

Implementation of the C4.5 Algorithm for Classifying the Causes of Diabetes at the Rejang Lebong District Hospital

Implementasi Algoritma C4.5 Untuk Klasifikasi Penyebab Penyakit Diabetes Di RSUD Kabupaten Rejang Lebong

Anisya Fatmala^{*1}, Murlena², Nurfitri Andayani³

¹ Universitas Pat Petulai

² Universitas Putra Indonesia YPTK Padang

³ Universitas Budi Luhur

Email: ¹annisafatmala0@gmail.com, ²ana.murlena@gmail.com, ³nurfitri.pietly@gmail.com

Abstrak

Rumah Sakit Umum Daerah Kabupaten Rejang Lebong beralamatkan di Jl. Dua Jalur, Durian Depun, Kec.Merigi, Kabupaten Rejang Lebong. Rumah sakit merupakan institusi pelayanan kesehatan yang menyelenggarakan pelayanan kesehatan perorangan secara paripurna yang menyediakan pelayanan rawat inap, rawat jalan, dan gawat darurat. Dalam penelitian, data sampel yang digunakan adalah 376 data pasien penderita penyakit diabetes dari RSUD Kabupaten Rejang Lebong. Setelah data yang diperlukan diperoleh, kemudian ditentukan atribut-atribut yang diperlukan pada penelitian ini antara lain : usia penderita penyakit diabetes, kadar gula (mg/dl), jumlah kolestrol total (mg/dl). Diabetes merupakan penyakit metabolis yang kronis yang mana penyakit diabetes tidak menghasilkan jumlah insulin yang cukup atau tubuh pasien tidak sanggup memanfaatkan insulin dengan baik sehingga menyebabkan glukosa dalam tubuh mengalami jumlah yang berlebihan. Tujuan penelitian ini untuk menganalisa data faktor penyebab penyakit diabetes pada RSUD Kabupaten Rejang Lebong dengan penerapan *data mining* menggunakan metode algoritma C4.5. Algoritma C4.5 adalah algoritma klasifikasi pohon keputusan yang banyak digunakan karena memiliki keunggulan yang luar biasa dibandingkan algoritma lainnya. Algoritma C4.5 dimulai dengan proses pemilihan atribut dengan *gain* tertinggi adalah akar pohon, kemudian melakukan percabangan untuk setiap nilai. Algoritma C4.5 yaitu salah satu metode untuk membentuk pohon keputusan berdasarkan data pelatihan yang sudah disediakan. Pada penelitian ini dataset yang digunakan sebanyak tujuh puluh enam dan atribut yang dipakai jenis kelamin, umur, glukosa, kolesterol dan jumlah kasus dengan satu kelas penentu ya atau tidak. Dengan hasil pengolahan data menggunakan aplikasi *Rapidminer Studio* dengan hasil nilai akurasi data 100%.

Kata Kunci: Data Mining, Klasifikasi, C4.5, Diabetes

Abstract

The Rejang Lebong District General Hospital is located at Jl. Two Lanes, Durian Depun, Merigi District, Rejang Lebong Regency. A hospital is a health service institution that provides complete individual health services, providing inpatient, outpatient and emergency services. In the research, the sample data used was 376 data from patients suffering from diabetes from the Rejang Lebong District Hospital. After the required data was obtained, the attributes required in this research were determined, including: age of the diabetes sufferer, sugar level (mg/dl), total cholesterol (mg/dl). Diabetes is a chronic metabolic disease in which diabetes does not produce sufficient amounts of insulin or the patient's body is unable to utilize insulin properly, causing excessive amounts of glucose in the body. The aim of this research is to analyze data on factors causing diabetes at the Rejang Lebong District Hospital by applying data mining using the C4.5 algorithm method. The C4.5 algorithm is a decision tree classification algorithm that is widely used because it has extraordinary advantages compared

to other algorithms. The C4.5 algorithm starts with the process of selecting the attribute with the highest gain, namely the root of the tree, then branching for each value. The C4.5 algorithm is a method for forming a decision tree based on the training data that has been provided. In this study, the dataset used was seventy-six and the attributes used were gender, age, glucose, cholesterol and number of cases with one class determining yes or no. With the results of data processing using the Rapidminer Studio application with a data accuracy value of 100%.

