

IMPLEMENTASI DECISION TREE UNTUK KLASIFIKASI OBAT PASIEN DIABETES DAN HIPERTENSI

Nindynar Rikatsih, Bagus Dadang Prasetyo, Wahyu Teja Kusuma

Informatika, Institut Teknologi, Sains, dan Kesehatan RS.DR. Soepraoen Kesdam V/BRW

Jl. S. Supriadi No. 22, Sukun, Kec. Sukun, Kota Malang, Indonesia

wtkusuma@itsk-soepraoen.ac.id

ABSTRAK

Diabetes dan hipertensi merupakan dua penyakit kronis yang sering terjadi bersamaan dan dapat memicu komplikasi serius seperti penyakit kardiovaskular, stroke, dan gangguan ginjal. Penanganan kedua penyakit ini seringkali bersifat subjektif, tergantung pada pengalaman tenaga medis, padahal kondisi fisiologis setiap pasien berbeda. Seiring perkembangan teknologi, data mining dapat dimanfaatkan untuk membantu pengambilan keputusan medis berbasis data. Penelitian ini bertujuan untuk mengklasifikasikan jenis obat yang diberikan kepada pasien diabetes dan hipertensi menggunakan algoritma C4.5. Hasil klasifikasi disusun dalam bentuk aturan “*if-then*” dan dievaluasi menggunakan teknik *cross-validation* untuk menghindari *overfitting*. Model yang dihasilkan menunjukkan akurasi tinggi sebesar 94,5% dengan presisi 94,6%, recall 94,5%, dan f-score 94,5%. Temuan ini menunjukkan bahwa pendekatan data mining berpotensi mendukung pengambilan keputusan yang lebih tepat dalam pelayanan medis terhadap pasien diabetes dan hipertensi.

Kata kunci : Diabetes, Hipertensi, Data Mining, C4.5, Cross-validation, Obat

1. PENDAHULUAN

Diabetes adalah kelainan metabolisme kronis dengan kompleks patogenesis dengan karakteristik terjadi gangguan sekresi insulin yang mengakibatkan peningkatan hiperglikemia yang berakibat pada disfungsi metabolisme lemak, karbohidrat, dan juga protein [1]. Disfungsi metabolisme protein dapat mempengaruhi kinerja organ tubuh yang membuat organ bekerja tidak normal dan komplikasi akut seperti ketoasidosis diabetes dan koma hiperosmolar nonketotik jika tidak ditangani dengan tepat [2]. Prevalensi penderita diabetes ditahun 2017 sebesar 425 juta dan akan meningkat ditahun 2045 menjadi 629 juta [3]. peningkatan ini sejalan dengan peningkatan tingkat obesitas dan gaya hidup tidak sehat secara global dan juga faktor eksternal lingkungan seperti transisi nutrisi [4]. diabetes dikategorikan dalam beberapa jenis diantaranya diabetes tipe 1 (T1DM), tipe 2 (T2DM), diabetes melitus gestasional (GDM), dan diabetes yang dipicu kondisi tertentu seperti kelainan dan patologi [1]. T1DM dipicu oleh gagalnya produksi insulin, sedangkan T2DM dipicu oleh resistensi pada insulin yang menyebabkan sel tidak dapat merespon insulin dengan baik [2]. GDM terjadi pada ibu hamil yang tidak memiliki riwayat diabetes, namun hiperglikemia meningkat saat masa kehamilan [5].

Gangguan metabolis seperti diabetes mempunyai prevalensi hipertensi yang tinggi yang memicu penyakit kronis lain seperti penyakit kardiovaskular (CVD), stroke, penyakit ginjal, dan komplikasi mikrovaskuler. Hipertensi dan diabetes memiliki keterhubungan dimana penderita hipertensi seringkali menderita diabetes, sehingga tidak hanya penderita diabetes yang mengalami hipertensi secara global [6]. sekitar 50% hingga 80% hipertensi dialami oleh pasien T2DM yang mendominasi populasi penderita diabetes sebesar 90%-95% dibanding dengan pasien T1DM

yang mengalami hipertensi yang hanya sekitar 30% secara global. Resistensi insulin memegang peran penting dalam patogenesis hipertensi yang ditunjukkan dengan sebagian besar hipertensi dialami oleh pasien T2DM [7]. Insulin berperan menyeimbangkan kadar glukosa dan ketika terjadi gangguan fungsi dan produksi dari insulin hingga menyebabkan kadar gula darah melebihi normal (hiperglikemia), hal ini yang memicu hipertensi [8].

