



Implementasi Algoritma C4.5 Untuk Deteksi Dini Penyakit Diabetes Mellitus Pada Manusia

Chandra^{1*}, David², Hendri²

¹Sistem Informasi, STMIK TIME, Medan, Indonesia

²Teknik Informatika, STMIK TIME, Medan, Indonesia

Email: ^{1*}chandra.wiejaya@gmail.com, ²davidyang1991@gmail.com, ³h4ndr7@hotmail.com

Email Penulis Korespondensi: chandra.wiejaya@gmail.com

Abstrak—Diabetes tergolong sebagai salah satu penyakit kronis yang mengancam jiwa dengan pertumbuhan tercepat yang telah mempengaruhi 422 juta orang di seluruh dunia menurut laporan World Health Organization (WHO), pada tahun 2018. Oleh sebab itu, sangat penting melakukan deteksi dini terhadap penyakit DM karena apabila penyakit tersebut dibiarkan terlalu lama tanpa adanya pengobatan maka dapat berakibat pada komplikasi yang berbahaya seperti gagal ginjal, kerusakan fungsi organ lainnya hingga serangan jantung. Pada penelitian ini akan dibangun sebuah sistem informasi dengan menerapkan algoritma data mining C4.5 untuk deteksi dini penyakit Diabetes Mellitus pada manusia. Dataset pada penelitian ini diambil dari Kaggle Diabetes Dataset. Hasil penelitian menunjukkan bahwa sistem informasi yang dibangun dapat membantu dunia medis dalam melakukan deteksi dini penyakit Diabetes Mellitus pada manusia melalui fitur prediksi yang mengimplementasikan algoritma C4.5. Selain itu, hasil pengujian algoritma C4.5 menunjukkan bahwa algoritma tersebut tergolong akurat dalam melakukan prediksi deteksi dini penyakit Diabetes Mellitus apabila mengikuti rule pohon keputusan yang terbentuk.

Kata Kunci: Sistem Prediksi; Diabetes Mellitus; Pohon Keputusan; C4.5

Abstract—Diabetes is classified as one of the fastest growing life-threatening chronic diseases which has affected 422 million people worldwide according to the World Health Organization (WHO) report, in 2018. Therefore, it is very important to carry out early detection of DM disease because if the disease is left for too long without treatment, it can result in dangerous complications such as kidney failure, damage to the function of other organs to heart attacks. In this research an information system will be built by applying the C4.5 data mining algorithm for early detection of Diabetes Mellitus in humans. The dataset in this study was taken from the Kaggle Diabetes Dataset. The results of the study show that the information system built can assist the medical world in early detection of Diabetes Mellitus in humans through a predictive feature that implements the C4.5 algorithm. In addition, the results of testing the C4.5 algorithm show that the algorithm is classified as accurate in predicting early detection of Diabetes Mellitus if it follows the decision tree rules that are formed.

Keywords: Prediction System; Diabetes Mellitus; Decision Tree; C4.5

1. PENDAHULUAN

Diabetes Mellitus (DM), menurut definisi World Health Organization (WHO), adalah penyakit degeneratif kronis yang disebabkan oleh produksi insulin yang tidak mencukupi di pankreas atau oleh ketidakmampuan tubuh untuk secara efektif menggunakan insulin yang diproduksi, mengambil hyperglycemia (peningkatan glukosa darah) sebagai indikator utama [1]. Diabetes tergolong sebagai salah satu penyakit kronis yang mengancam jiwa dengan pertumbuhan tercepat yang telah mempengaruhi 422 juta orang di seluruh dunia menurut laporan Organisasi Kesehatan Dunia (WHO), pada tahun 2018. Diabetes dianggap sebagai salah satu penyakit paling mematikan dan kronis yang menyebabkan peningkatan gula darah. Banyak komplikasi yang dapat terjadi jika diabetes tetap tidak diobati dan tidak teridentifikasi sejak awal [2]. Jumlah penderita penyakit diabetes meningkat dari tahun ke tahun, baik dari jumlah kasus maupun prevalensi. Pada tahun 2019, jumlah penderita diabetes di dunia sudah mencapai 463 juta orang dan diprediksi akan terus bertambah mencapai angka 700 juta orang pada tahun 2045 [3].

