Komputika: Jurnal Sistem Komputer
Volume 12, Nomor 2, April 2023, hlm. 73 - 79
Terakreditasi Peringkat 3, SK No. 148/M/KPT/2020
DOI: **10.34010/komputika.v12i1.9629**

ISSN: 2252-9039 (print)
ISSN: 2655-3198 (online)

Klasifikasi Penyakit Kanker Prostat Menggunakan Algoritma Naïve Bayes dan K-Nearest Neighbor

Adi Muzakir[1], Anita Desiani[2*], Ali Amran[3]

[1,2,3)]Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sriwijaya
Jl. Raya Palembang–Prabumulih Km. 32, Indralaya, Indonesia 30662
**email: anita_desiani@unsri.ac.id*

(Naskah masuk: *22 April 2023*; diterima untuk diterbitkan: *18 Mei 2023*)

ABSTRAK *– Deteksi dini terhadap kasus kanker prostat pada banyak pengidap atau pria yang rentan risiko kanker prostat penting dilakukan untuk memulai pengobatan dan perencanaan kebutuhan medis yang tepat. Salah satu cara yang dapat dilakukan dalam deteksi penyakit kanker prostat adalah dengan melakukan klasifikasi menggunakan pendekatan data mining dengan algoritma Naïve Bayes dan K-Nearest Neighbor (K-NN). Penelitian ini bertujuan untuk mendapatkan hasil klasifikasi terbaik untuk mendeteksi penyakit kanker prostat dengan membandingkan kedua algoritma tersebut. Hasil akurasi klasifikasi kanker prostat dengan menggunakan algoritma Naïve Bayes adalah 80% dan K-NN sebesar 90%. Sementara untuk rata-rata keseluruhan nilai presisi algoritma Naïve Bayes dan K-NN masing-masing berada pad angka 71,5% dan 93%. Nilai recall untuk algoritma Naïve bayes didapatkan sebesar 88% dan algoritma K-NN yaitu 87,5%. Berdasarkan nilai akurasi, presisi, dan recall kedua algoritma tersebut, algoritma K-NN memiliki nilai yang lebih tinggi dibandingkan dengan algoritma Naïve Bayes, sehingga dapat dikatakan bahwa algoritma K-NN bekerja dengan baik dalam melakukan klasifikasi penyakit kanker prostat. Meskipun algoritma Naïve Bayes memiliki nilai yang lebih rendah dibandingkan dengan algoritma K-NN, tetapi nilai rata-rata untuk performa presisi, recall, dan akurasinya masih berada di atas 70%. Dapat dikatakan bahwa algoritma Naïve Bayes cukup baik dalam mengklasifikasi penyakit kanker prostat.*

Kata Kunci *– Data Mining, K-Nearest Neighbor, Kanker Prostat, Klasifikasi, Naïve Bayes*

Classification of Prostate Cancer Using Naïve Bayes and K-Nearest Neighbor Algorithms

ABSTRACT *– Early detection of cases of prostate cancer in many people or men who are susceptible to prostate cancer risk is important to start treatment and planning proper medical needs. One way to detect prostate cancer is to classify using a data mining approach with the Naïve Bayes algorithm and K-Nearest Neighbor (K-NN). This research aims to get the best classification results for detecting prostate cancer by comparing the two algorithms. The result of prostate cancer classification accuracy using Naïve Bayes algorithm is 80% and K-NN by 90%. Meanwhile, the overall precision value of Naïve Bayes and K-NN algorithms was at 71.5% and 93% respectively. The recall value for the Naïve Bayes algorithm was 88% and the K-NN algorithm was 87.5%. Based on the accuracy, precision, and recall values of the two algorithms, the K-NN algorithm has a higher value compared to naïve Bayes' algorithm, so it can be said that the K-NN algorithm works well in classifying prostate cancer. Although Naïve Bayes' algorithm has a lower value compared to the K-NN algorithm, the average value for precision performance, recall, and accuracy is still above 70%. It can be said that Naïve Bayes' algorithm is quite good at classifying prostate cancer.*

Keywords *- Classification, Data Mining, K-Nearest Neighbor, Naïve Bayes, Prostate Cancer*

