



Perbandingan Performa Algoritma C4.5 dan Naive Bayes untuk Prediksi Kanker Serviks pada Wanita

Fathurrahman Nur Aziz¹

¹*Jurusan Informatika, Universitas Amikom Yogyakarta,
Jl. Ring Road Utara, Condongcatur, Kecamatan Depok, Kabupaten Sleman, Daerah Istimewa
Yogyakarta 55281, Indonesia*

*Email Penulis Koresponden: azizfath35@gmail.com

Abstrak:

Adanya peningkatan jumlah kasus kanker serviks di Indonesia berdasarkan data pada tahun 2020 menunjukkan perlunya upaya untuk menekan kenaikan melalui berbagai upaya pencegahan primer dan sekunder. Upaya primer yang dapat dilakukan di antaranya adalah seperti menanamkan pola hidup sehat serta melakukan vaksinasi HPV. Langkah ini tentunya perlu didukung dengan upaya pencegahan sekunder, yakni dengan melakukan skrining atau deteksi dini guna memastikan kesehatan leher rahim penduduk wanita Indonesia, sehingga pengembangan teknologi skrining perlu terus dilakukan demi menghasilkan teknologi skrining yang semutakhir mungkin. Pada penelitian ini, penulis berupaya membandingkan performa algoritma C4.5 dan Naive Bayes dari segi akurasi dan presisinya. Hasil penelitian dan pengujian algoritma C4.5 dan Naive Bayes akan dibandingkan kemudian dipilih yang terbaik untuk memprediksi kemungkinan terkena kanker serviks berdasarkan parameter diagnosis kanker serviks menunjukkan prediksi hasil yang akurat dengan kesalahan minimal. Pada akhir penelitian C4.5 Decision Tree mendapatkan hasil yang lebih unggul ketika kita melakukan 10-Fold Cross Validation sehingga algoritma C4.5 lebih tepat digunakan pada kasus prediksi kanker serviks pada wanita.

Kata Kunci:

*Klasifikasi,
Naive Bayes,
C4.5,
Akurasi*

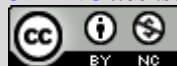
Riwayat Artikel:

Diserahkan 12 Juni, 2023
Direvisi 28 Jan, 2024
Diterima 19 April, 2024

DOI:

10.22441/incomtech.v14i1.20899

This is an open access article under the [CC BY-NC](https://creativecommons.org/licenses/by-nc/4.0/) license



1. PENDAHULUAN

Kanker merupakan sebuah tumor ganas yang disebabkan oleh pertumbuhan abnormal pada jaringan epitel sebagai akibat dari sebuah mutasi DNA[1]. Kanker serviks terjadi pada jaringan epitel leher rahim yang disebabkan oleh *Human Pappiloma Virus (HPV)*[2]. Berdasarkan data yang dipublikasikan oleh WHO 2021, terdapat 27 kasus kanker serviks per 100.000 wanita di Indonesia pada tahun 2020 dengan rata-rata umur wanita dari keseluruhan kasus kanker di Indonesia adalah 24,4 tahun[3]. Dari hal tersebut, dapat dilihat bahwa kanker serviks banyak menyerang wanita usia produktif di Indonesia.

Upaya pencegahan kanker serviks dapat dilakukan dengan meningkatkan kesadaran dalam mengenali risiko dan gejala kanker serviks. Salah satu faktor yang mendukung tindakan deteksi dini adalah akses informasi[4]. Kurangnya akses informasi dapat mengakibatkan risiko kanker serviks yang lebih tinggi. Tidak sedikit kasus kanker serviks yang pada saat diagnosis awal ternyata sudah dalam stadium lanjut yang meningkatkan risiko kematian[5].