Keywords: Data Mining, Classification, Decision Tree, Diabetes

PENDAHULUAN

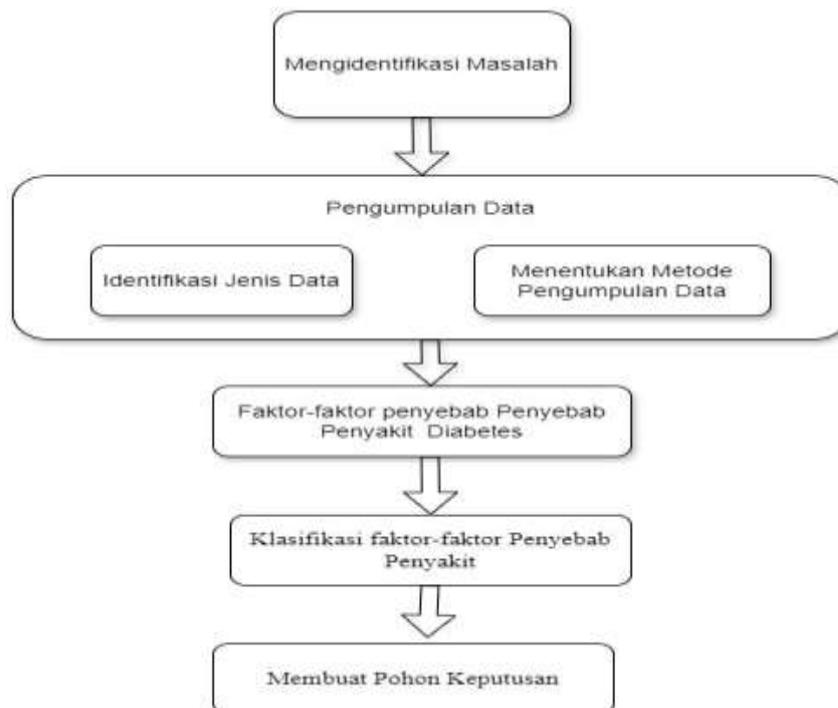
Diabetes merupakan penyakit metabolis yang kronis yang mana penyakit diabetes tidak menghasilkan jumlah insulin yang cukup atau tubuh pasien tidak sanggup memanfaatkan insulin dengan baik sehingga menyebabkan glukosa dalam tubuh mengalami jumlah yang berlebihan, kondisi ini sering kali dirasakan setelah komplikasi terjadi pada organ tubuh. Insulin merupakan hormon yang diciptakan oleh pankreas, yang mempunyai fungsi seperti menyeimbangkan glukosa dari makanan yang dimakan kemudian mengalir ke sel-sel mulai dari darah di dalam tubuh manusia yang menghasilkan energi. Diabetes merupakan penyakit ketika insulin tidak mampu diproduksi oleh pankreas atau tubuh manusia tidak dapat menggunakan insulin yang dibuat oleh pankreas.

