

# PERBANDINGAN METODE DECISION TREE DAN NAIVE BAYES UNTUK MEMPREDIKSI *THYROID CANCER RECURRENCE*

Fidya Hafizd<sup>1</sup>, Dian Rizky Julyani<sup>2</sup>, Hasna Yuliza<sup>3</sup>, Emely Nemy Agustine<sup>4</sup>,  
Kessya Immanuella Surbakti<sup>5</sup>, Iman Paryudi<sup>6</sup>

Program Studi Teknik Informatika, Universitas Pancasila, Jakarta<sup>1,2,3,4,5,6</sup>

fidyahafizd234@gmail.com<sup>1</sup>, rizkyjulyanidian@gmail.com<sup>2</sup>, <sup>3</sup>yulizahasna@gmail.com,  
nemyagustine31@gmail.com<sup>4</sup>, kessya.surbakti18@gmail.com<sup>5</sup>, iman.paryudi@univpancasila.ac.id<sup>6</sup>

**Abstrak** – Penelitian ini bertujuan untuk memprediksi kekambuhan kanker tiroid dengan membandingkan dua metode data mining, yaitu *Decision Tree* dan *Naive Bayes*. Data yang digunakan merupakan data klasifikasi yang telah melalui proses *preprocessing* dan pemodelan, kemudian diuji menggunakan uji *test and score* pada perangkat lunak analisis bernama *Orange*. Dengan menggunakan *Orange* sebagai alat analisis, dilakukan percobaan untuk menentukan metode mana yang memberikan akurasi paling baik. Hasil penelitian menunjukkan bahwa kedua metode tersebut memiliki perbandingan akurasi yang berbeda dalam memprediksi kekambuhan kanker tiroid. Penelitian ini diharapkan dapat membantu dalam mengidentifikasi gejala-gejala yang berisiko tinggi menyebabkan kekambuhan kanker tiroid serta memberikan wawasan berharga dalam analisis data.

**Kata Kunci** – Data Mining, Klasifikasi, *Naive Bayes*, *Decision Tree*, *Orange*

## I. PENDAHULUAN

Data mining adalah metode untuk mengidentifikasi nilai tambah dalam basis data dengan menganalisis pola-pola data, sehingga memperoleh informasi berharga yang tidak dapat ditemukan secara manual [1]. Data mining disebut juga sebagai ilmu yang bertujuan untuk mengekstraksi informasi yang berguna dari data besar melalui berbagai metode analisis. Salah satu aplikasi penting dari Data mining adalah proses klasifikasi yang bertujuan untuk mengelompokkan data ke dalam kategori yang sudah ditetapkan sebelumnya berdasarkan karakteristik tertentu. Klasifikasi memiliki berbagai aplikasi, termasuk dalam bidang kesehatan untuk prediksi penyakit dan hasil perawatan.

Penelitian ini akan membahas dan menguji dua metode klasifikasi utama, yaitu *Decision Tree* dan *Naive Bayes*, dalam prediksi kekambuhan kanker tiroid. Kedua metode ini dipilih karena masing-masing memiliki pendekatan yang unik dalam memproses dan menganalisis data. *Naive Bayes* adalah metode klasifikasi yang menggunakan pendekatan probabilitas dan statistik yang diperkenalkan oleh ilmuwan Inggris Thomas Bayes. Metode ini memprediksi kemungkinan di masa depan berdasarkan data historis, sehingga dikenal sebagai Teorema Bayes [2]. Algoritma ini sering digunakan dalam masalah klasifikasi teks dan dokumen. *Decision tree* merupakan algoritma yang sangat populer dan efektif untuk klasifikasi dan prediksi. Algoritma ini mampu

merepresentasikan aturan dari banyak fakta dalam bentuk pohon keputusan. Pohon keputusan adalah struktur yang memecah data dalam jumlah besar menjadi kelompok-kelompok data yang lebih kecil [3].

