

Prediksi Kanker Payudara Melalui Penerapan Algoritma C4.5

¹Jefri Junifer Pangaribuan, ²Romindo, ^{3*}Mirza Ilhami, ⁴Segar Napitupulu, ⁵Wenripin Chandra

^{1,2}Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pelita Harapan, Medan, Indonesia

^{3,4}Teknik Informatika, Fakultas Informatika, Universitas Mikroskil, Medan, Indonesia

⁵Informatika, Fakultas Ilmu Komputer, Universitas Pelita Harapan, Medan, Indonesia

¹jefri.pangaribuan@uph.edu, ^{2*}romindo@uph.edu, ³mirza.ilhami@mikroskil.ac.id,

⁴segarn70cs5@gmail.com, ⁵wenripin@lecturer.uph.edu

ABSTRACT – Cancer is one of the causes of death in both developed and developing countries. Breast cancer is cancer that originates from cells in the breast, either from milk duct cells, milk-producing gland cells, or other tissues. This disease was also the main cause of cancer deaths in Indonesia in 2012. Therefore, it is very important to detect breast cancer early so that treatment can be carried out immediately. In practice, in the medical world, medical record data such as breast cancer is often stored for various purposes. However, in reality, the process of storing and processing medical record data in several hospitals still does not utilize computer media so that data is often lost or damaged. In addition, medical record data that is recorded and collected is usually processed and used as knowledge to make predictions. Because of these problems, it is necessary to build an information system with the application of Data Mining in the world of health, especially managing medical record data. The aim of research using the C4.5 prediction algorithm is to produce a decision tree that has an acceptable level of accuracy and is efficient in handling discrete or numerical attributes. The research results show that the application of the Data Mining Decision Tree C4.5 algorithm which is used to predict breast cancer using 116 medical record training data can produce 4 rules. The accuracy level of the Data Mining Decision Tree C4.5 algorithm is quite accurate because out of 10 experiments there were 9 experiments that obtained correct prediction results according to existing rules.

Keywords: Data Mining; Breast Cancer Prediction Information System; C4.5 Decision Tree Algorithm; Medical Records.

ABSTRAK – Kanker merupakan salah satu penyebab kematian baik di negara maju maupun di negara yang sedang berkembang. Kanker payudara merupakan kanker yang berasal dari sel-sel yang terdapat di payudara, bisa dari sel-sel saluran air susu atau sel-sel kelenjar penghasil air susu atau jaringan lain. Penyakit ini juga menjadi penyebab kematian utama karena kanker di Indonesia pada tahun 2012. Oleh sebab itu sangat penting untuk dilakukan deteksi dini terhadap penyakit kanker payudara agar dapat segera dilakukan pengobatan. Pada praktiknya, di dunia kedokteran seringkali data rekam medis seperti penyakit kanker payudara disimpan untuk berbagai tujuan. Namun kenyataannya, proses penyimpanan dan pengolahan data rekam medis di beberapa rumah sakit masih belum memanfaatkan media komputer sehingga data sering hilang ataupun rusak. Selain itu, data rekam medis yang tercatat dan terkumpul biasanya diolah dan dimanfaatkan menjadi sebuah pengetahuan untuk melakukan prediksi. Oleh karena permasalahan tersebut, maka perlu dibangun sebuah sistem informasi dengan penerapan Data Mining dalam dunia kesehatan khususnya pengelolaan data rekam medis. Tujuan penelitian menggunakan penerapan algoritma prediksi C4.5 adalah dapat menghasilkan pohon keputusan yang memiliki tingkat akurasi yang dapat diterima dan efisien dalam menangani atribut yang bertipe diskret atau numerik. Hasil penelitian menunjukkan bahwa penerapan algoritma Data Mining Decision Tree C4.5 digunakan untuk melakukan prediksi penyakit kanker payudara menggunakan data training rekam medis sebanyak 116 data dapat menghasilkan rule sebanyak 4 rule. Tingkat keakuratan dari algoritma Data Mining Decision Tree C4.5 cukup akurat karena dari 10 percobaan terdapat 9 percobaan yang mendapatkan hasil prediksi yang tepat sesuai rule yang ada.

Kata Kunci: Algoritma Decision Tree C4.5; Data Mining; Sistem Informasi Prediksi Kanker Payudara; Rekam Medis.

1. PENDAHULUAN

Kanker merupakan salah satu penyebab kematian baik di negara maju maupun di negara yang sedang berkembang. Beban kanker meningkat di negara yang sedang berkembang akibat meningkatnya gaya hidup yang

berkaitan dengan kanker, misalnya merokok, aktivitas fisik yang kurang, dan konsumsi diet [1]. Berdasarkan data dari GLOBOCAN tahun 2012 dirilis oleh Badan Kesehatan Dunia (WHO) menyebutkan bahwa kanker paru-paru menempati peringkat pertama dalam jumlah kasus baru sebesar 2,094 juta kasus diseluruh dunia.



