

Implementasi Algoritma Decision Tree Untuk Klasifikasi Resiko Diabetes

Aisyah Putri Sabrina¹, Zaehol Fatah²

^{1,2} Sistem Informasi, Universitas Ibrahimy, Situbondo

Email: ^{1*}aisyahPutriSabrina31@gmail.com, ²zaeholfatah@gmail.com

Abstrak

Diabetes mellitus merupakan suatu kondisi yang ditandai oleh munculnya berbagai gejala akibat meningkatnya kadar gula darah (glukosa) yang disebabkan oleh kekurangan hormon insulin baik secara absolut maupun relatif. Penyakit ini berhubungan dengan gangguan metabolisme karbohidrat, protein, serta lemak yang disebabkan oleh berkurangnya sensitivitas jaringan terhadap insulin — kondisi ini dikenal sebagai resistensi insulin — atau akibat menurunnya produksi hormon insulin yang disebut defisiensi insulin. Penyakit ini tergolong berbahaya karena dapat menimbulkan berbagai komplikasi serius seperti gangguan ginjal, kerusakan penglihatan hingga kebutaan, penyakit pada kaki, serta meningkatkan risiko kematian. Indonesia sendiri menempati posisi keempat sebagai negara dengan jumlah penderita diabetes terbanyak setelah Amerika Serikat, Tiongkok, dan India. Jumlah penyandang diabetes di Indonesia diperkirakan akan mengalami peningkatan tajam hingga dua sampai tiga kali lipat pada tahun 2030 dibandingkan tahun 2000. Laporan dari World Health Organization juga menunjukkan bahwa pada tahun 2000 jumlah penderita diabetes di dunia telah mencapai 171 juta orang, dan angka tersebut diperkirakan meningkat menjadi 366 juta pada tahun 2030. Melalui pendekatan klasifikasi, berbagai pola dan hubungan tersembunyi dalam data kesehatan dapat diidentifikasi dengan lebih akurat. Klasifikasi sangat penting untuk mendeteksi faktor risiko sejak dini sehingga dapat mendukung pencegahan serta perencanaan tindakan medis yang lebih tepat. Salah satu algoritma yang banyak digunakan dalam klasifikasi adalah Decision Tree, karena mampu menyajikan informasi dalam bentuk visual dan aturan yang mudah dipahami. Pada penelitian ini, pengklasifikasian dilakukan untuk mengidentifikasi faktor-faktor penyebab diabetes mellitus dengan memanfaatkan algoritma Decision Tree. Dengan mengubah data ke dalam bentuk pohon keputusan serta menghasilkan rule (aturan) klasifikasi, proses analisis menjadi lebih terstruktur dan mudah dipahami. Hasil penelitian menunjukkan bahwa metode Decision Tree mampu memberikan tingkat akurasi sebesar 79,82%. Akurasi ini membuktikan bahwa teknik klasifikasi berbasis pohon keputusan cukup efektif dalam membantu proses deteksi dini penyakit diabetes. Dengan demikian, metode ini dapat menjadi salah satu solusi pendukung pengambilan keputusan di bidang kesehatan, terutama dalam upaya pencegahan dan pengendalian diabetes mellitus.

Kata Kunci: Penyakit Diabetes Melitus (DM), Data Mining, Klasifikasi, dan Algoritma Decision Tree.

PENDAHULUAN

Penyakit Diabetes mellitus merupakan kondisi medis yang muncul akibat peningkatan kadar glukosa dalam darah. Peningkatan ini disebabkan oleh ketidakseimbangan produksi insulin atau penurunan sensitivitas jaringan tubuh terhadap hormon tersebut. Ketidakseimbangan ini berdampak pada gangguan proses metabolisme utama

seperti karbohidrat, protein, dan lemak dalam tubuh (Lestari et al., 2021).

Diabetes mellitus sering dikaitkan dengan risiko serius terhadap kesehatan, seperti gangguan fungsi ginjal, kerusakan pada organ penglihatan, komplikasi pada ekstremitas, hingga meningkatkan angka kematian. Kondisi kadar gula darah yang tidak terkontrol menyebabkan pasien menghadapi risiko kesehatan setiap hari karena kadar glukosa dapat berubah-ubah

seiring aktivitas fisik dan pola makan (*DetikHealth, 2021*).

