

**PENERAPAN ALGORITMA KLASIFIKASI DATA MINING POHON
KEPUTUSAN UNTUK PREDIKSI PENYAKIT DIABETES**

Eka Sabna¹, Octavia Dewi²

Universitas Hang Tuah Pekanbaru

E-mail: es3jelita@yahoo.com¹, dewitavia@yahoo.com²

Abstrak

Penyakit Diabetes merupakan penyebab tertinggi kematian ketiga di Indonesia yang setiap tahun penderitanya semakin bertambah. Banyak faktor yang mempengaruhi orang menderita diabetes diantaranya yaitu atribut Jumlah Kehamilan (Pregnancies), Kadar Glukosa (Glucose), Tekanan Darah (Blood Pressure), Ketebalan Kulit (Skin Thickness), Indeks Massa Tubuh / Body Mass Record (BMI) dan Umur (Age). Algoritma Decision Tree (Pohon Keputusan) merupakan salah satu pemodelan yang dapat digunakan dalam pemodelan klasifikasi variabel dari penyakit diabetes. Untuk menghasilkan informasi baru maka digunakan algoritma C.45 dan pengujian yang menggunakan aplikasi Rapid Miner akan semakin memperkuat keputusan. Tujuan dari penelitian ini adalah untuk prediksi dengan mengklasifikasi variabel yang mempengaruhi penyakit diabetes apakah pasien menderita diabetes atau tidak. Hasilnya prediksi diabetes menggunakan algoritma C4.5 menghasilkan nilai akurasi (Accuracy) sebesar 71,43%. Pemodelan ini diharapkan dapat menjadi acuan untuk pengobatan penderita diabetes bagi dokter di rumah sakit dan di masyarakat untuk mengetahui cara menjaga pola hidup dan cara menghindari penyakit diabetes dilihat dari variabel yang mempengaruhi terjadinya penyakit.

Kata Kunci — *Algoritma C4.5, Penambangan Data, Ramalan, Penambang Cepat.*

1. PENDAHULUAN

Selama dua decade terakhir telah terjadi peningkatan yang dramatis terhadap jumlah informasi yang disimpan secara elektronik. Teknologi database saat ini memungkinkan untuk menyimpan sejumlah informasi dalam jumlah yang sangat besar dan terakumulasi. Data mining berhubungan erat dengan analisa data dan penggunaan perangkat lunak untuk mencari pola dan kesamaan dalam sekumpulan data. Salah satu metode data mining adalah klasifikasi. Pada persoalan klasifikasi data akan diprediksi dengan beberapa atribut dan satu atribut target. Terdapat algoritma untuk klasifikasi, salah satunya adalah Algoritma C4.5 yang menghasilkan sebuah pohon keputusan (Decision Tree).

Studi menggunakan Algoritma Klasifikasi yang diimplementasikan untuk penyakit telah banyak dilakukan, diantaranya adalah klasifikasi penyakit jantung, ISPA dan Diabetes. Penelitian ini dilakukan untuk prediksi penyakit Diabetes. Angka kejadian diabetes terus meningkat di

seluruh dunia. Saat ini penderita diabetes merupakan penyakit mematikan ketiga di Indonesia setelah stroke dan jantung. Jumlah penderita terus meningkat dan tercatat saat ini mencapai 422 juta orang di dunia, empat kali lebih banyak dari pada 30 tahun lalu, menurut Organisasi Kesehatan Dunia (WHO) [2]. Diabetes adalah kondisi di mana kandungan gula dalam darah melebihi normal dan cenderung tinggi. Diabetes mellitus merupakan salah satu penyakit metabolisme yang mampu menyerang siapa saja. Diabetes mellitus (atau kencing manis) adalah kondisi kronis dan berlangsung seumur hidup yang mempengaruhi kemampuan tubuh dalam menggunakan energi yang dari makanan. Ada dua jenis utama dari penyakit ini: Tipe 1 dan Tipe 2.