Prevalensi diabetes dinegara berkembang diperkirakan meningkat pada 2030 sebesar 184 juta dari 84 juta menjadi 228 juta dan ditahun 2025 sekitar 639 juta penderita hipertensi [9]. Jenis kelamin dan ras memiliki dampak terhadap hubungan resistensi insulin dengan diabetes yang memicu hipertensi. Prevalensi hipertensi lebih besar terjadi pada laki-laki daripada perempuan hingga batas usia mencapai 64 tahun pada individu non-diabetes. Sebaliknya, perempuan dengan gangguan glukosa memiliki risiko hipertensi lebih tinggi dibandingkan dengan laki-laki yang memiliki gangguan yang sama [6].

Hingga saat ini, kedua penyakit tersebut masih menjadi tantangan dalam penanganannya. Banyak hal yang harus dipertimbangkan memberikan penanganan terhadap kedua penyakit tersebut seperti tekanan darah pasien, usia, kadar kolesterol pasien, dan kadar elektrolit tubuh pasien. Sebagian besar keputusan tindakan medis mengacu pada pengalaman subjektif yang dimiliki oleh tenaga medis atau berpacu terhadap panduan. Hal ini tidak selalu sesuai dengan kondisi pasien secara umum. Perkembangan data yang semakin pesat telah mengambil peran penting dalam berbagai sektor, salah satunya kesehatan. Setiap harinya, industri kesehatan menghasilkan data dengan jumlah yang begitu besar dan telah berhasil memberi banyak perubahan di dunia kesehatan. Hingga saat ini, telah banyak industri kesehatan yang merekomendasikan bermacam sistem informasi medis

dengan tujuan untuk memberikan pelayanan medis terbaik untuk pasien. Pengembangan dan penerapan big data secara berkelanjutan dibidang medis mendorong untuk menerapkan big data pada penelitian medis modern dikalangan peneliti [10].

Big data ialah sekumpulan data dalam jumlah yang sangat besar yang tidak memungkinkan untuk dikelola secara tradisional yang berarti bahwa volume data melebihi kapasitas pengelolaan secara tradisional [11]. Menurut Douglas Laney, perkembangan data terjadi pada 3 dimensi yang berbeda yakni, volume, variety, dan velocity atau 3V [12]. Volume merupakan banyaknya data yang terkumpul dalam satu penyimpanan data, velocity menyatakan kecepatan data terkumpul, dan variety menyatakan variasi jenis data baik yang terorganisir maupun tidak seperti teks, audio, video, gambar, atau file [11]. Akses big data tidak bisa dilakukan secara perangkat tradisional karena kapasitas datanya yang begitu besar sehingga diperlukan suatu teknik untuk mengakses dan mengalisasinya agar big data dapat memberikan informasi dan wawasan yang berguna. Teknik data mining biasa digunakan untuk mengakses dan mengalisis big data hingga menghasilkan suatu informasi dan wawasan baru yang lebih bermanfaat. Data mining ialah suatu proses untuk menggali informasi dari big data dengan mempelajari pola yang terdapat pada kumpulan data yang tidak bisa dilihat secara eksplisit [13]. Data mining dapat digunakan untuk berbagai penugasan seperti klasifikasi juga regresi [14]. Data mining memiliki beberapa metode yang dapat digunakan untuk menggali informasi dari big data, salah satunya decision tree. Decision tree merupakan salah satu teknik dalam data mining yang digunakan untuk tugas klasifikasi dengan representasi pohon keputusan yang mudah dipahami dan diinterpretasikan [15]. keunggulan dari decision tree yakni kemampuannya dalam memvisualisasikan keterkaitan setiap data dalam bentuk pohon keputusan, sehingga mempermudah mengambil keputusan dari data kompleks.

Penelitian sebelumnya [16] menggunakan decision tree untuk memprediksi persediaan obat di apotek dengan menggunakan 30 data dan diperoleh hasil akurasi sebesar 89%. Implementasi decision tree untuk prediksi persediaan obat oleh Dwita [17] di apotek Franch Farma dan penelitiannya menghasilkan 12 rules decision tree dengan akurasi sebesar 80%. Penelitian lainnya juga dilakukan dengan menggunakan teknik decision tree untuk prediksi persediaan obat di rumah sakit PMC [18]. Penelitian ini menghasilkan akurasi yang bagus yakni sebesar 82,34%.

Dari penelitian-penelitian tersebut, akurasi yang dihasilkan decision tree bagus namun masih bisa dioptimalkan akurasinya. Tujuan penelitian ini untuk meningkatkan akurasi dan mengklasifikasikan obat-obatan penderita diabetes dan hipertensi dengan metode decision tree varian C4.5 dan cross-validation untuk mengevaluasi hasil dari model decision tree

yang dibangun dan menghindari terjadinya overfitting pada model. Penelitian ini juga bermaksud memberikan alternatif solusi bagi para dokter menentukan jenis obat yang tepat berdasarkan kondisi pasien melalui model yang dibangun kedepan. Penelitian ini memberikan wawasan baru dalam bidang penerapan data mining di bidang medis dan berguna untuk mendukung proses pengambilan keputusan medis berdasarkan kondisi pasien untuk diberikan tindakan yang tepat.

