

PREDIKSI KEKAMBUHAN KANKER PAYUDARA DENGAN ALGORITMA C4.5

Ai Rita Rizqiah¹; Agus Subekti²

¹ Ilmu Komputer

¹STMIK Nusa Mandiri Jakarta

^{1,2}www.nusamandiri.ac.id

¹airitarizqiah808@gmail.com

²Pusat Penelitian Elektronika dan Telekomunikasi

²Lembaga Ilmu Pengetahuan Indonesia

²www.lipi.go.id

²agus.subekti@lipi.go.id



Ciptaan disebarluaskan di bawah Lisensi Creative Commons Atribusi-NonKomersial 4.0 Internasional.

Abstract—Breast cancer is known as the fifth cause of death based on WHO data in 2015. The risk of developing breast cancer will increase with age, family medical history, personal medical history, caucasian descent, early menstruation, late menopause and others. This study aims to predict the use of Naïve Bayes and C4.5 data mining algorithms to classify the recurrence of cancer patients based on certain attributes in the breast cancer dataset. The data mining process will help identify the range or value of various attributes of what causes breast cancer. The results of this study indicate that the C4.5 algorithm has an accuracy value of 75.5% better than Naïve Bayes which only has an accuracy value of 72.7%.

Keywords: cancer, breast cancer, data mining, classification, Naïve Bayes, C4.5.

Intisari—Kanker payudara diketahui sebagai penyebab kematian ke lima berdasarkan data WHO pada tahun 2015. Resiko terkena kanker payudara akan semakin meningkat seiring bertambahnya usia, riwayat medis keluarga, riwayat medis personal, keturunan kaukasia, menstruasi awal, menopause terlambat dan lain-lain. Penelitian ini bertujuan untuk memprediksi penggunaan algoritma data mining Naïve Bayes dan C4.5 untuk mengklasifikasikan kekambuhan pasien kanker berdasarkan atribut tertentu pada dataset kanker payudara. Proses data mining akan membantu mengidentifikasi kisaran atau nilai berbagai atribut apa yang menyebabkan kanker payudara. Hasil dari penelitian ini menunjukkan bahwa algoritma C4.5 memiliki nilai akurasi 75.5% lebih baik dari

pada Naïve Bayes yang hanya memiliki nilai akurasi 72.7%.

Kata Kunci: kanker, kanker payudara, data mining, klasifikasi, Naïve Bayes, C4.5.

PENDAHULUAN

Kanker merupakan salah satu penyebab kematian utamadi seluruh dunia (Kristanto & Kahija, 2017). Kanker paru, hati, perut, serviks dan payudara adalah penyebab terbesar kematian akibat kanker setiap tahunnya. Kanker merupakan penyakit yang proses penyembuhannya yang sulit, selain itu memerlukan biaya yang cukup besar untuk pengobatan dan perawatannya. Seorang yang telah mengidap kanker payudara dapat disembuhkan dengan berbagai pengobatan untuk menghambat pertumbuhan sel kanker yaitu dengan operasi, kemoterapi, radio terapi, dan terapi hormonal. Meski demikian, pasien yang telah menjalani pengobatan tidak dapat sembuh sepenuhnya. Kekambuhan kanker payudara adalah manifestasi klinis dan penyebab utama kematian pada kanker payudara (Eveline, Purwanto, & Lestari, 2017). Terjadinya kekambuhan memerlukan penanganan lebih serius.

Tabel 1. Angka kematian di dunia berdasarkan jenis kanker tahun 2015

No	Jenis Kanker	Angka kematian
1	Paru	1.760.000
2	Liver	782.000
3	Kolorektal	862.000
4	Perut	783.000
5	Payudara	627.000

Sumber: (WHO, 2018)

Beberapa jurnal menyebutkan terjadinya kekambuhan lokal berhubungan dengan terjadinya metastasis di kemudian hari. Keberhasilan terapi kanker payudara dini dapat dinilai dengan memonitor terjadinya kekambuhan atau tidak yang disebut Disease Free Interval (Harris, Lippman, Morrow, & Osborne, 2014)

Penelitian ini fokus pada seleksi fitur menggunakan Ranking Attribute Information Gain, klasifikasi menggunakan algoritma Naïve Bayes dan C4.5 pada data *breast cancer* yang diperoleh dari Institute of Oncology, University Medical Center Ljubljana, Yugoslavia (1998).