Jumlah kasus baru tertinggi berikutnya adalah kanker payudara, kanker kolorektal, kanker prostat, dan kanker lambung [2].

Kanker payudara merupakan kanker yang berasal dari sel-sel yang terdapat di payudara, bisa dari sel-sel saluran air susu atau sel-sel kelenjar penghasil air susu atau jaringan lain. Kanker ini terjadi hampir seluruhnya pada wanita, tetapi dapat juga terjadi pada pria. Di Amerika Serikat, diperkirakan 231.840 wanita didiagnosis kanker payudara pada tahun 2015. Pada tahun 2012, kanker payudara menempati peringkat kedua setelah kanker paru sebagai penyebab kematian utama karena kanker di Amerika Serikat. Di Indonesia, diperkirakan 51.136 wanita didiagnosis kanker payudara pada tahun 2015. Penyakit ini juga menjadi penyebab kematian utama karena kanker di Indonesia pada tahun 2012. Oleh sebab itu sangat penting untuk dilakukan deteksi dini terhadap penyakit kanker payudara agar dapat segera dilakukan pengobatan [1].

Pada praktiknya, di dunia kedokteran seringkali data rekam medis seperti penyakit kanker payudara disimpan untuk berbagai tujuan. Namun kenyataannya, proses penyimpanan dan pengolahan data rekam medis di beberapa rumah sakit masih belum memanfaatkan media komputer sehingga data-data rekam medis yang tersimpan sering hilang ataupun rusak serta memakan waktu lama ketika akan melakukan proses pencarian. Selain itu, data rekam medis yang tercatat dan terkumpul biasanya diolah dan dimanfaatkan menjadi sebuah pengetahuan dimana pengetahuan tersebut dapat digunakan untuk melakukan prediksi. Namun untuk mengelola kumpulan data rekam medis menjadi pengetahuan tentunya tidak dapat dilakukan secara konvensional dikarenakan proses pengelolaan secara konvensional memiliki tingkat keakuratan yang kurang baik sehingga hasil kesimpulan prediksi yang diambil pun tentunya akan berbeda dan tidak begitu akurat.

Kanker payudara, penyebab kedua terbesar kematian pada wanita, sering terdeteksi pada fase yang sulit diobati. Biopsi memerlukan waktu lama, namun Machine Learning dengan Algoritma SVM mempercepat diagnosis dengan tingkat akurasi tinggi. Penelitian ini menggunakan SVM untuk aplikasi ML yang memprediksi sifat ganas atau jinak sel kanker payudara berdasarkan pola data masa lalu, menunjukkan hasil yang menggembirakan [3].

Kanker payudara adalah penyebab utama kematian global, termasuk di Indonesia. Penelitian ini bertujuan membuat model prediksi dengan akurasi di atas 99% untuk mendeteksi kanker payudara secara dini. Hal ini diharapkan dapat membantu dokter memberikan perawatan awal. Meskipun model ini mencapai 98,25% akurasi, target 99% belum tercapai karena beberapa parameter yang belum optimal. [4].

Oleh karena permasalahan tersebut, maka perlu dibangun sebuah sistem informasi dengan penerapan Data Mining dalam dunia kesehatan khususnya pengelolaan data rekam medis. Data Mining merupakan suatu proses penggalian data atau penyaringan data dengan memanfaatkan kumpulan data dengan ukuran yang cukup besar melalui serangkaian proses untuk

mendapatkan informasi yang berharga dari data tersebut. Data Mining itu sendiri disalurkan dalam bentuk algoritma-algoritma kompleks yang memiliki tingkat keakuratan yang cukup baik. Terdapat beberapa penelitian yang mencoba membangun sistem pengelolaan data rekam medis yaitu penelitian analisa data rekam medis menggunakan teknik Data Mining Association dengan algoritma Clustering [5]. Selanjutnya penelitian untuk membangun sebuah sistem klasterisasi data rekam medis dengan menggunakan algoritma K-Means Clustering [6]. Terdapat juga penerapan algoritma C4.5 dalam melakukan prediksi penyakit yaitu penelitian yang dilakukan oleh Handayani, Nurlalah, Raharjo, & Ramdani pada tahun 2019 membahas mengenai prediksi penyakit Liver menggunakan algoritma Decision Tree C4.5 dan Neural Network. Hasil penelitian menunjukkan bahwa algoritma Decision Tree C4.5 lebih akurat dengan akurasi sebesar 75,56% dan algoritma Neural Network dengan akurasi sebesar 74,17% [7].

Berdasarkan uraian tersebut, maka penulis tertarik untuk melakukan penelitian dengan mengambil judul "Prediksi Kanker Payudara Melalui Penerapan Algoritma C4.5".