Dalam penelitian ini, kami akan menggunakan dataset klasifikasi yang berisi data pasien kanker tiroid untuk memprediksi kemungkinan kekambuhan penyakit. Dataset ini mencakup berbagai fitur klinis dan demografis yang relevan. Percobaan akan dilakukan untuk membandingkan kinerja dari kedua metode klasifikasi tersebut berdasarkan matrik evaluasi seperti akurasi, presisi, dan *recall*. Dengan demikian, penelitian ini bertujuan untuk menetapkan metode klasifikasi yang lebih efektif dalam memprediksi kekambuhan kanker tiroid, yang diharapkan dapat memberikan kontribusi signifikan dalam bidang kesehatan, khususnya dalam upaya meningkatkan kualitas perawatan pasien kanker tiroid.

## II. DASAR TEORI

### 1. Data Mining

Definisi dasar data mining adalah mengidentifikasi informasi yang tersembunyi dalam basis data. Data mining adalah salah satu komponen dari Knowledge Discovery in Databases (KDD) yang bertujuan untuk menemukan informasi dan pola yang berguna dalam data [4]. Proses data mining juga dapat diartikan sebagai upaya menggali informasi berharga dari data yang besar dan rumit dengan menggunakan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin. Data mining juga memiliki beberapa tujuan utama, seperti mengidentifikasi pola, tren, dan hubungan tersembunyi dalam data yang dapat mendukung pengambilan keputusan dalam dunia bisnis, memprediksi kemungkinan yang akan terjadi, atau memberikan pemahaman yang mendalam tentang fenomena tertentu.

### 2. Decision Tree

*Decision tree* atau pohon keputusan adalah salah satu metode dari algoritma klasifikasi yang menggunakan struktur data berbentuk *tree* untuk menentukan kelas suatu data [5]. *Decision tree* dapat dijelaskan sebagai metode yang menggunakan struktur

berbentuk pohon untuk membuat keputusan, dengan setiap simpul internal merepresentasikan "percobaan" terhadap atribut, setiap cabang menunjukkan hasil dari percobaan tersebut, dan setiap daun merepresentasikan kelas atau nilai target (hasil keputusan). Terdapat tiga jenis simpul pada *decision tree* [6]:

- *Root node*: Simpul yang tidak memiliki input tetapi memiliki lebih dari satu output

- *Internal node*: Simpul yang memiliki satu input dan lebih dari dua output.

- *Leaf* atau *terminal node*: Simpul yang memiliki satu input dan tidak memiliki output.

Langkah pertama dalam membentuk metode *decision tree* adalah menentukan variabel apa yang akan digunakan untuk percabangan. Idealnya, setelah percabangan terbentuk distribusi kelas yang homogen.

### 3. Naive Bayes

*Naive bayes classifier* adalah sekumpulan algoritma yang kuat dan efektif untuk berbagai pengaplikasian berbasis probabilitas yang digunakan untuk klasifikasi. *Naive bayes* bekerja dengan memprediksi probabilitas bahwa suatu sampel data termasuk pada kelas tertentu [7]. Algoritma klasifikasi *Naive Bayes* merupakan metode klasifikasi yang sederhana dan didasarkan pada teori probabilitas yang dikenal dengan *Teorema Bayes*.

Terdapat beberapa keuntungan dari metode algoritma Naive Bayes, yaitu sebagai teknik klasifikasi yang sangat intuitif dan mudah digunakan. Algoritma ini tidak memerlukan parameter bebas yang kompleks seperti yang ditemukan dalam jaringan saraf, sehingga menyederhanakan proses desain. Selain itu, Naive Bayes mengembalikan probabilitas, yang membuat hasilnya mudah diterapkan pada berbagai macam tugas meskipun skala data berubah-ubah. Keuntungan lain adalah tidak memerlukan jumlah data yang besar sebelum pembelajaran dapat dimulai, dan secara komputasi cepat dalam mengambil keputusan [8].