Algoritma Klasifikasi yang digunakan pada penelitian-penelitian sebelumnya, diantaranya yang dilakukan oleh Sanni Ucha Putri (2021) menggunakan Data Mining Algoritma C4.5 yang menghasilkan sebuah pohon keputusan menggunakan Rapidminer agar pencegahan terhadap penyakit diabetes dapat dilakukan segera mungkin. Hasil dari penelitian ini akan dijadikan sebagai acuan untuk dapat melihat apakah seseorang beresiko terkena penyakit diabetes atau tidak berdasarkan atribut yang telah ditetapkan [4]. Penelitian untuk memprediksi penyakit diabetes telah dilakukan oleh Nining Nur Habibah, dkk (2023) dimana klasifikasi tersebut dapat mencapai ketelitian yang unggul. Berdasarkan hasil pengujian, akurasinya 76,67 persen, presisi 72 persen, dan recall 41,67 persen.

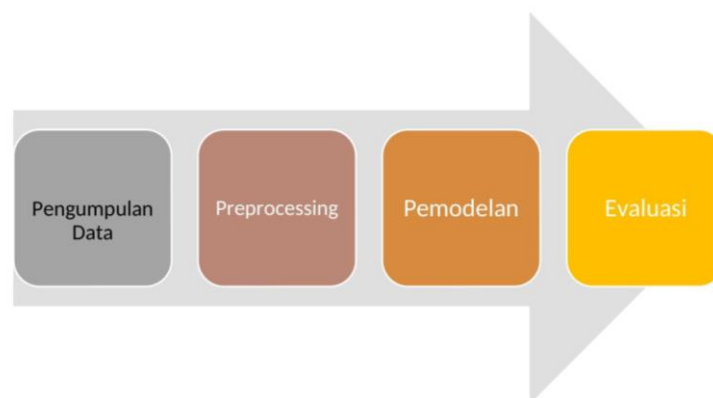
Atribut merupakan suatu hal yang penting pada keakuratan hasil proses, sehingga perlu diketahui atribut – atribut utama pada penyakit terkhususnya penyakit diabetes. Didalam proses pengambilan keputusan pada penyakit sering didapatkan hasil yang berbeda sehingga perlu mengetahui atribut – atribut utama didalam pengambilan keputusan. Menurut teori, ada banyak faktor risiko yang berhubungan dengan kejadian diabetes. Faktor-faktor tersebut yaitu: Jumlah Kehamilan (Pregnancies), Kadar Glukosa dalam darah (Glucose Blood Pressure, Tekanan Darah, Ketebalan Kulit (Skin Thickness), Index Massa Tubuh / Body Mass Index (BMI) dan Umur (Age).

Pada penelitian ini, dikembangkan suatu sistem klasifikasi dengan menggunakan 8 atribut. Penelitian ini menggunakan metode pohon keputusan untuk mengembangkan sistem klasifikasi yang mengklasifikasikan apakah seseorang menderita diabetes berdasarkan beberapa faktor pengalaman lainnya. Dan menggunakan dataset sebanyak 768 data, diharapkan hasil penelitian ini bermanfaat untuk peramalan seseorang menderita penyakit diabetes.

2. METODE PENELITIAN

A. Tahapan Penelitian

Metode pada penelitian ini terdiri dari 4 langkah yaitu pengumpulan data, preprocessing, klasifikasi dan evaluasi.



Gambar 1. Tahapan Penelitian

1. Pengumpulan Data

Pada penelitian ini dalam pengumpulan data dilakukan untuk mendapatkan sumber dataset yang akan digunakan dengan pengumpulan data sekunder yang berasal dari web penyedia datasets yaitu Kaggle dengan judul diabetes datasets. Data sekunder merupakan data yang bisa didapatkan oleh peneliti tanpa harus terjun langsung ke lapangan untuk pengambilan datanya karena datanya sudah diolah dan dikumpulkan oleh peneliti lain sebelumnya.