Makanan yang mengandung karbohidrat diproses menjadi glukosa dalam darah, *glukosa* dibantu insulin masuk ke dalam sel. Di waktu tubuh tidak mampu memproduksi insulin dan bahkan tidak mampu menggunakan dengan benar menyebabkan meningkatnya kadar *glukosa* dalam darah. Sehingga pada waktu jangka panjang, kadar *glukosa* tinggi dalam darah dapat merusak organ pada tubuh dan kegagalan fungsi organ dan jaringan. Gejala umum diabetes adalah penurunan berat badan mendadak, obesitas, penglihatan kabur, penyembuhan tertunda, dan lain-lain. Untuk menyikapi masalah ini perlu adanya pendeteksian penyakit diabetes, dengan adanya pendeteksian ini diharapkan dapat menurunkan resiko komplikasi pada pasien diabetes di waktu mendatang. Dengan kriteria yang digunakan untuk klasifikasi berdasarkan, tingkat kadar gula, usia, jenis kelamin. Pencatatan terhadap penyakit ini banyak dilakukan agar dapat dilakukan pencegahan, salah satu pencatatan yang bisa dilakukan adalah dengan memanfaatkan teknik klasifikasi dengan *data mining*.

Data mining adalah sebuah metode untuk melakukan akuisisi pengetahuan, dengan *data mining* informasi-informasi sebuah data dapat diekstrak. *Data mining* yaitu suatu prosedur menciptakan ikatan di mana memiliki arti pola, dan kecondongan dengan mengamati kelompok data besar yang berada dalam *storage* dengan menggunakan teknik identifikasi pola. Adapun metode yang biasanya dioperasikan pada data mining antara lain: deskripsi atau penggambaran, prediksi atau ramalan, *clustering*, klasifikasi dan asosiasi, dan estimasi. Klasifikasi merupakan sebuah proses untuk menciptakan fungsi atau model menjelaskan kelas pada data atau konsep guna untuk memprediksi kelas dari sebuah objek yang labelnya belum didapatkan. Teknik klasifikasi dimanfaatkan untuk meramal seseorang yang terjangkit penyakit diabetes dan tidak terjangkit. Beberapa algoritma yang dapat digunakan untuk perhitungan proses klasifikasi, algoritma klasifikasi diantaranya adalah *Decision Tree C4.5*, *naïve Bayes*, dan *K-Nearest Neighbor (KNN)*.

METODE PENELITIAN

Metode yang digunakan dalam penelitian ini adalah metode penelitian kuantitatif, untuk bahan analisis peneliti menggunakan data pasien penderita penyakit diabetes di RSUD Kabupaten Rejang Lebong. Adapun kerangka berpikir pada penelitian ini yaitu:



Gambar 1 kerangka berpikir

- a. Mengidentifikasi masalah
Identifikasi masalah adalah langkah pertama dalam pemecahan masalah dan pengambilan keputusan yang efektif. Langkah awal yang dilakukan peneliti adalah melakukan pengamatan di RSUD Kab. Rejang Lebong, dan mengumpulkan informasi.
- b. Mengumpulkan data
Mengumpulkan data adalah langkah penting dalam penelitian, analisis, dan pengambilan keputusan. Langkah yang dilakukan dalam mengumpulkan data adalah menentukan tujuan pengumpulan data, mengidentifikasi jenis data yang dibutuhkan, dan menentukan metode pengumpulan data, metode yang digunakan adalah algoritma C4.5.
- c. Mengelola data
Mengelola data adalah proses yang mencakup pengumpulan, penyimpanan, analisis, dan pengamanan data untuk memastikan keakuratan, keamanan, dan aksesibilitas data.
- d. Klasifikasi data adalah proses pengelompokan data berdasarkan karakteristik atau kategori tertentu untuk memudahkan pengelolaan, analisis, dan keamanan.
- e. Perhitungan akhir menggunakan Algoritma C4.5
- f. Membuat pohon keputusan adalah tahap untuk mengetahui apakah Algoritma C4.5 yang digunakan sudah sesuai untuk di gunakan.

HASIL DAN PEMBAHASAN

Dalam penelitian ini data yang digunakan sebagai bahan penelitian adalah data pasien penderita penyakit diabetes sebanyak 376 data yang diambil pada bulan Januari - Juli 2023. Data yang diperoleh kemudian dikelompokkan sesuai dengan atribut yang dibutuhkan, seperti jenis kelamin, umur, glukosa, kolesterol dan jumlah pasien berdasarkan tingkat kadar kolesterol dan glukosa.