1. PENDAHULUAN

Prostat merupakan bagian dari sistem reproduksi pria yang meliputi penis, prostat dan testikel. Prostat sebagai kelenjar asesorius terbesar pada pria, terletak tepat di bawah buli dan berada di sisi anterior dari rektum, berukuran sebesar buah kenari dan mengelilingi *uretra pars prostatikum*. Menurut Setiawan [1] kanker adalah penyakit yang ditandai dengan pembelahan sel yang tidak terkendali dan

kemampuan sel-sel tersebut untuk menyerang jaringan biologis yang lain. Kanker yang berkembang di prostat dalam sistem reproduksi pria merupakan kanker prostat, hal ini terjadi ketika sel prostat mengalami keterikatan pada reseptor androgen melalui proses *molecular docking* [2]. Terdapat beberapa faktor risiko yang dikaitkan dengan kejadian kanker prostat, antara lain usia, etnis, riwayat keluarga, diet dan gaya hidup, genetik, histopatologi, dan lokasi. Berdasarkan data *World Health Organization* (WHO), prevalensi kanker prostat menempati posisi pertama di dunia dan angka insidensinya merupakan peringkat kelima dari seluruh jenis kanker di dunia. Sekitar 95% pasien didiagnosa pada rentang usia 45 sampai 89 tahun (usia rerata 72 tahun). Insidensi kanker prostat meningkat seiring pertambahan usia. Risiko yang dimiliki pria untuk menderita kanker prostat dalam seumur hidupnya mendekati 10% [3].

Dengan tingginya jumlah kasus kanker prostat di dunia, perlu dilakukan deteksi dini terhadap kasus kanker prostat pada banyak pengidap atau pria yang rentan risiko kanker prostat. Salah satu cara yang dapat dilakukan dalam deteksi penyakit kanker prostat adalah dengan melakukan klasifikasi untuk mengetahui tingkat keganasannya, sehingga pengobatan dan perencanaan kebutuhan medis dapat dilakukan dengan efektif. Penyakit kanker prostat dapat dilakukan klasifikasi dengan menggunakan metode penyelesaian matematika, salah satunya adalah penggunaan pendekatan *data mining*. *Data mining* adalah proses penggalian data dari tumpukan *database* yang berukuran besar yang digunakan untuk menemukan *knowledge* berupa informasi penting dan bermanfaat [4]. Klasifikasi kanker prostat dengan menggunakan teknik *data mining* telah banyak dilakukan dalam beberapa penelitian sebelumnya. Diantaranya Saputra et al. [5] melakukan analisis klasifikasi risiko terhadap penderita prostat menggunakan metode *Naïve Bayes* dengan tingkat akurasi sebesar 70%. Peryoga et al. [6] melakukan deteksi kanker berdasarkan data *microarray* menggunakan metode *Naïve Bayes* dan *Hybrid Feature Selection*, di mana dengan menggunakan metode *Naïve Bayes* didapatkan tingkat akurasi kanker prostat sebesar 58,94%. Serta Arthawani [7] melakukan klasifikasi ekspresi genetika pada kanker prostat menggunakan metode *Support Vector Machine* yang mendapatkan tingkat akurasi sebesar 92%.

Dalam *data mining* banyak metode yang berkembang dalam klasifikasi seperti *Naïve Bayes*, *K-Nearest Neighbor* (K-NN), C4.5, dan lain-lain. Pada penelitian ini, akan digunakan dua algoritma untuk membandingkan algoritma mana yang terbaik yaitu *Naïve Bayes* dan *K-NN*. Hal ini dikarenakan kedua algoritma tersebut dianggap sebagai metode yang cepat, mudah, kuat, dan paling populer digunakan untuk klasifikasi [8]. *Naïve Bayes* merupakan sebuah algoritma pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan [9]. Algoritma *Naïve Bayes* memiliki beberapa keunggulan yaitu memiliki performa klasifikasi yang tinggi, mudah digunakan, hanya membutuhkan satu kali *scan* data *training*, mampu menangani data yang kosong (*missing value*), dan data kontinu [10]. Namun *Naïve Bayes* memiliki kekurangan pada pemilihan atribut yang ada di dalam data sehingga dapat mempengaruhi hasil akhir yang berupa nilai akurasi [11]. Beberapa contoh penelitian yang telah menggunakan algoritma *Naïve Bayes* dalam klasifikasi diantaranya Ari Bianto [12] merancang sistem klasifikasi penyakit jantung menggunakan *Naïve Bayes* yang memperoleh nilai akurasi sebesar 90,61%. Ridwan [13] menerapkan algoritma *Naïve Bayes* untuk klasifikasi penyakit *diabetes mellitus* dengan tingkat akurasi yang diperoleh sebesar 90,20%, dan Imandasari et al. [14] menggunakan algoritma *Naive Bayes* dalam klasifikasi lokasi pembangunan sumber air yang memperoleh tingkat akurasi sebesar 78,95%.

