

## Data Mining untuk Klasifikasi Diagnosa Kanker Payudara Dengan Menerapkan Algoritma C4.5

Yulia Eka Praptiningsih<sup>1</sup>, Winda Widya Ariestya\*<sup>2</sup>, Ida Astuti<sup>3</sup>, Sylvia Nurulita<sup>4</sup>

Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Gunadarma<sup>1,2,3,4</sup>

Jl. Margonda Raya No. 100 Depok

yulia\_eka@staff.gunadarma.ac.id<sup>1</sup>, winda\_widya@staff.gunadarma.ac.id<sup>2</sup>,

astuti@staff.gunadarma.ac.id<sup>3</sup>, sylvia.nurulita@gmail.com<sup>4</sup>

\*) Corresponding Author

(received: 15-04-23, revised: 03-07-23, accepted: 24-07-23)

### Abstract

Cancer is a health disorder in human organs or body tissues where abnormal cells multiply uncontrollably. Cancer is the second biggest cause of death, including breast cancer which affects most Indonesian women. The purpose of this study was to diagnose breast cancer in patients whether it is malignant or benign using the C4.5 algorithm so that it can help treat the cancer to prevent death. The research method used consisted of three stages, namely preprocessing, modeling, and evaluation. In the preprocessing stage, 570 clinical data records from the UCI (UC Irvine) Machine Learning Repository were used in this study and then split the data, namely train data and test data. The modeling stage implements the C4.5 algorithm as a method for classifying malignant and benign breast cancer. The final stage of evaluation of the classification results on 32 attributes obtained 8 attributes as determinants. The results of the performance evaluation show that the C4.5 algorithm can be used as an algorithm in the classification of breast cancer because the accuracy values obtained are quite large, namely 93.04%, 80.00% precision and 92.31% recall.

**Keyword:** Data Mining, Classification, C4.5, Breast Cancer

### Abstrak

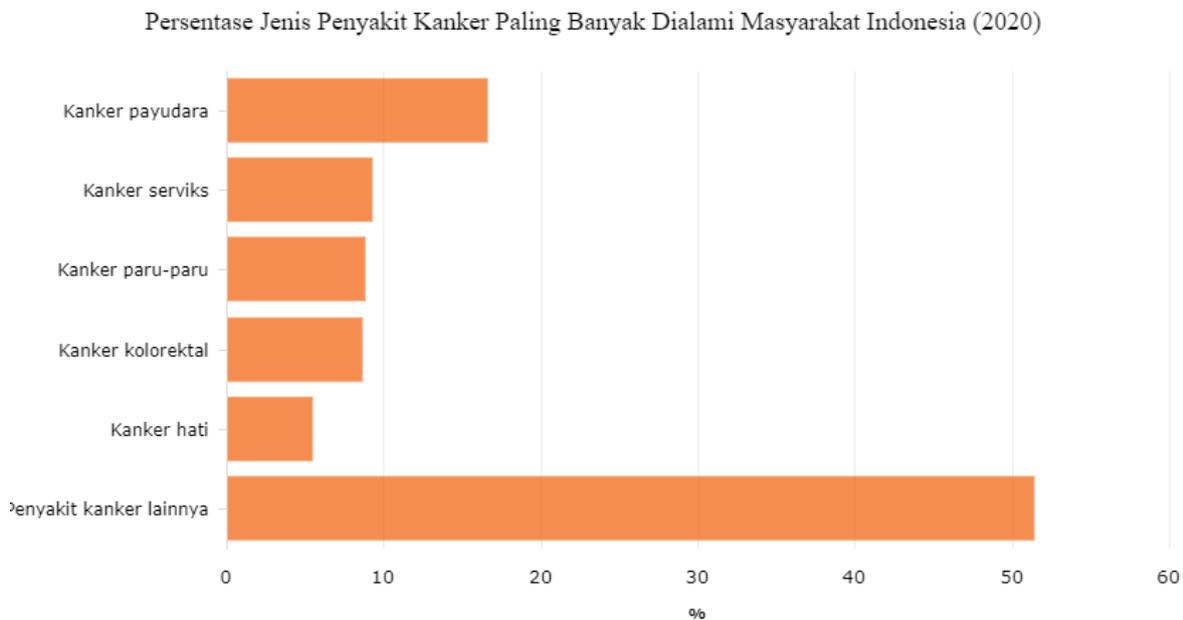
Penyakit kanker merupakan gangguan kesehatan pada organ tubuh manusia atau jaringan tubuh di mana sel-sel yang tidak normal berkembang biak dengan tidak terkendali. Kanker adalah penyebab kematian terbesar kedua, tak terkecuali kanker payudara yang diderita sebagian besar wanita Indonesia. Tujuan penelitian ini adalah untuk mendiagnosa penyakit kanker payudara yang diderita pasien apakah bersifat ganas atau jinak menggunakan algoritma C4.5 sehingga dapat membantu penanganan penyakit kanker tersebut untuk mencegah kematian. Metode penelitian yang digunakan terdiri dari tiga tahapan yaitu preprocessing, modeling, dan evaluation. Tahap preprocessing, 570 catatan data klinis dari UCI (UC Irvine) Machine Learning Repository digunakan dalam penelitian ini dan selanjutnya dilakukan split data yaitu data train dan data test. Tahap modeling (pembentukan model) mengimplementasikan algoritma C4.5 sebagai metode klasifikasi penyakit kanker payudara ganas dan jinak. Tahap akhir evaluation dari hasil klasifikasi pada 32 atribut diperoleh 8 atribut sebagai penentu. Hasil evaluasi performance menunjukkan algoritma C4.5 dapat digunakan sebagai algoritma pada klasifikasi penyakit kanker payudara karena nilai akurasi yang diperoleh cukup besar yaitu 93,04%, presisi 80,00% dan recall 92,31%.