2. DASAR TEORI

2.1. Penyakit Kanker Payudara

Kanker atau sering disebut dengan tumor secara umum dibedakan menjadi dua macam yaitu jinak (benign) dan ganas (malignant). Pada level jinak, tumor akan memiliki kondisi dan perkembangan yang tidak bersifat kanker, dimana penyakit ini dapat terdeteksi namun tidak menyebar dan merusak jaringan lain di sekitarnya. Sedangkan pada level ganas, tumor akan menyebar dan merusak jaringan dan organ di sekitarnya. Kanker ganas jika tidak cepat ditangani akan berdampak buruk bagi penderita sehingga deteksi dini pada penyakit kanker sangatlah penting bagi keselamatan penderita. Dengan mengetahui jinak atau ganasnya kanker yang diderita pasien sedini mungkin maka dapat segera ditangani dengan sesuai dan dapat diantisipasi karena bukan tidak mungkin dapat berujung pada kematian bagi penderita [8].

Kanker payudara, biasanya terbentuk di sel-sel payudara yang tumbuh tanpa kendali. Ini bisa terjadi pada semua jenis gender. Di Indonesia, jumlah kasus kanker payudara menempati peringkat pertama di antara jenis kanker lainnya, menjadi penyebab kematian utama. Dengan tingginya jumlah kasus kematian ini dan sifat kanker payudara yang tidak memandang gender, penting bagi pria dan wanita untuk meningkatkan kesadaran akan kesehatan mereka dengan melakukan deteksi dini dan mengurangi risiko terjadinya kanker [9].

Kanker payudara adalah kelompok kanker yang paling umum di seluruh dunia dan juga merupakan penyebab utama kematian pada wanita yang terkena penyakit tersebut. Dari data Global World Health Organization pada tahun 2018 menyatakan bahwa kasus kanker yang

terjadi di Indonesia 58,256 kasus atau 16,7% dari total 348,809 kasus kanker. Angka penyakit kanker payudara di dunia mencapai 42,1 orang per 100 ribu penduduk, rata-rata kematian akibat kanker payudara mencapai 17 orang per 100 ribu penduduk. Dari data tersebut dapat dikatakan bahwa kanker payudara merupakan penyebab pertama kematian pada wanita yang terkena penyakit tersebut [10].

2.2. Data Mining

Data mining merupakan suatu proses untuk mengekstraksi lalu mengidentifikasi informasi dengan cara statistik, kecerdasan buatan, machine learning, maupun matematika. Dengan adanya data mining maka didapatkan kemudahan untuk mencari informasi bisnis yang penting dari database yang jumlah datanya sangat besar. Adapun poin utama yang dapat dihasilkan dari data mining yaitu yang pertama prediksi tren, yang dapat berguna bagi proses bisnis yang dilakukan saat ini. Lalu selanjutnya data mining dapat digunakan untuk menemukan pola-pola unik dalam informasi. Terakhir, data mining dapat digunakan untuk membuat keputusan kritis dalam strategi [11].

2.3. Algoritma Decision Tree

Decision tree atau pohon keputusan merupakan salah satu teknik klasifikasi. Decision tree adalah top-down pohon rekursif dari algoritma induksi, yang menggunakan ukuran seleksi atribut untuk memilih atribut yang diuji. Dengan pohon keputusan, manusia dapat dengan mudah melihat hubungan antara faktor-faktor yang mempengaruhi suatu masalah. Pohon keputusan ini juga dapat menganalisa nilai suatu informasi yang terdapat dalam suatu alternatif pemecahan masalah. Pohon keputusan adalah salah satu metode klasifikasi yang mudah untuk diaplikasikan oleh manusia. Pohon keputusan bisa melakukan prediksi dengan menggunakan struktur pohon. Pohon keputusan mengubah data menjadi pohon keputusan serta menjadi beberapa aturan keputusan. Berikut ini akan ditampilkan gambar bentuk dari pohon keputusan [12].

2.4. Algoritma C4.5

Algoritma C4.5 adalah algoritma yang pada umumnya digunakan guna untuk membentuk Decision Tree. Kelebihan algoritma C4.5 adalah dapat menghasilkan pohon keputusan yang memiliki tingkat akurasi yang dapat diterima dan efisien dalam menangani atribut yang bertipe diskret atau numerik [13]. Menurut Bramer, secara global algoritma C4.5 untuk menciptakan pohon keputusan sebagai berikut yaitu [14]:

1. Pemilihan atribut yang akan dijadikan sebagai akar.
2. Pembuatan cabang untuk setiap nilai.
3. Pembagian kasus dalam cabang.
4. Repetisi proses sampai setiap cabang guna agar semua kasus pada cabang memiliki kelas yang sama.