Gaya hidup sehat yang beradaptasi akan membantu menjadi lebih sehat dan mencegah penyakit [4]. Secara umum, penyakit DM memiliki gejala yang hampir mirip dengan kondisi sakit biasa sehingga banyak orang yang tidak menyadari bahwa mereka mengidap penyakit tersebut dan bahkan sudah mengarah pada komplikasi. Oleh sebab itu, sangat penting melakukan deteksi dini terhadap penyakit DM karena apabila penyakit tersebut dibiarkan terlalu lama tanpa adanya pengobatan maka dapat berakibat pada komplikasi yang berbahaya seperti gagal ginjal, kerusakan fungsi organ lainnya hingga serangan jantung. Banyak cara yang dapat dilakukan dalam mendeteksi dini seseorang mengalami penyakit DM atau tidak. Jika dari sisi medis, diagnosa penyakit DM dapat dilakukan melalui pengecekan darah. Namun terkadang dari sisi medis, proses pengecekan darah hanya melihat dari satu parameter saja yaitu kadar gula darah pasien. Faktanya, terkadang kadar gula darah pasien rendah, namun pasien tersebut merupakan pasien DM sehingga perlu melibatkan parameter-parameter lain dari pasien agar menghasilkan hasil deteksi dini yang lebih akurat [5]. Pengetahuan tentang penyakit diabetes mellitus merupakan sarana yang dapat membantu pasien dalam melakukan penanganan terhadap diabetes mellitus [6].

Dari sisi sains komputer, proses deteksi dini penyakit DM pada manusia dapat dilakukan melalui sistem informasi dengan menerapkan teknik data mining. Data mining di gunakan untuk melakukan proses ekstraksi informasi yang tersembunyi dari dataset yang banyak dan terdapat beberapa teknik dalam data mining seperti klasifikasi, clustering, regresi dan asosiasi yang akan digunakan dalam data pada bidang medis [7]. Kelebihan dari pemanfaatan sistem informasi dan teknik data mining adalah dapat melakukan deteksi dini penyakit DM pada manusia secara cepat. Pengetahuan tentang penyakit diabetes mellitus merupakan sarana yang dapat membantu pasien dalam melakukan penanganan terhadap



diabetes mellitus. Selain itu, dengan data mining, maka dapat dilakukan prediksi dengan melibatkan parameter-parameter yang lebih banyak tidak hanya berdasarkan satu parameter saja namun juga melibatkan parameter-parameter lain seperti usia, berat badan, tekanan darah, dan parameter-parameter lainnya sehingga hasil deteksi lebih akurat [8]. Salah satu algoritma data mining yang memiliki akurasi yang cukup baik dalam deteksi dini penyakit adalah algoritma C4.5. Berdasarkan penelitian terdahulu yang menerapkan algoritma C4.5 untuk mengklasifikasi penyakit jantung memperoleh tingkat akurasi 79% [9]. Kemudian penelitian terdahulu yang menerapkan algoritma C4.5 untuk mendeteksi gangguan kepribadian memperoleh tingkat akurasi 73% [10]. Terdapat juga penelitian terdahulu yang dilakukan sebagai perbandingan algoritma C4.5 dan Naïve Bayes dalam memprediksi penyakit Cerebrovascular. Hasil penelitian yang dilakukan menyimpulkan bahwa algoritma C4.5 memperoleh tingkat akurasi yang lebih tinggi yaitu 95% sedangkan algoritma Naïve Bayes memperoleh tingkat akurasi 91% [11]. Berdasarkan uraian yang telah dipaparkan, maka akan dibangun sebuah sistem informasi dengan menerapkan algoritma data mining C4.5 untuk deteksi dini penyakit Diabetes Mellitus pada manusia.

2. METODOLOGI PENELITIAN

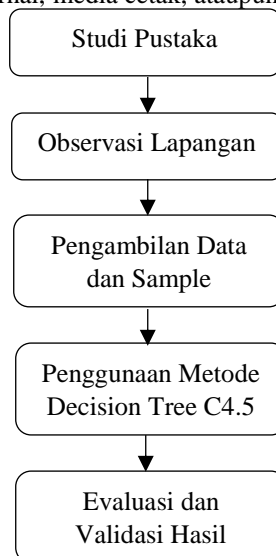
2.1 Tahapan Penelitian

Metodologi penelitian yang akan digunakan dalam penelitian ini adalah System Development Life Cycle (SDLC) [12].