**Kata Kunci:** Data Mining, Klasifikasi, Algoritma C4.5, Kanker Payudara

### I. Pendahuluan

Menurut Organisasi Kesehatan Dunia, kanker adalah kelompok gangguan kesehatan yang luas yang dapat dimulai di hampir semua organ atau jaringan tubuh ketika sel-sel tidak normal berkembang biak tak terkendali. Tahun 2018 data *World Health Organization* (WHO) melaporkan, kanker adalah penyebab kematian terbesar kedua di dunia terhitung sekitar 9,6 juta kematian atau dengan kata lain sebanyak satu dari enam kematian [1].

Laki-laki lebih mungkin terkena kanker paru-paru, prostat, kolorektal, perut, dan hati, sedangkan wanita lebih mungkin mengembangkan kanker payudara, kolorektal, paru-paru, leher rahim, dan tiroid dalam tubuhnya [1]. Menurut laporan *Global Burden of Cancer Study* (Globocan), terdapat 396.914 keganasan kanker yang menyerang penduduk Indonesia pada tahun 2020 [2].



Gambar 1. Data Penyakit Kanker di Indonesia Tahun 2020 [2]

Kanker payudara merupakan jenis penyakit yang paling sering terjadi di Indonesia, terhitung sebanyak 65.858 kejadian. Jumlah ini setara dengan 16,6% dari seluruh kasus kanker di Indonesia [2]. Untuk mengatasi masalah ini, diperlukan identifikasi dini kanker payudara. Identifikasi dini kanker payudara berkorelasi dengan penurunan kematian akibat kanker payudara dan dapat mengurangi efek keuangan yang signifikan [3].

Seiring perkembangan teknologi khususnya teknologi informasi, terdapat berbagai teknologi yang dapat dilakukan untuk mengatasi berbagai kesulitan dalam prediksi penyakit kanker. Data mining merupakan salah satu teknologi yang dapat digunakan untuk prediksi dan klasifikasi penyakit. Penambangan data, sering dikenal dengan KDD (*Knowledge Discovery in Database*) adalah proses penggalian informasi yang berguna dari data [4]. Dengan kata lain, Data Mining adalah proses untuk menemukan pengetahuan yang ada dalam kumpulan data berskala besar secara efektif dan efisien [5].