Menurut Elisa, terdapat formula ukuran atribut C4.5 antara lain [12]:

$$\text{Info}(D) = -\sum_{i=1}^m p_i \log_2(p_i) \quad (1)$$

Dengan Info (D) merupakan informasi yang dibutuhkan untuk mengklasifikasikan label kelas sebuah tuple di D. p_i adalah peluang bukan nol dengan sebuah tuple acak di D. Fungsi log menggunakan basis 2, karena informasi yang dikodekan dalam bit. Info (D) juga dikenal sebagai entropy. Nilai entropy yang dihasilkan untuk mengklasifikasi tuple dari D berdasarkan partisi oleh A adalah:

$$\text{Info}_A(D) = \sum_{j=1}^J \frac{|D_j|}{D} \times \text{Info}(D_j) \quad (2)$$

Untuk mendapatkan nilai Information Gain pada atribut A selanjutnya digunakan rumus:

$$\text{GAIN}(A) = \text{Info}(D) - \text{Info}(D_j) \quad (3)$$

Gain (A) menyatakan berapa banyak cabang yang akan diperoleh pada A. Atribut A dengan information gain tertinggi. Information Gain (A), dipilih sebagai atribut pada node N.

3. METODOLOGI

Metodologi penelitian yang digunakan dalam merancang sistem pada penelitian ini terdiri dari:

1. Pengumpulan Data

a. Studi Literatur

Penulis mengumpulkan informasi melalui buku, maupun bahan referensi lainnya yang berhubungan dengan sistem informasi serta bahan referensi yang berhubungan dengan kegiatan pencatatan data rekam medis pasien.

b. Studi Lapangan

Penulis secara langsung mengamati proses operasional yang dilakukan di klinik-klinik di kota Medan. Dalam mengumpulkan data dilakukan tanya jawab dengan beberapa manajer klinik.

2. Analisis Sistem

Pada tahap ini dilakukan analisis sistem dengan menguraikan cara kerja algoritma Decision Tree C4.5 dalam melakukan prediksi penyakit kanker payudara dan melakukan pemodelan analisis sistem dengan menggunakan Use Case Diagram.

3. Desain Sistem

Pada tahap ini dilakukan perancangan sistem dengan menggunakan software Balsamiq Mockup 3.

4. Pembangunan Sistem

Adapun beberapa bahas pemrograman yang digunakan dalam membangun sistem yaitu:

- a. Framework Laravel untuk Back End Programing.
- b. Framework Bootstrap untuk Front End Programing.

4. HASIL DAN PEMBAHASAN

4.1 Hasil Proses Prediksi dengan Decision Tree C4.5

Pada tahapan ini, akan dilakukan analisa proses berupa cara kerja dari metode Decision Tree dalam melakukan prediksi penyakit kanker payudara. Metode Decision Tree C4.5 akan mempelajari data-data kanker payudara dari dataset, kemudian melalui hasil

pembelajaran tersebut maka akan diolah dan diprediksi apakah data rekam medis pasien baru yang akan masuk merupakan Healthy Controls (1) atau Patiens (2).

Contoh kasus sederhana akan disediakan 10 data rekam medis pasien sebagai data training dan peserta ke 11 akan dilakukan prediksi.

Tabel 1. Contoh Data Rekam Medis Pasien Beserta Kriterianya

No	U	B	G	I	H	L	A	R	M	K	NP
1	34	23.50	70	10.56	0.74	8.81	13.11	11.79	733.80	1	Andi
2	29	20.69	92	16.64	4.47	8.84	26.72	4.30	1227.91	2	Budi
3	25	23.12	91	4.33	0.78	17.94	23.67	6.71	586.17	1	Cindy
4	24	21.37	77	41.61	15.29	9.88	36.06	4.50	887.16	2	Daddi
5	38	21.11	92	22.03	1.56	6.70	17.95	4.66	1102.11	2	Erdi
6	69	22.85	92	3.19	1.14	6.83	20.32	4.53	667.93	2	Santi
7	60	32.04	77	9.67	7.84	6.96	38.04	9.61	581.31	2	Gigi
8	77	23.80	118	28.68	2.63	4.31	7.78	8.49	864.97	2	Herry
9	76	22.00	97	10.40	3.78	4.47	5.46	11.77	695.75	2	Iriani
10	76	23.00	83	4.17	1.10	17.13	5.10	23.03	358.62	2	Jimmy
11	66	25.12	78	14.65	20.63	14.57	8.60	28.03	960.25	?	Nindy

Keterangan:

U = Umur (Tahun)

B = BMI (Body Mass Index) (Kg/M2)

G = Glucose (mg/dL)

I = Insulin (μ U/mL)

H = HOMA

L = Leptin (ng/mL)

A = Adiponectin (μ g/mL)

R = Resistin (ng/mL)

K = Klasifikasi

1 = Healthy Controls (pasien diprediksi sehat dan hanya perlu pengontrolan saja)

2 = Patients (pasien diprediksi terindikasi penyakit kanker payudara dan perlu penanganan lebih lanjut ke dokter spesialis)

NP = Nama Pasien

Tentukan Nindy diklasifikasikan/diprediksi sebagai Healthy Controls (1) atau Patiens (2)?