K-NN adalah suatu metode untuk mengklasifikasikan objek berdasarkan data *training* yang jaraknya paling dekat dengan objek tersebut [15]. Beberapa kelebihan dari metode K-NN adalah sangat non linier, mudah dipahami serta diimplementasikan, tangguh terhadap *training* data yang *noise*, dan efektif apabila data *training* besar [16]. Kekurangan K-NN yaitu pada pemilihan nilai k perlu mempertimbangkan ukuran data, apabila ukuran data terlalu kecil maka dimensi data dan variasi jarak dalam tabel jarak antara data latih menjadi lebih kecil sehingga peluang suatu data uji dikenali masuk ke kelas lain menjadi lebih besar [17]. Beberapa penelitian yang menggunakan algoritma K-NN mendapatkan hasil yang sangat baik di antaranya sistem klasifikasi penyakit *diabetes mellitus* menggunakan metode K-NN yang dilakukan oleh Yunita [18] dengan tingkat akurasi sebesar 96% dan Sharma et al. [19] melakukan klasifikasi penyakit kanker serviks menggunakan K-NN yang memperoleh akurasi maksimum sebesar 82,9%.

Pada penelitian ini akan dilakukan perbandingan penggunaan algoritma *Naïve Bayes* dan K-NN untuk mendapatkan hasil klasifikasi terbaik dalam mendeteksi penyakit kanker prostat. Dataset yang digunakan dalam penelitian ini yaitu dataset kanker prostat yang didapatkan dari situs *Kaggle* yang berupa data laboratorium. Di mana akan menghasilkan dua kelas klasifikasi yaitu kanker prostat ganas (*Malignant*) dan kanker prostat jinak (*Benign*). Dalam mengukur kinerja algoritma *Naïve Bayes* dan K-NN akan dilihat berdasarkan nilai akurasi, presisi, dan *recall* sehingga didapatkan algoritma terbaik dalam melakukan klasifikasi

kanker prostat.

2. METODE DAN BAHAN

2.1. Deskripsi Data

Data yang digunakan dalam penelitian ini yaitu dataset hasil laboratorium untuk penyakit kanker prostat yang diperoleh dari situs *Kaggle* (https://www.kaggle.com/datasets/sajidsaifi/prostate-cancer) dengan format csv. Di mana data tersebut terdiri dari 9 atribut, 8 di antaranya sebagai atribut prediksi yaitu *radius*, *texture*, *perimeter*, *area*, *smoothness*, *compactness*, *symmetry*, dan *fractal dimension* serta satu atribut target yaitu *diagnosis result* yang terdiri dari 2 kelas yaitu *Malignant* yang berarti penderita kanker prostat ganas dan *Benign* yang berarti penderita kanker prostat jinak. Data yang tersedia berjumlah 100 data di mana 62 data masuk ke dalam kelas kanker prostat ganas (*Malignant*) dan 38 data masuk ke dalam kelas kanker prostat jinak (*Benign*). Atribut data yang digunakan dalam penelitian ini dapat dilihat pada tabel 1.

Tabel 1. Atribut Data

Atribut	Keterangan	Variabel	Rentang Nilai
diagnosis result	*Malignant, Benign*	Kategorik	-
radius	radius	Nominal	[9,25]
texture	tekstur	Nominal	[11,27]
perimeter	perimeter	Nominal	[52,172]
area	Area	Nominal	[202,1878]
smoothness	kelancaran	Nominal	[0,1]
compactness	Kekompak-an	Nominal	[0,1]
symmetry	simetri	Nominal	[0,1]
fractal dimension	dimensi fraktal	Nominal	[0,1]