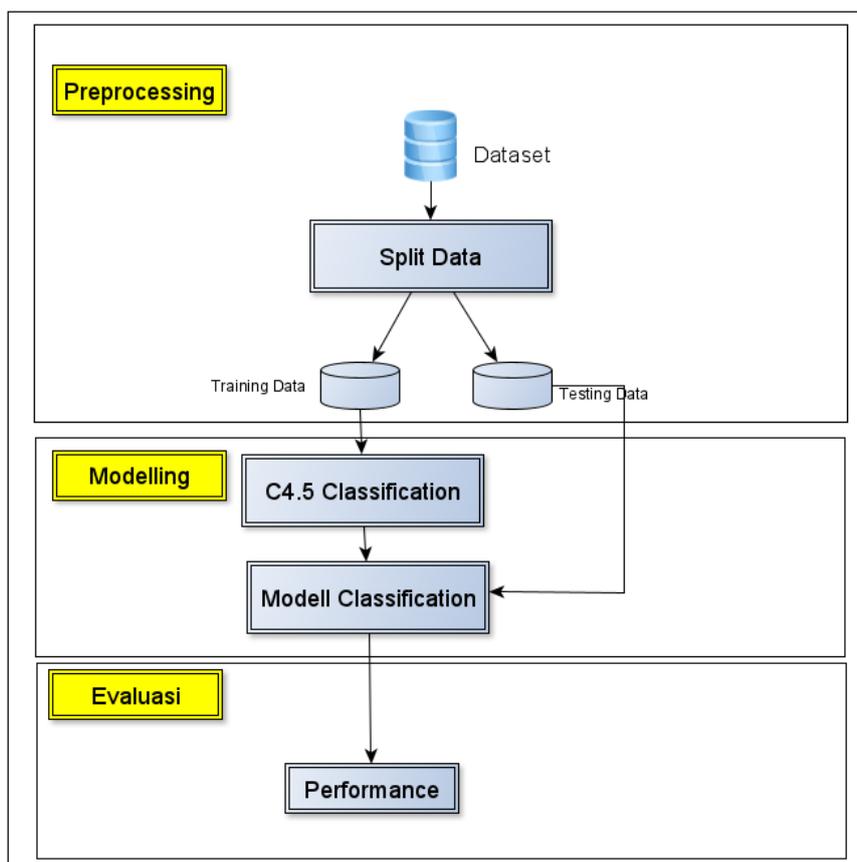
Pendekatan *Decision Tree* dengan algoritma C4.5 merupakan salah satu metode data mining yang dapat digunakan untuk klasifikasi. *Decision Tree* merupakan algoritma yang paling populer dan umum digunakan karena merupakan salah satu metode klasifikasi yang praktis [6]. Algoritma C4.5 merupakan algoritma pengambilan keputusan yang dapat meniru penalaran manusia. Hal ini berdasarkan penelitian terdahulu yang membandingkan algoritma *Decision Tree* seperti C&RT, CHAID, QUEST, C4.5, dan ID3, yang menemukan bahwa algoritma C4.5 memiliki tingkat akurasi yang tinggi, unggul dari pendekatan *Decision Tree* lainnya [7].

Penelitian ini merujuk pada beberapa penelitian yang telah dilakukan sebelumnya. Pada suatu penelitian, metode *Naïve Bayes* berhasil melakukan klasifikasi jenis kanker berdasarkan struktur protein dengan akurasi yang diperoleh sebesar 79,17% [8]. Pada penelitian lain, metode *Support Vector Machine* (SVM) diimplementasikan dalam deteksi kanker paru dengan hasil akurasi sebesar 87,10% [9]. Penelitian lain menemukan aturan untuk mendeteksi stadium kanker serviks menggunakan pendekatan *Decision Tree* dan algoritma C4.5 dengan akurasi 85,5% [10]. Penelitian lain yang menggunakan pendekatan *Decision Tree* dengan algoritma C4.5 untuk mendeteksi kanker serviks menghasilkan akurasi sebesar 98,61% [11]. Berdasarkan uraian penelitian, metode

*Decision Tree* dengan algoritma C4.5 cukup baik dalam klasifikasi kanker dengan menghasilkan akurasi yang baik; sehingga pada penelitian ini diterapkan metode *Decision Tree* dengan algoritma C4.5 untuk memprediksi kanker payudara dan menentukan atribut penentu pasien kanker payudara jinak atau ganas.