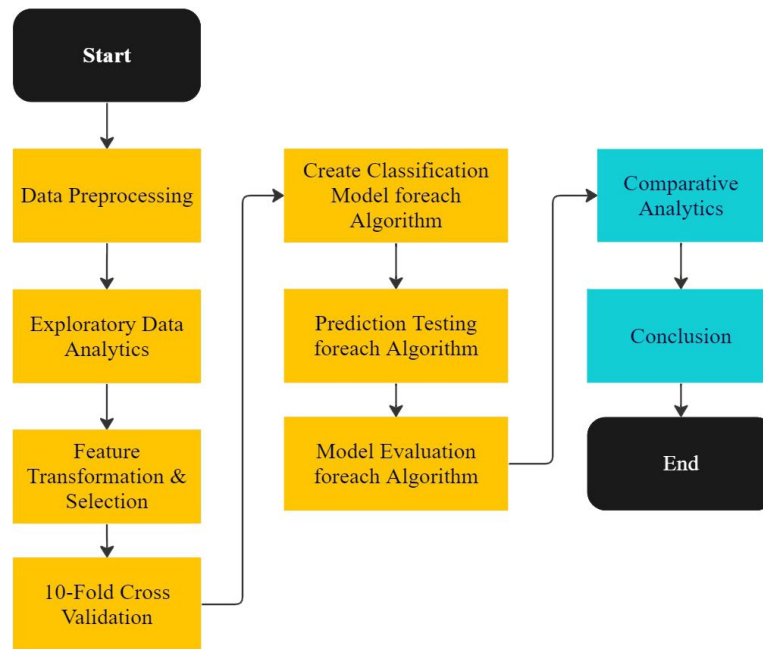
Fakta tersebut kemudian mendorong penulis untuk mendukung pengembangan instrumen deteksi dan prediksi kanker serviks dengan mencoba mempercepat proses dan meningkatkan akurasi diagnosis kanker serviks dengan teknik *data mining*, tepatnya dengan mempercepat klasifikasi data dalam proses prediksi tersebut. Adapun dalam penelitian ini, dilakukan klasifikasi dengan dua algoritma yaitu C4.5 dan Gaussian Naïve Bayes. Penelitian sebelumnya mengenai prediksi kanker serviks dilakukan dengan cara klasifikasi menggunakan Naïve Bayes dan menghasilkan akurasi sebesar 87%[6]. Penelitian lain mengenai diagnosis kanker serviks tahap awal berdasarkan survei dan dataset dari UCI Machine Learning menggunakan algoritma C4.5 menghasilkan akurasi sebesar 98,61%[7]. Perbedaan yang signifikan terjadi karena dataset yang digunakan pada penelitian kedua hanya memiliki jumlah data sebanyak 72. Perbedaan tersebut mendorong penulis untuk melakukan perbandingan performa algoritma Naïve Bayes dan C4.5 untuk dataset yang setara. Penelitian sebelumnya terkait perbandingan performa kedua algoritma untuk studi kasus lulusan tepat waktu mahasiswa menghasilkan akurasi sebesar 69,54% untuk C4.5 dan 68,38% untuk Naïve Bayes dengan selisih akurasi sebesar 1,16%[8]. Penelitian lain terkait Prediksi Ketepatan Waktu Studi Mahasiswa menghasilkan akurasi dengan C4.5 sebesar 61,99% dan untuk Naïve Bayes sebesar 69,97%[9]. Perbedaan konsistensi akurasi dari dua kasus yang berbeda menghasilkan sebuah perdebatan terkait mana yang lebih baik untuk melakukan klasifikasi. Oleh karena itu, penulis melakukan penelitian yang serupa dengan kasus yang berbeda yaitu prediksi kanker serviks pada Wanita. Hasil yang lebih baik dari segi presisi dan akurasi akan digunakan sebagai algoritma yang digunakan dalam melakukan prediksi kanker serviks pada wanita dan dapat digunakan sebagai metode deteksi dini kanker serviks untuk menurunkan risiko kematian akibat kanker serviks.