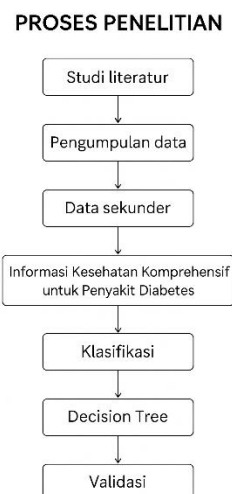
Secara global, jumlah penderita diabetes mengalami peningkatan yang cukup signifikan. Menurut data WHO, pada tahun 2000 terdapat sekitar 171 juta penderita diabetes di seluruh dunia. Angka ini diprediksi melonjak menjadi 366 juta pada tahun 2030. Indonesia sendiri menduduki peringkat keempat sebagai negara dengan jumlah penderita diabetes terbanyak setelah Tiongkok, Amerika Serikat, dan India (*DetikHealth, 2021*).

Peningkatan jumlah penderita ini menunjukkan pentingnya upaya deteksi dini untuk menekan risiko komplikasi yang lebih berat. Salah satu pendekatan yang dapat digunakan untuk mendeteksi risiko diabetes adalah melalui analisis data menggunakan teknik data mining. Teknik ini memungkinkan pengolahan dan analisis data dalam jumlah besar untuk menemukan pola tertentu yang relevan dengan penyakit tersebut.

Metode Decision Tree menjadi salah satu teknik klasifikasi yang sering digunakan dalam data mining karena penyajiannya yang sederhana dan mudah dipahami. Metode ini membentuk struktur pohon keputusan yang dapat membantu dalam proses prediksi berdasarkan atribut-atribut tertentu (*Baharudin & Nuryana, 2022*).

Hasil prediksi yang dihasilkan metode Decision Tree sangat dipengaruhi oleh beberapa faktor, seperti jumlah data pelatihan, data uji, serta atribut yang digunakan dalam pemodelan. Dengan metode ini, diharapkan analisis risiko terhadap diabetes dapat dilakukan lebih akurat, membantu proses pengambilan keputusan medis, dan mendukung pencegahan komplikasi penyakit di masa mendatang (*Pratama & Bhakti, 2023*).

METODE



Gambar 1. Diagram penelitian

1. Penampilan data

Bagian ini menjelaskan bagaimana peneliti mendapatkan data untuk penelitian ini. Dengan metode studi literatur untuk pengumpulan data, dan menggunakan jenis data sekunder yang digunakan adalah Informasi Kesehatan Komprehensif untuk Penyakit diabetes, yang diakses melalui situs web Kaggle. Kumpulan data ini berisi informasi kesehatan yang luas untuk 769 pasien, dengan menggunakan beberapa faktor penyebab DM untuk melakukan klasifikasi apakah pasien tersebut beresiko terkena diabetes atau tidak. Kumpulan data mencakup data jumlah kehamilan, Tingkat glukosa, tekanan darah, ketebalan lipatan kulit, kadar inulin dalam darah, berat badan, faktor keturunan serta usia. Data ini ideal untuk peneliti dan ilmuwan data yang ingin mengeksplorasi faktor-faktor penyebab DM, mengembangkan model prediktif, dan melakukan analisis statistik. Berikut ini data set beserta penjelasan dari atribut-atributnya disajikan pada gambar 1 di bawah ini.