2.1.1 Metode Pengumpulan Data

Metode pengumpulan data yang digunakan pada penelitian ini antara lain sebagai berikut:

- a. Metode Observasi
Melakukan survei ke beberapa klinik untuk mengumpulkan data-data yang akan diolah pada penelitian yaitu data pasien penderita penyakit Diabetes Mellitus.
- b. Metode Studi Pustaka
Mengumpulkan data-data teori melalui jurnal, media cetak, ataupun sumber-sumber referensi dari internet.



Gambar 1. Prosedur Penelitian

2.1.2 Analisis Sistem

Analisis sistem pada penelitian ini terbagi menjadi 3 tahapan proses yaitu:

- a. Analisis sistem berjalan yaitu melakukan analisis sistem berjalan yang digunakan saat ini khususnya sistem berjalan yang digunakan dalam melakukan deteksi dini penyakit Diabetes Mellitus.
- b. Analisis algoritma yang digunakan yaitu algoritma Decision Tree C4.5 dengan memaparkan contoh kasus sederhana implementasi algoritma tersebut dalam membentuk pohon keputusan [13] yang digunakan untuk deteksi dini penyakit Diabetes Mellitus. Pada tahap ini akan diuraikan secara lengkap cara kerja dari algoritma Decision Tree C4.5 dalam bentuk contoh kasus sederhana berupa 10 data training dan 1 data testing serta menunjukkan hasil deteksi yang didapatkan. Analisis data ini melihat hubungan antara variable yang satu dan variable yang lain [14].

2.1.3 Algoritma C4.5

Penelitian ini mengimplementasikan algoritma C4.5 untuk deteksi dini penyakit Diabetes Mellitus pada manusia. Berikut ini Tabel 3.1 menunjukkan contoh 10 dataset yang digunakan diambil dari <https://www.kaggle.com/datasets/mathchi/diabetes-data-set>.



Tabel 1. Contoh Dataset Penelitian

Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
6	148	72	35	0	33.6	0.627	50	1
1	85	66	29	0	26.6	0.351	31	0
8	183	64	0	0	23.3	0.672	32	1
1	89	66	23	94	28.1	0.167	21	0
0	137	40	35	168	43.1	2.288	33	1
5	116	74	0	0	25.6	0.201	30	0
3	78	50	32	88	31.0	0.248	26	1
10	115	0	0	0	0.0	0.261	30	0
2	197	70	45	543	30.5	0.158	53	1
8	125	96	0	0	0.0	0.232	54	1

Deskripsi variabel dataset :

- a. Pregnancies: Berapa kali hamil
- b. Glucose: Konsentrasi glukosa plasma 2 jam dalam tes toleransi glukosa oral
- c. BloodPressure: Tekanan darah diastolik (mm Hg)
- d. Skin Thickness: Ketebalan lipatan kulit trisep (mm)
- e. Insulin: Insulin serum 2 jam (mu U/ml)
- f. BMI: Indeks massa tubuh (berat badan dalam kg/(tinggi badan dalam m)^2)
- g. DiabetesPedigreeFunction: Fungsi silsilah Diabetes
- h. Age: Usia (tahun)
- i. Outcome: Variabel kelas (0 atau 1), 0 = non diabetes, 1 = diabetes

Langkah-langkah dalam membuat model C4.5 adalah sebagai berikut:

- a. Menghitung entropy dari label (Outcome) di seluruh dataset. Jumlah total sampel = 10

Jumlah sampel dengan Outcome 0 = 5 Jumlah sampel dengan Outcome 1 = 5

Entropy untuk seluruh dataset dapat dihitung dengan menggunakan rumus berikut:

$$\text{Entropy} = -(p_1 * \log_2(p_1)) - (p_2 * \log_2(p_2)) \quad [15]$$

p1 dan p2 masing-masing merupakan proporsi sampel dengan Outcome 0 dan Outcome 1.

$$p_1 = 5 / 10 = 0.5$$

$$p_2 = 5 / 10 = 0.5$$

$$\text{Entropy} = -(0.5 * \log_2(0.5)) - (0.5 * \log_2(0.5)) = 1.0$$

Jadi, entropi dari label (Outcome) di seluruh dataset adalah 1.0.

- b. Membagi dataset menjadi subset yang lebih kecil, berdasarkan nilai dari setiap atribut. Pada langkah ini, akan diperiksa setiap atribut dan membagi dataset menjadi subset yang lebih kecil berdasarkan nilai dari atribut tersebut.

Selanjutnya akan dipilih atribut dengan gain informasi tertinggi sebagai atribut pemisah pada node (akar).