Setelah dilakukan beberapa proses *selection* dan *cleaning* maka total data pasien penderita penyakit diabetes tersebut menjadi 76 data berdasarkan keterangan ya= 43 dan Tidak = 33. Kemudian mengelompokkan dan menghitung semua total atribut dan disimpan dalam kolom "jumlah" dan total "atribut". Berdasarkan dari banyak data set peminjaman yang terdapat dalam kolom "normal" dan "tinggi". seperti tabel data set di bawah ini:

Tabel 1 data set

No	Jenis Kelamin	Umur	Glukosa	Kolesterol	Jumlah	Ya/Tidak
1	laki-laki	Lansia	200	190	6	Ya
2	laki-laki	Lansia	200	200	1	Ya
3	laki-laki	Lansia	200	210	3	Ya
4	laki-laki	Lansia	200	220	1	Ya
5	laki-laki	Dewasa	200	210	4	Ya
6	laki-laki	Dewasa	205	210	2	Ya
7	laki-laki	Lansia	210	190	2	Ya
8	laki-laki	Lansia	210	200	5	Ya
9	laki-laki	Lansia	210	220	3	Ya
10	laki-laki	Lansia	210	225	1	Ya
11	laki-laki	Dewasa	210	200	5	Ya
12	laki-laki	Dewasa	215	200	1	Ya
13	laki-laki	Lansia	220	200	5	Ya
14	laki-laki	Lansia	220	210	5	Ya
15	laki-laki	Dewasa	220	200	2	Ya
16	laki-laki	Dewasa	220	230	2	Ya
17	laki-laki	Dewasa	225	200	1	Ya
18	laki-laki	Lansia	230	200	7	Ya
19	laki-laki	Dewasa	230	210	7	Ya
20	laki-laki	Dewasa	235	210	1	Ya
...
70	Perempuan	Lansia	275	240	1	Tidak
71	Perempuan	Dewasa	275	240	5	Tidak
72	Perempuan	Lansia	280	230	5	Tidak
73	Perempuan	Dewasa	280	230	4	Tidak
74	Perempuan	Lansia	300	250	3	Tidak
75	Perempuan	Lansia	300	250	6	Tidak
76	Perempuan	Lansia	370	240	5	Tidak

Selanjutnya yaitu menghitung total data yang terdiri atas 76 total data berdasarkan keterangan ya = 43 dan Tidak = 33. Kemudian mengelompokkan dan menghitung semua total atribut dan disimpan dalam kolom "jumlah" dan total "atribut". Berdasarkan dari banyak data set pasien penderita penyakit diabetes yang terdapat dalam kolom "ya" dan tidak".

Tabel 2 total data, atribut dan kelas

Keterangan		Jumlah kasus (S)	Tidak (S1)	Ya (S2)
Total		76	33	43
Jenis kelamin	Laki-laki	48	20	28
	Perempuan	28	13	15
Umur	Lansia	40	18	22
	Dewasa	36	15	21
Glukosa	200-240 mg/dl	43	0	43
	> 240 mg/dl	33	33	0
Kolesterol	190-210 mg/dl	33	3	30
	220-235 mg/dl	30	18	12
	240-320 mg/dl	13	12	1

Proses Pembentukan *Decesion Tree* Menggunakan Algoritma C4.5

Data yang sudah diklasifikasikan kemudian diolah menggunakan *Rapidminer* menggunakan *performance* yang berfungsi sebagai validasi dan rehabilitasi data untuk mencari keakuratan data. Data yang akurat maka akan dilakukan pengolahan data untuk mencari hasil dari penyakit diabetes dengan menggunakan *Rapidminer* dan mengambil keputusan dari hasil pengolahan data menggunakan *Rapidminer* yang dilakukan.