Berdasarkan beberapa jurnal penelitian metode yang digunakan untuk mengklasifikasikan data kekambuhan adalah metode C4.5 atau C4.5 (Bhargava, Sharma, Purohit, & Rathore, 2017). Selain itu metode lain yang digunakan yaitu C4.5, SVM dan Naïve Bayes (Pritom, Munshi, Sabab, & Shihab, 2016). Penelitian tersebut dilakukan dengan membagi dataset menjadi 66% sebagai data training dan 34% sebagai data testing.

Tujuan dari penelitian ini yaitu untuk memperoleh knowledge dari sekumpulan data kekambuhan pasien kanker payudara, mengidentifikasi variable/ atribut mana saja yang memiliki kontribusi besar pada data dan mengklasifikasikan data kekambuhan kanker payudara secara kasat mata/ metode manual.

BAHAN DAN METODE

Data kekambuhan kanker payudara terdiri dari 10 atribut termasuk atribut kelas. Sebagai berikut:

Tabel 2. Atribut data kekambuhan kanker payudara

Atribut	Definisi	Record
Age	Umur pasien pengidap kanker payudara	{'10-19','20-29','30-39','40-49','50-59','60-69','70-79','80-89','90-99'}
Menopause	Masa berakhirnya siklus menstruasi.	{'lt40','ge40','preme no'}
Tumor Size	Ukuran tumor payudara dalam ukuran millimeter (mm).	{'0-4','5-9','10-14','15-19','20-24','25-29','30-34','35-39','40-44','45-49','50-54','55-59'}
Inv-nodes	Kelenjar getah bening aksila (ketiak).	{'0-2','3-5','6-8','9-11','12-14','15-17','18-20','21-23','24-26','27-29','30-32','33-35','36-39'}
Node-caps	Penyebaran ke kelenjar getah bening.	{'yes','no'}

Deg-Malig	Tingkat keganasan	{'1','2','3'}
Breast	Lokasi kanker payudara	{'left','right'}
Breast-Quad	Kuadran lokasi kanker	{'left_up', 'left_low', 'right_up', 'right_low', 'central'}
Irradiat	Riwayat terapi menggunakan Radiasi(x-rays)	{'yes','no'}
Class	Kelas kanker payudara	{'no-recurrence-events','recurrence-events'}

Sumber: (Rizqiah & Subekti, 2018)

Penelitian dimulai dengan melakukan tahap *preprocessing*, untuk menghilangkan *missing values*. Setelah itu dilakukan proses imputasi untuk menghilangkan *missing values*. Kemudian dilakukan seleksi fitur untuk melihat atribut mana yang memiliki pengaruh besar terhadap data. Tahap terakhir dilakukan klasifikasi dengan dua metode, yaitu Naive Bayes dan C4.5. Di akhir penelitian dilakukan perbandingan metode yang paling baik untuk mengklasifikasi data kekambuhan pasien kanker payudara.

Data yang digunakan berupa data kekambuhan pasien kanker payudara yang berasal dari Institute of Oncology, University Medical Center Ljubljana, Yugoslavia (1998) berjumlah 286 data terdiri dari 10 atribut termasuk atribut kelas.

Teknik pengumpulan data dilakukan dengan metode studi literatur dari berbagai sumber.

HASIL DAN PEMBAHASAN

Missing Values merupakan tahanan dalam *pre-processing*, tahap ini digunakan untuk mencari apakah terdapat *value* (nilai) data atribut yang kosong dalam setiap baris data. Hal ini dapat terjadi baik pada data bertipe numerik dan kategorik. Dalam hal ini penelitian dilakukan pada data bertipe kategorik.

Pada data kekambuhan pasien kanker payudara ini terdapat beberapa atribut yang memiliki *missing values*, sehingga harus dilakukan imputation. Berikut daftar *missing values* (nilai yang hilang) setiap atribut:

Tabel 3. Jumlah *missing values*

Atribut	Jumlah <i>Missing Values</i>
node-caps	8
breast-quad	1

Sumber: (Rizqiah & Subekti, 2018)

Untuk mengisi nilai yang hilang tersebut dapat menggunakan means/ median/ modus. Jika

berupa data kategorikal maka dapat diterapkan menggunakan modus.