Pregnancies	Glucose	BloodPress...	SkinThicke...	Insulin	BMI	DiabetesPe...	Age
1	85	66	29	0	26.600	0.351	31
8	183	64	0	0	23.300	0.672	32
1	89	66	23	94	28.100	0.167	21
0	137	40	35	168	43.100	2.288	33
5	116	74	0	0	25.600	0.201	30
3	78	50	32	88	31	0.248	26
10	115	0	0	0	35.300	0.134	29
2	197	70	45	543	30.500	0.158	53
8	125	96	0	0	0	0.232	54
4	110	92	0	0	37.600	0.191	30
10	168	74	0	0	38	0.537	34
10	139	80	0	0	27.100	1.441	57
1	189	60	23	846	30.100	0.398	59
5	166	72	19	175	25.800	0.587	51

Gambar 2. Data Set

Penjelasan Gambar 1 :

1. Pregnancies: Jumlah kehamilan
2. Glucose: jumlah glukosa dalam tubuh
3. Blood Preaa : tekanan darah
4. Skin Thicness: jumlah lipatan kulit
5. Insulin: kadar inulin dalam darah
6. BMI: berat badan
7. Diabetes Pedifree Function: faktor keturunan
8. Age: usia

2. Data Mining

Metode data mining memberi kemampuan kepada pengguna untuk mengolah dan mengakses kumpulan data berukuran besar dalam rentang waktu yang relatif singkat. Secara konseptual, data mining juga dipahami sebagai proses mengekstraksi informasi baru yang diperoleh dari tumpukan data masif sehingga membantu pengambilan keputusan yang lebih tepat. Istilah ini kerap pula disebut sebagai *knowledge discovery* karena fokusnya pada penemuan pengetahuan baru dari data mentah (Cakrawala University, 2024).

Dengan kata lain, data mining merupakan seperangkat alat dan aplikasi yang memadukan teknik statistik dan algoritma komputasi untuk mengekstraksi atau menggali potensi informasi yang sebelumnya tersembunyi dalam data. Pada praktiknya, proses penggalian data bertujuan menghasilkan temuan-temuan baru melalui identifikasi pola, aturan, atau korelasi tertentu dalam volume data yang besar, sebagaimana dijelaskan pada

referensi yang berjudul *Data Mining dan Aplikasinya* (Mustika et al., 2021).

Salah satu tujuan utama teknik dalam data mining adalah membangun model dari data yang tersedia; model tersebut kemudian dipakai untuk mengenali pola pada kumpulan data lain yang belum tersimpan dalam basis data sebelumnya. Proses ini melibatkan langkah-langkah seperti seleksi atribut, pembentukan model, dan penerapan model untuk melakukan prediksi atau klasifikasi terhadap data baru (Haryati, Sudarsono, & Suryana, 2015).

a. Decision Tree

Salah satu teknik klasifikasi yang sangat populer dalam proses data mining adalah *Decision Tree*. Metode ini menggunakan representasi struktur pohon, di mana setiap simpul (*node*) menggambarkan atribut, cabang mewakili nilai dari atribut, dan daun menunjukkan kelas (Partogi & Pasaribu, 2022).

Proses *Decision Tree* dimulai dengan mengubah data berbentuk tabel menjadi struktur pohon, kemudian membentuk aturan (*rule*) yang selanjutnya disederhanakan untuk memudahkan interpretasi. Dalam penerapannya, algoritma ini memerlukan atribut kelas karena termasuk dalam kategori *supervised learning*. Oleh karena itu, data latih (*training data*) harus cukup besar dan beragam untuk memperoleh hasil yang optimal (Google Books, 2020).

Model *Decision Tree* banyak dipilih karena mampu mengelola data dengan berbagai variabel serta menghasilkan interpretasi yang jelas dan mudah dipahami, sehingga sangat berguna dalam proses analisis maupun pengambilan keputusan (Assauqi, 2024).

b. Klasifikasi

Klasifikasi merupakan salah satu teknik *data mining* yang sangat efektif dan banyak dimanfaatkan dalam pengolahan data, khususnya di bidang kesehatan (Hendrian, 2018). Dalam metode ini, terdapat variabel target berbentuk kategori yang menjadi acuan utama proses analisis. Secara umum, klasifikasi dapat diartikan

sebagai proses pencarian dan pembentukan suatu model atau fungsi yang mampu menggambarkan serta membedakan kelas-kelas data atau konsep yang ada. Model tersebut kemudian digunakan untuk memperkirakan kelas dari suatu objek yang kelasnya belum diketahui sebelumnya (Rani, 2016).