## II. Metodologi Penelitian

Untuk menjawab permasalahan tersebut, penelitian ini menggunakan metodologi eksperimental. Gambar 2 menggambarkan langkah-langkah penelitian yang meliputi tiga tahapan dalam penelitian ini yaitu *preprocessing*, *modelling*, dan *evaluation*. Data didistribusikan untuk tujuan pelatihan dan pengujian selama tahap *preprocessing*. Selanjutnya menerapkan algoritma C4.5 untuk klasifikasi selama tahap pemodelan.



Gambar 2 Tahapan Penelitian

### Tahap *Preprocessing*

Sebelum data diolah, terlebih dahulu data dipersiapkan, proses ini dikenal dengan *preprocessing* [12]. Tahap ini meliputi proses penyiapan dataset dan pembagian data (*split data*). Dataset penelitian ini didasarkan pada database UCI Machine Learning Repository [13]. Dataset berisi 570 *record* data klinis.

Tabel 1. Dataset Penelitian [13]

id	diagnos is	radius_m_ean	texture_m_ean	perimete_r_mean	area_me_an	smoothnes_mean	compactnes_mean	concavity_mean	concave_m_ean	symmetry_mean	fractal_dime nsion mean
84302	M	17,99	10,38	122,8	1001	0,1184	0,2776	0,3001	0,1471	0,2419	0,07871
842517	M	20,57	17,77	132,9	1326	0,08474	0,07864	0,0869	0,07017	0,1812	0,05667
84300903	M	19,69	21,25	130	1203	0,1096	0,1599	0,1974	0,1279	0,2069	0,05999
...	...	...	...	...	...	...	...	...	...	...	...
926954	M	16,6	28,08	108,3	858,1	0,08455	0,1023	0,09251	0,05302	0,159	0,05648
927241	M	20,6	29,33	140,1	1265	0,1178	0,277	0,3514	0,152	0,2397	0,07016
92751	B	7,76	24,54	47,92	181	0,05263	0,04362	0	0	0,1587	0,05884

radius_se	texture_se	perimeter_se	area_se	smoothnes_se	compactnes_se	concavity_se	concave_se	symmetry_se	fractal_dimension_se
1,095	0,9053	8,589	153,4	0,006399	0,04904	0,05373	0,01587	0,03003	0,006193
0,5435	0,7339	3,398	74,08	0,005225	0,01308	0,0186	0,0134	0,01389	0,003532
0,7456	0,7869	4,585	94,03	0,00615	0,04006	0,03832	0,02058	0,0225	0,004571
...	...	...	...	...	...	...	...	...	...
0,4564	1,075	3,425	48,55	0,005903	0,03731	0,0473	0,01557	0,01318	0,003892
0,726	1,595	5,772	86,22	0,006522	0,06158	0,07117	0,01664	0,02324	0,006185
0,3857	1,428	2,548	19,15	0,007189	0,00466	0	0	0,02676	0,002783

radius_worst	texture_worst	perimeter_worst	area_worst	Smoothnes_worst	Compactnes_worst	concavity_worst	Concave_worst	symmetry_worst	fractal_dimension_worst
25,38	17,33	184,6	2019	0,1622	0,6656	0,7119	0,2654	0,4601	0,1189
24,99	23,41	158,8	1956	0,1238	0,1866	0,2416	0,186	0,275	0,08902
23,57	25,53	152,5	1709	0,1444	0,4245	0,4504	0,243	0,3613	0,08758
...	...	...	...	...	...	...	...	...	...
18,98	34,12	126,7	1124	0,1139	0,3094	0,3403	0,1418	0,2218	0,0782
25,74	39,42	184,6	1821	0,165	0,8681	0,9387	0,265	0,4087	0,124
9,456	30,37	59,16	268,6	0,08996	0,06444	0	0	0,2871	0,07039

Tabel 1 menunjukkan dataset yang digunakan dalam penelitian ini, yang meliputi 32 atribut dengan tiga jenis pengukuran: rata-rata (*mean*), kesalahan standar (*standard error*), dan terburuk (*worst*). Atribut diagnosis dijadikan sebagai atribut penentu apakah seseorang menderita kanker payudara ganas atau jinak.