2. METODE

Penelitian ini menerapkan algoritma *machine learning* yaitu klasifikasi prediksi kanker serviks dengan algoritma Naïve Bayes dan C4.5. Dataset yang digunakan diambil dari Kaggle[10].

2.1. Alur Penelitian

Gambar 1 merupakan diagram alir pada penelitian yang dilakukan.



Gambar 1. Diagram alir penelitian

Penelitian diawali dengan melakukan *preprocessing* dataset sehingga dapat meminimalisasi kesalahan dalam proses klasifikasi. Selanjutnya dilakukan *Exploratory Data Analytics* untuk melihat isi dan fitur dataset agar nantinya dapat ditentukan fitur yang menjadi *dependent variable* dan *independent variable*. Kemudian, data *training* dan *testing* dibagi menggunakan metode *10-Fold Cross Validation* yang selanjutnya dijadikan model klasifikasi pada masing-masing algoritma. Setelah model ditentukan, klasifikasi dilakukan beserta evaluasi modelnya untuk masing-masing algoritma. Dari hasil evaluasi model, dilakukan perbandingan sehingga dapat diambil kesimpulan terkait algoritma yang terbaik untuk kasus prediksi kanker serviks pada wanita.

2.2. Tinjauan Pustaka

a. Naïve Bayes

Naïve Bayes adalah salah satu cara melakukan klasifikasi sederhana dengan menghitung kemungkinan dan menggabungkan suatu kombinasi dan frekuensi sebuah nilai[11] Naïve Bayes memperkirakan seluruh atribut yang bersifat independen yaitu bebas dan saling lepas. Atribut yang independen berarti tidak ada keterkaitan antara nilai suatu atribut dengan atribut yang lain[12]. Pada penelitian ini, perhitungan probabilitas dari seluruh fitur yang diambil akan menghasilkan klasifikasi hasil diagnosis kanker serviks dalam kategori positif dan negatif.

Rumus dasar Gaussian Naïve Bayes adalah sebagai berikut:

$$p(x_i|y_j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{(x_i-\mu_j)^2}{2\sigma_j^2}}$$

Dimana,

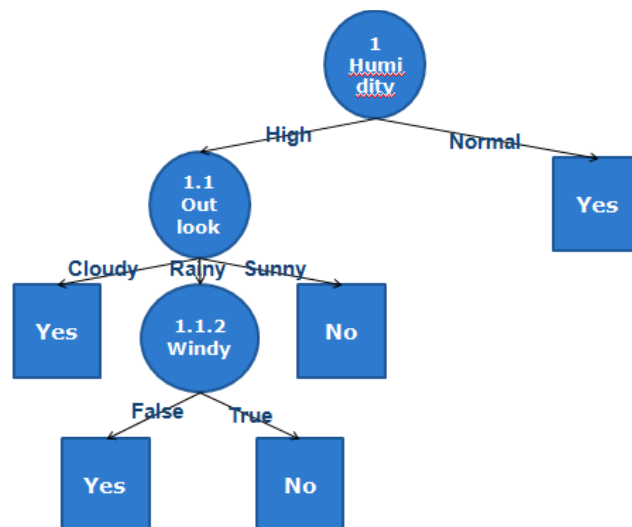
- P = peluang dari data yang dihitung
- $\pi = 3.142 \dots$
- $\sigma =$ standar deviasi
- $\mu =$ rata-rata data dari kelas yang sama

b. C4.5

C4.5 adalah salah satu algoritma yang digunakan untuk klasifikasi dalam *machine learning*. Algoritma C4.5 berdasar pada algoritma yang lain yaitu pohon keputusan atau *decision tree* [13]. Kelebihan dari algoritma C4.5 adalah dapat melakukan klasifikasi pada data yang memiliki *missing value*. Cara kerja algoritma C4.5 adalah sebagai berikut [14]:

- Memilih akar atau *root* berdasarkan nilai *gain ratio* tertinggi.
- Menghitung ulang *gain ratio* tanpa atribut root pada perhitungan sebelumnya.
- Memilih cabang atau *branches* berdasarkan *gain ratio* tertinggi pada perhitungan terakhir.
- Mengulangi langkah kedua hingga mendapatkan nilai *entropy* = 0.