Pada tabel 3.1, akan dibentuk klasifikasi pohon keputusan melalui penerapan algoritma C4.5, namun tahap awal adalah melakukan klasifikasi berdasarkan kriteria penilaian. Klasifikasi tersebut ditunjukkan pada tabel 2.

Tabel 2. Klasifikasi Berdasarkan Kriteria Penilaian

Atribut	Nilai Atribut	Total (Jumlah Orang)	Klasifikasi (Jumlah Orang)	
			2	1
Total	Total	10	8	2
Umur (Tahun)	20-50	5	3	2
	>50	5	5	0
BMI (Kg/M2)	<17.00	0	0	0
	17.00-22.99	5	5	0
	23.00-27.00	4	2	2
	>27.00	1	1	0
Glucose (mg/dL)	50 - 109	9	7	2
	110 - 144	1	1	0
	145 - 180	0	0	0
	>180	0	0	0
Insulin (μ U/mL)	\leq 25	8	6	2
	>25	2	2	0
HOMA	<1.0	2	0	2
	1.0-2.9	4	4	0
	>2.9	4	4	0
Leptin (ng/mL)	<5	2	2	0
	5-20	8	6	2
	>20	0	0	0
Adiponectin (μ g/mL)	<15	4	3	1
	15-30	4	3	1
	>30	2	2	0
Resistin (ng/mL)	<15	9	7	2

	15-30	1	1	0
	>30	0	0	0
MCP-1 (pg/dL)	<500	1	1	0
	500-1500	9	7	2
	>1500	0	0	0

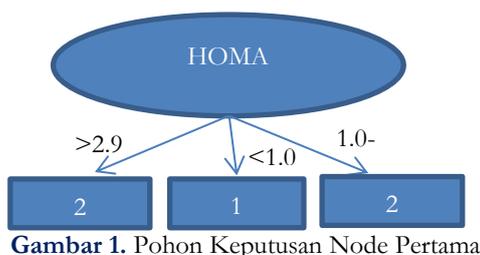
Selanjutnya akan dilakukan perhitungan Entropy dan Information Gain berdasarkan tabel 2 tersebut untuk menghasilkan node pertama sesuai dengan rumus dari metode Decision Tree C4.5.

Tabel 3. Perhitungan Node Pertama

Atribut	Nilai Atribut	Entropy	Gain	Information Gain
Total	Total	$-\frac{8}{10} * \log_2(\frac{8}{10}) - \frac{2}{10} * \log_2(\frac{2}{10})$ 0.721928		
Umur (Tahun)	20-50	$-\frac{3}{5} * \log_2(\frac{3}{5}) - \frac{2}{5} * \log_2(\frac{2}{5})$ 0.970950	$(\frac{5}{10} * 0.485475 + \frac{5}{10} * 0.29726469)$ 0.970950594	$0.721928 - 0.970950594$ -0.249022594
	>50	$-\frac{5}{5} * \log_2(\frac{5}{5})$ 0	$(\frac{5}{10} * 0)$ 0	$0.721928 - 0$ 0.721928
				0.236452
BMI (Kg/M2)	<17.0	0	$(\frac{0}{10} * 0)$ 0	$0.721928 - 0$ 0.721928
	17.00-22.99	$-\frac{5}{5} * \log_2(\frac{5}{5})$ 0	$(\frac{5}{10} * 0)$ 0	$0.721928 - 0$ 0.721928
	23.00-27.00	$-\frac{2}{4} * \log_2(\frac{2}{4}) - \frac{2}{4} * \log_2(\frac{2}{4})$ 1.000000	$(\frac{4}{10} * 0.400000 + \frac{4}{10} * 0.400000)$ 0.800000	$0.721928 - 0.800000$ -0.078072
	>27.0	$-\frac{1}{1} * \log_2(\frac{1}{1})$ 0	$(\frac{1}{10} * 0)$ 0	$0.721928 - 0$ 0.721928
				0.321928
Glucose (mg/dL)	50 - 109	$-\frac{7}{9} * \log_2(\frac{7}{9}) - \frac{2}{9} * \log_2(\frac{2}{9})$ 0.764204	$(\frac{9}{10} * 0.687784 + \frac{1}{10} * 0.5591069)$ 0.764204506	$0.721928 - 0.764204506$ -0.042276506
	110 - 144	$-\frac{1}{1} * \log_2(\frac{1}{1})$ 0	$(\frac{1}{10} * 0)$ 0	$0.721928 - 0$ 0.721928
	145 - 180	0	$(\frac{0}{10} * 0)$ 0	$0.721928 - 0$ 0.721928
	>180	0	$(\frac{0}{10} * 0)$ 0	$0.721928 - 0$ 0.721928
Insulin (µU/mL)	<=25	$-\frac{6}{8} * \log_2(\frac{6}{8}) - \frac{2}{8} * \log_2(\frac{2}{8})$ 1.245215	$(\frac{8}{10} * 0.649022 + \frac{2}{10} * 0.49961725)$ 0.811278124	$0.721928 - 0.811278124$ -0.089350124
	>25	$-\frac{2}{2} * \log_2(\frac{2}{2})$ 0	$(\frac{2}{10} * 0)$ 0	$0.721928 - 0$ 0.721928
				0.072905
HOMA	<1.0	$-\frac{2}{2} * \log_2(\frac{2}{2})$ 0	$(\frac{2}{10} * 0)$ 0	$0.721928 - 0$ 0.721928
	1.0-2.9	$-\frac{4}{4} * \log_2(\frac{4}{4})$ 0	$(\frac{4}{10} * 0)$ 0	$0.721928 - 0$ 0.721928
	>2.9	$-\frac{4}{4} * \log_2(\frac{4}{4})$ 0	$(\frac{4}{10} * 0)$ 0	$0.721928 - 0$ 0.721928
			0.721928	
			0.094943	
			Tertinggi 1	

