

Breast Cancer Classification Model Using Decision Tree Algorithm

Model Klasifikasi Kanker Payudara Menggunakan Algoritma Decision Tree

Ismi Asmawati¹, Syamsul Bahri², Nuzla Af'idatur Robbaniyah^{3*}

^{1,2,3} *Departement of Mathematics, Mataram University*

Email: ismiasmawati09@gmail.com, syamsul.math@unram.ac.id, nuzla@unram.ac.id

**Corresponding author*

Received: 19 January 2026, revised: 6 April 2026, accepted: 13 April 2026

Abstract

Cancer is a disease characterized by the presence of abnormal cells or tissues that grow rapidly, uncontrollably, can spread to other parts of the patient's body and it can also sometimes be malignant. According to the International Agency for Research on Cancer, in 2024 breast cancer will rank second in terms of the highest number of cases and fourth as the leading cause of death globally. The objective of this study is to apply the Classification and Regression Tree (CART) decision tree algorithm to a breast cancer classification model based on patient medical records obtained from the Kaggle website using the Wisconsin method. The model developed has a specificity of 95.77%, recall, precision, and F1-Score of 93.02%, and accuracy of 94.74%. The model was evaluated using a confusion matrix to measure its performance. Thus, the CART algorithm can be applied in classification models, and the resulting model is considered optimal as it achieves percentages within the 90%-100% range for all performance evaluation metrics.

Keywords: Breast Cancer, Decision Tree Algorithm, CART.

Abstrak

Kanker merupakan penyakit yang ditandai dengan adanya sel atau jaringan abnormal, tumbuh cepat tidak terkendali, serta dapat menyebar ke tempat lain dalam tubuh penderita dan terkadang dapat bersifat ganas. Menurut *International Agency for Research on Cancer* menyatakan bahwa, pada tahun 2024 kanker payudara menempati urutan kedua dengan kasus terbanyak dan berada pada urutan keempat sebagai penyebab utama kematian secara *global*. Tujuan dari penelitian ini yaitu menerapkan algoritma *decision tree* jenis *Classification And Regression Tree* (CART) pada model klasifikasi kanker payudara berdasarkan data rekam medis pasien yang didapatkan dari *Website Kaggle* yang menggunakan metode *wisconsin*. Model yang telah dibuat memiliki *specificity* sebesar 95,77%, *recall*, *precision*, serta F1-Score sebesar 93,02% dan *accuracy* sebesar 94,74%. Evaluasi model menggunakan *confusion matrix* untuk mengukur kinerja model yang telah dibuat. Dengan demikian, algoritma CART dapat diterapkan dalam model klasifikasi dan model yang dihasilkan bersifat sempurna karena memiliki persentase yang berada pada *range* 90%-100% untuk semua komponen metrik evaluasi kinerja model.

Kata kunci: Kanker Payudara, Algoritma *Decision Tree*, CART.



1. PENDAHULUAN

Kanker payudara adalah tumor ganas yang tumbuh di dalam jaringan payudara bersifat ganas, tumbuh cepat tidak terkendali dan dapat menyebar [11]. Kanker payudara merupakan salah satu jenis kanker yang mematikan di dunia, penyakit ini umumnya terjadi pada kalangan wanita akan tetapi tidak menutup kemungkinan menyerang pria dengan perbandingan 1:100 [7]. Kanker dibagi menjadi dua jenis yaitu jinak dan ganas. Kanker jinak tidak menyerang jaringan yang berdekatan dan tidak menyebarkan benih tetapi tumbuh secara bertahap dan menjadi besar serta tidak akan timbul kembali setelah pelaksanaan operasi [14]. Sedangkan kanker ganas mampu menyebar melalui pembuluh darah ke organ lain meskipun telah dilakukan operasi sisa sell kanker akan tetap ada dan memicu pertumbuhan sel menjadi lebih besar [12]. Menurut *International Agency for Research on Cancer* (IARC) pada tahun 2024 kanker payudara menjadi jenis kanker dengan kasus terbanyak pada urutan kedua dan penyebab utama kematian akibat kanker pada urutan ke-empat [5]. Menurut Elvida Sariwati selaku Plt Direktur Pencegahan dan Pengendalian Penyakit Tidak Menular memaparkan bahwa “kasus kanker di Indonesia 70% diantaranya sudah berada pada tahap lanjut, dimana 43% dari kematian akibat kanker yang dapat dikalahkan jika pasien melakukan deteksi dini dan menghindari faktor resiko penyebab kanker”.