Gambar 2 merupakan salah satu penerapan Algoritma C4.5 dalam menentukan cuaca.



Gambar 2. Contoh penerapan Algoritma C4.5

c. *K-Fold Cross Validation*

K-Fold Cross Validation merupakan teknik validasi data dengan cara membagi data ke dalam *k* bagian dengan masing-masing bagian akan dilakukan proses *training* dan *testing*. *Cross Validation* memungkinkan percobaan *training* dan *testing* sebanyak *k*. Umumnya, *K-Fold Cross Validation* dilakukan sebanyak 10 kali untuk mendapatkan performa yang terbaik dari klasifikasi yang dilakukan. [15]

Gambar 3 merupakan cara kerja *10-Fold Cross Validation*.

K-1	Test	Train	Train	Train	Train	Train	Train	Train	Train	Train
K-2	Train	Test	Train	Train	Train	Train	Train	Train	Train	Train
K-3	Train	Train	Test	Train	Train	Train	Train	Train	Train	Train
K-4	Train	Train	Train	Test	Train	Train	Train	Train	Train	Train
K-5	Train	Train	Train	Train	Test	Train	Train	Train	Train	Train
K-6	Train	Train	Train	Train	Train	Test	Train	Train	Train	Train
K-7	Train	Train	Train	Train	Train	Train	Test	Train	Train	Train
K-8	Train	Train	Train	Train	Train	Train	Train	Test	Train	Train
K-9	Train	Train	Train	Train	Train	Train	Train	Train	Test	Train
K-10	Train	Train	Train	Train	Train	Train	Train	Train	Train	Test

Gambar 3. *10-Fold Cross Validation*

d. *Model Evaluation*

Salah satu metode untuk evaluasi model adalah *confusion matrix*. *Confusion matrix* merupakan tabel matrix yang memiliki komponen berupa kelas yang sesungguhnya dan kelas yang diprediksi. Dari dua komponen tersebut maka didapatkan 4 nilai yaitu [15]:

- *True positive* = kondisi dimana kelas sesungguhnya *True* dan diprediksi *True*
- *True negative* = kondisi dimana kelas sesungguhnya *False* dan diprediksi *False*
- *False positive* = kondisi dimana kelas sesungguhnya *False* dan diprediksi *True*
- *False negative* = kondisi dimana kelas sesungguhnya *True* dan diprediksi *False*

Confusion Matrix

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

Gambar 4. *confusion matrix*.

2.3. Data set

Data set yang digunakan diambil dari Kaggle [10] dengan jumlah kasus sebanyak 858 baris dan 31 fitur. Beberapa data tersebut masih berupa data *unknown* yang diwakili dengan simbol “?” yang dapat mengacaukan proses klasifikasi. Untuk mengatasinya, penulis mengubah data tersebut menjadi *null values*. Terdapat pula dua fitur yang memiliki terlalu banyak *null values*, yaitu “*STDs: Time since first diagnosis*” dan “*STDs: Time since last diagnosis*” sehingga perlu di-drop.

Berikut adalah deskripsi statistik dari data set yang digunakan.