Dalam teknik algoritma C4.5 memilih atribut sebagai *node* (akar) dengan mencari terlebih dahulu jumlah kasus keseluruhan, jumlah kasus untuk keputusan diabetes normal dan jumlah kasus untuk keputusan diabetes tinggi erta *entropy* dari semua kasus yang dibagi berdasarkan atribut jenis kelamin, usia, glukosa, dan kolesterol. Setelah itu dilakukan perhitungan *Gain* untuk masing-masing atribut. Hasil perhitungan dapat dilihat pada tabel dibawah ini.

Tabel 3 Perhitungan *Node 1*

Keterangan		Jumlah kasus (S)	Tidak (S1)	Ya (S2)	Entropy	Gain
Total		76	33	43	0.987475	
Jenis kelamin	Laki-laki	48	20	28	0.979869	0.73567451
	Perempuan	28	13	15	0.996317	
Umur	Lansia	40	18	22	0.992774	0.9291105
	Dewasa	36	15	21	0.979869	
Glukosa	200-240 mg/dl	43	0	43	0	0
	> 240 mg/dl	33	33	0	0	
Kolesterol	190-210 mg/dl	33	3	30	0.439497	0.82229046
	220-235 mg/dl	30	18	12	0.970951	
	240-320 mg/dl	13	12	1	0.391244	

Kolom *Entropy* pada baris total dihitung dengan persamaan sebagai berikut :

$$\text{Entropy (S)} = \sum_{i=1}^n - p_i * \log_2 p_i$$

$$\text{Entropy (total)} = ((-33/76)*\log_2(33/76)+(-43/76)*\log_2(43/76)) \\ = 0.987475$$

Kemudian pada nilai *Gain* dihitung dengan persamaan sebagai berikut :

$$\text{Gain(S,A)} = \text{Entropy(S)} - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy(S}_i)$$

$$\text{Gain (total, jenis kelamin)} = 0.98748 - (48/76 * 0.97987) + (28/76 * 0.99632) \\ = 0.7356745$$

Dari hasil perhitungan pada tabel 3 atribut yang menjadi *node* (akar) adalah umur dimana nilai *gain* dari umur lebih tinggi dari atribut-atribut lainnya. Berdasarkan nilai *entropy* dari atribut diatas, hanya sub atribut rendah yang memperoleh satu keputusan, maka perlu dilakukan perhitungan lebih lanjut. Untuk hasil perhitungan elanjutnya dapat dilihat pada table berikut ini :

Tabel 4. Hasil Perhitungan *node* 1.1

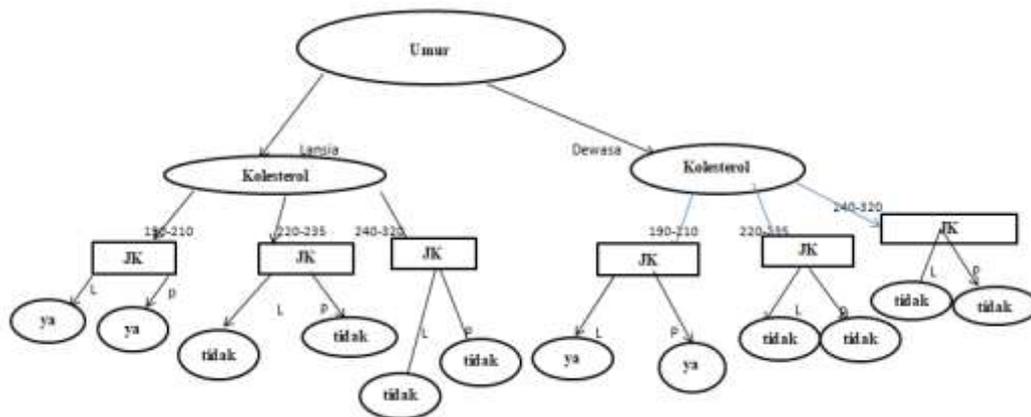
Keterangan		Jumlah kasus (S)	Tidak (S1)	Ya (S2)	Entropy	Gain
Total		76	33	43	0.987475	
Jenis kelamin						0.735675
	Laki-laki	48	20	28	0.979869	
	Perempuan	28	13	15	0.996317	
Glukosa						0
	200-240 mg/dl	43	0	43	0	
	> 240 mg/dl	33	33	0	0	
Kolesterol						0.82229
	190-210 mg/dl	33	3	30	0.439497	
	220-235 mg/dl	30	18	12	0.970951	
	240-320 mg/dl	13	12	1	0.391244	