- a) Atribut Pregnancies

- 1. Subset 1: Sampel dengan Pregnancies <= 1

Jumlah total sampel = 2

Jumlah sampel dengan Outcome 0 = 1

Jumlah sampel dengan Outcome 1 = 1

- 2. Subset 2: Sampel dengan Pregnancies > 1

Jumlah total sampel = 8

Jumlah sampel dengan Outcome 0 = 4

Jumlah sampel dengan Outcome 1 = 4

- b) Atribut Glucoses

- 1. Subset 1: Sampel dengan Glucose <= 137 \

Jumlah total sampel = 4

Jumlah sampel dengan Outcome 0 = 1

Jumlah sampel dengan Outcome 1 = 3

- 2. Subset 2: Sampel dengan Glucose > 137

Jumlah total sampel = 6

Jumlah sampel dengan Outcome 0 = 4 Jumlah sampel dengan Outcome 1 = 2

- c) Atribut BloodPressure

- 1. Subset 1: Sampel dengan BloodPressure <= 70

Jumlah total sampel = 3

Jumlah sampel dengan Outcome 0 = 1

Jumlah sampel dengan Outcome 1 = 2

- 2. Subset 2: Sampel dengan BloodPressure > 70

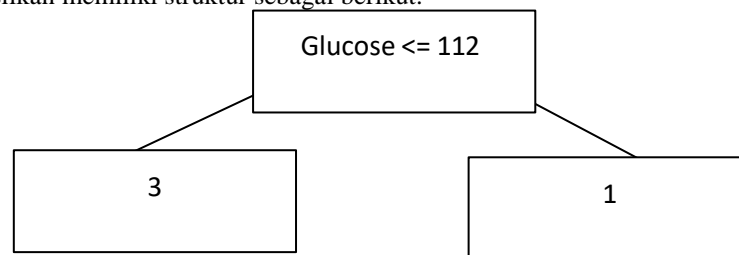
Jumlah total sampel = 7

Jumlah sampel dengan Outcome 0 = 4



- Jumlah sampel dengan Outcome 1 = 3
- d) Atribut SkinThickness
1. Subset 1: Sampel dengan SkinThickness ≤ 23
 Jumlah total sampel = 5
 Jumlah sampel dengan Outcome 0 = 3
 Jumlah sampel dengan Outcome 1 = 2
 2. Subset 2: Sampel dengan SkinThickness > 23
 Jumlah total sampel = 5
 Jumlah sampel dengan Outcome 0 = 2
 Jumlah sampel dengan Outcome 1 = 3
- e) Dan seterusnya.
- c. Menghitung gain informasi dari setiap atribut dan memilih atribut dengan gain informasi tertinggi sebagai atribut pemisah pada node (akar).
- a) Atribut Pregnancies
 $Gain = 1.0 - (2 / 10 * - (0.5 * \log_2(0.5)) - (0.5 * \log_2(0.5))) - (8 / 10 * - (0.5 * \log_2(0.5)) - (0.5 * \log_2(0.5))) = 0.0$
 - b) Atribut Glucose
 $Gain = 1.0 - (4 / 10 * - (0.25 * \log_2(0.25)) - (0.75 * \log_2(0.75))) - (6 / 10 * - (0.67 * \log_2(0.67)) - (0.33 * \log_2(0.33))) = 0.066$
 - c) Atribut BloodPressure
 $Gain = 1.0 - (3 / 10 * - (0.33 * \log_2(0.33)) - (0.67 * \log_2(0.67))) - (7 / 10 * - (0.43 * \log_2(0.43)) - (0.57 * \log_2(0.57))) = 0.002$
 - d) Atribut SkinThickness
 $Gain = 1.0 - (5 / 10 * - (0.4 * \log_2(0.4)) - (0.6 * \log_2(0.6))) - (5 / 10 * - (0.4 * \log_2(0.4)) - (0.6 * \log_2(0.6))) = 0.0$
 - e) Atribut Insulin
 $Gain = 1.0 - (3 / 10 * - (1 / 3 * \log_2(1 / 3)) - (2 / 3 * \log_2(2 / 3))) - (7 / 10 * - (0.57 * \log_2(0.57)) - (0.43 * \log_2(0.43))) = 0.089$
 - f) Dan seterusnya.
- Atribut dengan gain informasi tertinggi adalah Glucose dengan gain sebesar 0.066.
- d. Membuat node baru dengan atribut pemisah Glucose.
- e. Membuat cabang-cabang baru pada node Glucose berdasarkan nilai-nilai yang dimilikinya, yaitu ≤ 112 dan > 112 .
- f. Mengulangi proses untuk setiap subset yang terbentuk pada cabang-cabang baru sampai mencapai kondisi terminasi.
- Pada contoh ini, terdapat 6 sampel pada subset Glucose ≤ 112 dan 4 sampel pada subset Glucose > 112 . Karena setiap subset telah mencapai kondisi terminasi (semua sampel pada subset memiliki nilai Outcome yang sama), maka proses pembentukan cabang-cabang baru dihentikan dan Decision Tree selesai dibentuk.