2.2. Praproses Data

Praproses data dilakukan untuk mendapatkan hasil yang lebih akurat, pengurangan waktu perhitungan untuk *large scale problem*, dan membuat nilai data menjadi lebih kecil tanpa merubah informasi yang ada di dalamnya [20]. Praproses data dapat berupa *cleaning* data, integrasi data, reduksi data, dan transformasi data. Pada tabel 1 dapat diperhatikan bahwa terdapat rentang nilai yang berbeda di beberapa atribut. Perbedaan rentang nilai pada atribut tersebut menyebabkan tidak berfungsinya atribut yang memiliki nilai yang jauh lebih kecil dibandingkan dengan atribut-atribut lainnya. Untuk mengatasinya maka diperlukan adanya transformasi data dengan normalisasi untuk menyamakan rentang nilai setiap atribut dengan skala tertentu, sehingga dapat menghasilkan klasifikasi yang lebih baik. Normalisasi data yang dilakukan dalam penelitian ini yaitu dengan menggunakan teknik *min-max normalization* dengan persamaan (1) [20].

$$normalized\ (x) = \frac{min\ Range\ x + (x\ minValue)(maxRange - minRange)}{maxValue - minValue} \quad (1)$$

Dalam penelitian ini, teknik pembagian data yang digunakan yaitu *percentage split*, di mana dataset keseluruhan dibagi menjadi dua bagian yakni data *training* dan data *testing*. Data *training* yang diambil sebesar 80% data dari keseluruhan data sedangkan 20% sebagai data *testing* secara acak.

2.3. Algoritma *Naïve Bayes*

Naïve Bayes adalah salah satu metode dalam *data mining* yang menerapkan konsep peluang *Bayes*. Semua atribut diperlakukan sama dan bebas antara satu atribut dengan atribut lainnya. Metode ini menggunakan *Naïve Bayes Classifier* untuk untuk menghitung bobot peluang setiap atribut. Menurut Ratniasih [21] langkah-langkah melakukan klasifikasi menggunakan algoritma *Naïve Bayes* adalah sebagai berikut:

1. Menghitung jumlah kategori dari setiap variabel;
2. Menghitung peluang pada setiap kategori;
3. Menentukan frekuensi atau jumlah kemunculan pada setiap kategori;
4. Menentukan kategori dengan nilai maksimal.

Perhitungan algoritma *Naïve Bayes* dilakukan dengan menggunakan persamaan (2) berikut [21]:

$$P(H|X) = \frac{P(X|H)\ P(H)}{P(X)} \quad (2)$$

Keterangan
X : Data dengan kelas yang belum diketahui
H : Hipotesis data X merupakan kelas spesifik
$P(H)$: Probabilitas hipotesis H
$P(X)$: Probabilitas hipotesis X
$P(H|X)$: Probabilitas hipotesis H berdasarkan kondisi X
$P(X|H)$: Probabilitas hipotesis X berdasarkan kondisi H

2.4. Algoritma *K-Nearest Neighbor*

Algoritma *K-Nearest Neighbor* (K-NN) adalah suatu metode yang digunakan untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut [22]. Kelas yang paling banyak muncul dalam algoritma ini akan menjadi kelas hasil klasifikasi. Langkah-langkah untuk menghitung algoritma K-NN yaitu sebagai berikut:

1. Menentukan banyaknya tetangga k;
2. mengitung jarak dari data untuk dibandingkan dengan dataset *training*, dapat dihitung dengan

persamaan jarak *Euclidean* pada persamaan (3) [23];

$$d(x,y) = \sqrt{\sum_{i=0}^{n}(x_i - y_i)^2} \quad (3)$$

Keterangan:
$d(x,y)$: Jarak *Euclidean*
x : Data 1
y : Data 2
i : Index Fitur
n : Index Fitur

3. mengatur urutan menaik dari jarak (mengurutkan dari kecil ke besar) dan pilih himpunan k paling sedikit dari dataset terkecil;
4. Menentukan bahwa jawaban dengan data yang akan diprediksi adalah kelompok data yang memiliki jumlah k pertama dari kumpulan data terbesar;
5. Menetapkan kelas kelas terdekat dengan titik pertimbangan.