<i>Leptin</i> (ng/mL)	<5	$-2/2 * \log_2(2/2)$	0	$(2/10 * 0)$	0	0.721928	0.072905 5953256 7
			0.811278	$(8/10 * 0.649022)$	0.649022	0.9494292	
		$-6/8 * \log_2(6/8)$	1245215	0.811278124	49961725	5 - $(2/10 * 0)$	
	5-20	$2/8 * \log_2(2/8)$	73	521573	84	$(8/10 * 0.811278)$	
						124521573 - $(0/10 * 0)$	
	>20	= 0	0	$(0/10 * 0)$	0	0.721928	
<i>Adiponectin</i> (µg/mL)			0.811278	$(4/10 * 0.324511)$	0.324511	0.9494292	0.072905 5953256 7
		$-3/4 * \log_2(3/4)$	1245215	0.811278124	24980862	0.9494292	
	<15	$1/4 * \log_2(1/4)$	73	521573	92	5 - $(4/10 * 0.811278)$	
	15-30	$-3/4 * \log_2(3/4)$	1245215	0.811278124	24980862	$(4/10 * 0.324511)$	
		$1/4 * \log_2(1/4)$	73	521573	92	3) - $(4/10 * 0.811278)$	
	>30	$2/2 * \log_2(2/2)$	0	$(2/10 * 0)$	0	0.721928	
<i>Resistin</i> (ng/mL)			0.764204	$(9/10 * 0.687784)$	0.687784	0.9494292	0.034144 0390322 3
		$-7/9 * \log_2(7/9)$	5065674	0.764204506	0.5591069	0.9494292	
	<15	$2/9 * \log_2(2/9)$	37	567437	33	5 - $(9/10 * 0.764204)$	
	15-30	$-1/1 * \log_2(1/1)$	0	$(1/10 * 0)$	0	$(9/10 * 0.764204)$	
						7) - $(1/10 * 0.5065674)$	
	>30	= 0	0	$(0/10 * 0)$	0	0.721928	
MCP-1 (pg/dL)	<500	$-1/1 * \log_2(1/1)$	0	$(1/10 * 0)$	0	0.721928	0.034144 0390322 3
			0.764204	$(9/10 * 0.687784)$	0.687784	0.9494292	
	500-1500	$-7/9 * \log_2(7/9)$	5065674	0.764204506	0.5591069	5 - $(1/10 * 0.764204)$	
		$2/9 * \log_2(2/9)$	37	567437	33	$(9/10 * 0.764204)$	
						7) - $(0/10 * 0.5065674)$	
	>1500	0	0	$(0/10 * 0)$	0	0.721928	

Pada tabel 3, didapatkan atribut HOMA memiliki nilai Information Gain terbesar dibandingkan atribut lainnya yaitu 0.721928094943, sehingga dijadikan root pertama. Dikarenakan seluruh nilai atribut bernilai 0, maka pohon keputusan selesai. Berikut ini gambaran pohon keputusannya. Gambar 1 Pohon Keputusan Node Pertama.

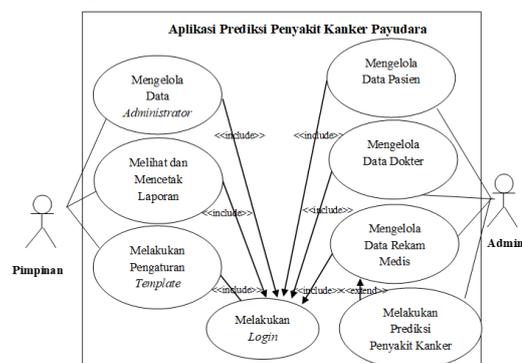


Gambar 1. Pohon Keputusan Node Pertama

Maka berdasarkan pohon keputusan yang terbentuk pada gambar 3.1, Nindy memiliki HOMA sebesar 20.63, maka Nindy akan diprediksi sebagai Patients (2).