### III. PEMROSESAN DATA

#### A. Deskripsi data

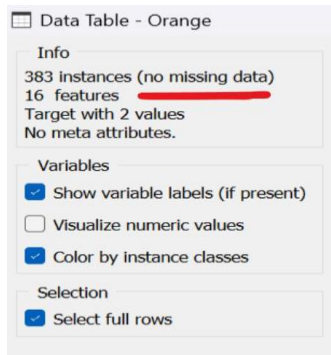
Data yang akan digunakan dalam pemrosesan ini dipilih dari salah satu *public repository* data, yaitu Kaggle. Dataset berisi mengenai *Thyroid Cancer Recurrence Prediction* yang didalamnya terdapat 383 data dengan 17 atribut (16 atribut dan 1 atribut target). Data merupakan data klasifikasi dengan 16 atribut bertipe kategorikal dan 1 atribut bertipe numerik.

Beberapa atribut kategorikal antara lain, *Gender, Smoking, Hx Smoking, Hx Radiotherapy, Thyroid Function, Physical Examination, Adenopathy, Focality, Risk, T, M, N, Stage, Response, dan Recurred*. Sedangkan, untuk atribut numerik adalah *Age*.

#### B. Data cleaning

##### 1. Imputation

Pada dataset ini, tidak dilakukan proses *imputation* karena tidak terdapat *missing data*. Seperti yang terlihat pada gambar 1 di bawah ini:

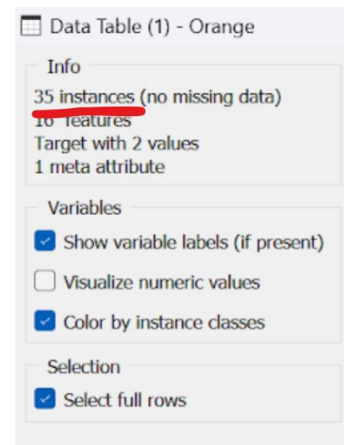


Gambar 1. *Missing Data*

Gambar 1 Menunjukkan bahwa pada data yang digunakan tidak terdapat *missing data*.

##### 2. Deteksi *Outliers*

Deteksi *outlier* menjadi taha pertama yang dilakukan untuk pemrosesan data ini. Hasil dari deteksi *outlier* yang telah dilakukan dapat dilihat pada gambar 2 berikut ini:

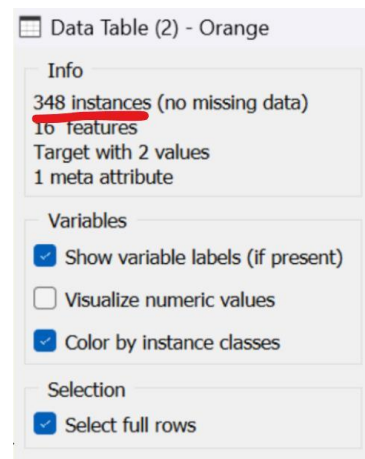


Gambar 2. *Outlier Data*

Gambar 2 menunjukkan terdapat 35 data dengan 16 atribut dengan 1 target dan 1 atribut meta setelah dideteksi *outliers*.

##### 3. Data *Inliers*

Dalam pemrosesan data, hanya yang bukan *outlier* saja yang akan dipakai atau dengan kata lain data *inliers*. Pada gambar 3 berikut ini adalah *inliers* yang didapatkan dari dataset yang telah diproses:

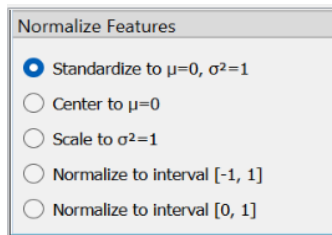


Gambar 3. Data *inliers* dengan 348 data dengan 16 atribut, 1 target, dan 1 atribut meta

#### C. *Data transformation*

##### 1. Normalisasi

Pada data ini dilakukan proses normalisasi untuk mengubah atribut *Age* yang bertipe numerik menjadi lebih sederhana. Metode untuk normalisasi ini dapat dilihat pada gambar 4 berikut.