Jumlah data sampel yang dipakai pada penelitian ini ialah sebanyak 786 data dan terdiri dari 8 atribut.

2. Preprocessing
Adalah langkah pembersihan data sebelum diklasifikasikan. Preprocessing data tidak sempurna dan data yang tidak konsisten. Pengolahan data yaitu pembersihan data mentah yang diperoleh .
3. Klasifikasi
Proses untuk mengkategorikan atribut dimana proses klasifikasi menggunakan algoritma C4.5 (Decision Tree) . Data dibagi menjadi dua set data yaitu training dan data testing.
4. Evaluasi
Evaluasi termasuk untuk memastikan bahwa tes itu benar. Tes dilakukan untuk mengukur keakuratan hasil dan menemukan hasil tes yang terbaik.

B. Dataset

Data yang digunakan pada penelitian ini berupa dataset yang berasal dari web yaitu Kaagle. Keterangan Atribut atau variable pada data set dapat dilihat pada Tabel 2.

Tabel 2. Keterangan Dataset penyakit Diabetes

Variabel (Atribut)	Deskripsi Variabel
Jumlah Kehamilan (The Number of pregnancies)	Jumlah Kehamilan pada Wanita
Kadar Glukosa dalam Darah (The Glucose level in blood).	Kadar glukosa darah adalah jumlah glukosayangseseorang miliki dalam aliran darahnya pada waktu tertentu.
Usia (Age)	Usia Pasien
Ketebalan Kulit (Skin Thickness)	Ketebalan Kulit Pasien
Insulin	Serum Insulin dalam 2 jam
BMI	Body Mass Index (Indeks Massa Tubuh)
Tekanan Darar (Blood Presure)	Tekanan Darah Diastolic
Diabetes Prediksi	Persentase Diabetes
Outcome (Final Result)	Hasil = 1 (Yes), Hasil = 0 (No)

3. HASIL DAN PEMBAHASAN

A. Data Acquisition

Dari keterangan variabel pada Tabel 2, lakukan analisis terhadap nilai-nilai pada setiap variable, hasilnya dapat dilihat pada gambar berikut :

Outcome	Pregnancies	Glucose	BloodPress...	SkinThickne...	Insulin	BMI	DiabetesPe...	Age
1	6	148	72	35	0	33.600	0.627	50
0	1	85	66	29	0	26.600	0.351	31
1	8	183	64	0	0	23.300	0.672	32
0	1	89	66	23	94	28.100	0.167	21
1	0	137	40	35	168	43.100	2.288	33
0	5	116	74	0	0	25.600	0.201	30
1	3	78	50	32	88	31	0.248	26
0	10	115	0	0	0	35.300	0.134	29
1	2	197	70	45	543	30.500	0.158	53
1	8	125	96	0	0	0	0.232	54
0	4	110	92	0	0	37.600	0.191	30
1	10	168	74	0	0	38	0.537	34
0	10	139	80	0	0	27.100	1.441	57

ExampleSet (768 examples, 1 special attribute, 8 regular attributes)

Gambar 2. Tampilan Data pada Rapidminer

B. Preprocessing

Data yang telah didapatkan dari situs kaggle perlu dibersihkan terlebih dahulu dengan pengecekan data duplikat, missing value, dan outlier, hasilnya sebagai berikut :