2. TINJAUAN PUSTAKA

2.1. Diabetes

Diabetes merupakan suatu kelainan metabolik kronis kompleks ditandai dengan tingginya glukosa darah akibat resistensi insulin dimana klasifikasi dan diagnosis penyakit ini telah menjadi fokus utama dalam beberapa dekade [15], [19]. Diabetes dikategorikan dalam beberapa jenis diantaranya diabetes tipe 1 (T1DM), tipe 2 (T2DM), diabetes gestasional (GDM), dan diabetes yang disebabkan faktor tertentu seperti genetik [15]. T1DM disebabkan kegagalan tubuh dalam memproduksi insulin yang cukup bagi tubuh dan sebelumnya dikenal dengan sebutan Insulin-Dependent Diabetes Mellitus (IDDM). T2DM diawali dengan terjadinya resistensi insulin dalam tubuh dimana tubuh tidak dapat merespon insulin dengan baik atau disebut juga Non-Insulin-Dependent Diabetes Mellitus (NIDDM). Gestasional diabetes terjadi pada perempuan yang mengalami masa kehamilan tanpa memiliki riwayat diabetes, namun pada masa kehamilan tingkat glukosa darah meningkat drastis.

2.2. Hipertensi

Hipertensi ialah suatu kondisi dimana tekanan darah sistolik dan diastolik arteri melebihi ambang batas normal yakni diatas 140 mmHg untuk tekanan sistolik dan diatas 90 mmHg untuk tekanan diastolik [20]. Hipertensi yang tidak segera ditangani dengan tepat dapat menyebabkan sindrom klinis multisistem yakni hipertensi akselerasi atau dikenal dengan sebutan maligna dan pendarahan pada otak. Hingga saat ini, hipertensi akselerasi ini jarang terjadi, namun hipertensi juga memicu munculnya penyakit kardiovaskular (CVD) [21].

2.3. Data Mining

Data mining digunakan untuk menggali informasi berharga yang tersimpan secara implisit dari kumpulan data besar. Data mining adalah multidisiplin teknologi yang menggabungkan teknologi berbasis data, statistik, machine learning dan juga pengenalan pola yang memberikan banyak kontribusi positif dalam segala bidang ilmu [13]. Data mining memiliki kemampuan untuk mengeneralisasikan pola potensial yang saling berasosiasi pada data. Data mining memiliki beberapa metode yang dapat digunakan untuk tugas penggalian informasi pada data salah satunya decision tree [13].

2.4. Decision Tree C4.5

Salah satu teknik klasifikasi dalam data mining ialah decision tree. Decision tree menginterpretasikan hasil klasifikasi dalam bentuk visualisasi pohon keputusan yang mudah untuk dipahami [2]. Decision tree memiliki varian seperti ID3 dan C4.5. ID3 merupakan decision tree yang paling sederhana sedangkan C4.5 merupakan versi perbaikan dari ID3 yang hanya mengklasifikasikan data dalam bentuk kategori. Kelebihan dari versi C4.5 diantaranya menggunakan rasio gain dalam interpretasi pohon keputusan untuk menentukan percabangan, kemungkinan untuk pemangkasan cabang pada pohon keputusan untuk menghindari overfitting pada model, kemampuannya untuk menangani atribut fitur berkelanjutan, dan kemampuan menangani data hilang [22]. Untuk menentukan rasio gain, C4.5 menggunakan entropy guna mengukur kerandoman data dengan nilai yang selalu diantara 0 dan 1 [15]. Pohon keputusan yang dibangun C4.5 digunakan untuk membuat suatu aturan atau rules yang digunakan untuk menggali informasi dari kumpulan data. Aturan tersebut menggunakan aturan “if-then” [23]. Decision tree memiliki struktur pohon keputusan yang diawali dengan root yang kemudian membagi atribut data menjadi cabang dan daun atau leaf yang merupakan hasil akhir klasifikasi yang nilainya berupa kelas-kelas data [24].