Gambar 4. Metode normalisasi data.

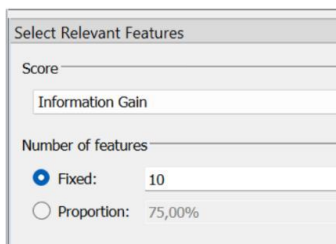
Age	Pathology	M
-0.976	icropapillary	M0
-0.487	icropapillary	M0
-0.767	icropapillary	M0
1.470	icropapillary	M0
1.470	icropapillary	M0
0.771	icropapillary	M0
0.002	icropapillary	M0
0.352	icropapillary	M0
0.701	icropapillary	M0
-0.067	icropapillary	M0
2.379	icropapillary	M0
1.261	icropapillary	M0
0.562	icropapillary	M0
0.632	icropapillary	M0
2.449	icropapillary	M0
0.072	icropapillary	M0
-0.067	icropapillary	M0

Gambar 5. Hasil normalisasi pada atribut Age.

Pada Gambar 5 di atas, dapat dilihat bahwa nilai atribut age telah berubah menjadi angka yang lebih kecil.

#### D. Data reduction

Proses untuk mereduksi data menggunakan *Select Relevant Feature* yang terdapat pada aplikasi *Orange*, dapat dilihat pada Gambar 6.



Gambar 6. *Select Relevant Feature* dengan *Information Gain* dengan memilih 10 atribut

Rank	Attribute	#	Info. gain
2	Risk	3	0.571
3	Adenopathy	6	0.257
4	T	7	0.256
5	N	3	0.238
6	Stage	5	0.155
7	Focality	2	0.102
8	Age	4	0.077
9	Pathology	4	0.066
10	M	2	0.060
11	Gender	2	0.056
12	Smoking	2	0.046
13	Physical Examination	5	0.036
14	Hx Radiotherapy	2	0.014
15	Hx Smoking	2	0.013
16	Thyroid Function	5	0.013

Gambar 7. Rank atribut berpengaruh

Gambar 7 adalah tampilan hasil *Rank* untuk menentukan atribut berpengaruh. Pada gambar 7, didapatkan beberapa atribut berpengaruh yang dipilih untuk direduksi berdasarkan nilai *information gain*.

#### E. Hasil Preprocessing

Dari hasil preprocessing setelah dilakukan proses *data cleaning*, *data transformation*, dan *data reduction*, didapatkan sejumlah 348 data dan 10 atribut serta satu atribut target. Dengan atribut yang tersisa diantaranya adalah *Recurred*, *Response*, *Risk*, *Adenopathy*, *T*, *N*, *Stage*, *Focality*, *Age*, *Pathology*, dan *M*.

### IV. TAHAPAN PEMODELAN

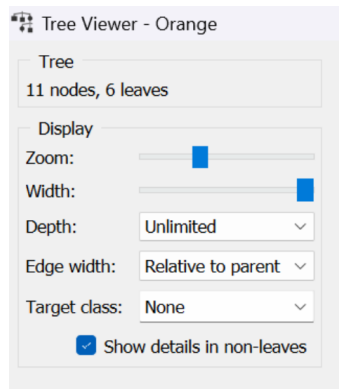
#### A. Deskripsi Data

Data yang digunakan merupakan data yang telah melalui tahap preprocessing, yaitu mengenai *Thyroid Cancer Recurrence Prediction*. Data berjumlah 348 dengan 10 atribut dan 1 target atribut. Data didapatkan dari *data repository* publik, yaitu *Kaggle*.