Setelah data telah dilakukan imputasi *missing values*, maka tahap selanjutnya adalah mengukur atribut-atribut mana saja yang memiliki kontribusi besar pada data dengan cara meranking atribut menggunakan metode tertentu. Pada penelitian ini, ranking atribut menggunakan metode Information Gain. Information Gain merupakan salah satu metode yang digunakan untuk mencari seberapa potensial informasi pada atribut masing-masing dalam data. Output dari metode tersebut adalah berupa *score* setiap atribut yang merepresentasikan urutan atribut berkontribusi besar sampai terkecil. Berikut ini adalah cara mengevaluasi atribut yakni berdasarkan perolehan frekuensi “tiap nilai atribut”.

Menghitung nilai *Expected Information (Entropy)*
 $H(X) = - \sum_{i=1}^n p(X_i) \log_b p(X_i) \dots \dots \dots (1)$

Tabel 4. Frekuensi Nilai Atribut deg-malig

Atribut	Nilai atribut	Kelas	Freq.	Jumlah
deg-malig	'1'	'no-recurrence-events'	59	71
		'recurrence-events'	12	
	'2'	'no-recurrence-events'	102	130
		'recurrence-events'	28	
	'3'	'no-recurrence-events'	40	85
		'recurrence-events'	45	
Total			286	286

Sumber: (Rizqiah & Subekti, 2018)

Jika salah satu hasil dari perhitungan $p \log_2 p = 0$ atau pada Excel #NUM! atau #DIV/0!, maka hasil perhitungan nilai q adalah 0.

Tabel 5. Entropy atribut deg-malig

Q	$p \log_2 p$	$p \log_2 p + p \log_2 p$
q1	$(59/71) * \log_2(59/71) = -0,22196$	$-((-0,22196)$
	$(12/71) * \log_2(12/71) = -0,43348$	$+(-0,43348)) = 0,655444445$
q2	$(102/130) * \log_2(102/130) = -0,27457$	$-((-0,27457)$
	$(28/130) * \log_2(28/130) = -0,47708$	$+(-0,47708)) = 0,751649946$
q3	$(40/85) * \log_2(40/85) =$	$-((-0,51175)$

$$\frac{-0,51175}{(45/85) * \log_2(45/85) = -0,48576} + (-0,48576) = 0,997502546$$

Sumber: (Rita & Subekti, 2018)

Menghitung nilai *Information Attribute*

$$\sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v) \dots \dots \dots (2)$$

Berikut contoh perhitungan pada atribut deg-malig:

$$\sum \frac{|S_v|}{|S|} \text{Entropy}(S_v) = (71/286 * 0,655444445) + (130/286 * 0,751649946) + (85/286 * 0,997502546) = 0,800834843$$

Menghitung Gain All (total kelas)

$$p \log_2 p + p \log_2 p = (201/286) * \log_2(201/286) + 85/286 * \log_2(85/286) = 0,877844695$$

Menghitung *Information Gain (Gain All - Information Attribute)*

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v) \dots (3)$$

Gain atribut deg-malig

$$0,877844695 - 0,800834843 = 0,002001615$$

Tabel 6. Hasil seleksi fitur *ranking attribute* dengan Information Gain

Atribut	Skor
deg-malig	0.07701
inv-nodes	0.069
tumor-size	0.05717
node-caps	0.05136
Irradiat	0.02582
Age	0.01061
breast-quad	0.00934
Breast	0.00249
Menopause	0.002

Sumber: (Rizqiah & Subekti, 2018)

Setelah melakukan seleksi fitur, selanjutnya dilakukan klasifikasi dengan algoritma Naïve Bayes. Naïve Bayes merupakan sebuah metode untuk mengklasifikasikan beberapa atribut ke dalam sebuah label kelas menggunakan pendekatan probabilitas. Dalam penelitian ini berarti menargetkan tiap kondisi atribut-atribut ke label kelas *no-recurrence-events* dan *recurrence-events*. Berikut ini merupakan langkah-langkah dalam algoritma Naïve Bayes:

- a. Menghitung probabilitas jumlah data pada tiap kelas

$$P(C_i) \dots\dots\dots (4)$$

Tabel 7. Frekuensi data (data training) pada tiap kelas

Frekuensi data pada kelas masing-masing (data training)	'no-recurrence-events'	200
	'recurrence-events'	85
Total		285