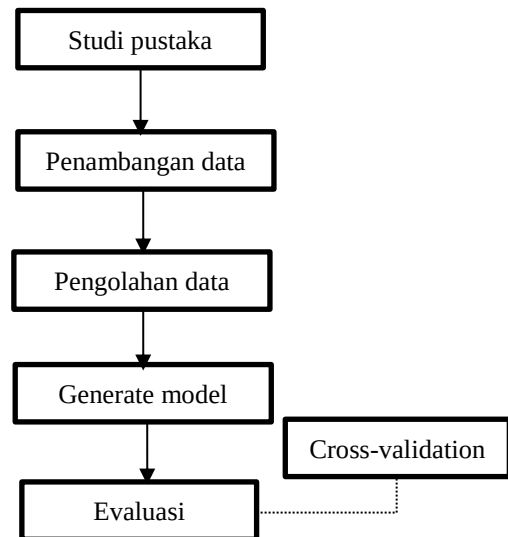
Beberapa penelitian telah menerapkan algoritma data mining untuk berbagai tugas klasifikasi. Penelitian yang dilakukan Faris (2024), menggunakan algoritma decision tree untuk memprediksi pengobatan yang lebih efektif dalam perawatan kucing. Penelitian ini menggunakan 100 data yang dibagi menjadi data latih dan training dengan proporsi 8:2 dan menghasilkan akurasi yang tinggi yakni 90% dengan presisi 83%, dan recall 90%. Penelitian lainnya dilakukan untuk memprediksi persediaan obat di Apotek Az-Zikra Bengkulu menggunakan algoritma C4.5 dan menghasilkan akurasi yang cukup baik yakni 72% [26]. Kedua penelitian ini telah menghasilkan akurasi yang sangat bagus, namun akurasi ini masih dapat ditingkatkan dengan memodifikasi penelitian. Penelitian lain juga menggunakan C4.5 untuk mengklasifikasikan jenis obat yang dilakukan oleh Rika (2021). Penelitian ini menggunakan 200 data yang dibagi menjadi data latih dan data uji dengan proporsi 7:3 dan menghasilkan akurasi sempurna yakni 100. Hasil ini menunjukkan bahwa terjadi overfitting pada model yang dibangun. Untuk mengatasi permasalahan tersebut.

2.5. Cross-Validation

Cross-validation merupakan teknik yang banyak digunakan untuk meminimalisir error pada model dan menghindari terjadinya overfitting pada model. Cross-validation meminimalisir error dengan k-fold cross-validation yang membagi data kedalam subset dengan besaran sama dengan nilai k-fold cross-validation [15]. Secara general, nilai k yang selalu digunakan

penelitian sebesar 10 atau 20, sehingga penelitian ini juga akan menggunakan nilai k-10 untuk menstabilkan model dan meminimalisir terjadinya overfitting pada model. Model yang dibangun, setiap bagian subset yang dibangun secara bergantian digunakan sebagai data uji, sehingga hasil yang diperoleh lebih general dan mewakili seluruh karakteristik data [28].

3. METODE PENELITIAN



Gambar 1. Tahapan Penelitian

Gambar 1 menjelaskan tentang tahapan penelitian yang dilakukan. Tahapan dari penelitian ini terdiri dari studi pustaka, pengumpulan data, pengolahan data, generate model, dan evaluasi. Setiap tahapannya dijelaskan lebih detail pada paragraf berikut:

- Studi pustaka dilakukan untuk mencari gap dan peluang yang dapat dikembangkan dari penelitian terdahulu yang relevan dengan penelitian ini.
- Pengumpulan data dilakukan dengan mencari kumpulan data dari situs publik Kaggle.
- Pengolahan data dilakukan untuk memastikan data yang akan digunakan untuk menguji model telah sesuai dengan kriteria kualitas yang ditetapkan pada penelitian ini yakni datanya lengkap, tidak ada data terduplikasi, dan tipe data yang digunakan telah sesuai.
- Generate model yakni melatih model C4.5 dengan menggunakan data yang telah diolah dan dipastikan telah sesuai dengan kriteria kualitas data penelitian. melatih model C4.5 untuk klasifikasi obat pasien diabetes dan hipertensi dilakukan dengan memanfaatkan tools WEKA.
- Evaluasi merupakan tahap akhir yang dilakukan pada penelitian ini dengan mengevaluasi model menggunakan teknik k-fold cross validation dengan nilai k-10. Evaluasi yang dilakukan mencakup nilai akurasi yang dihasilkan, presisi dan recall dari model, dan yang terakhir f-score.

- Evaluasi akurasi dilakukan untuk mengetahui perbandingan ketepatan model dalam memprediksi data dengan benar terhadap seluruh data baik yang benar maupun tidak. Evaluasi akurasi menggunakan rumus.

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

- Evaluasi presisi dilakukan untuk mengetahui persentase data yang diprediksi positif dengan tepat terhadap semua data yang diprediksi positif. Evaluasinya dilakukan dengan rumus.

$$\text{Presisi} = \frac{TP}{TP+FP} \quad (2)$$

- Evaluasi recall dilakukan untuk mengetahui persentase yang diprediksi dengan tepat terhadap data yang sebenarnya bernilai positif. evaluasinya dilakukan dengan rumus.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

- Evaluasi f-score berguna untuk menyeimbangkan nilai presisi dan recall dan dievaluasi dengan menggunakan rumus.

$$\text{F1-score} = \frac{\text{presisi} \times \text{recall}}{\text{presisi} + \text{recall}} \quad (4)$$

TP ialah true positive yang bermakna nilai prediksi benar tepat sesuai dengan nilai aktualnya yang bernilai positif, TN ialah true negative yang bermakna nilai prediksi benar menyatakan negatif sesuai dengan nilai sebenarnya, FP ialah false positive yang bermakna model memprediksi positif namun sebenarnya nilainya adalah negatif, dan yang terakhir yakni FN ialah false negative yang bermakna model memprediksi negatif namun kenyataannya data bernilai positif.