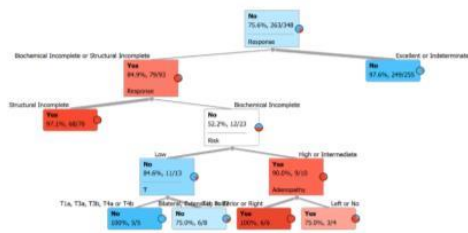
#### B. Pemodelan Data

##### 1. Decision Tree

Pemodelan selanjutnya adalah dengan menggunakan metode *decision tree*, seperti yang terlihat pada gambar 8.



Gambar 8. Pengaturan yang digunakan untuk *decision tree*



Gambar 9. Bentuk *Decision Tree* yang dihasilkan.

Berdasarkan gambar 9 di atas, didapatkan *decision tree* 6 level dengan 11 nodes dan 6 leaves yang jika dikonversikan menjadi aturan, maka didapatkan hasil sebagai berikut:

1. Jika *Response = Excellent or Intermediate*, maka *No*
2. Jika *Response = Biochemical Incomplete or Structural Incomplete, Response = Structural Incomplete*, maka *Yes*
3. Jika *Response = Biochemical Incomplete or Structural Incomplete, Response = Biochemical Incomplete, Risk = Low, T= T1a, T3a, T3b, T4a, or T4b*, maka *No*
4. Jika *Response = Biochemical Incomplete or Structural Incomplete, Response = Biochemical Incomplete, Risk = Low, T = T1b, T2*, maka *No*
5. Jika *Response = Biochemical Incomplete or Structural Incomplete, Response = Biochemical Incomplete, Risk = High or Intermediate*,

*Adenopathy = Bilateral, Extensive, Posterior, or Right*, maka *Yes*

6. Jika *Response = Biochemical Incomplete or Structural Incomplete, Response = Biochemical Incomplete, Risk = High or Intermediate, Adenopathy = Left or No*, maka *Yes*
7. Jika *Response = Biochemical Incomplete or Structural Incomplete, Response = Biochemical Incomplete, Risk = Low*, maka *No*
8. Jika *Response = Biochemical Incomplete or Structural Incomplete, Response = Biochemical Incomplete, Risk = High or Intermediate*, maka *Yes*
9. Jika *Response = Biochemical Incomplete or Structural Incomplete, Response = Biochemical Incomplete*, maka *No*
10. Jika *Response = Biochemical Incomplete or Structural Incomplete*, maka *Yes*

#### C. Hasil Pemodelan Data

##### 1. Cross Validation

Dilakukan uji *test and score* dengan metode *cross validation* menggunakan *fold* sebanyak 10 dengan bentuk *stratified*, pengujian *test and score* tersebut dapat dilihat pada gambar 10.

Evaluation results for target (None, show average over classes)						
Model	AUC	CA	F1	Prec	Recall	MCC
Naive Bayes	0.976	0.914	0.915	0.917	0.914	0.775
Tree	0.941	0.963	0.962	0.963	0.963	0.897

Gambar 10. CA dari hasil uji *test and score*.

Dari hasil uji pada gambar 10 di atas, menunjukkan besar CA (*Classification Accuracy*) berdasarkan evaluasi model:

- a. *Naive Bayes* dengan CA sebesar 0.914
- b. *Tree* dengan CA sebesar 0.963

Compare models by: Area under ROC curve		
	Naive Bayes	Tree
Naive Bayes		0.953
Tree	0.047	

Gambar 11. Hasil perbandingan model dengan AUC.

Dari hasil uji pada Gambar 11 di atas, menunjukkan hasil perbandingan model sebagai berikut:

- a. *Naive Bayes* memiliki performa terbaik secara keseluruhan karena

memiliki nilai AUC tertinggi dibandingkan dengan *Tree* (0,953).