Dari hasil perhitungan pada tabel 4 atribut yang menjadi *node* (akar) adalah kolesterol dimana nilai *gain* dari kolesterol lebih tinggi dari atribut-atribut lainnya. Berdasarkan nilai *entropy* dari atribut diatas, hanya sub atribut rendah yang memperoleh satu keputusan, maka perlu dilakukan perhitungan lebih lanjut. Untuk hasil perhitungan selanjutnya dapat dilihat pada tabel berikut ini :

Tabel 5 hasil perhitungan *node* 1.2

Keterangan		Jumlah kasus (S)	Tidak (S1)	Ya (S2)	Entropy	Gain
Total		76	33	43	0.987475	
Jenis kelamin						0.735675
	Laki-laki	48	20	28	0.979869	
	Perempuan	28	13	15	0.996317	
Glukosa						0
	200-240 mg/dl	43	0	43	0	
	> 240 mg/dl	33	33	0	0	

Dari hasil perhitungan pada tabel 5 atribut yang menjadi *node* (akar) adalah jenis kelamin dimana nilai *gain* dari jenis kelamin lebih tinggi dari nilai *gain* glukosa. Masing-masing dari sub atribut sudah menghasilkan klaifikasi. Dari hasil tersebut dapat digambarkan pohon keputusan akhir seperti



gambar berikut ini :

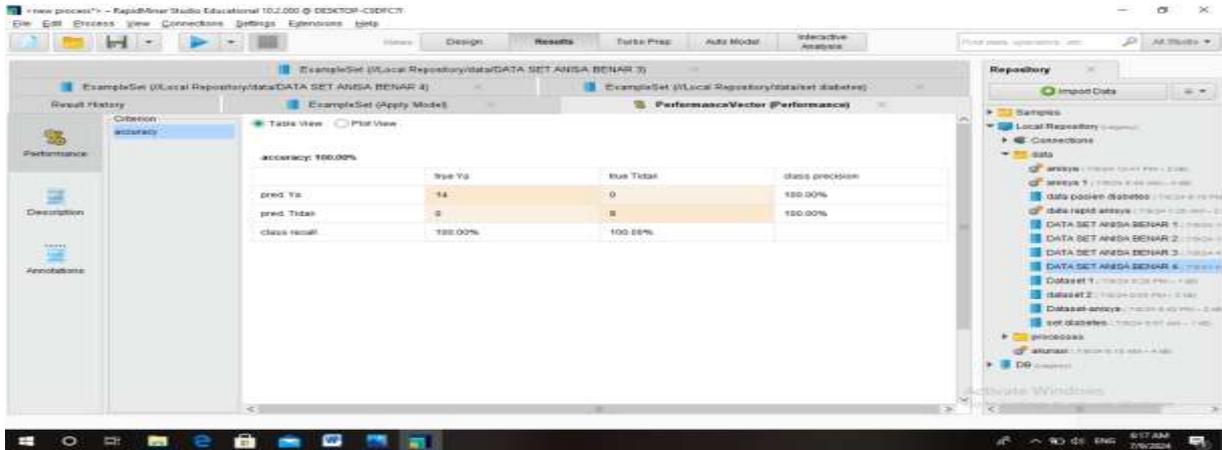
Gambar 2. Pohon Keputusan Hasil *Node* Akhir