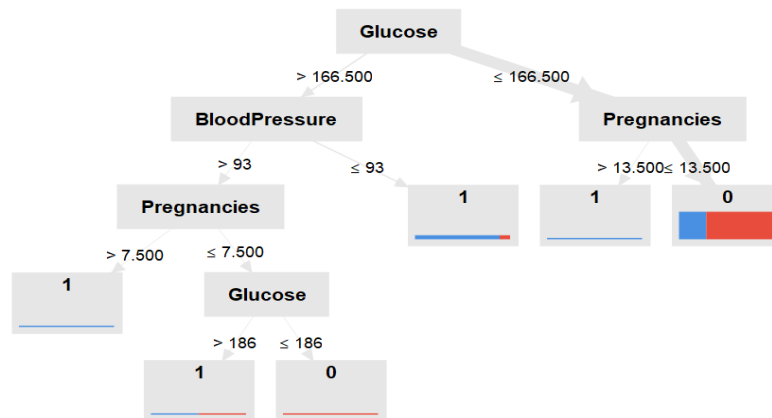
Label	Binomial	0	Negative	1	Positive	0	Value
Outcome	Binomial	0	1	0	0	0	5
Pregnancies	Integer	0	0	Min	Max	17	Average 3.84
Glucose	Integer	0	0	Min	Max	199	Average 120
BloodPressure	Integer	0	0	Min	Max	122	Average 69.1
SkinThickness	Integer	0	0	Min	Max	99	Average 20.5
Insulin	Integer	0	0	Min	Max	846	Average 79.7
BMI	Real	0	0	Min	Max	67.100	Average 31.5

Showing attributes 1 - 9 Examples: 768 Special Attributes: 1 Regular Attributes:

Gambar 3. Tampilan Data pada Rapidminer

C. Pemodelan dan Evaluasi

Pemodelan dilakukan pada data testing, data dipisahkan (split) menjadi data latih (training) dan data testing (Uji) dengan rasio 80:20. Hasil pemodelan adalah sebagai berikut :



Gambar 4. Tampilan Model Pohon Keputusan (Decision Tree)

Berikut dilakukan evaluasi terhadap model dengan menggunakan Performance Vector dimana diperoleh nilai akurasi (accuracy) sebesar 71,43%.

accuracy: 71,43%

	true 1	true 0	class precision
pred. 1	17	7	70.83%
pred. 0	37	93	71.54%
class recall	31.48%	93.00%	

Gambar 5. Hasil Performance Vector dalam bentuk Tabel

PerformanceVector

```
PerformanceVector:  
accuracy: 71.43%  
ConfusionMatrix:  
True:    1    0  
1:      17    7  
0:      37   93
```

Gambar 6. Performance Vector dalam bentuk deskripsi

Performance hasil diatas bahwa ketepatan akurasi decision tree dengan hasil uji mencapai 71.43% . Performance vector dapat di ketahui pada kategori mana saja data yang sesuai atau tidak sesuai antara hasil (prediksi) dan data.

Row No.	Outcome	prediction(O...	confidence(1)	confidence(0)	Pregnancies	Glucose	BloodPress...	SkinThick
1	0	1	0.923	0.077	10	115	0	0
2	1	0	0.061	0.939	8	125	96	0
3	1	1	0.871	0.129	1	189	60	23
4	1	0	0.419	0.581	7	147	76	0
5	0	0	0.419	0.581	5	109	75	26
6	0	0	0	1	3	88	58	11
7	0	0	0.419	0.581	10	122	78	31
8	0	0	0.419	0.581	11	138	76	0
9	0	1	0.871	0.129	3	180	64	25
10	0	0	0.061	0.939	7	105	0	0
11	0	0	0.419	0.581	0	100	88	60
12	1	0	0.419	0.581	7	114	66	0
13	0	0	0	1	1	96	66	17

ExampleSet (154 examples, 4 special attributes, 8 regular attributes)

Gambar 7. Perbandingan Hasil Prediksi dengan Data

4. KESIMPULAN

Prediksi penyakit diabetes menggunakan algoritma klasifikasi C4.5 menghasilkan tingkat akurasi sebesar 71,43 %. Dengan demikian, pemodelan dapat digunakan sebagai Prediksi penyakit diabetes dengan Baik dan diharapkan dapat menjadi acuan bagi penderita diabetes, dokter dan masyarakat . Model ini sebagai pengetahuan agar Masyarakat menjaga pola hidup dan cara menghindari penyakit diabetes dilihat dari variabel yang mempengaruhi terjadinya penyakit. Selain itu, disarankan untuk melakukan penelitian tentang prediksi penyakit diabetes menggunakan algoritma lain agar mendapatkan kinerja model yang lebih tinggi.