4.2 Hasil Pemodelan Kebutuhan Sistem

Pemodelan kebutuhan sistem akan dimodelkan dengan menggunakan salah satu tools UML yaitu Use Case Diagram. Kebutuhan sistem akan menunjukkan fitur-fitur yang akan tersedia dalam sistem usulan yang akan dibangun.



Gambar 2. Use Case Diagram Sistem Usulan

Terdapat 2 aktor yang terlibat dalam sistem yaitu Pimpinan dan Admin. Gambar 2 adalah definisi 2 aktor yang terlibat dalam sistem.

Tabel 4. Definisi Aktor

No.	Aktor	Deskripsi
1.	Pimpinan	Aktor yang merupakan pimpinan rumah sakit yang berfokus dalam memantau data-data rumah sakit serta mencetak laporan

No.	Aktor	Deskripsi
2.	Admin	Aktor yang merupakan admin rumah sakit yang bertugas memasukkan data-data master dan rekam medis

Selanjutnya, adalah membuat skenario dari setiap butir Use Case agar dapat dilihat secara lengkap interaksi antara aktor dengan sistem usulan yang dibangun.

Tabel 5. Skenario Use Case: Melakukan Login

Melakukan <i>Login</i>		Nama Use Case
Pimpinan dan Admin		Aktor
Melakukan <i>login</i> agar dapat menggunakan fitur-fitur yang terdapat pada aplikasi		Deskripsi
Aksi Aktor	Respon Sistem	
1. Masuk ke halaman utama aplikasi		
	2. Menampilkan halaman utama berupa <i>form</i> masuk/ <i>login</i>	
3. Melakukan pengisian <i>form</i> masuk/ <i>login</i> dan menekan tombol proses pada sistem yang tersedia		Normal Flow Event
	4. Sistem akan melakukan pengecekan basis data dan jika sesuai akan ditampilkan tampilan beranda aktor dan jika tidak sesuai akan ditampilkan pesan kesalahan kombinasi kata sandi tidak tepat	

4.3 Pembahasan

Selanjutnya akan dilakukan pembahasan terkait kelebihan dan kekurangan dari sistem informasi prediksi penyakit kanker payudara yang telah selesai dibangun. Kelebihan dari sistem yang dibangun antara lain:

1. Sistem yang dibangun memiliki fitur-fitur yang cukup lengkap dalam melakukan penyimpanan data rekam medis, baik data pasien, data dokter, maupun data rekam medis.
2. Terdapat adanya fitur cetak laporan yang memudahkan admin dan pimpinan dalam melihat dan mencetak seluruh laporan seperti laporan pasien, laporan dokter, laporan rekam medis, dan catatan history rekam medis pasien.
3. Sistem yang dibangun menerapkan algoritma Data Mining Decision Tree C4.5 dalam melakukan prediksi

penyakit kanker payudara serta hasil prediksi dapat ditampilkan secara cepat dan akurat.

Kekurangan dari sistem yang dibangun antara lain:

1. Penelitian ini masih menggunakan satu dataset bersumber dari UCI *Machine Learning Repository* dan belum dibandingkan dengan sumber dataset lainnya.
2. Variabel kriteria (atribut) penyakit kanker payudara yang digunakan dalam prediksi dengan algoritma C4.5 masih terbatas.

Pada penelitian ini, diterapkan metode Data Mining Decision Tree C4.5 dalam membangun sebuah sistem informasi prediksi penyakit kanker payudara. Proses prediksi dilakukan ketika melakukan penambahan data rekam medis pasien baru pada form penambahan data seperti yang terlihat pada gambar 3.

RS. XYZ Beranda Data Master Data Rekam Medis Admin Randy Otton

Tambah Data Rekam Medis

Nama Pasien *
Kevin Tin. Krakatau No. 51 (29 Mar 1946)

Nama Dokter *
Dr. dr. Andi Huang, Sp.Pd/In. Sarabaya No. 54 (03 Feb 1954)

Informasi Tes

Body Mass Index (BMI) (kg/m²) *
24.00

Glucose (mg/dL) *
200

Insulin (µU/mL) *
2.61

HOMA *
1.17

Leptin (ng/mL) *
21.23

Adiponectin (µg/mL) *
7.60

Resistin (ng/mL) *
11.51

MCP-1 (pg/dL) *
417.12

Hasil Prediksi Yang Didapat Adalah Pasien (Pasien Kanker Payudara)
Hasil Prediksi *
Pasien (Pasien Kanker Payudara)

PREDIKSI KANKER PAYUDARA
SIMPAN
BATALKAN

Gambar 4. Tampilan Form Penambahan Data Rekam Medis Baru

Pada gambar 4 terdapat informasi rekam medis pasien dimana glucose pasien tersebut adalah 200 sehingga pasien diklasifikasi adalah seorang pasien sesuai dengan rule metode C4.5. Berdasarkan hasil pengolahan dataset didapatkan bahwa atribut glucose merupakan node awal sehingga nilai atribut glucose menjadi penentu utama klasifikasi. Atribut glucose diasumsikan sebagai berikut:

$$\begin{aligned} 50-109 &= 0 \\ 110-144 &= 1 \\ 145-180 &= 2 \\ >180 &= 3 \end{aligned}$$

Berdasarkan gambar 4.28 glucose pasien adalah di atas >180 sebesar 200 sehingga data rekam medis tersebut diprediksi menjadi patients (2).