- b. *Decision Tree* memiliki performa lebih rendah dibandingkan dengan *Naive Bayes* karena memiliki nilai AUC yang paling rendah dibandingkan keduanya (0,047)

## 2. Confusion Matrix Decision Tree

Di bawah ini merupakan hasil dari confusion matriks untuk decision tree, dapat dilihat pada gambar 12.

		Predicted		Σ
		No	Yes	
Actual	No	260	3	263
	Yes	10	75	85
Σ		270	78	348

Gambar 12. Confusion Matrix dari Decision Tree

TP: 260  
 FP: 10  
 TN: 75  
 FN: 3

Menghitung *Precision*, *Recall*, *F1 Score*, *Accuracy*, dan *AUC*: *Precision*:

$$\frac{TP}{(TP + FP)} = \frac{260}{(260 + 10)}$$

$$= \frac{260}{270} = 0,962 = 96,2\%$$

*Recall:*

$$\frac{TP}{(TP + FN)} = \frac{260}{(260 + 3)}$$

$$= \frac{260}{263} = 0,988 = 98,8\%$$

*F1:*

$$\frac{2 \times (Precision \times Recall)}{(Precision + Recall)}$$

$$= \frac{2 \times (0,962 \times 0,988)}{(0,962 + 0,988)} = \frac{1,9}{1,95} = 0,974$$

$$= 97,4\%$$

*Accuracy:*

$$\frac{TP+TN}{(TP+TN+FP+FN)} = \frac{260+75}{(260+75+10+3)}$$

$$= \frac{335}{348} = 0,962 = 96,2\%$$

*AUC:*

- *Specificity:*

$$\frac{TN}{(TN + FP)} = \frac{75}{(75 + 10)} = \frac{75}{85}$$

$$= 0,882 = 88,2\%$$

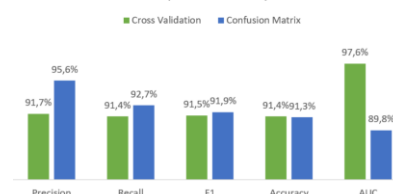
- *AUC:*

$$\frac{Recall + Specificity}{2} = \frac{0,988 + 0,882}{2}$$

$$= \frac{1,87}{2} = 0,935 = 93,5\%$$

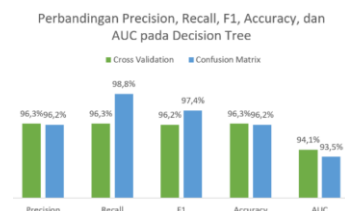
## D. Perbandingan

Perbandingan Precision, Recall, F1, Accuracy, dan AUC pada Naive Bayes



Gambar 13. Grafik perbandingan Cross Validation dan Confusion Matrix pada Naive Bayes.

Berdasarkan perbandingan yang dapat dilihat pada gambar 13 di atas, terlihat bahwa *Accuracy* yang dihitung menggunakan *cross validation* dan *confusion matrix* memiliki nilai paling dekat, yaitu 91,4% dan 91,3% (0,914 dan 0,913) dengan perbedaan hanya sebesar 0,001 atau 0,1%. Selain itu nilai F1 juga terlihat hampir mendekati satu sama lain dengan nilai 91,5% dan 91,9% (0,915 dan 0,919) dengan perbedaan nilai sebesar 0,004 atau 0,4%.



Gambar 14. Grafik perbandingan *Cross Validation* dan *Confusion Matrix* pada *Decision Tree*.

Berdasarkan perbandingan yang dapat dilihat pada gambar 14 di atas, terlihat bahwa *Precision* dan *Accuracy* yang dihitung menggunakan *cross validation* dan *confusion matrix* memiliki nilai paling dekat, yaitu untuk *precision* sebesar 96,3% dan 96,2% (0,963 dan 0,962), sedangkan untuk *accuracy* sebesar 96,3% dan 96,2% (0,963 dan 0,962) dengan perbedaan keduanya sama-sama sebesar 0,001 atau 0,1%.