Decision Tree yang dihasilkan memiliki struktur sebagai berikut.



Gambar 2. Decision Tree

Keterangan: Decision Tree hanya memiliki satu node (akar) dan dua daun. Nilai 3 pada daun kiri mengindikasikan bahwa terdapat 3 sampel pada subset Glucose ≤ 112 yang memiliki nilai Outcome = 0 (tidak terdiagnosis Diabetes Mellitus), sedangkan nilai 1 pada daun kanan mengindikasikan bahwa terdapat 1 sampel pada subset Glucose > 112 yang memiliki nilai Outcome = 1 (terdiagnosis Diabetes Mellitus).

Decision Tree yang dihasilkan dapat digunakan untuk melakukan klasifikasi pada data baru. Misalnya, untuk seorang pasien dengan nilai Glucose sebesar 100, Decision Tree akan menentukan bahwa pasien tersebut tidak terdiagnosis Diabetes Mellitus (Outcome = 0) karena nilai Glucose ≤ 112 . Sedangkan untuk seorang pasien dengan nilai Glucose sebesar 120, Decision Tree akan menentukan bahwa pasien tersebut terdiagnosis Diabetes Mellitus (Outcome = 1) karena nilai Glucose > 112 . Misalkan contoh terdapat sebuah data pasien yang akan diprediksi sebagai berikut:

Pregnancies: 8

Glucose: 125

Blood Pressure: 96

Skin Thickness: 0

Insulin: 0

BMI: 23,5

Diabetes Pedigree Function: 0.234

Age : 54

Maka berdasarkan pohon keputusan dimana akar utama dimulai dari Glucose sehingga untuk memprediksi data pasien dengan melihat nilai Glucose pasien tersebut yaitu sebesar 125. Berdasarkan pohon keputusan, Glucose > 112 adalah diprediksi sebagai pasien terdiagnosis Diabetes Mellitus (Outcome = 1) sehingga pasien tersebut diprediksi pasien terdiagnosis Diabetes Mellitus.

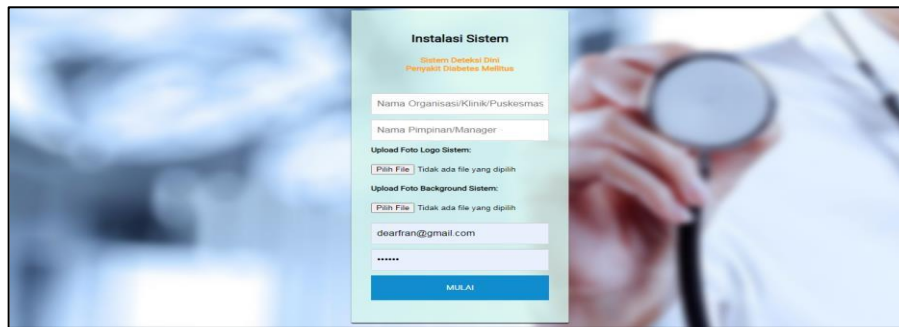
2.2 Perancangan Sistem

Perancangan sistem pada penelitian ini terbagi menjadi 2 tahapan yaitu:

1. Perancangan prototype tampilan dengan menggunakan software Balsamiq Mockup 3.
2. Perancangan basis data yang menunjukkan relasi antar tabel dari setiap basis data yang dimodelkan dengan menggunakan tools Entity Relationship Diagram (ERD).