2.5. Analisis Hasil

Untuk analisis hasil, digunakan tabulasi silang (*confusion matrix*) yang menampilkan visualisasi kinerja dari algoritma klasifikasi menggunakan data dalam matriks yang membandingkan klasifikasi dalam bentuk *True Positive* (TP), *False Positive* (FP), *True Negative* (TN), dan *False Negative* (FN). Pada Tabel 2 dapat diperhatikan *confusion matrix* untuk dua kelas.

Tabel 2. *Confusion Matrix* Untuk Dua Kelas

Actual	***Prediction***	
	Positive	*Negative*
Positive	*True Positive* (TP)	*False Negative* (FN)
Negative	*False Positive* (FP)	*True Negative* (TN)

True positive menunjukkan data positif yang diklasifikasikan sebagai positif dan *false positive* diklasifikasikan sebagai negatif. Sementara *true negative* menunjukkan data negatif yang diklasifikasikan sebagai negatif dan *false negative* diklasifikasikan sebagai positif [24]. Dari hasil *confusion matrix* dapat diukur kinerja metode yang digunakan dalam penelitian ini dengan menggunakan akurasi, presisi, dan *recall*.

a. Akurasi

Akurasi atau tingkat kesalahan merupakan angka prediksi yang benar atau salah yang dibuat oleh model melalui kumpulan dari data. Akurasi biasanya dihitung dengan menggunakan tes independen yang tidak selalu digunakan dalam proses pembelajaran. Untuk menghitung nilai akurasi digunakan persamaan (4) [25]:

$$Akurasi = \frac{TP+TN}{TP+TN+FN+FP} \quad (4)$$

b. Presisi

Presisi merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. Persamaan (5) digunakan untuk menghitung nilai presisi [25]

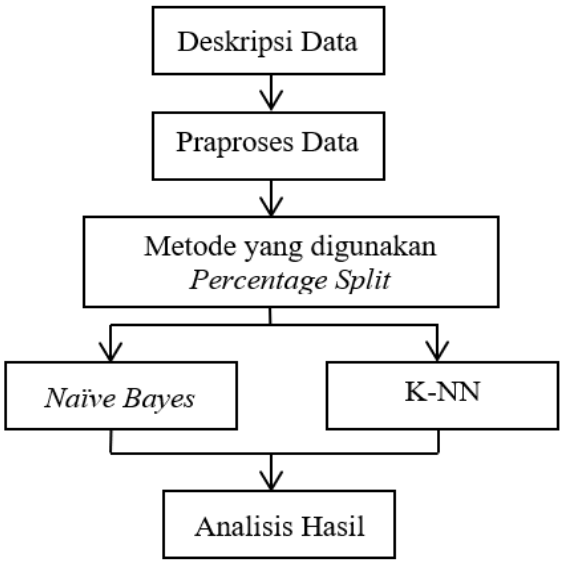
$$Presisi = \frac{TP}{TP+FP} \quad (5)$$

c. *Recall*

Recall atau sensitivitas merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. Nilai *recall* dapat dihitung dengan menggunakan persamaan (6) [25]

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

Hasil yang didapatkan oleh pengujian menggunakan algoritma *Naïve Bayes* dan K-NN akan dibandingkan untuk mengetahui algoritma mana yang memiliki keakuratan lebih baik dalam menciptakan model. Secara umum, alur metode yang digunakan dalam penelitian ini dapat dilihat pada gambar 1.



Gambar 1. Alur Metode Penelitian

3. HASIL DAN PEMBAHASAN

3.1. Hasil

1. Praproses Data

Normalisasi data menggunakan teknik *min-max normalization* yang diterapkan pada penelitian ini yaitu mengubah setiap nilai atribut menjadi rentang nilai antara 0 sampai 1. Normalisasi data diterapkan pada atribut *radius*, *texture*, *perimeter*, dan *area* yang memiliki rentang yang jauh berbeda dengan atribut lainnya (tabel 1). Dengan demikian, data akan memiliki skala yang seragam dan mempertahankan proporsi relatif antar atribut dan memungkinkan perbandingan yang lebih adil antara atribut-atribut tersebut.

2. Algoritma *Naïve Bayes*

Pada metode *Naïve Bayes* yang menggunakan teknik pembagian data *percentage split* dengan data *training* sebesar 80% dan data *testing* sebesar 20% didapatkan *confusion matrix* pada tabel 3.