Sebagai ilustrasi, dalam konteks ekonomi, pendapatan masyarakat dapat dikelompokkan ke dalam tiga kategori utama, yaitu berpendapatan tinggi, menengah, dan rendah. Sementara dalam dunia bisnis dan penelitian, klasifikasi sering diterapkan untuk berbagai tujuan, seperti mendeteksi apakah suatu transaksi kartu kredit termasuk transaksi curang atau tidak, menilai kelayakan pengajuan hipotek oleh nasabah, hingga mendiagnosis jenis penyakit yang diderita pasien berdasarkan gejala tertentu (Mustika et al., 2021).

c. Rapid Miner

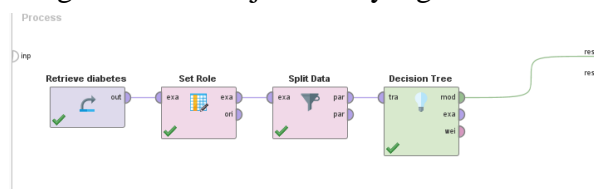
RapidMiner merupakan salah satu perangkat lunak *open source* yang dapat diakses secara bebas oleh siapa saja, sehingga sangat populer di kalangan peneliti, akademisi, maupun praktisi industri. Perangkat lunak ini dirancang khusus untuk membantu proses analisis dan pengolahan data (*data processing*) dalam berbagai bidang, mulai dari riset akademik, bisnis, kesehatan, hingga teknologi informasi (Elvitaria, 2017).

RapidMiner menyediakan beragam teknik analisis, baik yang bersifat deskriptif maupun prediktif, yang memungkinkan pengguna memahami pola data dan membuat prediksi berdasarkan informasi historis. Salah satu keunggulan utamanya adalah kemudahan penggunaan melalui antarmuka grafis (*graphical user interface*), sehingga pengguna tidak harus menguasai bahasa pemrograman tingkat lanjut untuk melakukan analisis yang kompleks.

HASIL DAN PEMBAHASAN

Gambar dibawah ini menunjukkan Antarmuka pengguna perangkat lunak RapidMiner Studio, sebuah alat yang

populer untuk pemrosesan data dan pembelajaran mesin, digambarkan di sini. Tampilan ini menunjukkan sebuah proses yang sedang dibangun. Proses ini terdiri dari beberapa bagian atau operator yang saling terhubung, yang membentuk alur kerja. Komponen-komponen ini menunjukkan berbagai tahapan dalam proses analisis data, mulai dari membaca data (menggunakan Microsoft Excel), melakukan transformasi data (menggunakan numerik ke nominal), dan membangun model klasifikasi (menggunakan klasifikasi). Koneksi antar komponen menunjukkan aliran data dari satu tahap ke tahap berikutnya. Selain itu, komponen evaluasi, juga dikenal sebagai *cross-validation*, digunakan untuk mengevaluasi kinerja model yang dibuat.



Gambar 3. Proses Rapid Miner

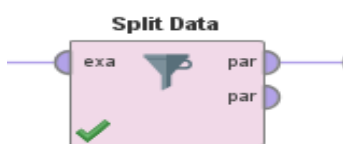


Gambar 4. Set Role

Fungsi "Set Role" di RapidMiner digunakan untuk menentukan peran atau atribut dari kolom dalam dataset. Ini sangat penting dalam konteks persiapan data dan pemodelan. Berikut adalah beberapa fungsi utama dari "Set Role":

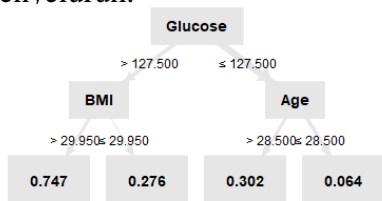
1. Menentukan Target Variabel yaitu dapat menetapkan kolom tertentu sebagai target atau label yang ingin diprediksi oleh model. Ini penting dalam proses klasifikasi atau regresi.
2. Mengatur Atribut Fitur yaitu Kolom-kolom lain dalam dataset dapat ditetapkan sebagai atribut input (features), yang akan digunakan oleh model untuk membuat prediksi.