Sumber: (Rizqiah & Subekti, 2018)

Tabel 8. Probabilitas jumlah data training keseluruhan

$P(C \text{ no-recurrence-events})$	$= 200/85 = 0,701754$
$P(C \text{ recurrence-events})$	$= 85/285 = 0,298246$

Sumber: (Rita & Subekti, 2018)

- b. Menghitung probabilitas jumlah data pada tiap kasus (nilai atribut) per kelas.

$$P(X|C_i) \dots\dots\dots (5)$$

- c. Mengalikan semua antara probabilitas semua jumlah data pada tiap kasus (nilai atribut) per kelas dengan probabilitas jumlah data pada kelas-nya

$$P(C_i|X) = P(X|C_i) * P(C_i) \dots\dots\dots (6)$$

- X : Data dengan class yang belum diketahui
- C : Hipotesis data merupakan suatu class spesifik
- P(C|X) : Probabilitas hipotesis berdasar kondisi (posteriori probability)
- P(C) : Probabilitas hipotesis (prior probability)
- P(X|C) : Probabilitas berdasarkan kondisi pada hipotesis
- P(X) : Probabilitas X

$$P(C_i|X) = P(X|C \text{ no-recurrence-events}) * P(C \text{ no-recurrence-events})$$

$$= 0,35 * 0,465 * 0,075 * 0,83 * 0,87 * 0,195 * 0,51 * 0,095 * 0,815 * 0,701754$$

$$= 4,76267E-05$$

$$P(C_i|X) = P(X | C \text{ recurrence-events}) * P(C \text{ recurrence-events})$$

$$= 0,294117647 * 0,411764706 * 0,070588235 * 0,529411765 * 0,635294118 * 0,517647059 * 0,576470588 * 0,152941176 * 0,623529412 * 0,298246$$

$$= 2,44027E-05$$

- d. Membandingkan hasil nilai probabilitas kelas masing-masing

Jika $P(X | \text{Class} = \text{'no-recurrence-events'}) > P(X | \text{Class} = \text{'recurrence-events'})$, Maka label prediksi nya = 'no-recurrence-events' atau sebaliknya.

Jika $P(X | \text{Class} = \text{'no-recurrence-events'}) < P(X | \text{Class} = \text{'recurrence-events'})$, Maka label prediksi nya = 'recurrence-events'.

Maka dapat dibandingkan, $4,76267E-05 > 2,44027E-05$, Jadi label prediksi nya adalah 'no-recurrence-events'.

Berikut ini merupakan hasil klasifikasi data kekambuhan pasien kanker payudara dengan algoritma Naïve Bayes menggunakan tools Weka.

```

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      208          72.7273 %
Incorrectly Classified Instances     78          27.2727 %
Kappa statistic                     0.2996
Mean absolute error                  0.3268
Root mean squared error              0.4534
Relative absolute error              77.8451 %
Root relative squared error          98.8513 %
Total Number of Instances           286

=== Detailed Accuracy By Class ===
          TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
          0.856   0.576   0.778     0.856   0.815     0.699   no-recurrence-events
          0.424   0.144   0.554     0.424   0.48      0.699   recurrence-events
Weighted Avg.   0.727   0.448   0.712     0.727   0.716     0.699

=== Confusion Matrix ===
  a  b  <-- classified as
172 29 | a = no-recurrence-events
 49 36 | b = recurrence-events
    
```

Sumber: (Rizqiah & Subekti, 2018)

Gambar 1. Hasil Klasifikasi dengan algoritma Naïve Bayes

Dari gambar 1, di atas terlihat bahwa akurasi pada data kekambuhan kanker payudara menggunakan klasifikasi algoritma Naïve Bayes sebesar 72.7%, dengan penjelasan sebagai berikut:

- a. Weka Confusion Matrix, jika diasumsikan a sebagai kelas negative (no-recurrence-events / tidak kambuh) :

Tabel 9. Weka Confusion Matrix asumsi a kelas negative (no-recurrence-events / tidak kambuh)

	a	b	
TN	172	29	a = no-recurrence-events
FN	49	36	b = recurrence-events
	FN	TP	

Sumber: (Rizqiah & Subekti, 2018)