Tabel 3. *Confusion Matrix Naïve Bayes*

Actual	*Prediction*	
	Benign	***Malignant***
Benign	3	0
Malignant	4	13

Dari tabel 3 dapat dilihat bahwa terdapat 3 data diprediksi secara benar sebagai kanker prostat jinak (*Benign*). 13 data pada *testing* ini berhasil diklasifikasikan sebagai kanker prostat ganas (*Malignant*), namun 4 data yang harusnya diklasifikasikan sebagai *Malignant* dikenali sebagai *Benign*. Dari *confusion matrix* dapat dihitung nilai akurasi yang diperoleh dari total data yang berhasil diprediksi secara benar sebesar 80%. Hasil ini menunjukkan algoritma *Naïve Bayes* cukup baik dalam mengklasifikasi kanker prostat. Hasil pengukuran *recall* untuk masing-masing label adalah 100% untuk label *Benign* dan 76% untuk label *Malignant*. Hasil pengukuran presisi untuk masing-masing label adalah sebagai berikut, label *Benign* adalah 43% dan 100% untuk label *Malignant*. Secara lebih lengkap, hasil presisi, akurasi, dan *recall* dengan menggunakan metode *Naïve Bayes* dapat dilihat pada tabel 4.

Tabel 4. Hasil Pemodelan Algoritma *Naïve Bayes*

Target Class	**Presisi**	***Recall***	**Akurasi**
Benign	43%	100%	80%
Malignant	100%	76%	

3. Algoritma *K-Nearest Neighbor* (K-NN)

Pada implementasi algoritma *K-Nearest Neighbor* (K-NN) ini akan dilakukan percobaan dengan nilai k sebanyak 7 nilai, hal ini dilakukan untuk mendapatkan hasil pemodelan yang terbaik. Pada tabel 5 dapat diperhatikan nilai *confusion matrix* dengan menggunakan nilai k = 1.

Tabel 5. *Confusion Matrix* K-NN dengan Nilai K = 1

Actual	*Prediction*	
	Benign	***Malignant***
Benign	5	3
Malignant	1	11

Dari tabel 5 dapat dilihat terdapat 16 data diprediksi secara benar, tetapi 4 data diprediksi dalam label yang salah. Akurasi yang diperoleh pada algoritma K-NN dengan nilai k = 1 adalah 80%, sama seperti hasil akurasi algoritma *Naïve Bayes*. Nilai *recall* yang diperoleh pada K-NN untuk masing-masing label adalah 62% untuk label *Benign* dan 92% untuk label *Malignant*. Nilai presisi untuk label *Benign* yaitu sebesar 83%, sementara untuk label *Malignant* adalah sebesar 79%. Dapat diperhatikan bahwa nilai akurasi yang didapatkan dengan menggunakan nilai k = 1 memiliki hasil yang sama persis seperti pada algoritma *Naïve Bayes*, maka dari itu untuk mendapatkan hasil akurasi yang terbaik dari kedua algoritma tersebut akan dilakukan percobaan dengan nilai k yang berbeda. Pada tabel 6 dapat dilihat hasil pemodelan algoritma K-NN dengan menggunakan 7 nilai k yang berbeda.

Tabel 6. Hasil Pemodelan Algoritma K-NN

K	***Target Class***	**Presisi**	***Recall***	**Akurasi**
1	*Benign*	83%	62%	80%
	Malignant	79%	92%	
2	*Benign*	71%	62%	75%
	Malignant	77%	83%	
3	*Benign*	100%	62%	85%
	Malignant	80%	100%	
4	*Benign*	86%	75%	85%
	Malignant	85%	92%	
5	*Benign*	100%	75%	90%
	Malignant	86%	100%	
6	*Benign*	86%	75%	85%
	Malignant	85%	92%	
7	*Benign*	100%	62%	85%
	Malignant	80%	100%	

Dari tabel 6 dapat diperhatikan bahwa baik menggunakan nilai akurasi, presisi, dan *recall* didapatkan nilai terbesar pada k = 5 yaitu akurasi sebesar 90%, presisi masing-masing untuk label *Benign* sebesar 100% dan untuk label *Malignant* 86%. Untuk nilai *recall* didapatkan untuk label *Benign* adalah 75% dan label *Malignant* sebesar 100%. Hasil pemodelan menggunakan metode K-NN tersebut menunjukkan bahwa algoritma ini sangat baik dalam mengklasifikasikan kanker prostat pada dataset *Kaggle*.