4. HASIL DAN PEMBAHASAN

Pengumpulan data dari dataset publik Kaggle (<https://www.kaggle.com/datasets/prathamtripathi/drug-classification>) diperoleh sejumlah 200 total data, dimana data tersebut memiliki 6 atribut, 5 diantaranya sebagai atribut input dan 1 sebagai atribut yang berisi kelas masing-masing data yang ditetapkan sebagai target klasifikasi. Atribut pada dataset tersebut meliputi usia, jenis kelamin, tekanan darah, kolesterol, Na_to_Ka, dan jenis obat sebagai kelas data. Adapun penjelasan rincinya disajikan dalam Tabel 1.

Tabel 1. Penjelasan rinci dataset

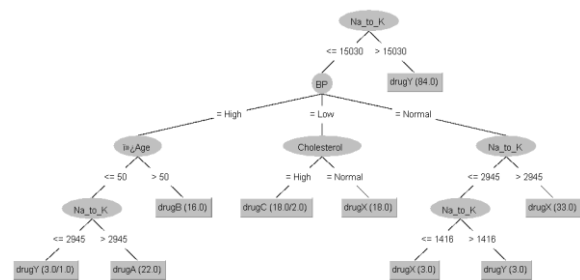
No	Atribut	Nilai	Penjelasan
1	Usia	Usia 15-74	Usia pasien
2	Jenis kelamin	Laki-laki, perempuan	Jenis kelamin pasien
3	Tekanan darah	Rendah, sedang, tinggi	Kadar tekanan darah pasien
4	Kolesterol	Normal, tinggi	Kadar kolesterol pasien
5	Na_to_Ka	734-38247	Perubahan natrium ke potasium
6	Obat	obatY, obatC, obatX, obatA, obatB	Jenis obat yang diperuntukkan pada pasien.

Hasil pengumpulan data yang diperoleh dari Kaggle, kemudian dilakukan olah data dengan

melakukan pengecekan secara menyeluruh terhadap 200 data yang diperoleh. Pengolahan ini dimulai dengan mengecek kelengkapan data. Pengecekan ini dilakukan untuk mengetahui apakah ada data yang kosong atau tidak lengkap. Selanjutnya dilakukan pengecekan kesesuaian data untuk memastikan dataset yang digunakan relevan dengan keseluruhan data. Selanjutnya dilakukan pengecekan apakah pada data terdapat data yang terduplikasi atau tidak, jika ditemukan data yang terduplikasi maka data akan dihapus. Dataset yang digunakan pada penelitian ini hanya ditemukan 1 kesalahan yakni terdapat 1 data di Na_to_Ka pada baris ke 151, datanya tidak relevan dengan seluruh data Na_to_Ka, sehingga dilakukan perbaikan dengan menghitung total rata-rata data Na_to_Ka dan mengganti nilai 87 ke rata-rata kolom yakni 15030. Berdasarkan hasil pengolahan data secara keseluruhan, data telah siap untuk dilakukan pelatihan menggunakan model C4.5.

4.1. Generate Model C4.5

Pelatihan model C4.5 penelitian ini menghasilkan aturan pengambilan keputusan dari interpretasi visualisasi pohon keputusan dari model yang dibangun seperti yang disajikan pada Gambar 2.



Gambar 2. Interpretasi visualisasi pohon keputusan

Dari Gambar 2 tersebut, keputusan dapat diambil berdasarkan aturan yang dibuat dengan rules “if-then” sebagaimana disajikan dalam Tabel 2.

Tabel 2. Aturan “If-Then”

No	Aturan
1	Jika Na_to_K > 15030, Maka drugY
2	Jika Na_to_K ≤ 15030 ^ tekanan darah = Normal ^ Na_to_K > 2945, Maka drugX
3	Jika Na_to_K ≤ 15030 ^ tekanan darah = Normal ^ Na_to_K ≤ 2945 ^ Na_to_K > 1416, Maka drugY
4	Jika Na_to_K ≤ 15030 ^ tekanan darah = Normal ^ Na_to_K ≤ 2945 ^ Na_to_K ≤ 1416, Maka drugX
5	Jika Na_to_K ≤ 15030 ^ tekanan darah = Low ^ Kolesterol = Normal, Maka drugX
6	Jika Na_to_K ≤ 15030 ^ tekanan darah = Low ^ Kolesterol = High, Maka drugC
7	Jika Na_to_K ≤ 15030 ^ tekanan darah = High ^ usia > 50, Maka drugB
8	Jika Na_to_K ≤ 15030 ^ tekanan darah = High ^ usia ≤ 50 ^ Na_to_K > 2945, Maka drugA
9	Jika Na_to_K ≤ 15030 ^ tekanan darah = High ^ usia ≤ 50 ^ Na_to_K ≤ 2945, Maka drugY

Hasil visualisasi menghasilkan 9 aturan yang ketika dijabarkan Tabel 2 sebagai berikut.