Dari confusion matrix di atas, diperoleh pengukuran berdasarkan akurasi kelas kondisi pasien kanker payudara yang tidak kambuh adalah:

Tabel 10. Pengukuran asumsi kelas negative (no-recurrence-events / tidak kambuh)

Measurement	Formula
Accuracy TP+TN/ TP+TN+FP+FN	$(36+172)/ (36+172+29+49) = 72.7\%$
Precision TP / TP+FP	$36 / (36+29) = 55.3\%$
Recall TP/TP+FN	$36 / (36+49) = 42.3\%$
F-Measure 2*Precision*Recall / Precision + Recall	$2*55.3\%*42.3\% / (55.3\%+42.3\%) = 48\%$

Sumber: (Rizqiah & Subekti, 2018)

b. Weka Confusion Matrix, jika diasumsikan a sebagai kelas positive (recurrence-events / kambuh) :

Tabel 11. Weka Confusion Matrix asumsi a kelas positive (recurrence-events / kambuh)

a	b	
TP	FN	
172	29	a = no-recurrence-events
49	36	b = recurrence-events
FP	TN	

Sumber: (Rizqiah & Subekti, 2018)

Dari confusion matrix di atas, diperoleh pengukuran berdasarkan akurasi kelas kondisi pasien kanker payudara yang tidak kambuh adalah:

Tabel 12. Pengukuran asumsi kelas positive (recurrence-events / kambuh)

Measurement	Formula
Accuracy TP+TN/ TP+TN+FP+FN	$(172+36)/ (172+36+49+29) = 72.7\%$
Precision TP / TP+FP	$172 / (172+49) = 77.8\%$
Recall TP/TP+FN	$172 / (172+29) = 85.5\%$
F-Measure 2*Precision*Recall / Precision + Recall	$2*77.8\%*85.5\% / (77.8\%+85.5\%) = 81.5\%$

Sumber: (Rizqiah & Subekti, 2018)

Metode klasifikasi C4.5 dilakukan dengan perhitungan nilai Gain, seperti dalam persamaan di bawah ini:

$$Gain(S, A) = Entropy \sum_i^n = 1 \frac{|S_i|}{|S|} * Entropy(S_i) \dots \dots \dots (7)$$

S : Himpunan kasus

A : Atribut
n : Jumlah partisi himpunan A
|S_i| : Jumlah kasus pada partisi ke-i
|S| : Jumlah kasus dalam S

Tabel 13. Perhitungan gain pada atribut node-caps

Nilai fitur	Kelas	Freq.	Sum	Entropy
'yes'	'no-recurrence-events'	25	56	0.9917
	'recurrence-events'	31		
'no'	'no-recurrence-events'	176	230	0.78626
	'recurrence-events'	54		

Sumber: (Rizqiah & Subekti, 2018)

$$Gain(Total, node-caps) = 0,877844695 - ((25/286)*0,9917) + ((230/286)*0,78626) = 0.051361454$$

Untuk menghitung nilai Entropy dapat dilihat pada persamaan berikut ini:

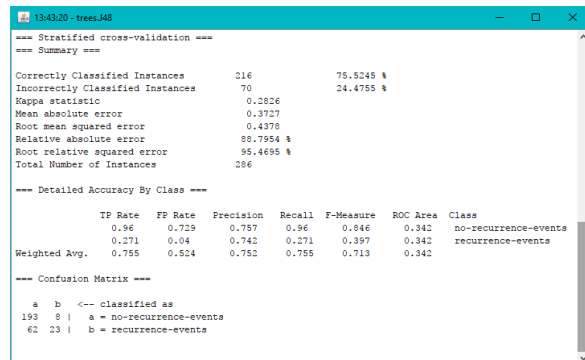
$$Entropy(S) = \sum_i^n = 1 - p_i * \log_2(p_i) \dots \dots \dots (8)$$

n : Jumlah partisi S
p_i : Proporsi dari S_i terhadap S

$$Entropy(total) = -((\frac{201}{286} * \log_2(\frac{201}{286})) + (\frac{85}{286} * \log_2(\frac{85}{286}))) = 0,877844695$$

Entropy total adalah menghitung nilai total keputusan no-recurrence-events (201) dan recurrence-events (85), sedangkan 286 adalah jumlah keseluruhan kasus.