Meningkatnya kasus kanker di tiap tahunnya menjadi perhatian banyak pihak terlebih tenaga medis. Salah satu cara yang dapat dimanfaatkan dalam menekan kasus peningkatan kanker payudara yaitu dengan memanfaatkan perkembangan teknologi salah satunya *machine learning*. Pemanfaatan teknologi digunakan untuk meningkatkan kualitas pelayanan seperti diagnosis yang lebih akurat, pengelolaan data pasien yang efisien hingga pengembangan terapi yang lebih terpersonalisasi [2]. *Machine learning* adalah sebuah mesin yang dibuat untuk dapat belajar dan melakukan pekerjaan tanpa arahan dari penggunanya. Teknologi ini berkaitan erat dengan algoritma yang dijadikan sebagai pondasi, salah satu algoritma yang dapat digunakan ialah algoritma *decision tree* [15]. Algoritma *decision tree* merupakan struktur pohon yang terdiri dari *node* yang merepresentasikan keputusan dan cabang-cabang yang merepresentasikan konsekuensi dari sebuah keputusan [4]. *Decision tree* merupakan model prediksi yang memerlukan *training dataset* dimana perannya untuk menggantikan pengalaman manusia di masa lalu dalam membuat keputusan [9].

Beberapa penelitian sebelumnya yang meneliti tentang penerapan algoritma *decision tree* pada model klasifikasi yang dibentuk dengan objek penelitian berupa kanker payudara berbasis *machine learning*. Penelitian Ibrahim *et al* tahun 2023 menemukan hasil bahwa algoritma *decision tree* dan *ensambel* memiliki tingkat akurasi yang tinggi jika dibandingkan dengan beberapa algoritma klasik. Proses perhitungan akurasi menggunakan algoritma pemilihan fitur, dan dipatkan hasil sebesar 98,6%. Adapun algoritma klasik yang dimaksud yaitu *decision tree*, *linear discriminant*, *logistic regression*, *support vector machine*, dan *ensambel* [6].

Penelitian Tarawaneh *et al* tahun 2022 mendapatkan kesimpulan bahwa algoritma *decision tree* memiliki tingkat akurasi sebesar 100% sedangkan *random forest* memiliki tingkat akurasi sebesar 50%. Oleh sebab itu, algoritma *decision tree* dinilai lebih baik karena tingkat akurasinya tinggi. Perhitungan akurasi menggunakan beberapa metrik seperti *F-Measure*, *Receiver Operating Characteristic Area (ROC Area)*, *Precision*, *Recall*, *TP Rate* dan *FP Rate* [13].

Penelitian Keerthika *et al* tahun 2021 menunjukkan bahwa algoritma *decision tree* memiliki akurasi yang sama dengan algoritma K-NN yakni sebesar 97%. Akan tetapi, jika dilihat dari waktu yang diperlukan dalam memprediksi. Algoritma *decision tree* hanya memerlukan sekitar 0,07 detik sedangkan K-NN memerlukan waktu lebih lama. Selain itu, dalam pelatihan modelnya algoritma K-NN tidak banyak memberikan solusi. Oleh sebab itu, algoritma *decision tree* dinilai lebih optimal [8].

Pada penelitian kali ini peneliti akan menggunakan algoritma *decision tree* jenis *Classification And Regression Tree (CART)*, jenis algoritma ini lebih bagus jika dibandingkan dengan jenis

lainnya seperti 1D3, C4.5 atau C5.0 [1]. Sebagian besar penelitian sebelumnya menggunakan algoritma C4.5 atau ID3 karena strukturnya yang lebih sederhana dan mudah untuk diterapkan. Selain itu, pada penelitian kali ini peneliti akan menggunakan *dataset* sebanyak 569 data dan variabel inputan yang lebih bervariasi serta akan diberikan matriks korelasi untuk mengetahui variabel mana yang memiliki pengaruh terbesar terhadap diagnosis pasien kanker payudara. Tujuan dari penelitian ini untuk mengetahui bagaimana algoritma yang digunakan pada model dan bagaimanakah performa dari model tersebut.