	Age	Number of sexual partners	First sexual intercourse	Num of pregnancies	Smokes	Smokes (years)	Smokes (packs/year)	Hormonal Contraceptives	Hormonal Contraceptives (years)	IUD	...	STDs:HPV	STDs: Number of diagnosis	Dx:Cancer
count	668.000000	668.000000	668.000000	668.000000	668.000000	668.000000	668.000000	668.000000	668.000000	668.000000	...	668.000000	668.000000	668.000000
mean	27.264970	2.523952	17.142216	2.323353	0.143713	1.235524	0.458953	0.643713	2.290037	0.112275	...	0.002994	0.092814	0.025449
std	8.727432	1.640299	2.852046	1.465319	0.351061	4.193611	2.336308	0.479260	3.724400	0.315942	...	0.054677	0.310355	0.157603
min	13.000000	1.000000	10.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000
25%	21.000000	2.000000	15.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000
50%	26.000000	2.000000	17.000000	2.000000	0.000000	0.000000	0.000000	1.000000	0.500000	0.000000	...	0.000000	0.000000	0.000000
75%	33.000000	3.000000	18.000000	3.000000	0.000000	0.000000	0.000000	1.000000	3.000000	0.000000	...	0.000000	0.000000	0.000000
max	84.000000	28.000000	32.000000	11.000000	1.000000	37.000000	37.000000	1.000000	22.000000	1.000000	...	1.000000	3.000000	1.000000

Gambar 5. Deskripsi Statistik *Cervical Cancer Risk Classification*

Dari deskripsi statistik tersebut, terdapat beberapa poin yang dapat diambil, yakni:

- Nilai maksimum dari “*Age*” adalah 84, dimana nilai ini terlalu besar jika dibandingkan dengan nilai maksimum pada kolom lain. Sehingga perlu dilakukan standarisasi skala seluruh kolom agar tidak mengganggu proses klasifikasi.
- Nilai maksimal dari “*Num of pregnancies*” adalah 11, dimana rata-ratanya hanya 2. Maka nilai 11 tersebut termasuk outlier yang dapat mengganggu performa klasifikasi.
- Kolom “*STDs:cervical condylomatosis*” dan “*STDs:AIDS*” hanya berisi nilai 0 yang membuat fitur tersebut tidak memiliki arti sehingga perlu dihilangkan.

- Kolom "*Dx:CIN*", "*Dx:HPV*" tidak berhubungan dengan diagnosis kanker serviks dan kolom "*Dx*" hanya merupakan jumlah diagnosis sehingga tidak berkaitan dan perlu dihilangkan.

2.4. Perbandingan Algoritma

Setelah melalui tahapan *model evaluation*, nilai *precision*, *recall*, dan *accuracy* dari Algoritma Naïve Bayes dan C4.5 dilakukan rekapitulasi. Kemudian, hasil rekapan tersebut akan dibandingkan sehingga dapat diambil kesimpulan terkait algoritma terbaik untuk kasus prediksi kanker serviks pada wanita.

3. HASIL DAN PEMBAHASAN

Data set yang didapatkan belum memiliki label target tetapi sudah memiliki 4 fitur yang cukup kuat menjadi hasil akhir diagnosis yaitu pada fitur jenis tes dengan kolom "*Hinselmann*", "*Schiller*", "*Citology*", dan "*Biopsy*". Namun, fitur tersebut memiliki rentang nilai dari 0-4 sehingga membuat kelas terlalu banyak. Dari 858 baris yang terdapat di data set, didapatkan kelas yang belum disederhanakan sebagai berikut:

Tabel 1. Jumlah Data Pada Setiap Kelas yang Belum Disederhanakan

Kelas	Jumlah Data	Persentase
0	756	88,1 %
1	41	4,8 %
2	33	3,8 %
3	22	2,6 %
4	6	0,7 %
Total	858	100 %

Diagnosis kanker serviks hanya memiliki dua nilai, yaitu positif dan negatif. Oleh karena itu, penulis mengklasifikasikan keempat fitur itu menjadi negatif dan positif dengan parameter jika nilai 0 berarti negatif dan jika nilai lebih dari 1 berarti positif. Berikut ini adalah jumlah data pada kelas yang disederhanakan:

Tabel 2. Jumlah Data Pada Setiap Kelas yang Telah Disederhanakan

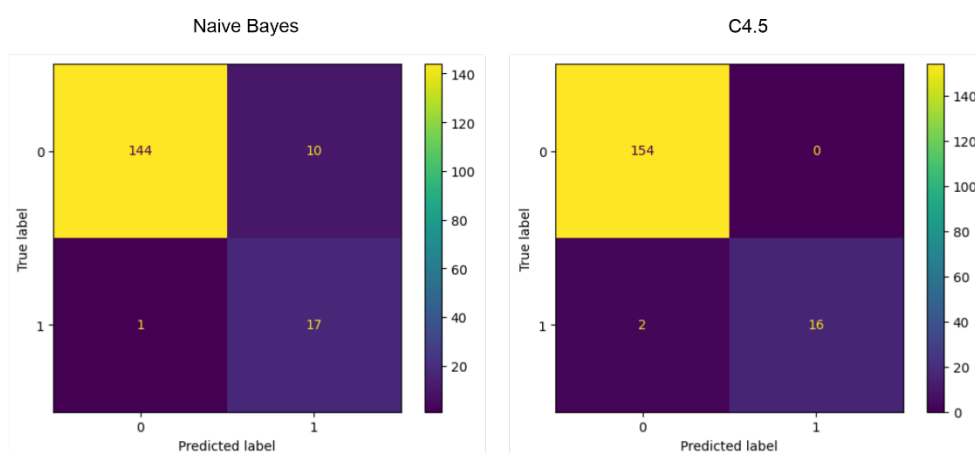
Kelas	Jumlah Data	Persentase
0	756	88,1 %
1	102	11,9 %
Total	858	100 %

Setelah penyederhanaan kelas selesai, dilanjutkan dengan melakukan *feature selection*. Fitur yang akan digunakan adalah fitur yang memiliki nilai selain ya dan tidak (0 atau 1) sehingga fitur-fitur seperti "*Smokes*", "*IUD*", dan "*STDs*" akan dibuang dan tersisa fitur-fitur yang lainnya. Total fitur yang digunakan adalah 25 fitur dengan rincian yaitu: "*Age*", "*Number of sexual partners*", "*First*

sexual intercourse", *"Num of pregnancies"*, *"Smokes (years)"*, *"Smokes (packs/year)"*, *"Hormonal Contraceptives (years)"*, *"IUD (years)"*, *"STDs:condylomatosis"*, *"STDs:cervical condylomatosis"*, *"STDs:vaginal condylomatosis"*, *"STDs:vulvo-perineal condylomatosis"*, *"STDs:syphilis"*, *"STDs:pelvic inflammatory disease"*, *"STDs:genital herpes"*, *"STDs:molluscum contagiosum"*, *"STDs:AIDS"*, *"STDs:HIV"*, *"STDs:Hepatitis B"*, *"STDs:HPV"*, *"Hinselmann"*, *"Schiller"*, *"Citology"*, *"Biopsy"*, dan *"result"*.

1. Percobaan pertama dengan metode *Splitting Data*:

Dari 858 baris, dilakukan *splitting data* untuk *data training* dan *data testing* dengan perbandingan 80%:20% yang berarti terdapat 686 *data training* dan 172 *data testing*. Percobaan ini menghasilkan *confusion matrix* seperti pada gambar 5:



Gambar 5. *Confusion Matrix* antara Naïve Bayes dan C4.5 Classifier

Model evaluasi dari percobaan pertama adalah sebagai berikut:

Tabel 3. Model Evaluasi dari Naïve Bayes dan C4.5 Classifier

Evaluation Item	Naïve Bayes	C4.5
Average Precision	81 %	99 %
Average Recall	0.94	0.94
Average F1-Score	0.86	0.97
Accuracy	94 %	99 %

Dari hasil model evaluasi pada tabel 3, nilai tertinggi dari keseluruhan poin didapatkan oleh Algoritma C4.5 dengan akurasi 99% dibandingkan dengan Algoritma Gaussian Naïve Bayes yang hanya mendapatkan akurasi 94%.