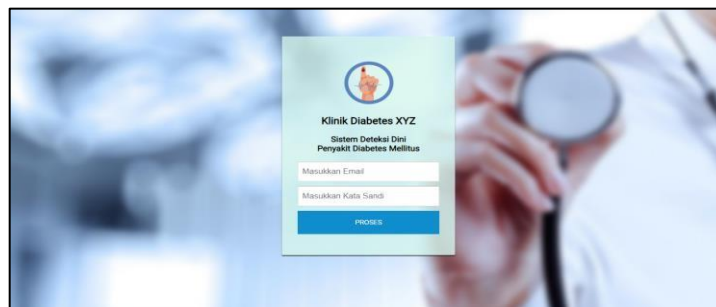
3. HASIL DAN PEMBAHASAN

Hasil penelitian berupa dibangunnya sebuah sistem deteksi dini penyakit Diabetes Mellitus pada manusia dengan mengimplementasikan algoritma C4.5. Berikut ini akan disajikan keseluruhan hasil tampilan dari sistem usulan yang akan dibangun. Tampilan paling awal adalah berupa tampilan untuk melakukan proses instalasi sistem. Setiap pelayanan medis yang hendak menggunakan sistem harus melakukan instalasi awal seperti yang ditunjukkan pada Gambar 3



Gambar 3. Tampilan Instalasi Sistem

Untuk melakukan instalasi awal sistem cukup mengisikan informasi yang tersedia pada form. Jika semua form telah terisi, maka pengguna cukup menekan tombol mulai untuk memproses form yang diisikan dan mendaftarkan informasi ke dalam basis data sistem. Setelah proses instalasi selesai dilakukan, maka selanjutnya akan ditampilkan tampilan login awal bagi user pimpinan dan admin. Gambar 4 menunjukkan hasil tampilan login.



Gambar 4. Tampilan Login

Berikut ini akan ditampilkan hasil tampilan yang akan digunakan oleh pimpinan:

a. Tampilan Dashboard Pimpinan

Tampilan dashboard pimpinan merupakan tampilan awal ketika pimpinan berhasil melakukan proses login berisikan fitur yang dapat digunakan. Gambar 5 menunjukkan hasil tampilan dashboard pimpinan.



Gambar 5. Tampilan Dashboard Pimpinan



b. Tampilan Tambah Data Rekam Medis

Tampilan tambah data rekam medis merupakan tampilan yang berisikan form untuk melakukan penambahan data rekam medis. Pada tampilan ini algoritma C4.5 akan dijalankan untuk menentukan prediksi data rekam medis yang dimasukkan apakah tergolong pasien sehat atau pasien yang didiagnosa kemungkinan menderita penyakit Diabetes Mellitus berdasarkan atribut-atribut rekam medis pasien yang dimasukkan. Algoritma C4.5 akan belajar dari data-data sebelumnya dan kemudian menentukan kecocokan data tersebut. Gambar 6 menunjukkan hasil tampilan tambah data rekam medis.



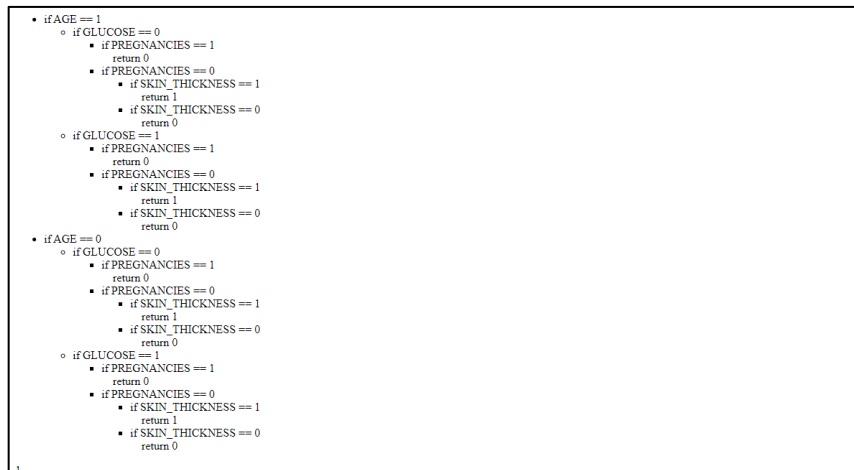
Gambar 6. Tampilan Tambah Data Rekam Medis

Untuk menjalankan hasil prediksi deteksi dini dapat dilakukan dengan menekan tombol training prediksi algoritma C4.5 maka sistem otomatis akan menampilkan hasil prediksi pada combobox yang tersedia serta teks hasil prediksi. Setelah hasil penelitian diuraikan, maka tahapan berikutnya adalah melakukan pembahasan mengenai hasil penelitian yang didapatkan. Pembahasan dibagi menjadi 2 yaitu pembahasan mengenai algoritma C4.5 yang diimplementasikan untuk deteksi dini penyakit Diabetes Mellitus serta pembahasan mengenai kelebihan dan kekurangan dari hasil penelitian yang didapatkan. Pada penelitian ini diimplementasikan algoritma C4.5 untuk deteksi dini penyakit Diabetes Mellitus pada manusia dengan menggunakan jumlah data training sebanyak 177 data rekam medis pasien seperti yang ditunjukkan pada Gambar 7 berikut.