3.2 Perbandingan Hasil Kedua Algoritma

Hasil klasifikasi dua algoritma *Naïve Bayes* dan K-NN, terlihat bahwa kedua metode tersebut bekerja baik dalam melakukan klasifikasi kanker prostat pada dataset *Kaggle*. Perbandingan hasil pengukuran kedua algoritma tersebut dapat dilihat pada tabel 7 dan 8.

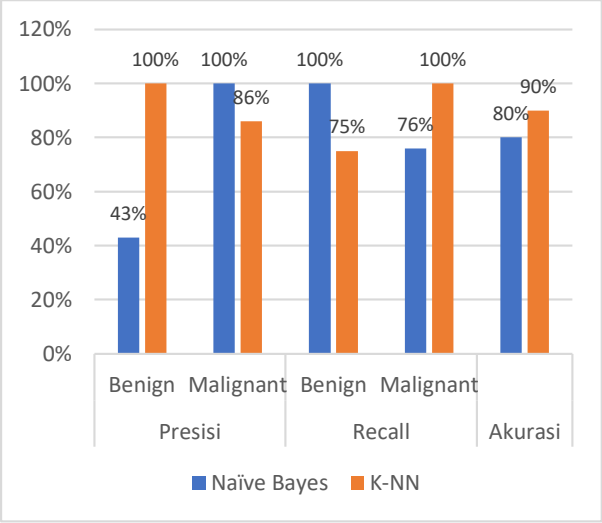
Tabel 7. Nilai Akurasi dan Presisi Algoritma *Naïve Bayes* dan K-NN

Algoritma	Presisi		Akurasi
	Benign	*Malignant*	
Naïve Bayes	43%	100%	80%
K-NN	100%	86%	90%

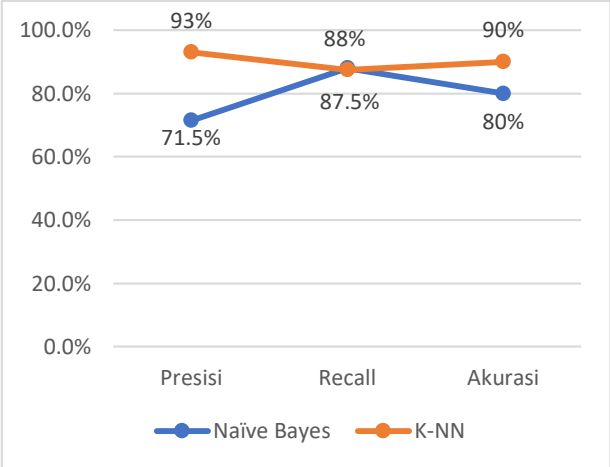
Tabel 8. Nilai *Recall* Algoritma *Naïve Bayes* dan K-NN

Algoritma	*Recall*	
	Benign	*Malignant*
Naïve Bayes	100%	76%
K-NN	75%	100%

Dari tabel 7 dilihat bahwa algoritma K-NN nilai akurasi klasifikasinya lebih baik dibandingkan algoritma *Naïve Bayes*. Hasil akurasi, presisi, dan *recall* dari algoritma *Naïve Bayes* dan K-NN dapat dilihat secara ringkas pada gambar 2, dan untuk nilai rata-rata dari seluruh label dapat dilihat pada gambar 3.



Gambar 2. Nilai Presisi, *Recall*, dan Akurasi Untuk Hasil Klasifikasi



Gambar 3. Rata-rata Keseluruhan Nilai Presisi, *Recall*, dan Akurasi

Pada gambar 3 dapat dilihat bahwa nilai presisi yang dihasilkan oleh algoritma *Naïve Bayes* yaitu sebesar 71,5% atau lebih kecil 21,5% dibandingkan dengan algoritma K-NN yang memiliki nilai presisi sebesar 93%. Untuk nilai *recall* baik yang dihasilkan algoritma *Naïve Bayes* dan K-NN tidak terlalu jauh berbeda yaitu masing-masing di angka 88% dan 87,5%. Sementara untuk nilai akurasi yang dihasilkan oleh algoritma *Naïve Bayes* dan K-NN masing-masing sebesar 80% dan 90% atau memiliki selisih sebesar 10%.