- Selain itu juga dapat menandai kolom tertentu sebagai ID (identifikasi) untuk menunjukkan bahwa kolom tersebut tidak perlu dipertimbangkan dalam proses pelatihan model. Kolom ini biasanya tidak memengaruhi prediksi, seperti ID pelanggan.
- Dengan menetapkan peran yang sesuai untuk setiap kolom dapat lebih mudah menganalisis data dan menghasilkan visualisasi yang relevan. Berikut gambar dibawah ini adalah hasil pohon decision tree dan persentase klasifikasi



Gambar 5. Split Data

- Split data digunakan untuk beberapa tujuan penting dalam proses analisis data dan machine learning:
- Dengan membagi dataset menjadi bagian pelatihan dan pengujian, dan membantu mengukur kinerja model secara lebih akurat.
- Pembagian data membantu mengurangi risiko overfitting, di mana model terlalu cocok dengan data pelatihan dan tidak dapat generalisasi dengan baik pada data baru.
- Dengan menggunakan data pengujian yang terpisah dapat menilai kinerja model menggunakan metrik evaluasi seperti akurasi, presisi, dan recall. Ini memberikan gambaran yang lebih jelas tentang seberapa baik model akan bekerja di dunia nyata.
- Selain itu juga dapat menggunakan teknik lain seperti cross-validation untuk lebih lanjut mengoptimalkan dan menilai model dengan cara menyeluruh.



Gambar 6. Pohon Keputusan Glucose

RegressionTree

```

Glucose > 127.500
| BMI > 29.950: 0.747 {count=166}
| BMI <= 29.950: 0.276 {count=58}
Glucose <= 127.500
| Age > 28.500: 0.302 {count=172}
| Age <= 28.500: 0.064 {count=218}
    
```

Gambar 7. Description

Perhitungan :

- Jika glucose > 127.500
 BMI > 29.950 prediksi 0.747 (mendekati 1.0)
 Maka beresiko diabetes (jumlah) : 166
 - Jika glucose > 127.500
 BMI <= 29.950 prediksi 0.276 (mendekati 0)
 Maka tidak beresiko diabetes (jumlah) : 58
 - Jika glucose <= 127.500
 Age > 28.500 prediksi 0.302 (mendekati 1)
 Maka beresiko diabetes (jumlah) : 172
 - Jika glucose <= 127.500
 Age <= 28.500 prediksi 0.64 (mendekati 0)
 Maka tidak beresiko diabetes (jumlah) : 218
- Total data = 166 + 58 + 172 + 218 = 614
 Akurasi = 769 / 614 ≈ 79.82%
 Akurasi dari penelitian ini adalah = 79.82%

Berdasarkan hasil pengujian model klasifikasi menggunakan algoritma Decision Tree, diperoleh tingkat akurasi yang cukup baik. Perhitungan akurasi dilakukan dengan membandingkan jumlah data yang berhasil diklasifikasikan dengan benar terhadap jumlah total data.

Rumus Akurasi:

$$\text{Akurasi} = \frac{\text{Jumlah Data Terklasifikasi benar}}{\text{Jumlah Total Data}} \times 100\%$$

Perhitungan:

$$\text{Akurasi} = \frac{614}{769} \times 100\% \approx 79,83\%$$

Dengan demikian, tingkat akurasi model klasifikasi dalam mendeteksi risiko diabetes adalah sebesar **79,82%**. Hal ini menunjukkan bahwa algoritma Decision Tree mampu mengklasifikasikan data pasien secara efektif dengan ketepatan yang cukup tinggi. Dari total 769 data pasien, sebanyak 614 data berhasil diklasifikasikan dengan benar.

SIMPULAN (PENUTUP)

Dalam penelitian dapat disimpulkan bahwa pendekatan data mining, khususnya metode klasifikasi dengan algoritma decision tree dapat berkontribusi dalam mengetahui resiko seseorang untuk terkena penyakit DM.