2. Percobaan kedua dengan *10-Fold Cross Validation*:

Percobaan kedua dilakukan menggunakan metode *K-Fold Cross Validation* untuk mendapatkan hasil yang lebih acak dan tidak berpola dibandingkan metode *splitting data*. Jumlah *fold* yang digunakan adalah 10.

Tabel 4. Model Evaluasi dengan 10-Fold Cross Validation

Fold	Evaluation Item	Naïve Bayes	C4.5
1	Precision	66 %	100 %
	Recall	0.94	1
	F1-Score	0.78	1
	Accuracy	93 %	100 %
2	Precision	67 %	100 %
	Recall	0.95	1
	F1-Score	0.78	1
	Accuracy	93 %	100 %
3	Precision	68 %	100 %
	Recall	0.94	1
	F1-Score	0.79	1
	Accuracy	94 %	100 %
4	Precision	54%	100 %
	Recall	0.95	1
	F1-Score	0.69	1
	Accuracy	90 %	100 %
5	Precision	73 %	100 %
	Recall	0.94	1
	F1-Score	0.82	1
	Accuracy	95 %	100 %
6	Precision	66 %	100 %
	Recall	0.94	1
	F1-Score	0.78	1
	Accuracy	93 %	100 %
7	Precision	65 %	100 %
	Recall	0.94	1
	F1-Score	0.77	1
	Accuracy	93 %	100 %
8	Precision	72 %	100 %
	Recall	0.96	1
	F1-Score	0.82	1
	Accuracy	95 %	100 %
9	Precision	66 %	100 %
	Recall	0.94	1
	F1-Score	0.78	1
	Accuracy	93 %	100 %
10	Precision	86 %	100 %
	Recall	0.96	1
	F1-Score	0.90	1
	Accuracy	98 %	100 %
	Average Precision	68 %	100 %
	Average Recall	0.95	1
	Average F1-Score	0.79	1
	Average Accuracy	94 %	100 %

Dari hasil model evaluasi pada tabel 4, nilai tertinggi dari rata-rata yang didapatkan dalam 10 kali operasi *cross validation* didapatkan oleh Algoritma C4.5 dengan akurasi 100% dibandingkan dengan Algoritma Gaussian Naïve Bayes yang hanya mendapatkan akurasi 94%.

Terdapat peningkatan akurasi pada algoritma C4.5 dengan menggunakan metode *10-Fold Cross Validation* menunjukkan bahwa metode ini dapat digunakan untuk memperbaiki akurasi dalam klasifikasi *data mining* karena menghasilkan *data training* dan *data testing* yang lebih bervariasi.

4. KESIMPULAN

Percobaan pertama dan kedua konsisten menghasilkan Algoritma C4.5 yang lebih unggul dari Algoritma Gaussian Naïve Bayes. Pada percobaan pertama, C4.5 mendapatkan akurasi 99% sedangkan Naïve Bayes mendapatkan akurasi 94%. Untuk percobaan kedua, C4.5 mendapatkan akurasi 100% sedangkan Naïve Bayes tetap mendapatkan akurasi 94%. Dari kedua percobaan tersebut, dapat disimpulkan bahwa Algoritma C4.5 lebih baik dan lebih unggul untuk digunakan sebagai prediksi kanker serviks pada wanita.