2. METODE PENELITIAN

Pada penelitian ini terdapat beberapa proses langkah dalam penyelesaian penelitian. Langkah yang pertama adalah *preprocessing* data. Pada tahap ini dilakukan proses penghapusan kolom id dan transformasi data menggunakan standarisasi. Penghapusan kolom id dilakukan karena *input* data tidak mempengaruhi model yang akan dibuat. Standarisasi digunakan karena untuk mempertahankan *outlier* hasil pemeriksaan pasien karena hasil pemeriksaan pasien seringkali berada pada satuan yang berbeda serta memiliki perbedaan hasil pemeriksaan yang amat jauh antara pasien yang satu dengan lainnya [10]. Adapun persamaan yang digunakan ialah Z-Score yakni sebagai berikut:

$$z = \frac{x - \bar{x}}{s} \quad (2.1)$$

dengan

- x : Nilai asli
- \bar{x} : Rata-rata (mean) dari fitur
- s : Standar deviasi dari fitur.

Berikut persamaan dari mean dan standar deviasi dari fitur:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (2.2)$$

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (2.3)$$

dengan x_i merupakan nilai data ke-i.

Setelah melakukan *preprocessing data* maka tahap berikutnya ialah organisasi data dengan *output* berupa matriks korelasi, analisis korelasi menggunakan *point biserial* karena variabel yang akan dihubungkan memiliki tipe yang berbeda. Inputannya bersifat numerik kontinu sedangkan *output* bersifat biner, berikut persamaan korelasi *point biserial*:

$$r_{pb} = \frac{M_{ganas} - M_{jinak}}{s} \times \sqrt{\frac{n_{ganas} \times n_{jinak}}{n^2}} \quad (2.4)$$

dengan

- r_{pb} : Korelasi *point biserial*
- M_{ganas} : Rata-rata dari kelompok ganas
- M_{jinak} : Rata-rata dari kelompok jinak

- s : Standar deviasi
 n_{ganas} : Jumlah sampel kelompok ganas
 n_{jinak} : Jumlah sampel kelompok jinak
 n : Jumlah total sampel

Untuk mengukur kinerja model peneliti menggunakan *confusion matrix* dengan komponen evaluasi kinerja model yang terdiri dari *specificity*, *recall*, *precision*, *accuracy*, dan *f1-score* [3].

Tabel 2.1. Bentuk *confusion matrix*

Kelas	Kelas hasil klasifikasi	
	Jinak	Ganas
Jinak	TN	FP
Ganas	FN	TP

Berikut persamaan dari beberapa komponen evaluasi kerja model:

$$specificity = \frac{TN}{TN + FP} \times 100\% \quad (2.5)$$

$$recall = \frac{TP}{TP + FN} \times 100\% \quad (2.6)$$

$$precision = \frac{TP}{TP + FP} \times 100\% \quad (2.7)$$

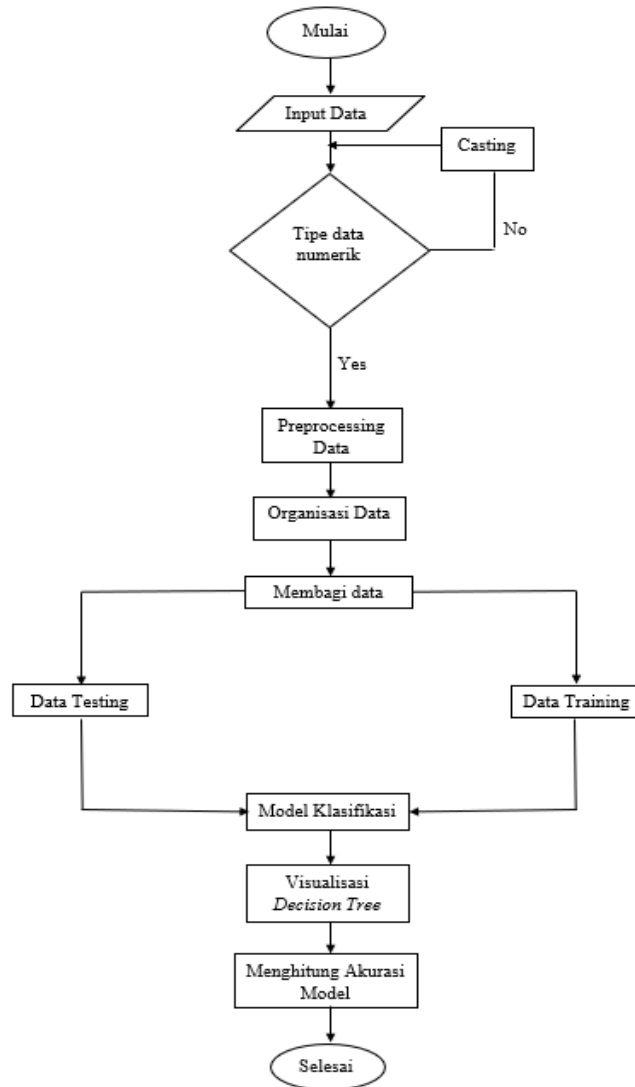
$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (2.8)$$

$$f1 - score = \frac{2(precision \times recall)}{precision + recall} \times 100\% \quad (2.9)$$

dengan

- TN : Jumlah data dengan nilai sebenarnya jinak dan hasil prediksi jinak
 FP : Jumlah data dengan nilai sebenarnya jinak dan hasil prediksi ganas
 FN : Jumlah data dengan nilai sebenarnya ganas dan hasil prediksi jinak
 TP : Jumlah data dengan nilai sebenarnya ganas dan hasil prediksi ganas

Gambar 2.1 merupakan diagram alir untuk memudahkan mengetahui proses langkah pada penelitian ini.