- **Aturan 1:** jika jumlah perubahan natrium ke potassium melebihi 15030 mg, maka obat yang direkomendasikan diberikan kepada pasien yakni drugY.
- **Aturan 2:** jika jumlah perubahan natrium ke potassium melebihi 2945 mg atau sama dengan 15030 mg dan tekanan darah pasien dalam ambang batas normal, maka obat yang direkomendasikan diberikan kepada pasien yakni drugX.
- **Aturan 3:** jika jumlah perubahan natrium ke potassium melebihi 1416 mg atau sama dengan 2945 mg dan tekanan darah pasien dalam ambang batas normal, maka obat yang direkomendasikan yakni drugY.
- **Aturan 4:** jika jumlah perubahan natrium ke potassium dibawah atau sama dengan 1416 mg dan tekanan darah pasien dalam ambang batas normal, maka obat yang direkomendasikan yakni drugX.
- **Aturan 5:** jika jumlah perubahan natrium ke potassium dibawah atau sama dengan 15030 mg dan tekanan darah pasien tergolong rendah namun kadar kolesterol pasien masih normal, maka obat yang direkomendasikan yakni drugX.
- **Aturan 6:** jika jumlah perubahan natrium ke potassium dibawah atau sama dengan 15030 mg dan tekanan darah pasien tergolong rendah dan kolesterol pasien tergolong tinggi, maka obat yang direkomendasikan yakni drugC.
- **Aturan 7:** jika jumlah perubahan natrium ke potassium dibawah atau sama dengan 15030 mg dan tekanan darah pasien tergolong tinggi dan usia pasien diatas 50 tahun, maka obat yang direkomendasikan yakni drugB.
- **Aturan 8:** jika jumlah perubahan natrium ke potassium dibawah atau sama dengan 2945 mg dan tekanan darah pasien tergolong tinggi sedangkan usia pasien dibawah atau sama dengan 50 tahun, maka obat yang direkomendasikan yakni drugA.
- **Aturan 9:** jika jumlah perubahan natrium ke potassium diatas 2945 mg dan kurang dari atau sama dengan 15030 mg dan tekanan darah pasien tergolong tinggi sedangkan usia pasien dibawah atau sama dengan 50 tahun, maka obat yang direkomendasikan yakni drugY.

Aturan-aturan yang diperoleh tersebut dapat digunakan sebagai bahan pendukung keputusan dalam memberikan tindakan pemberian obat kepada pasien diabetes dan hipertensi. Untuk memastikan keakuratan aturan yang diperoleh tersebut, k-fold cross-validation dengan nilai k-10 diterapkan.

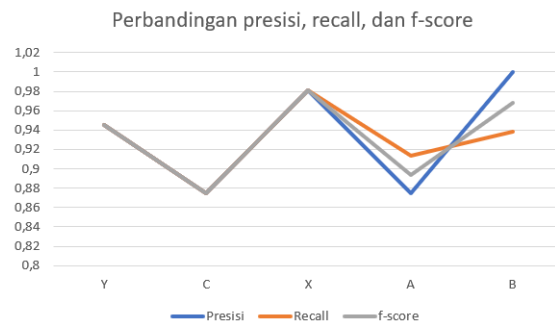
4.2. Evaluasi Model C4.5

Tahap evaluasi terhadap model juga dilakukan dengan memanfaatkan tools WEKA. Penentuan Teknik evaluasi ini dilakukan dengan memilih cross-validation sebagai test options dan menentukan nilai fold sebesar 10. Hasil yang diperoleh pada tahap evaluasi ini disajikan dalam bentuk tabel dan diagram

untuk memudahkan analisis. Tabel 3 menyajikan hasil evaluasi model C4.5 dan Gambar 3 menyajikan diagram hasil evaluasi dari masing-masing kelas klasifikasi.