Data terpisah diterapkan ke kumpulan data untuk memisahkannya menjadi data pelatihan dan pengujian. Persentase yang digunakan dalam distribusi dataset adalah 80% data *training* sebagai pembentuk pola atau model dan 20% data *testing* sebagai model *tester*.

### Tahap Modelling

Algoritma C4.5 digunakan dalam penelitian ini untuk melakukan klasifikasi data mining. Secara singkat, proses pembuatan pohon keputusan dengan algoritma C4.5 adalah sebagai berikut: [14]

- Langkah pertama dalam membuat pohon keputusan adalah membuat *root* (berada paling atas). Selanjutnya dilakukan pendistribusian data berdasarkan fitur yang sesuai dengan pembentukan daun.
- Tahap selanjutnya adalah pemangkasan pohon yang dibuat, juga dikenal sebagai pemangkasan pohon, yang melibatkan pencarian dan pemangkasan cabang yang tidak dibutuhkan pada pohon yang terbentuk. Pemangkasan pohon dilakukan untuk mengurangi tingkat kesalahan dalam peramalan kasus baru dari hasil pemecahan dengan cara membagi dan menaklukkan, selain itu untuk memperkecil ukuran pohon. Pemangkasan memiliki dua pendekatan: pra-pemangkasan dan pasca-pemangkasan.
- Langkah selanjutnya adalah membuat aturan keputusan berdasarkan pohon yang telah dibuat. Menelusuri dari akar ke daun pohon keputusan menghasilkan aturan.

Rapidminer dimanfaatkan sebagai *tools* dalam mengimplementasikan algoritma C4.5. Rapidminer merupakan aplikasi yang menyediakan prosedur *preprocessing*, *modelling* sampai dengan visualisasi pada data mining dan *machine learning* [15].

### Tahap Evaluasi

Metode *Confusion Matrix* digunakan sebagai metode evaluasi untuk menghitung nilai akurasi penelitian ini. *Confusion Matrix* dapat digunakan untuk mengevaluasi kinerja algoritma *Machine Learning* (ML). Ini mewakili prediksi serta keadaan aktual dari data yang dibuat oleh algoritma ML [16].

Terdapat tiga *performance metrics* populer yang digunakan dalam mengukur *performance* pada *Confusion Matrix*, yaitu: akurasi, *precision*, dan *recall* [17]. Akurasi didefinisikan sebagai proporsi prediksi yang benar (baik positif maupun negatif) terhadap total data. Presisi didefinisikan sebagai rasio prediksi positif yang benar terhadap keseluruhan hasil yang diharapkan positif. Rasio prediksi positif yang benar untuk semua data positif dan negatif yang salah dikenal sebagai *Recall* (sensitivitas). Persamaan berikut digunakan untuk menghitung nilai akurasi, presisi, dan *recall*:

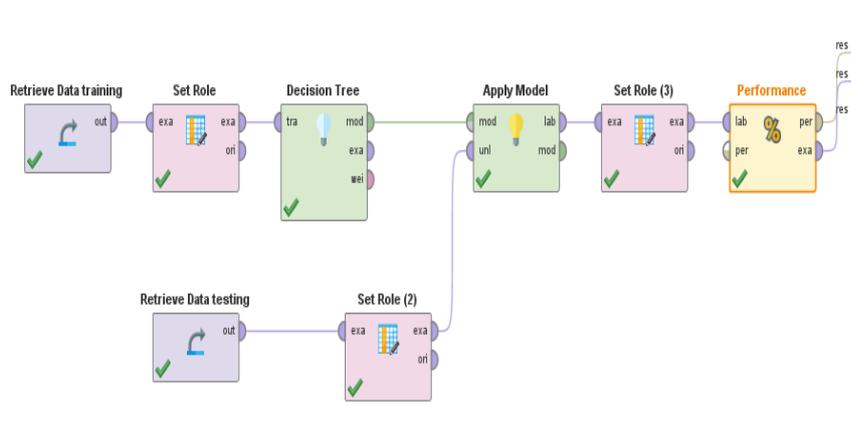
$$Akurasi = \frac{tp + tn}{tp + tn + fp + fn} \quad (1)$$