Tahapan Implementasi *Decision tree*

Pengujian terhadap hasil perhitungan manual menggunakan *software RapidMiner* dilakukan melalui beberapa proses sebagai berikut :

1. Awal menjalankan aplikasi rapidminer, install terlebih dahulu aplikasi di laptop masing-masing setelah terinstall maka open aplikasi setelah klik import data untuk memproses data yang akan dianalisis menggunakan aplikasi rapidminer.
2. Setelah tampil menu awal aplikasi rapidminer klik import data yang terletak sebelah kiri pada aplikasi rapidminer.
3. Setelah mengimport data maka akan tampil data yang akan diimport untuk diproses.
4. Setelah memasukkan data untuk diproses lalu klik button *next* yang ada di kanan bawah pada aplikasi *rapidminer*.
5. Langkah selanjutnya *drag* data yang telah disimpan, pada menu operators yang ada dibagian kiri aplikasi ketikkan *set role* lalu *drag* pada layar proses, kembali pada menu operators ketikkan *decision tree* lalu *drag* pada layar proses.
6. Langkah selanjutnya klik ikon *run* atau ikon *pause* yang ada di kiri atas aplikasi.
7. Untuk melihat *decission tree* maka klik *tree (decission tree)* pada bagian kanan atas.

Dari proses pengujian data menggunakan *software RapidMiner* diatas didapat hasil akurasi penerapan algoritma C4.5 yaitu 100,00% untuk melihat nilai akurasi data dilihat pada gambar berikut ini :



Gambar 3. Nilai Akurasi Algoritma C4.5 Menggunakan *RapidMiner*

KESIMPULAN

Adapun kesimpulan yang dapat diambil pada implementasi algoritma C4.5 untuk klaifikasi penyebab penyakit diabetes di RSUD kabupaten rejang lebong adalah sebagai berikut :

1. Implementasi *data mining* dengan menggunakan algoritma C4.5 dapat membantu peneliti dalam mengklaifikasikan penyebab penyakit diabet di RUD kabupaten rejang lebong berdaarkan variable-variabel yang diperlukan yaitu umur, jenis kelamin, glukosa, dan kolesterol. Pada penelitian ini menggunakan data sebanyak 376 data yang diklasifikasikan menjadi 76 data dengan keputusan "Ya" = Normal sebanyak 43 data dan keputusan "Tidak" = Tinggi sebanyak 33 data.
2. Dengan menerapkan algoritma C4.5 pada *software RapidMiner* menghasilkan nilai perhitungan yang sama pada perhitungan manual. Berdasarkan pengolahan data menggunakan *software RapidMiner* didapat nilai akurasi istem ebear 100,%.

DAFTAR PUSTAKA

- Akhiri, S., & Windarto, T. (2019). *Data mining: Konsep, teknik, dan aplikasinya*. Yogyakarta: Pustaka Pelajar.
- Az-Zahra, L. (2023). *Klasifikasi Faktor-Faktor Penyebab Penyakit Diabetes Mellitus di RSUD Kota Pariaman Tahun 2022 Menggunakan Algoritma C4. 5* (Doctoral dissertation, Universitas Negeri Padang).
- Carudin, C., Marisa, M., Murnawan, M., Reba, F., Koibur, M. E., Thantawi, A. M., ... & Wattimena, F. Y. (2024). *Buku Ajar Data Mining*. PT. Sonpedia Publishing Indonesia.
- Ente, D. R., Thamrin, S. A., Arifin, S., Kuswanto, H., & Andreza, A. (2020). *Klasifikasi Faktor-Faktor Penyebab Penyakit Diabetes Melitus Di Rumah Sakit Unhas*