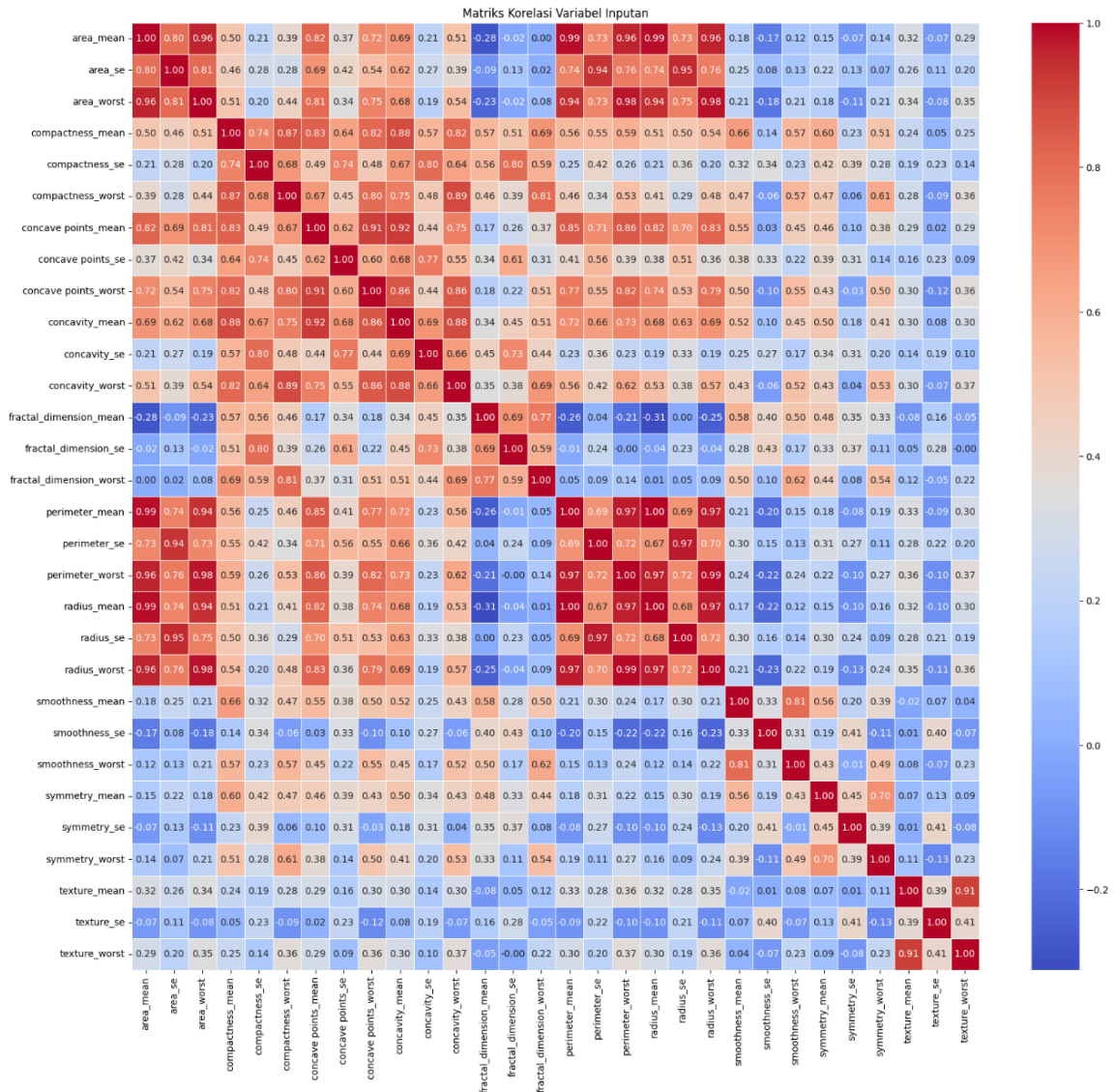
Gambar 2.1. Diagram alir penellitian

3. HASIL DAN PEMBAHASAN

Pada bagian hasil dan pembahasan akan memaparkan hasil penelitian yang diperoleh dari proses pengolahan dan analisis data. Setelah melewati tahap *preprocessing* data, tahap berikutnya adalah organisasi data dengan menggunakan semua *dataset* dan *output* berupa matriks korelasi. Pada penelitian ini matriks korelasi dijadikan sebagai output agar pembaca atau analisis dapat melihat korelasi antara satu dengan lainnya dengan jelas. Sehingga dapat diketahui variable mana yang perlu dijadikan sebagai titik perhatian utama untuk mengurangi kemungkinan terburuk dari pasien tersebut. Matriks korelasi memiliki dua arah hubungan yaitu positif dan negatif, hubungan positif bergerak searah sedangkan negatif bergerak berlawanan arah. Selain dari nilai yang dihasilkan, warna pada matriks korelasi dapat merepresentasikan hubungan. Semakin merah maka korelasi semakin positif dan semakin biru maka korelasi semakin negatif. Perhitungan korelasi menggunakan Persamaan 2.4, berikut matriks korelasi yang didapatkan:

JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI

Ismi Asmawati, Syamsul Bahri, Nuzla Af'idatur Robbaniyah



Gambar 3.1. Matriks korelasi dari variabel inputan

Matriks pada Gambar 3.1 hanya melihat hubungan variabel satu terhadap variabel lainnya bukan terhadap output atau diagnosis yang dihasilkan. Berikut urutan dari variabel dengan pengaruh terbesar terhadap diagnosis pasien:

Tabel 3.1. Urutan variabel yang memiliki pengaruh terbesar

No.	Variabel	Nilai korelasi	No.	Variabel	Nilai korelasi
1.	<i>concave points_worst</i>	0.794	16.	<i>texture_worst</i>	0.457
2.	<i>perimeter_worst</i>	0.783	17.	<i>smoothness_worst</i>	0.421
3.	<i>concave points_mean</i>	0.777	18.	<i>symmetry_worst</i>	0.416