Tabel 3. Hasil evaluasi model C4.5

Drug	Presisi	Recall	f-score
Y	0,945	0,945	0,945
C	0,875	0,875	0,875
X	0,981	0,981	0,981
A	0,875	0,913	0,894
B	1,000	0,938	0,968
Mean	0,946	0,945	0,945



Gambar 3. Diagram hasil evaluasi

Dapat dilihat pada Gambar 3 bahwa presisi tertinggi sebesar 0,981 pada drugX, Recall tertinggi pada drugX sebesar 0,981, dan f-score tertinggi pada drugX. drugC memiliki presisi, recall, dan f-score terendah yakni 0,875. Rata-rata yang diperoleh pada evaluasi model untuk presisi, recall, dan f-score berturut-turut yakni sebesar 0,946;0,945;0,945 dengan akurasi sebesar 0,945 atau 94,5%.

Akurasi yang dihasilkan sangat bagus, namun tidak bisa sepenuhnya dijadikan acuan karena akurasi dipengaruhi oleh banyak faktor mulai dari kualitas data yang digunakan untuk melatih model atau bisa juga dipengaruhi oleh banyaknya variasi dari data, sehingga presisi, recall, dan f-score juga digunakan untuk evaluasi. Hasil presisi dan recall menyatakan ketepatan model yang digunakan untuk klasifikasi obat.

5. KESIMPULAN DAN SARAN

Penerapan metode C4.5 untuk mengklasifikasikan jenis obat pasien diabetes dan hipertensi memberikan hasil aturan yang diperoleh dari hasil analisa terhadap visualisasi pohon keputusan hasil model yang dibangun. Aturan ini menggunakan rules "If-Then" untuk menggali informasi dari kumpulan data. Hasil dari pelatihan model, kemudian dianalisis menggunakan teknik evaluasi k-fold cross-validation dengan nilai k-10 untuk menghindari terjadinya overfitting pada model sehingga diperoleh akurasi yang sangat bagus yakni 94,5% dan rata-rata presisi 94,6%, recall dan f-score 94,5%.

DAFTAR PUSTAKA

- [1] M. Z. Bandy, A. S. Sameer, and S. Nissar, "Pathophysiology of diabetes: An overview,"