Algoritma decision tree menjadi pilihan yang untuk penelitian karena kemudahan untuk menjabarkan hasilnya, yang dapat membantu dalam pengambilan keputusan terkait diagnosis dan perawatan. Secara keseluruhan, penelitian ini memberikan wawasan berharga mengenai potensi penggunaan teknik klasifikasi dalam menentukan apakah seseorang beresiko terkena dm atau tidak

UCAPAN TERIMA KASIH

Ucapan terima kasih diberikan kepada Bapak Zaehol Fatah, M.Kom sebagai dosen pembimbing kami yang senantiasa memberikan kritikan yang memotivasi dan membangun demi keberlangsungan keberhasilan kami. Tidak lupa pula ucapan terima kasih yang paling utama di tujukan kepada seseorang di balik figur yaitu orang tua kami yang tidak pernah lelah dalam menasehati dan memberikan motivasi yang bersifat membangun juga tentunya, Kalian semua hebat Jazakumullahu Khairon Ahsanal jaza’.

DAFTAR PUSTAKA

Assauqi, M. F. R. (2024). *Decision tree berdasarkan gejala klinis dan data laboratorium*, 2(4), 42–53.

Baharudin, M. N., & Nuryana, I. K. D. (2022). Implementasi algoritma decision tree untuk klasifikasi surat pada aplikasi mobile e-surat Dinas Komunikasi dan Informatika Kota Kediri berbasis Android. *Journal of Informatics and Computer Science*, 4(1), 76–85. <https://doi.org/10.26740/jinacs.v4n01.p76-85>

Cakrawala University. (2024, February 22). *Data mining: Definisi, fungsi, metode, dan contoh penerapannya*. Retrieved from <https://www.cakrawala.ac.id/berita/data-mining-adalah>

DetikHealth. (2021, December 6). *Gawat! Indonesia masuk negara diabetes terbanyak di dunia*. Retrieved from <https://health.detik.com/berita-detikhealth/d-5842320/gawat-indonesia-masuk-negara-kasus-diabetes-terbanyak-di-dunia>

Elvitaria, M. H. L. (2017). *SMK analisis kesehatan Abdurrah menggunakan algoritma*. *Jurnal Teknologi dan Sistem Informasi Univrab*, 2(2), 220–233.

Google Books. (2020, October 6). *Data mining menggunakan Android, Weja, dan SPSS*.

Haryati, S., Sudarsono, A., & Suryana, E. (2015). Implementasi data mining untuk memprediksi masa studi mahasiswa menggunakan algoritma C4.5 (studi kasus: Universitas Dehasen Bengkulu). *Jurnal Media Infotama*, 11(2), 130–138.

Hendrian, S. (2018). Algoritma klasifikasi data mining untuk memprediksi siswa dalam memperoleh bantuan dana pendidikan. *Faktor Exacta*, 11(3), 266–274. <https://doi.org/10.30998/faktorexacta.v11i3.2777>

Lestari, Z., Zulkarnain, S., & Aisyah, S. (2021). *Diabetes melitus: Review etiologi, patofisiologi, gejala, penyebab, cara pemeriksaan, cara pengobatan dan cara pencegahan*.

- UIN Alauddin Makassar*, 1(2), 237–241. Retrieved from <http://journal.uin-alauddin.ac.id/index.php/psb>
- Mustika, N., et al. (2021). *Data mining dan aplikasinya*. Retrieved from <https://repository.penerbitwidina.com/uk/publications/351768/data-mining-dan-aplikasinya>
- Partogi, Y., & Pasaribu, A. (2022). Perancangan metode decision tree terhadap sistem perpustakaan STMIK Kuwera. *Jurnal Sistem Informasi dan Teknologi*, 1(2), 20–25. <https://doi.org/10.56995/sintek.v1i2.4>
- Pratama, F. D., & Bhakti, H. D. (2023). Implementasi aplikasi prediksi ketepatan pembayaran customer perusahaan dengan metode decision tree. *Indexia*, 5(1), 72. <https://doi.org/10.30587/indexia.v5i01.5082>
- Rani, L. N. (2016). Klasifikasi nasabah menggunakan algoritma C4.5 sebagai dasar pemberian kredit. *INOVTEK Polbeng – Seri Informatika*, 1(2), 126. <https://doi.org/10.35314/isi.v1i2.131>