4.	<i>radius_worst</i>	0.776	19.	<i>texture_mean</i>	0.415
5.	<i>perimeter_mean</i>	0.743	20.	<i>concave points_se</i>	0.408
6.	<i>area_worst</i>	0.734	21.	<i>smoothness_mean</i>	0.359
7.	<i>radius_mean</i>	0.730	22.	<i>symmetry_mean</i>	0.330
8.	<i>area_mean</i>	0.709	23.	<i>fractal_dimension_worst</i>	0.324
9.	<i>concavity_mean</i>	0.696	24.	<i>compactness_se</i>	0.293
10.	<i>concavity_worst</i>	0.660	25.	<i>concavity_se</i>	0.254
11.	<i>compactness_mean</i>	0.597	26.	<i>fractal_dimension_se</i>	0.078
12.	<i>compactness_worst</i>	0.591	27.	<i>smoothness_se</i>	-0.067
13.	<i>radius_se</i>	0.567	28.	<i>fractal_dimension_mean</i>	-0.013
14.	<i>perimeter_se</i>	0.556	29.	<i>texture_se</i>	-0.008
15.	<i>area_se</i>	0.548	30.	<i>symmetry_se</i>	-0.007

Berdasarkan Tabel 3.1 dapat diketahui bahwa *concave point_worst* sangat mempengaruhi diagnosis pasien kanker payudara diikuti oleh variabel lainnya seperti *perimeter_worst*, *concave point_mean*, *radius_worst*, dan lain sebagainya. Tahap berikutnya ialah membangun model. Sebelum membangun model langkah awal yang dilakukan adalah dengan membagi data menjadi dua, yaitu data *training* dan data *testing*. Pembagian data dilakukan secara acak, data *training* digunakan sebesar 80% dan data *testing* sebesar 20%. Data *training* digunakan untuk membuat model dan data *testing* digunakan untuk mengukur kinerja model yang telah dibuat. Model yang dihasilkan berupa struktur pohon yang meliputi cabang, nilai gini, *threshold*, dan fitur yang digunakan di tiap simpulnya. Berikut deskripsi algoritma CART yang digunakan pada penelitian ini:

1. Memisahkan data *input* dan *output*, dataset dengan fitur X_1, X_2, \dots, X_n sebagai input serta label Y_1 (jinak) dan label Y_2 (ganas) sebagai output.
2. Jika semua data dalam dataset berada pada satu kelas maka dibentuk *leaf* dengan label kelas tersebut.
3. Jika tidak, untuk setiap fitur dalam dataset akan mengalami proses sebagai berikut:
 - a. Menghitung nilai gini berdasarkan karakteristik pada node sebelumnya (sebelum dilakukannya *split*) menggunakan persamaan

$$G_{fitur} = 1 - \sum_{i=1}^k (p_i)^2 \quad (3.1)$$

- b. Mencari semua kemungkinan nilai *threshold* dengan mengambil nilai tengah antara sampel yang satu dengan yang lainnya pada fitur tersebut.
- c. Berdasarkan nilai *threshold* yang didapatkan, sampel pada *node* tersebut akan terbagi menjadi dua yang memenuhi kondisi *true* dan *false*. Kondisi *true* didapatkan ketika nilai sampel $\leq threshold$, sedangkan kondisi *false* terpenuhi ketika nilai sampel $> threshold$.
- d. Berikutnya ialah menghitung nilai gini split untuk setiap nilai *threshold* dimana nilai yang didapatkan berdasarkan karakteristik *node* dari tiap *threshold* menggunakan persamaan

$$G_{split} = \frac{n_{true}}{n} \times G_{true} + \frac{n_{false}}{n} \times G_{false} \quad (3.2)$$

- e. Selanjutnya ialah menghitung nilai gain dimana nilai *threshold* terbaik merupakan *threshold* dengan nilai gain terbesar menggunakan persamaan

$$Gain = Gini(node\ yang\ berkaitan) - G_{split} \quad (3.3)$$

- f. Pilih fitur dan *threshold* berikutnya untuk dijadikan sebagai *node*.
4. Setelah didapatkan *node* terbaru dengan mengulang langkah 3 secara rekursif untuk masing-masing subset. Dimana subset kiri berisi data yang memenuhi kondisi *true* dan subset kanan berisi data yang memenuhi kondisi *false*.