REFERENSI

- [1] Sp. B. dr. Nur Rahman, "Selayang Pandang Tentang Kanker," <https://rs.uns.ac.id/selayang-pandang-tentang-kanker/>, Feb. 04, 2021.
- [2] U. Khabibah, K. Adyani, and A. Rahmawati, "Faktor Risiko Kanker Serviks: Literature Review," *Faletehan Health Journal*, vol. 09, no. 3, pp. 270–277, 2022, [Online]. Available: www.journal.lppm-stikesfa.ac.id/ojs/index.php/FHJ
- [3] "Cervical cancer Indonesia 2021 country profile," *World Health Organization*, Accessed: Apr. 03, 2023. [Online]. Available: <https://www.who.int/publications/m/item/cervical-cancer-idn-country-profile-2021>
- [4] K. Adyani and F. Realita, "Factors that influence the participation among women in Inspection Visual Acetic acid (IVA) test," *Jurnal Aisyah : Jurnal Ilmu Kesehatan*, vol. 5, no. 2, pp. 115–121, Dec. 2020, doi: 10.30604/jika.v5i2.289.
- [5] K. Wilayah and I. Ramadini, "Hubungan Deteksi Dini (Pap Smear) Dengan Kejadian Kanker Serviks Di Poli Obgyn," *Jurnal Endurance*, vol. 3, no. 1, 2018, doi: 10.22216/jen.v3i1.1885.
- [6] N. Ayyanar, G. Thavasi Raja, M. Sharma, and D. Sriram Kumar, "Photonic Crystal Fiber-Based Refractive Index Sensor for Early Detection of Cancer," *IEEE Sens J*, vol. 18, no. 17, pp. 7093–7099, Sep. 2018, doi: 10.1109/JSEN.2018.2854375.
- [7] T. G. Pratama, A. Ridwan, and A. Prihandono, "Penerapan Algoritma C4.5 untuk Klasifikasi Kanker Serviks Tingkat Awal," *Urecol Journal. Part E: Engineering*, vol. 1, no. 1, pp. 1–6, Mar. 2021, doi: 10.53017/uje.4.
- [8] M. Kamil, W. Cholil, D. Palembang Ji Jendral Ahmad Yani No, and S. Selatan, "Perbandingan Algoritma C4.5 dan Naive Bayes Pada Lulusan Tepat Waktu Mahasiswa," *Jurnal Informatika*, vol. 7, no. 2, pp. 97–106, 2020, [Online]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/ji>
- [9] J. Nata Permana, R. Goejantoro, and S. Prangga, "Comparison Of C4.5 Algorithm and Naïve Bayes for Prediction Of Student Study Timeliness (Case Study: Departement of Statistics Mulawarman University)," *Jurnal Eksponensial*, vol. 13, no. 2, 2022.
- [10] UCI Repository and Gokagglers, "Cervical Cancer Risk Classification," <https://www.kaggle.com/datasets/loveall/cervical-cancer-risk-classification>.
- [11] C. Anam and H. B. Santoso, "Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes untuk Klasifikasi Penerima Beasiswa," 2018.

- [12] R. Rachman, R. N. Handayani, and I. Artikel, "Klasifikasi Algoritma Naive Bayes Dalam Memprediksi Tingkat Kelancaran Pembayaran Sewa Teras UMKM," *Jurnal Informatika*, vol. 8, no. 2, 2021, [Online]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/ji>
- [13] S. W. Siahaan, K. D. R. Sianipar, P. P. P. A. N. W. F. I. R.H Zer, and D. Hartama, "Penerapan Algoritma C4.5 Dalam Meningkatkan Kemampuan Bahasa Inggris Pada Mahasiswa," *PETIR*, vol. 13, no. 2, pp. 229–239, Sep. 2020, doi: 10.33322/petir.v13i2.1029.
- [14] N. Yahya and A. Jananto, "Komparasi Kinerja Algoritma C.45 Dan Naive Bayes Untuk Prediksi Kegiatan Penerimaanmahasiswa Baru (Studi Kasus : Universitas Stikubank Semarang)," 2019.
- [15] E. Fitriani, R. Aryanti, A. Saepudin, and D. Ardiansyah, "Penerapan Algoritma C4.5 Untuk Klasifikasi Penempatan Tenaga Marketing," vol. 22, no. 1, 2020, doi: 10.31294/p.v21i2.