- Avicenna J. Med.*, vol. 10, no. 04, pp. 174–188, 2020, doi: 10.4103/ajm.ajm_53_20.
- [2] Saha et al, “A Review on Diabetes Mellitus : Type1 & Type2,” *World J. Pharm. Pharm. Sci.*, vol. 9, no. 10, pp. 838–850, 2020, doi: 10.20959/wjpps202010-17336.
- [3] I. Smokovski, *Managing diabetes in low income countries: Providing sustainable diabetes care with limited resources*. 2020. doi: 10.1007/9783030514693.
- [4] S. Klein, A. Gastaldelli, H. Yki-Järvinen, and P. E. Scherer, “Why does obesity cause diabetes?,” *Cell Metab.*, vol. 34, no. 1, pp. 11–20, 2022, doi: 10.1016/j.cmet.2021.12.012.
- [5] W. Ye, C. Luo, J. Huang, C. Li, Z. Liu, and F. Liu, “Gestational diabetes mellitus and adverse pregnancy outcomes: systematic review and meta-analysis,” *BMJ*, 2022, doi: 10.1136/bmj-2021-067946.
- [6] G. Jia and J. R. Sowers, “Hypertension in Diabetes: An Update of Basic Mechanisms and Clinical Disease,” *Hypertension*, vol. 78, no. 5, pp. 1197–1205, 2021, doi: 10.1161/HYPERTENSIONAHA.121.17981.
- [7] M. Yildiz, K. Esenboğa, and A. A. Oktay, “Hypertension and diabetes mellitus: Highlights of a complex relationship,” *Curr. Opin. Cardiol.*, vol. 35, no. 4, pp. 397–404, 2020, doi: 10.1097/HCO.0000000000000748.
- [8] U. Galicia-garcia, A. Benito-vicente, S. Jebari, and A. Larrea-sebal, “Costus ignus: Insulin plant and it’s preparations as remedial approach for diabetes mellitus,” *Int. J. Mol. Sci.*, pp. 1–34, 2020.
- [9] I. W. Suryasa, M. Rodríguez-Gámez, and T. Koldoris, “Health and Treatment of Diabetes Mellitus,” *Int. J. Health Sci. (Qassim)*, vol. 5, no. 1, pp. I–V, 2021, doi: 10.53730/IJHS.V5N1.2864.
- [10] J. Yang et al., “Brief introduction of medical database and data mining technology in big data era,” *J. Evid. Based. Med.*, vol. 13, no. 1, pp. 57–69, 2020, doi: 10.1111/jebm.12373.
- [11] J. Wang, Y. Yang, T. Wang, R. Simon Sherratt, and J. Zhang, “Big data service architecture: A survey,” *J. Internet Technol.*, vol. 21, no. 2, pp. 393–405, 2020, doi: 10.3966/160792642020032102008.
- [12] A. Z. Faroukhi, I. El Alaoui, Y. Gahi, and A. Amine, “Big data monetization throughout Big Data Value Chain: a comprehensive review,” *J. Big Data*, vol. 7, no. 1, 2020, doi: 10.1186/s40537-019-0281-5.
- [13] W. T. Wu et al., “Data mining in clinical big data: the frequently used databases, steps, and methodological models,” *Mil. Med. Res.*, vol. 8, no. 1, pp. 1–12, 2021, doi: 10.1186/s40779-021-00338-z.
- [14] D. Papakyriakou and I. S. Barbounakis, “Data Mining Methods: A Review,” *Int. J. Comput. Appl.*, vol. 183, no. 48, pp. 5–19, 2022, doi: 10.5120/ijca2022921884.
- [15] A. M. A. Bahzad Taha Jijo, “Classification Based on Decision Tree Algorithm for Machine Learning,” *J. Appl. Sci. Technol. Trends*, vol. 2, no. 01, pp. 20–28, 2021, doi: 10.38094/jastt20165.
- [16] F. P. Dewanti, S. Setiyowati, and S. Harjanto, “Prediksi Persediaan Obat Untuk Proses Penjualan Menggunakan Metode Decision Tree Pada Apotek,” *J. Teknol. Inf. dan Komun.*, vol. 10, no. 1, 2022, doi: 10.30646/tikomsin.v10i1.604.
- [17] Dwita Elisa Sinaga, Agus Perdana Windarto, and Rizki Alfadillah Nasution, “Analisis Data Mining Algoritma Decision Tree Pada Prediksi Persediaan Obat (Studi Kasus : Apotek Franch Farma),” *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 2, no. 4, pp. 123–131, 2022, doi: 10.30865/klik.v2i4.328.
- [18] S. Dirgantara and E. Asril, “Analisis Data Mining Untuk Prediksi Persediaan Obat Di Rumah Sakit Pmc,” *J-Com (Journal Comput.*, vol. 4, no. 2, pp. 178–184, 2024, doi: 10.33330/j-com.v4i2.3339.
- [19] J. B. Cole and J. C. Florez, “Genetics of diabetes mellitus and diabetes complications,” *Nat. Rev. Nephrol.*, vol. 16, no. 7, pp. 377–390, 2020, doi: 10.1038/s41581-020-0278-5.
- [20] D. Lukitaningtyas and E. A. Cahyono, “HIPERTENSI; ARTIKEL REVIEW,” *Pengemb. Ilmu dan Prakt. Kesehat.*, vol. 13, no. 1, pp. 104–116, 2023.
- [21] G. Gabb, “What is hypertension?,” *Aust. Prescr.*, vol. 43, no. 4, pp. 108–109, 2020, doi: 10.18773/austprescr.2020.025.
- [22] B. A. C. Permana, R. Ahmad, H. Bahtiar, A. Sudianto, and I. Gunawan, “Classification of diabetes disease using decision tree algorithm (C4.5),” *J. Phys. Conf. Ser.*, vol. 1869, no. 1, 2021, doi: 10.1088/1742-6596/1869/1/012082.
- [23] H. Sulistiani and A. A. Aldino, “Decision Tree C4.5 Algorithm for Tuition Aid Grant Program Classification (Case Study: Department of Information System, Universitas Teknokrat Indonesia),” *Edutic - Sci. J. Informatics Educ.*, vol. 7, no. 1, pp. 40–50, 2020, doi: 10.21107/edutic.v7i1.8849.
- [24] P. B. Silva, M. Andrade, and S. Ferreira, “Machine learning applied to road safety modeling: A systematic literature review,” *J. Traffic Transp. Eng. (English Ed.)*, vol. 7, no. 6, pp. 775–790, 2020, doi: 10.1016/j.jtte.2020.07.004.
- [25] F. Hidayat, M. Z. Sarwani, and R. Hariyanto, “Rekomendasi Pengobatan Pada Penyakit Kucing Menggunakan Metode Decision Tree (Studi Kasus : Klinik Drh Panti Absari),” pp. 155–164, 2024.
- [26] F. E. Aziz and D. Abdullah, “Prediksi Persediaan Obat Menggunakan Algoritma Decision Tree (

- C4 . 5) Pada Apotek Az-Zikra Bengkulu,” pp. 784–792, 2024.
- [27] R. Nursyahfitri, A. N. Maharadja, R. A. Farissa, and Y. Umaidah, “Klasifikasi Penentuan Jenis Obat Menggunakan Algoritma Decision Tree,” *J. Inform. Polinema*, vol. 7, no. 3, pp. 53–60, 2021, doi: 10.33795/jip.v7i3.629.
- [28] S. M. Malakouti, M. B. Menhaj, and A. A. Suratgar, “The usage of 10-fold cross-validation and grid search to enhance ML methods performance in solar farm power generation prediction,” *Clean. Eng. Technol.*, vol. 15, no. July, p. 100664, 2023, doi: 10.1016/j.clet.2023.100664.