## V. PEMBAHASAN

### 1. Akurasi dari Metode Klasifikasi

Dalam penelitian ini, digunakan dua metode klasifikasi utama, yaitu *Naive Bayes* dan *Decision Tree*. Berikut adalah hasil akurasi dari setiap metode setelah dilakukan pemodelan pada dataset *Thyroid Cancer Recurrence Prediction*.

- *Naive Bayes*

Metode *Naive Bayes* memberikan akurasi sebesar 0.913. Meskipun lebih sederhana dan cepat dalam komputasi, metode ini berasumsi bahwa fitur-fitur bersifat independen, yang mungkin tidak selalu berlaku dalam data yang kompleks.

- *Decision Tree*

Model *Decision Tree* menunjukkan akurasi sebesar 0.962. *Decision Tree* memiliki keunggulan dalam kemampuan sistem untuk bertukar data secara aman dan otomatis tanpa memandang batas-batas tertentu, karena struktur pohon keputusan mudah dipahami dan diikuti.

### 2. Perbandingan AUC Pemodelan

*Naive Bayes*:

- Dibandingkan dengan CN2 Rule Induction: 0.559
- Dibandingkan dengan *Decision Tree*: 0.953

*Decision Tree*:

- Dibandingkan dengan *Naive Bayes*: 0.047
- Dibandingkan dengan CN2 Rule Introduction: 0.036

Dari keterangan di atas, *Naive Bayes* yang memiliki AUC merupakan klasifikasi yang lebih baik secara keseluruhan dibandingkan dengan *Decision Tree*.

### 3. Percobaan *Cross-Validation*

Untuk memastikan hasil akurasi tidak dipengaruhi oleh overfitting atau pemilihan dataset yang tidak representatif, dilakukan k-Fold *Cross-Validation* dengan beberapa nilai k yang berbeda.

- k-Fold = 5:

- *Naive Bayes*: Akurasi klasifikasi = 0.914
- *Decision Tree*: Akurasi klasifikasi = 0.960

- k-Fold = 10:

- *Naive Bayes*: Akurasi klasifikasi = 0.914
- *Decision Tree*: Akurasi klasifikasi = 0.963

- k-Fold = 20:

- *Naive Bayes*: Akurasi klasifikasi = 0.917
- *Decision Tree*: Akurasi klasifikasi = 0.957

Dari hasil *Cross-Validation*, terlihat bahwa nilai k-Fold yang lebih besar memberikan hasil akurasi yang lebih stabil. k-Fold = 10 dengan nilai

akurasi sebesar 0.963 memberikan hasil optimal dengan akurasi klasifikasi tertinggi untuk ketiga metode.

Berdasarkan hasil pemodelan didapatkan aturan dari metode decision tree yaitu:

1. Respon Pengobatan Sangat Baik atau Sedang (*Excellent or Intermediate*):
  - Jika respon pengobatan adalah sangat baik atau sedang, maka penyakit tidak akan kambuh.
2. Respon Pengobatan Biokimia Tidak Lengkap atau Struktural Tidak Lengkap (*Biochemical Incomplete or Structural Incomplete*):
  - Jika respon adalah struktural tidak lengkap, maka penyakit akan kambuh.
  - Jika respon adalah biokimia tidak lengkap:
    - Jika risiko rendah dan tingkat tumor adalah T1a, T3a, T3b, T4a, atau T4b, maka penyakit tidak akan kambuh.
    - Jika risiko rendah dan tingkat tumor adalah T1b atau T2, maka penyakit tidak akan kambuh.
    - Jika risiko tinggi atau sedang dan adenopati bilateral, ekstensif, posterior, atau kanan, maka penyakit akan kambuh.
    - Jika risiko tinggi atau sedang dan adenopati di kiri atau tidak ada, maka penyakit akan kambuh.
    - Jika risiko rendah, maka penyakit tidak akan kambuh.
    - Jika risiko tinggi atau sedang, maka penyakit akan kambuh.
    - Jika hanya dinyatakan biokimia tidak lengkap tanpa kondisi tambahan, maka penyakit tidak akan kambuh.