$$Presisi = \frac{tp}{tp + fp} \quad (2)$$

$$Recall = \frac{tp}{fn + tp} \quad (3)$$

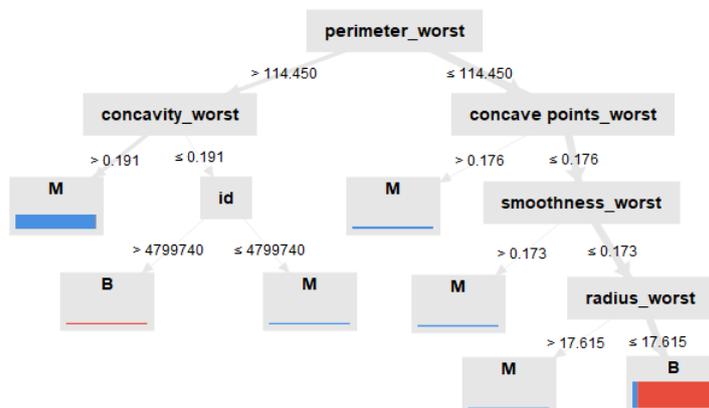
### III. Hasil dan Pembahasan

Dari dataset sebanyak 570 record yang dipisahkan menjadi dua bagian. Data *training* terdiri dari 455 record (80%), sedangkan data *testing* terdiri dari 105 record (20%). Proses klasifikasi menggunakan algoritma C4.5 dilakukan dengan bantuan Rapidminer seperti pada Gambar 3.



Gambar 3. Proses Klasifikasi Menggunakan Rapidminer

Selanjutnya setelah proses klasifikasi akan menghasilkan pohon keputusan. Pohon keputusan ini digunakan sebagai *rule* dalam klasifikasi prediksi penyakit kanker payudara. Gambar 4 merupakan pohon keputusan yang dihasilkan dari pengolahan data menggunakan alat bantu Rapidminer.



Gambar 4. Pohon Keputusan

Berdasarkan gambar 4, terlihat bahwa dari 32 atribut pada dataset, ada 8 atribut yang memiliki penentu seseorang masuk kriteria penderita penyakit kanker payudara ganas atau jinak, yaitu *Id* (nomor identitas pasien), *Perimeter\_worst* (keliling), *Concavity\_worst* (tingkat keparahan dari kontur), *Concave points\_worst* (jumlah konkav), *Smoothness\_worst* (variasi lokal dari nilai radius), *Radius\_worst*, *M* (diagnosis kanker payudara ganas) dan *B* (diagnosis kanker payudara jinak).

Atribut diagnosis (M/B) dijadikan sebagai label penentu yang menentukan hasil akhir penderita kanker payudara gabas atau jinak. M (*malignant*) mewakili prediksi seseorang menderita kanker payudara ganas, sedangkan B (*benign*) mewakili prediksi seseorang menderita kanker payudara jinak.

### Tree

```

perimeter_worst > 114.450
| concavity_worst > 0.191: M {M=147, B=1}
| concavity_worst ≤ 0.191
| | id > 4799740: B {M=0, B=2}
| | id ≤ 4799740: M {M=2, B=0}
perimeter_worst ≤ 114.450
| concave points_worst > 0.176: M {M=12, B=0}
| concave points_worst ≤ 0.176
| | smoothness_worst > 0.173: M {M=5, B=0}
| | smoothness_worst ≤ 0.173
| | | radius_worst > 17.615: M {M=2, B=0}
| | | radius_worst ≤ 17.615: B {M=18, B=265}
    
```

Gambar 5. Rules yang Dihasilkan

Terdapat dua *rules* yang dihasilkan pada gambar 5 yaitu nilai *perimeter\_worst* >114.450 dan nilai *perimeter\_worst* <=114.450. Berikut ini uraian dari rule yang dihasilkan:

- Nilai *perimeter\_worst* >114.450

Jika nilai *perimeter\_worst* >114.450 dan *concavity\_worst* > 0.191. Terdeteksi sebanyak 147 penderita penyakit kanker payudara ganas dan 1 penderita kanker payudara jinak.