2. *Recall* menggunakan Persamaan 2.6 didapatkan hasil sebesar 93,02%. Dengan contoh kasus 200 kasus kanker, 100 diantaranya merupakan kasus kanker jinak dan 100 lainnya merupakan kasus kanker ganas. Model mendeteksi 93 kasus kanker ganas dengan benar dan 7 kasus kanker ganas dideteksi sebagai kanker jinak. Sehingga oleh model mendapatkan hasil 93 kasus kanker ganas dan 107 kasus kanker jinak.
3. *Precision* menggunakan Persamaan 2.7 didapatkan hasil sebesar 93,02%. Dengan contoh kasus 200 kasus kanker, 100 diantaranya merupakan kasus kanker jinak dan 100 lainnya merupakan kasus kanker ganas. Model mendeteksi 93 kasus kanker jinak dengan benar dan 7 kasus kanker jinak dideteksi sebagai kanker ganas. Sehingga oleh model mendapatkan hasil 93 kasus kanker jinak dan 107 kasus kanker ganas.
4. *Accuracy* menggunakan Persamaan 2.8 didapatkan hasil sebesar 94,74%. Dengan contoh 100 kasus kanker dengan berbagai kemungkinan seperti jinak diprediksi ganas dan ganas diprediksi jinak. Dari 100 kasus, 94 diantaranya diprediksi dengan benar.
5. *F1-score* menggunakan persamaan 2.9 didapatkan hasil sebesar 93,02% yang menunjukkan bahwa performa model secara keseluruhan menunjukkan kinerja yang baik.

Berikutnya ialah menghitung kompleksitas algoritma yang terdiri dari kompleksitas algoritma *training* dan *testing*, kompleksitas *training* sebesar $O\left(24n \times \log_{\frac{4}{5}} n\right)$ sedangkan kompleksitas *testing* yaitu $O\left(\frac{7}{5}n\right)$. Total kompleksitas dari model yang telah dibuat sebesar $O\left(24n \times \log_{\frac{4}{5}} n\right) + O\left(\frac{7}{5}n\right)$ kompleksitas tersebut didapatkan berdasarkan hasil penjumlahan antara proses *training* dan *testing* pada model.

Berdasarkan hasil dan analisis yang telah dilakukan, didapatkan bahwa algoritma decision tree jenis CART dapat digunakan dalam model klasifikasi kanker payudara. Beberapa komponen metrik evaluasi kinerja model yang diuji antara lain specificity, recall, precision, accuracy dan f1score. Model yang telah dibuat tergolong pada kategori model sempurna, ditunjukkan dengan nilai persentase dari semua komponen evaluasi kinerja model yang berada pada 90% - 100%. Hal tersebut menunjukkan bahwa metode decision tree jenis CART bagus digunakan untuk mengklasifikasikan kasus pada kondisi medis yang berfokus pada kanker payudara.

4. KESIMPULAN

Algoritma *decision tree* dapat diterapkan pada model klasifikasi untuk memprediksi jenis kanker pada penderita kanker payudara. Penerapan algoritma *decision tree* jenis CART dilakukan melalui beberapa tahapan seperti persiapan data, pemisahan data, konstruksi model, evaluasi model dan visualisasi. Hasil dari proses permodelan tersebut berupa sebuah pohon keputusan yang memiliki 15 simpul dan 16 daun. Model klasifikasi kanker payudara yang dibuat menunjukkan kinerja yang sangat baik berdasarkan hasil evaluasi dimana hasil dari *specificity*, *recall*, *precision*, *accuracy* dan *F1-Score* secara berturut-turut sebesar 95,77%, 93,02%, 93,02%, 94,74% dan 93,02%.

Pengembangan sistem pada model yang telah dibuat dapat dilakukan dengan menerapkan penyimpanan data menggunakan basis data yang terstruktur. Basis data digunakan untuk menyimpan data pasien, data hasil *preprocessing*, serta hasil prediksi dan evaluasi model. Dengan adanya penyimpanan data yang terintegrasi, informasi tersebut dapat dimanfaatkan dikemudian hari untuk melakukan analisa lanjutan, pelatihan ulang model dengan data baru, serta memantau kinerja model secara berkelanjutan. Selain itu, penerapan basis data juga mendukung pengolahan data yang lebih efisien.