DAFTAR PUSTAKA

- [1] Binus, "Mengenal Konsep Dasar, Tujuan, dan Teknik Data Mining – School of Information Systems," Apr. 2021. <https://sis.binus.ac.id/2021/04/23/mengenal-konsep-dasar-tujuan-dan-teknik-data-mining/> (accessed Jan. 31, 2024).
- [2] WHO, "Diabetes," 2024. https://www.who.int/health-topics/diabetes?gclid=CjwKCAiAtt2tBhBDEiwALZuhAK22nBpwTaliHRvLyph-LfJiSu4JBCCaKJuWckZJRMdL9HKAYw5uRxoCymcQAvD_BwE#tab=tab_1 (accessed Jan. 30, 2024).
- [3] M. Keluarga, "Ciri-Ciri Diabetes di Usia Muda, Penyebab, & Cara Menanganinya - Mitra Keluarga," Jan. 2024. <https://www.mitrakeluarga.com/artikel/penyebab-diabetes-melitus-di-usia-muda> (accessed Jan. 30, 2024).
- [4] S. Ucha Putri, E. Irawan, F. Rizky, S. Tunas Bangsa, P. A. -Indonesia Jln Sudirman Blok No, and S. Utara, "Implementasi Data Mining Untuk Prediksi Penyakit Diabetes Dengan

- Algoritma C4.5,” *Kesatria J. Penerapan Sist. Inf. (Komputer dan Manajemen)*, vol. 2, no. 1, pp. 39–46, Jan. 2021, doi: 10.30645/KESATRIA.V2I1.56.
- [5] N. N. Habibah, A. Nazir, I. Iskandar, F. Syafria, L. Oktavia, and I. Syurfi, “Pemodelan Klasifikasi Untuk Menentukan Penyakit Diabetes dengan Faktor Penyebab Menggunakan Decision Tree C4.5 Pada Wanita,” *J. Sist. Komput. dan Inform.*, vol. 4, no. 4, pp. 654–661, Jun. 2023, doi: 10.30865/JSON.V4I4.6202.
- [6] R. Puspitasari *et al.*, “Gambaran Disparitas Diabetes Melitus Tipe 2 Ditinjau Dari Faktor Body Mass Index (Bmi),” *JIMPK J. Ilm. Mhs. Penelit. Keperawatan*, vol. 3, no. 6, pp. 251–256, 2023, doi: 10.35892/JIMPK.V3I6.1094.
- [7] “Apa Saja Faktor-Faktor yang Mempengaruhi Harga Minyak Dunia? - Ajaib.” <https://ajaib.co.id/apa-saja-faktor-faktor-yang-mempengaruhi-harga-minyak-dunia/> (accessed May 16, 2022).
- [8] A. Pahlawati and P. S. Nugroho, “Hubungan Tingkat Pendidikan dan Usia dengan Kejadian Diabetes Melitus di Wilayah Kerja Puskesmas Palaran Kota Samarinda Tahun 2019,” *Borneo Stud. Res.*, vol. 1, no. 1, pp. 1–5, Dec. 2019, Accessed: Jan. 31, 2024. [Online]. Available: <https://journals.umkt.ac.id/index.php/bsr/article/view/479>.
- [9] M. J. Lede *et al.*, “PENGARUH KADAR GULA DARAH TERHADAP PENYEMBUHAN LUKA DIABETES MELLITUS DI PUSKESMAS DINOYO MALANG,” *Nurs. News J. Ilm. Keperawatan*, vol. 3, no. 1, Mar. 2018, doi: 10.33366/NN.V3I1.826.
- [10] A. of D. American, “Diabetes: 10 warning signs that can appear on your skin,” Jan. 2024. <https://www.aad.org/public/diseases/a-z/diabetes-warning-signs> (accessed Jan. 30, 2024).