Sedangkan jika nilai *concavity\_worst* <= 0.191 pada id > 4799740 terdapat 2 penderita kanker payudara jinak. Untuk id <4799740 terdapat 2 penderita kanker payudara.

- Nilai *perimeter\_worst* <=114.450

Jika nilai *perimeter\_worst* <=114.450 dan *concavity\_worst* >0.176. Terdeteksi sebanyak 12 penderita penyakit kanker payudara ganas.

Jika nilai *perimeter\_worst* <=114.450, *Concavity\_worst* <=0.176, *smoothness\_worst* >0.173, terdeteksi 5 penderita penyakit kanker payudara ganas.

Jika nilai *perimeter\_worst* <=114.450 *Concavity\_worst* <=0.176, *smoothness\_worst* <=0.173, radius\_worstnya > 17.615, terdapat 2 penderita penyakit kanker payudara ganas.

Sedangkan jika nilai *perimeter\_worst* <=114.450 *Concavity\_worst* <=0.176, *smoothness\_worst* <=0.173, radius\_worstnya <= 17.615, terdapat 18 penderita penyakit kanker payudara ganas dan 265 penyakit kanker payudara jinak.

accuracy: 93.04%			
	true B	true M	class precision
pred. B	83	2	97.65%
pred. M	6	24	80.00%
class recall	93.26%	92.31%	

Gambar 6. Hasil Evaluasi

Tahap terakhir yaitu dilakukan evaluasi dengan melihat nilai *performance* yang dihasilkan. Gambar 6 menggambarkan hasil akurasi, *class precision* dan *class recall* dari model yang dibentuk. Nilai akurasi yang dihasilkan sebesar 93,04%, *precision* sebesar 80,00% dan *recall* sebesar 92,31%. Hasil yang diperoleh terdapat peningkatan nilai akurasi dibanding penelitian sebelumnya yang memperoleh akurasi sebesar 85.5% [10].

#### IV. Kesimpulan

Algoritma C4.5 digunakan dalam penelitian ini untuk mengklasifikasikan pasien kanker payudara. Dataset yang digunakan berisi 570 *record*, 455 diantaranya adalah data *training* dan 105 diantaranya adalah data *testing*. Nomor identitas pasien, keliling, keparahan kontur, jumlah cekungan, variasi lokal dari nilai radius, rata-rata jarak rata-rata dari titik tengah ke tepi, diagnosis kanker payudara ganas, dan diagnosis kanker payudara jinak adalah delapan dari 32 atribut yang diperoleh dari kumpulan data yang menentukan apakah seseorang diklasifikasikan sebagai penderita kanker payudara ganas atau jinak.

Berdasarkan hasil evaluasi model klasifikasi diagnosis kanker payudara yang diterapkan, diperoleh nilai akurasi yang cukup signifikan sebesar 93,04%, presisi sebesar 80,00%, dan *recall* sebesar 92,31%, menunjukkan bahwa metode C4.5 merupakan algoritma yang baik dalam model klasifikasi.