id_trainingrekam medis	umur	pregnancies	glucose	bloodpressure	skintickness	insulin	bmi	dpf	klasifikasi	created_at
158	23	1	109	56	21	135	25.20	0.833	0	2023-05-28
159	22	2	88	74	19	53	29.00	0.229	0	2023-05-28
160	47	17	163	72	41	114	40.90	0.817	1	2023-05-28
161	36	4	151	90	38	0	29.70	0.294	0	2023-05-28
162	45	7	102	74	40	105	37.20	0.204	0	2023-05-28
163	27	0	114	80	34	285	44.20	0.167	0	2023-05-28
164	21	2	100	64	23	0	29.70	0.368	0	2023-05-28
165	32	0	131	88	0	0	31.60	0.743	1	2023-05-28
166	41	6	104	74	18	156	29.90	0.722	1	2023-05-28
167	22	3	148	66	25	0	32.50	0.256	0	2023-05-28
168	34	4	120	68	0	0	29.60	0.709	0	2023-05-28
169	29	4	110	66	0	0	31.90	0.471	0	2023-05-28
170	29	3	111	90	12	78	28.40	0.495	0	2023-05-28
171	36	6	102	82	0	0	30.80	0.180	1	2023-05-28
172	29	6	134	70	23	130	35.40	0.542	1	2023-05-28
173	25	2	87	0	23	0	28.90	0.773	0	2023-05-28
174	23	1	79	60	42	48	43.50	0.678	0	2023-05-28
175	33	2	75	64	24	55	29.70	0.370	0	2023-05-28
176	36	8	179	72	42	130	32.70	0.719	1	2023-05-28
177	42	6	85	78	0	0	31.20	0.382	0	2023-05-28

Gambar 7. Tampilan Data Rekam Medis di Database

Berdasarkan data training yang diolah menghasilkan pohon keputusan C4.5 seperti yang ditunjukkan pada Gambar 8.



Gambar 8. Pohon Keputusan Yang Terbentuk

Berdasarkan Gambar 8 dapat dilihat bahwa node tertinggi adalah pada atribut umur yang artinya umur memiliki atribut yang cukup penting dalam menentukan seseorang kemungkinan menderita penyakit Diabetes Mellitus atau tidak. Mengingat penyakit ini lebih sering menyerang orang dengan usia yang sudah lansia. Kemudian atribut kedua yang menjadi pertimbangan dalam prediksi deteksi dini penyakit Diabetes Mellitus adalah Glucose atau kadar gula dalam darah. Glucose yang tinggi tentu memiliki kemungkinan yang lebih besar untuk menderita penyakit Diabetes Mellitus. Selanjutnya atribut pendukung berikutnya adalah jumlah pasien melahirkan juga menjadi pertimbangan serta Skin Thickness. Terdapat atribut-atribut yang tidak masuk dalam pohon keputusan berdasarkan data training yang digunakan yakni Insulin, BMI, dan Diabetes Pedigree Function. Berdasarkan hasil pengujian algoritma C4.5 tergolong akurat dalam melakukan prediksi deteksi dini penyakit Diabetes Mellitus apabila mengikuti rule pohon keputusan yang terbentuk.

4. KESIMPULAN

Tahapan dari penelitian ini adalah menguraikan kesimpulan yang didapatkan dari penelitian yang dilakukan antara lain, Sistem informasi yang dibangun dapat membantu dunia medis dalam melakukan deteksi dini penyakit Diabetes Mellitus pada manusia melalui fitur prediksi yang mengimplementasikan algoritma C4.5. Hasil pengujian algoritma C4.5 menunjukkan bahwa algoritma tersebut tergolong akurat dalam melakukan prediksi deteksi dini penyakit Diabetes Mellitus apabila mengikuti rule pohon keputusan yang terbentuk.

