 2023. Indonesia Society Of Hypertension Perhimpunan Dokter Hipertensi Indonesia, 50–51.
- Maulani, G., Rito, C. S. H., & Tania, D. K. (2023). Penerapan Data Mining Di Berbagai Bidang.
- Mohi, N. Y., Irwan, & Ahmad, Z. F. (2024). Faktor-Faktor Yang Berhubungan Dengan Kejadian Puskesmas Wonggarasi I The Factors Associated With Hypertension In Elderly In The Working Area Of Puskesmas (Public Helath Center) Wonggarasi I. *Journal Health & Science: Gorontalo Journal Health And Science Community*, 8.
- Muhyadi, M. (2022). Teknik Pengambilan Keputusan. Efisiensi - Kajian Ilmu Administrasi, 3(2). <https://doi.org/10.21831/Efisiensi.V3i2.3796>
- Novianti, N., Alkadri, S. P. A., & Fakhruzi, I. (2024). Klasifikasi Penyakit Hipertensi Menggunakan Metode Random Forest. *Progresif: Jurnal Ilmiah Komputer*, 20(1), 380. <https://doi.org/10.35889/Progresif.V20i1.1663>
- Paransi, P. I., Antu, Y., Otto, S., Nur, M., Yusuf, S., & Helma, N. (2025). Hubungan Indeks Massa Tubuh Dengan Gula Darah Puasa Pada Peserta Prolanis Correlation Between Body Mass Index And Fasting Blood Sugar In Prolanis Participants. *Jurnal Kolaboratif Sains*, 8(1), 481–490. <https://doi.org/10.56338/Jks.V8i1.6739>
- Permenkes Ri Nomor 19 Tahun 2024, 4 7823 (2024).
- Permenkes Ri Nomor 24, Kementerian Kesehatan Republik Indonesia 1 (2022). <https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders>
- Prasetyo, M. Z., Susanto, E., & Wantoro, A. (2023). Sistem Informasi Rekam Medis Pasien Thalasemia (Studi Kasus : Popti Cabang Bandar Lampung). *Jurnal Teknologi Dan Sistem Informasi*, 4(3), 349–355. <https://doi.org/10.33365/Jtsi>
- Putra, J. W. G. (2020). Pengenalan Konsep Pembelajaran Mesin Dan Deep Learning (Vol. 4).
- Ramadhon, R. N., Ogi, A., Agung, A. P., Putra, R., Febrihartina, S. S., & Firdaus, U. (2024). Implementasi Algoritma Decision Tree Untuk Klasifikasi Pelanggan Aktif Atau Tidak Aktif Pada Data Bank. *Karimah Tauhid*, 3(2), 1860–1874. <https://doi.org/10.30997/Karimahtauhid.V3i2.11952>
- Riyada, F., Fauziah, S. A., Liana, N., & Hasni, D. (2023). Faktor Yang Mempengaruhi Terjadinya Resiko Hipertensi Pada Lansia. *Scientific Journal*, 3.
- Rumaf, F., Tutu, C. G., & Talamati, B. H. (2025). Pemetaan Penderita Hipertensi Pada Lansia (Studi Pada Wilayah Kerja Puskesmas Sangkub Kabupaten Bolaang Mongondow Utara). *Jurnal Ilmiah Mahasiswa*, VI, 27–40.
- Sari, N. N. (2024). Hubungan Pengetahuan Dan Sikap Dalam Upaya Pencegahan Komplikasi Hipertensi Pada Lansia. *Jurnal Kesehatan Jompa*, 3(2).
- Sholikhah, M. N., Rahmalia, D., & Pradana, M. S. (2023). Penerapan Algoritma ID3 Dan Algoritma C4 .5. 9, 21–28.
- Siregar, R. A. (2024). Penerapan Permenkes Nomor 24 Tahun 2022 Tentang Rekam Medis Terhadap Efektivitas Pelayanan Kesehatan. *Jihk*, 5(2), 1–11. <https://doi.org/10.46924/Jihk.V5i2.182>
- Soraya, R., Hayati, M. N., & Goejantoro, R. (2023). Klasifikasi Status Hipertensi Pasien UPTD Puskesmas Sempaja, Kota Samarinda Menggunakan Metode K-Nearest Neighbor. *Ekspansional*, 14(2), 67. <https://doi.org/10.30872/Ekspansional.V14i2.1009>
- Utama, M. S., & Nugroho, H. W. (2023). Kajian Algoritma C4 . 5 Dan K-NN Untuk Memprediksi Penduduk Miskin. Darmajaya.
- WHO. (2025). Hypertension. <https://www.who.int/news-room/fact-sheets/detail/hypertension>