- Menggunakan Algoritma C4. 5. *Indonesian Journal of Statistics and Its Applications*, 4(1), 80-88.
- Fadhillah, R. P., Rahma, R., Sepharni, A., Mufidah, R., Sari, B. N., & Pangestu, A. (2022). Klasifikasi Penyakit Diabetes Mellitus Berdasarkan Faktor-Faktor Penyebab Diabetes menggunakan Algoritma C4. 5. *Jurnal Ilmiah Penelitian dan Pembelajaran Informatika*.
- Fadlan, C., Ningsih, S., & Windarto, A. P. (2018). Penerapan Metode Naïve Bayes Dalam Klasifikasi Kelayakan Keluarga Penerima Beras Rastra. *JUTIM (Jurnal Teknik Informatika Musirawas)*, 3(1), 1-8.
- Gultom, J. P., & Rikki, A. (2020). Implementasi Data Mining menggunakan Algoritma C-45 pada Data Masyarakat Kecamatan Garoga untuk Menentukan Pola Penerima Beras Raskin. *Kumpul. Artik. Karya Ilm. Fak. Ilmu Komput*, 2(01), 11-19.
- Habibah, N. N., Nazir, A., Iskandar, I., Syafria, F., Oktavia, L., & Syurfi, I. (2023). Pemodelan Klasifikasi Untuk Menentukan Penyakit Diabetes dengan Faktor Penyebab Menggunakan Decision Tree C4. 5 Pada Wanita. *Jurnal Sistem Komputer dan Informatika (JSON)*, 4(4), 654-661.
- Larose, D. T., & Larose, C. D. (2020). *Discovering knowledge in data: an introduction to data mining*. Wiley..
- Murlena, D. A., & Apriana, D. (2022). Penerapan Data Mining Untuk Memprediksi Ketersediaan Stok Produk HNI HPAI Menggunakan Algoritma C4. 5. *Arcitech: Journal of Computer Science and Artificial Intelligence*, 2(1), 19-32.
- Murlena, M., & Syahindra, W. (2020). Data Mining Pengolahan Penempatan Library Books Menggunakan Metode Association Rule Dengan Algoritma Apriori. *Jurnal INSTEK (Informatika Sains Dan Teknologi)*, 5(2), 199-208.
- Priyanda, R., Agustina, T. S., Ariantini, N. S., Rusmayani, N. G. A. L., Aslindar, D. A., Ningsih, K. P., ... & Wicaksono, D. (2022). *Metodologi Penelitian Kuantitatif*. Pradina Pustaka.
- Ridwan, A. (2022). Penerapan Algoritma C4. 5 Untuk Klasifikasi Penyakit Diabetes Mellitus. *Jurnal Bisnis Digital Dan Sistem Informasi*, 3(2), 41-48.
- Rizal, M., Syahaf, M. Z., Priyambodo, S. R., & Ramdhani, Y. (2023). Optimasi algoritma Naive Bayes Menggunakan Forward Selection untuk klasifikasi penyakit ginjal kronis. *Naratif: Jurnal Nasional Riset, Aplikasi dan Teknik Informatika*, 5(1), 71-80.
- Safitri, M., & Praba, A. D. (2024). PREDIKSI PENYAKIT DIABETES DENGAN MENGGUNAKAN ALGORITMA C4. 5. *JIKA (Jurnal Informatika)*, 8(1), 74-81.
- Santoso, M., Al-Akbar, B., Nurjaya, H., Ramadhan, S. A., Rizky, N. A., & Fadillah, A. (2023). Klasifikasi Potensi Penyakit Jantung Menggunakan Algoritma C4. 5. *Jurnal INSAN Journal of Information System Management Innovation*, 3(2), 96-103.
- Sari, Z. D. R., & Jasmir, J. (2024). Penerapan Data Mining Untuk Prediksi Penyakit Diabetes Menggunakan Algoritma C4. 5 Zudyanti Dwi Rahma Sari1, Ja. *Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM)*, 4(1), 827-834.

Sholeh, M., Nurnawati, E. K., & Lestari, U. (2023). Penerapan Data Mining dengan Metode Regresi Linear untuk Memprediksi Data Nilai Hasil Ujian Menggunakan RapidMiner. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 8(1), 10-21.