## VI. KESIMPULAN

Telah dilakukan pemodelan menggunakan metode *Naive Bayes* dan *Decision Tree* pada Dataset *Thyroid Cancer Recurrence Prediction* yang terdapat 348 data dengan 11 atribut (10 atribut dan 1 atribut target).

Hasil dari percobaan ini menunjukkan bahwa kedua metode tersebut memiliki kinerja yang berbeda dalam memprediksi kekambuhan kanker tiroid. Evaluasi menunjukkan bahwa

*Decision Tree* sedikit lebih unggul dalam akurasi, sementara *Naive Bayes* lebih cepat dan konsisten. Hal tersebut dapat dilihat melalui hasil pemodelan dari metode decision tree menghasilkan dua aturan yaitu respon pengobatan sangat baik atau sedang (*Excellent or Intermediate*) dan respon pengobatan biokimia tidak lengkap atau struktural tidak lengkap (*Biochemical Incomplete or Structural Incomplete*).

## DAFTAR PUSTAKA

- [1] A. A. Dewayanti, "Penerapan Data Mining Pada Data Nilai Siswa Dengan Menggunakan Algoritma Association Rule Metode Apriori (Studi Kasus di SMP N 36 Semarang)," dalam *Prosiding Seminar Nasional Matematika Dan Pendidikan Matematika*, 2018.
- [2] K. Kartarina, N. K. Sriwinarti, dan N. luh P. Juniarti, "Analisis Metode K-Nearest Neighbors (K-NN) Dan Naive Bayes Dalam Memprediksi Kelulusan Mahasiswa," *JTIM : Jurnal Teknologi Informasi dan Multimedia*, vol. 3, no. 2, 2021, doi: 10.35746/jtim.v3i2.159.
- [3] D. Septhya dkk., "Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 1, 2023, doi: 10.57152/malcom.v3i1.591.
- [4] P. M. S. Tarigan, J. T. Hardinata, H. Qurniawan, M. Safii, dan R. Winanjaya, "Implementasi Data Mining Menggunakan Algoritma Apriori Dalam Menentukan Persediaan Barang," *Jurnal Janitra Informatika dan Sistem Informasi*, vol. 2, no. 1, hlm. 9–19, Apr 2022, doi: 10.25008/janitra.v2i1.142.
- [5] C. A. Sari, A. Sukmawati, R. P. Aprilli, P. S. Kayaningtias, dan N. Yudistira, "PERBANDINGAN METODE NAÏVE BAYES, SUPPORT VECTOR MACHINE DAN DECISION TREE DALAM KLASIFIKASI KONSUMSI OBAT," *Jurnal Litbang Edusaintech*, vol. 3, no. 1, 2022.
- [6] D. E. Sinaga, A. P. Windarto, dan R. A. Nasution, "Analisis Data Mining Algoritma Decision Tree Pada Prediksi Persediaan Obat



(Studi Kasus : Apotek Franch Farma),” *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 2, no. 4, 2022, doi: 10.30865/klik.v2i4.328.

[7] A. Nugroho dan Y. Religia, “Analisis Optimasi Algoritma Klasifikasi Naive Bayes menggunakan Genetic Algorithm dan Bagging,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 3, 2021, doi: 10.29207/resti.v5i3.3067.

[8] I. Paryudi, E. Winarko, S. Priyanta, dan S. R. C. Nursari, “Modeling of Personality Traits based on Demographic Data on Multi-Races Samples of Ages from 13 to 50 Years Old: Investigating the Effect of Race on Model,” dalam *Proceedings - 2020 6th International Conference on Science and Technology, ICST 2020*, Institute of Electrical and Electronics Engineers Inc., 2020. doi: 10.1109/ICST50505.2020.9732792.