REFERENCES

- [1] A. Vilorio, Y. Herazo-Beltran, D. Cabrera, and O. B. Pineda, "Diabetes Diagnostic Prediction Using Vector Support Machines," in *Procedia Computer Science*, Elsevier B.V., 2020, pp. 376–381. doi: 10.1016/j.procs.2020.03.065.
- [2] W. Apriliah, I. Kurniawan, M. Baydhowi, and T. Haryati, "Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest," *SISTEMASI*, vol. 10, no. 1, p. 163, Jan. 2021, doi: 10.32520/stmsi.v10i1.1129.
- [3] Y. Nora Marlim, L. Suryati, and N. Agustina, "Deteksi Dini Penyakit Diabetes Menggunakan Machine Learning dengan Algoritma Logistic Regression," 2022.
- [4] I. W. Karmana, "Penerapan Pola Hidup Sehat untuk Mencegah Diabetes Mellitus," *Panthera : Jurnal Ilmiah Pendidikan Sains dan Terapan*, vol. 3, no. 1, pp. 1–6, Jan. 2023, doi: 10.36312/pjipst.v3i1.141.
- [5] W. Yusnaeni, "Penerapan Algoritma C4.5 Dalam Prediksi Resiko Diabetes Tahap Awal (Early Stage Diabetes)," *Jurnal Teknik Komputer AMIK BSI*, vol. III, no. 1, pp. 56–60, 2022, doi: 10.31294/jtk.v4i2.
- [6] M. Dewi, Y. Yellyanda, and D. Ulfa, "Edukasi Penatalaksanaan Diabetes terhadap Manajemen Perawatan Diri Pasien Diabetes Mellitus Tipe II," *Jurnal Keperawatan Silampari*, vol. 5, no. 2, pp. 981–990, Apr. 2022, doi: 10.31539/jks.v5i2.3583.
- [7] N. Sunanto and G. Falah, "PENERAPAN ALGORITMA C4.5 UNTUK MEMBUAT MODEL PREDIKSI PASIEN YANG MENGIDAP PENYAKIT DIABETES," *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 7, no. 2, pp. 208–216, Jul. 2022, doi: 10.36341/rabit.v7i2.2435.
- [8] I. Made, S. Ramayu, F. Susanto, and G. S. Mahendra, "PENERAPAN DATA MINING DENGAN ALGORITMA C4.5 DALAM PEMESANAN OBAT GUNA MENINGKATKAN KEUNTUNGAN APOTEK," 2022. [Online]. Available: <http://senada.idbbali.ac.id>
- [9] A. Novriandy, "Implementasi Algoritma Naive Bayes dan Algoritma C4.5 dalam Klasifikasi Kelayakan Bantuan UMKM," *Media Online*, vol. 4, no. 1, pp. 208–217, 2023, doi: 10.30865/klik.v4i1.1099.
- [10] A. Sephami, I. E. Hendrawan, C. Rozikin, and S. Karawang, "ANALISIS DATA HASIL DIAGNOSA UNTUK KLASIFIKASI GANGGUAN KEPERIBADIAN MENGGUNAKAN ALGORITMA C4.5," 2021. [Online]. Available: <https://www.kaggle.com/fedesoriano/heart>
- [11] K. L. Kohsasih and Z. Situmorang, "Analisis Perbandingan Algoritma C4.5 dan Naive Bayes Dalam Memprediksi Penyakit Cerebrovascular," *Jurnal Informatika*, vol. 9, no. 1, pp. 13–17, Apr. 2022, doi: 10.31294/inf.v9i1.11931.
- [12] D. Gifu, "The use of decision trees for analysis of the epilepsy," in *Procedia Computer Science*, Elsevier B.V., 2021, pp. 2844–2853. doi: 10.1016/j.procs.2021.09.055.



- [13] V. A. Zulaiha, N. Rahaningsih, O. Nurdiawan, and A. Rinaldi Dikananda, "Implementasi Algoritma C.45 dalam Klasifikasi Bahan Baku," vol. 3, no. 1, pp. 41–45, 2022.
- [14] A. Purwanto, M. Asbari, and T. I. Santoso, "Analisis Data Penelitian Marketing: Perbandingan Hasil," *Journal of Industrial Engineering & Management Research*, vol. 2, no. 4, doi: 10.7777/jiemar.
- [15] I. Putu et al., "REKOMENDASI PENGAMBILAN MATA KULIAH PILIHAN UNTUK MAHASISWA SISTEM INFORMASI MENGGUNAKAN ALGORITME DECISION TREE," vol. 6, no. 3, 2019, doi: 10.25126/jtiik.2019.6892.