5. KESIMPULAN

Kesimpulan dari hasil penelitian yang telah dilakukan adalah sistem informasi prediksi penyakit kanker payudara yang dibangun memiliki fitur untuk melakukan pencatatan data rekam medis pasien sehingga mencegah informasi data rekam medis rusak ataupun hilang serta sistem yang dibangun memiliki fitur untuk melakukan pencarian data rekam medis secara cepat. Penerapan algoritma Data Mining Decision Tree C4.5 digunakan untuk melakukan prediksi penyakit kanker payudara dengan keakuratan yang baik apabila jumlah data training yang digunakan tepat dan banyak.

Saran untuk penelitian selanjutnya diharapkan dapat menambah jumlah dataset agar lebih menguji keakuratan pada hasil prediksi yang diberikan algoritma C4.5.

DAFTAR PUSTAKA

- [1] Yuliana, "Risiko dan Deteksi Dini Kanker Payudara," *CDK-261*, vol. XLV, no. 2, pp. 144-149, 2018.
- [2] GLOBOCAN, "Global Cancer Statistics 2012," 2012. [Online]. Available: <https://acsjournals.onlinelibrary.wiley.com/doi/pdf/10.3322/caac.21262#:~:text=Based%20on%20GLOBOCAN%20estimates%2C%20about,65%25%20of%20cancer%20deaths%20worldwide..> [Accessed 11 Maret 2021].
- [3] C. Chazar and W. Widhiaputra, "Machine Learning Diagnosis Kanker Payudara Menggunakan Algoritma Support Vector Machine," *INFORMASI (Jurnal Informatika dan Sistem Informasi)*, vol. XII, no. 1, pp. 67-80, 2020.
- [4] I. Setiawan and E. K. Nurnawati, "Penerapan Algoritma Decision Tree Untuk Memprediksi Kanker Payudara menggunakan Data Mining dan Machine Learning," *Jurnal Dinamika Informatika*, vol. XI, no. 2, pp. 103-112, 2022.
- [5] E. Kurniawan, "Analisa Data Rekam Medis Menggunakan Teknik Data Mining Association

- Rules Dengan Algoritma Clustering," Semarang, 2017.
- [6] A. Ali, "Klasifikasi Data Rekam Medis Pasien Menggunakan Metode K-Means Clustering Di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo," *Jurnal Matrik*, vol. XIX, no. 1, pp. 186-195, 2019.
- [7] P. Handayani, E. Nurlelah, M. Raharjo and P. M. Ramdani, "Prediksi Penyakit Liver Dengan Menggunakan Metode Decision Tree dan Neural Network," *Journal of Computer Engineering System and Science (CESS)*, vol. IV, no. 1, pp. 55-59, 2019.
- [8] B. A. Farahdiba and Y. S. Nugroho, "Klasifikasi Kanker Payudara Menggunakan Algoritma Gain Ratio," *Jurnal Teknik Elektro*, vol. 8, no. 2, pp. 43-46, 2016.
- [9] F. Ghalib and W. Wasilah, "Deteksi Dini Kanker Payudara Menggunakan Algoritma K-Nearest Neighbour (KNN) dan Decision Tree C-45," *Jurnal TEKNIKA*, vol. XVII, no. 2, pp. 427-434, 2023.
- [10] P. Singhal and S. Pareek, "Artificial Neural Network for Prediction of Breast Cancer," India, 2018.
- [11] I. P. P. Iswara, F. Farhan, W. Kumara and A. A. Supianto, "Rekomendasi Pengambilan Mata Kuliah Pilihan Untuk Mahasiswa Sistem Informasi Menggunakan Algoritma Decision Tree," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 6, no. 3, pp. 341-348, 2019.
- [12] E. Elisa, "Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti," *Jurnal Online Informatika (JOIN)*, vol. 2, no. 1, pp. 36-41, 2017.
- [13] J. J. Pangaribuan, C. Tedja and S. Wibowo, "Perbandingan Metode Algoritma C4.5 dan Extreme Learning Machine Untuk Mendiagnosis Penyakit Jantung Koroner," *Informatics Engineering Research and Technology*, vol. I, no. 1, pp. 1-7, 2019.
- [14] M. Bramer, *Principles of Data Mining*, London: Springer, 2007.