#### Daftar Pustaka

- [1].WHO, "Cancer," 2016. [https://www.who.int/health-topics/cancer#tab=tab\\_1](https://www.who.int/health-topics/cancer#tab=tab_1) (accessed Jan. 12, 2023).
- [2].Global Cancer Observatory, "Cancer Over Time," 2020. [Online]. Available: <https://gco.iarc.fr/>
- [3].Agung, "Kalkulasi Faktor Risiko Kanker Payudara Perlu Dilakukan," *Universitas Gadjah Mada*, 2020. <https://www.ugm.ac.id/id/berita/20137-kalkulasi-faktor-risiko-kanker-payudara-perlu-dilakukan> (accessed Feb. 10, 2023).
- [4].W. W. Ariestya, W. Supriyatin, and I. Astuti, "Marketing Strategy for the Determination of Staple Consumer Products Using Fp-Growth and Apriori Algorithm," *J. Ilm. Ekon. Bisnis*, vol. 24, no. 3, pp. 225–235, 2019, doi: 10.35760/eb.2019.v24i3.2229.
- [5].U. Ependi and A. Putra, "Solusi Prediksi Persediaan Barang dengan Menggunakan Algoritma Apriori (Studi Kasus: Regional Part Depo Auto 2000 Palembang)," *JEPIN (Jurnal Edukasi dan Penelit. Inform.)*, vol. 5, no. 2, pp. 139–145, Aug. 2019, doi: 10.26418/JP.V5I2.32648.
- [6].F. Elfaladonna and A. Rahmadani, "Analisa Metode Classification-Decission Tree Dan Algoritma C.45 Untuk Memprediksi Penyakit Diabetes Dengan Menggunakan Aplikasi Rapid Miner," *SINTECH (Science Inf. Technol. J.)*, vol. 2, no. 1, pp. 10–17, Apr. 2019, doi: 10.31598/SINTECHJOURNAL.V2I1.293.
- [7].M. Ture, F. Tokatli, and I. Kurt, "Using Kaplan–Meier analysis together with decision tree methods (C&RT, CHAID, QUEST, C4.5 and ID3) in determining recurrence-free survival of breast cancer patients," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 2017–2026, Mar. 2009, doi: 10.1016/J.ESWA.2007.12.002.
- [8].T. Wulandari, M. Marji, and L. Muflikkah, "Klasifikasi Jenis Kanker Berdasarkan Struktur Protein Menggunakan Algoritma Naive Bayes," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 10, pp. 3738–3743, Feb. 2018, Accessed: Jul. 07, 2023. [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/2718>
- [9].M. A. E. Setyorini, "Analisis perbandingan metode machine learning: random forest dan support vector machine untuk deteksi kanker paru-paru," Universitas Jember, 2020. Accessed: Jul. 07, 2023. [Online]. Available: <https://repository.unej.ac.id/handle/123456789/101771>
- [10].M. Maryam and H. W. Ariono, "Sistem Pakar Pengklasifikasi Stadium Kanker Serviks Berbasis Mobile Menggunakan Metode Decision Tree," *J. Kaji. Ilm.*, vol. 22, no. 3, pp. 267–278, Sep. 2022, doi: 10.31599/JKI.V22I3.1368.
- [11].A. R. Pratama, TG, "Application of the C4. 5 Algorithm for Early Cervical Cancer Classification," *Urecol J.*, vol. 1, no. 1, p. 1, 2021, doi: 10.53017/uje.4.
- [12].D.-K. I. J. I. K. Gunawan, "Evaluasi Performa Pemecahan Database dengan Metode Klasifikasi Pada Data

- Preprocessing Data mining,” *J. Ilmu Komput. dan Inform. Khazanah Inform.*, vol. 2, no. 1, pp. 10–13, 2016, Accessed: Feb. 02, 2023. [Online]. Available: <https://journals.ums.ac.id/index.php/khif/article/view/1749>
- [13].UCI, “Breast Cancer Coimbra - UCI Machine Learning Repository,” 2018. <https://archive.ics.uci.edu/dataset/451/breast+cancer+coimbra> (accessed Dec. 12, 2022).
- [14].F. M. Hana, “Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4. 5,” *J. Sist. Komput. dan Kecerdasan Buatan*, vol. IV, no. 1, pp. 32–39, 2020, doi: <https://doi.org/10.47970/siskom-kb.v4i1.173>.
- [15].D. Novianti, “Implementasi Algoritma Naïve Bayes Pada Data Set Hepatitis Menggunakan Rapid Miner,” *J. Inform. dan Komput.*, vol. 21, no. 1, pp. 49–54, 2019, doi: 10.31294/p.v20i2.
- [16].I. Astuti, W. Ariestya, and B. Solehudin, “Deteksi Objek Daun Semanggi Secara Real Time Menggunakan CNN-Single Shot Multibox Detector (SSD),” *J. Ilm. FIFO*, vol. XIV, no. 1, pp. 47–58, 2022, doi: 10.22441/fifo.2022.v14i1.005.
- [17].Kuncahyo Setyo Nugroho, “Confusion Matrix untuk Evaluasi Model pada Supervised Learning,” 2019. <https://ksnugroho.medium.com/confusion-matrix-untuk-evaluasi-model-pada-unsupervised-machine-learning-bc4b1ae9ae3f> (accessed Jan. 02, 2023).