UCAPAN TERIMAKASIH

Terima kasih kami ucapkan kepada Program studi Matematika, Fakultas MIPA, Universitas Mataram atas dukungannya sehingga penelitian ini dapat terselesaikan dengan baik.

REFERENSI

- [1] Arrahimi, A. R., Ihsan, M. K., Kartini, D., Faisal, M. R., & Indriani, F., 2019. Teknik Bagging Dan Boosting Pada Algoritma CART Untuk Klasifikasi Masa Studi Mahasiswa. *Jurnal Sains Dan Informatika*, 5(1), 21–30.
- [2] Br.Sinulingga, P. S., & Nasution, M. I. P., 2024. *ANALISIS TANTANGAN DAN PELUANG DALAM PERKEMBANGAN TEKNOLOGI INFORMASI DAN KOMUNIKASI DI ERA DIGITAL: PERSPEKTIF MASA DEPAN*. 2(12), 25–35.
- [3] Breiman, L., Friedman, J., Olshen, R. A., & Stone, C. J., 1984. *Classification and Regression Trees*
- [4] Feby., 2023. *Apa Itu Decision Tree di Machine Learning Model?* <https://dqlab.id/apa-itu-decision-tree-di-machine-learning-model>
- [5] IARC., 2024. GLOBAL CANCER OBSERVATORY. *International Agency for Research on Cancer, 2022*, 9–10.
- [6] Ibrahim, M., Sedky, A. A. H., & Mesbah, S. (2023). Accuracy Assessment of Machine Learning Algorithms Used to Predict Breast Cancer. *Data*.
- [7] Imaduddin, H., Hermansyah, B. A., & Salsabilla B, F. A., 2021. ARISON OF SUPPORT VECTOR MACHINE AND DECISION TREE METHODS IN THE CLASSIFICATION OF BREAST CANCER. *Cyberspace: Jurnal Pendidikan Teknologi Informasi*, 5(1), 22.
- [8] Keerthika, J., Sruthi, D., Swathi, D., Swetha, S., & Vinupriya, R. (2021). Diagnosis of Breast Cancer using Decision Tree Data Mining Tecnique. *2021 7th International Conference on Advanced Computing and Communication Systems, ICACCS 2021*, 1530-1535
- [9] Kurniawan, D., 2020. *Pengenalan Machine Learning dengan Python*. PT. Elex Media Komputindo.
- [10] Lopez, M. A., Duarte, O. C. M. B., & Pujolle, G., 2019. *A Monitoring and Threat Detection System Using Stream Processing as a Virtual Function for Big Data*.
- [11] Novianti, F. A., & Purnami, S. W., 2012. Analisis Diagnosis Pasien Kanker Payudara Menggunakan Regresi Logistik dan Support Vector Machine (SVM) Berdasarkan Hasil Mamografi. *Jurnal SAINS Dan Seni ITS*, 1(1), D147–D152.
- [12] Putri, S. I., Pratamaningtyas, S., Fajriah, A. S., & Widiyanto, A., 2024. Nutrition Knowledge and Food Frequency Among People with Diabetes Mellitus. *Journal of Local Therapy*, 3(1), 18.
- [13] Tarawneh, O., Otair, M., Husni, M., Abuaddous, H. Y., Tarawneh M., & Almomani, M. A. (2022). Breast Cancer Classification using Decision Tree Algorithms. *International Journal of Advanced Computer Science and Applications*, 13(4), 676-680.
- [14] Ung, E. J., Olofsson, A. C., Björkman, I., Hallén, T., Olsson, D. S., Ragnarsson, O., Skoglund, T., Jakobsson, S., & Johannsson, G., 2019. The pre-and postoperative illness trajectory in patients with pituitary tumours. *Endocrine Connections*, 8(7), 878–886.
- [15] Wijoyo A, Saputra A, Ristanti S, Sya'ban S, Amalia M, & Febriansyah R., 2024. Pembelajaran Machine Learning. *OKTAL (Jurnal Ilmu Komputer Dan Science)*, 3(2), 375–380.