Pada tahap ini cross-validation dengan 'Folds = 286' (sesuai dengan jumlah data), dikarenakan validasi sampling yang digunakan adalah metode Leave-one-out (LOO) yakni 1 per 1 data di ujikan secara bergantian.



Sumber: (Rizqiah & Subekti, 2018)

Gambar 2. Hasil klasifikasi algoritma C4.5

a. Weka Confusion Matrix, jika diasumsikan *a* sebagai kelas negative (no-recurrence-events / tidak kambuh) :

Tabel 14. Weka Confusion Matrixasumsi *a* kelas negative(no-recurrence-events / tidak kambuh)

a	b	
TN	FP	
193	8	a = no-recurrence-events
62	23	b = recurrence-events
FN	TP	

Sumber: (Rizqiah & Subekti, 2018)

Dari confusion matrix di atas, diperoleh pengukuran berdasarkan akurasi kelas kondisi pasien kanker payudara yang tidak kambuh adalah:

Tabel 15. Pengukuran asumsi *a* kelas negative (no-recurrence-events / tidak kambuh)

Measurement	Formula
Accuracy	$(23+193) / (23+193+8+62) = 75.5\%$
Precision	$23 / (23+8) = 74.2\%$
Recall	$23 / (23+62) = 27\%$
F-Measure	$2 * 74.2\% * 27\% / (74.2\% + 27\%) = 39,59\%$

Sumber: (Rizqiah & Subekti, 2018)

b. Weka Confusion Matrix, jika diasumsikan *a* sebagai kelas positive (recurrence-events / kambuh) :

Tabel 16. Weka Confusion Matrixasumsi *a* kelas positive(recurrence-events / kambuh)

a	b	
TP	FN	
193	8	a = no-recurrence-events
62	23	b = recurrence-events
FP	TN	

Sumber: (Rizqiah & Subekti, 2018)

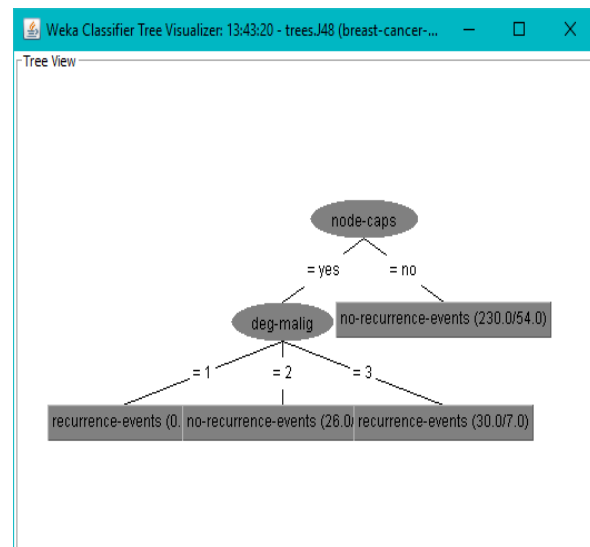
Dari confusion matrix di atas, diperoleh pengukuran berdasarkan akurasi kelas kondisi pasien kanker payudara yang tidak kambuh adalah:

Tabel 17. Pengukuran asumsi *a* kelas positive(recurrence-events / kambuh)

Measurement	Formula
Accuracy	$(193+23) / (193+23+62+8) = 75.5\%$
Precision	$193 / (193+62) = 75.6\%$
Recall	$193 / (193+8) = 96\%$
F-Measure	$2 * 75.6\% * 96\% / (75.6\% + 96\%) = 84.6\%$

Sumber: (Rizqiah & Subekti, 2018)

Berdasarkan klasifikasi kekambuhan pasien kanker payudara dengan algoritma C4.5 diperoleh pohon keputusan sebagai berikut:



Sumber: (Rizqiah & Subekti, 2018)

Gambar 3. Pohon keputusan C4.5

Dapat disimpulkan bahwa atribut node-caps (penyebaran kanker ke kelenjar getah bening) menjadi *root node*, sedangkan atribut deg-malig menjadi *internal node* yang memiliki satu inputan dan memiliki 3 output sesuai dengan nilai fituranya. Pohon keputusan C4.5 di atas memiliki ukuran pohon 6 dan 4 buah daun. Dari pohon keputusan diatas diperoleh rule sebagai berikut:

1. Jika node-caps (penyebaran kanker ke kelenjar getah bening)='yes' and deg-malig (tingkat keganasan kanker)='1' (rendah), maka 'recurrence-events' (kambuh)
2. Jika node-caps (penyebaran kanker ke kelenjar getah bening)='yes' and deg-malig (tingkat keganasan kanker)='2' (sedang), maka 'no-recurrence-events' (tidak kambuh)
3. Jika node-caps (penyebaran kanker ke kelenjar getah bening)='yes' and deg-malig (tingkat

keganasan kanker)=‘3’ (sedang), maka ‘reccurrence-events’ (kambuh)

4. Jika node-caps (penyebaran kanker ke kelenjar getah bening)=‘no’ and deg-, maka ‘no-reccurrence-events’ (tidak kambuh)

Setelah dataset kekambuhan kanker payudara diolah dan berhasil diklasifikasikan dengan kedua metode (Naïve Bayes dan C4.5), hasil klasifikasi tersebut disimpulkan pada tabel 18.

Tabel 18. Perbandingan nilai akurasi algoritma Naïve Bayes dan C.45

Metode	Akurasi	Presisi	Recall	Cross Valid.	Ket.
Naïve Bayes	71.70%	77.80%	83.60%	286	Tanpa Imputasi
Naïve Bayes	72.72%	77.80%	85.60%	286	Setelah Imputasi
C4.5	75.50%	75.70%	96%	286	Tanpa Imputasi
C4.5	75.50%	75.70%	96%	286	Setelah Imputasi

Sumber: (Rizqiah & Subekti, 2018)

KESIMPULAN

Berdasarkan uraian-uraian pada pembahasan sebelumnya, dapat ditarik kesimpulan bahwa klasifikasi kekambuhan pasien kanker payudara menggunakan algoritma C4.5 menghasilkan akurasi yang lebih baik yaitu sebesar 75,5% dibandingkan menggunakan Naïve Bayes 72,7%. Proses imputasi tidak terlalu berdampak pada hasil akurasi pada algoritma C4.5, dan berdampak baik pada algoritma Naïve Bayes yaitu penambahan nilai akurasi sebesar 1,02%.

REFERENSI

Bhargava, N., Sharma, S., Purohit, R., & Rathore, P. S. (2017). Prediction of Recurrence Cancer Using J48 Algorithm. In *2017 2nd International Conference on Communication and Electronics Systems (ICCES)* (pp. 386–390). Coimbatore: IEEE. <https://doi.org/10.1109/CESYS.2017.8321306>

Eveline, K., Purwanto, H., & Lestari, P. (2017). Faktor Klinis dan Histopatologi serta Hubungannya dengan Kekambuhan Pascaoperasi pada Pasien Kanker Payudara di RSUD Dr. Soetomo, Januari–Juni 2015. *Indonesianjournalofcancer.or.Id*, *11*(2), 55–60.

Harris, J. R., Lippman, M. E., Morrow, M., & Osborne, C. K. (2014). Evaluation After Primary Therapy and Management of Reccurent Breast Cancer. In *Diseases of the breast: Fifth*

edition (1st ed., p. 871). London: Wolters Kluwer Health Adis (ESP).

Kristanto, A. D., & Kahija, Y. F. La. (2017). PENGALAMAN COPING TERHADAP DIAGNOSIS KANKER PADA PENDERITA USIA KERJA DI RUMAH SAKIT MARGONO SOEKARJO PURWOKERTO. *Jurnal Empati*, *6*(2), 1–9.

Pritom, A. I., Munshi, M. A. R., Sabab, S. A., & Shihab, S. (2016). Predicting Breast Cancer Recurrence Using Effective Classification and Feature Selection Technique. In *2016 19th International Conference on Computer and Information Technology (ICCIT)* (pp. 310–314). Dhaka: IEEE. <https://doi.org/10.1109/ICCITECHN.2016.7860215>

Rizqiah, A. R., & Subekti, A. (2018). *PREDIKSI KEKAMBUHAN KANKER PAYUDARA DENGAN ALGORITMA C4.5*. Jakarta.

WHO. (2018). Cancer. Retrieved from <https://www.who.int/news-room/fact-sheets/detail/cancer>