4. KESIMPULAN

Hasil klasifikasi penyakit kanker prostat dengan menggunakan algoritma K-NN lebih baik dibandingkan dengan algoritma *Naïve Bayes* yang dapat dilihat dari nilai presisi, *recall*, dan akurasinya yang lebih tinggi. Hal tersebut menjelaskan bahwa algoritma K-NN bekerja dengan sangat baik dalam melakukan klasifikasi penyakit kanker prostat pada dataset *Kaggle*. Meskipun algoritma *Naïve Bayes* memiliki nilai yang lebih rendah dibandingkan dengan algoritma K-NN, tetapi nilai rata-rata untuk performa presisi, *recall*, dan akurasinya masih berada di atas 70%. Dapat disimpulkan bahwa algoritma *Naïve Bayes* cukup baik dalam mengklasifikasi penyakit kanker prostat pada dataset *Kaggle*. Untuk meningkatkan akurasi, beberapa saran yang dapat diberikan penulis yaitu dapat menggunakan teknik pengolahan data lainnya seperti reduksi dimensi atau penghilangan atribut yang tidak relevan untuk memperbaiki representasi data. Selain itu dapat menggunakan metode penanganan *imbalance* atau ketidakseimbangan data seperti *oversampling* atau *undersampling* untuk membuat distribusi kelas menjadi lebih seimbang.

DAFTAR PUSTAKA

[1] D. Setiawan, "The Effect of Chemotherapy in Cancer Patient To Anxiety," *J. Major.*, vol. 4, no. 4, pp. 94–99, 2015.

[2] A. Arwansyah, L. Ambarsari, and T. I. Sumaryada, "Simulasi *Docking* Senyawa Kurkumin dan Analognya Sebagai Inhibitor Reseptor Androgen pada Kanker Prostat," *Curr. Biochem.*, vol. 1, no. 1, pp. 11–19, 2014.

[3] A. F. Indarti and S. M. Sekarutami, "Tatalaksana Kanker Prostat," *Radioter. Onkol. Indones.*, vol. 6, no. 1, pp. 19–28, 2015.

[4] S. Wahyuni, K. S. S, and M. I. Perangin-Angin, "Implementasi Rapidmaner dalam Menganalisa Data Mahasiswa Drop Out," *J. Abdi Ilmu*, vol. 10, no. 2, pp. 1899–1902, 2017.

[5] I. Saputra, "Analisis Klasifikasi Risiko Terhadap Penderita Prostat Menggunakan Metode Naive Bayes," *Universitas Muhammadiyah Jember*, 2018.

[6] B. Peryoga, A. Adiwijaya, and W. Astuti, "Deteksi Kanker Berdasarkan Data Microarray Menggunakan Metode Naïve Bayes dan Hybrid

Feature Selection," *J. Media Inform. Budidarma*, vol. 4, no. 3, p. 486, 2020.

[7] G. Arthawani, "Klasifikasi Ekspresi Genetika Pada Kanker Prostat Menggunakan Metode Support Vector Machine," *Digit. Repos. Univ. Jember*, no. September 2019, pp. 2019–2022, 2021.

[8] D. Sartika and D. I. Sensuse, "Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian," *J. Tek. Inform. Dan Sist. Inf.*, vol. 1, no. 2, pp. 151–161, 2017.

[9] A. Saleh, "Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga," *Creat. Inf. Technol. J.*, vol. 2, no. 3, pp. 207–217, 2015.

[10] A. Riza Khadafy and R. Satria Wahono, "Penerapan Naive Bayes untuk Mengurangi Data Noise pada Klasifikasi Multi Kelas dengan Decision Tree," *J. Intell. Syst.*, vol. 1, no. 2, 2015.

[11] T. Arifin and D. Ariesta, "Prediksi Penyakit Ginjal Kronis Menggunakan Algoritma Naive Bayes Classifier Berbasis Particle Swarm Optimization," *J. Tekno Insentif*, vol. 13, no. 1, pp. 26–30, 2019.

[12] M. Ari Bianto, "Perancangan Sistem Klasifikasi Penyakit Jantung Mengunakan Naïve Bayes Designing a Heart Disease Classification System Using Naïve Bayes," *Citec J.*, vol. 6, no. 1, 2019.

[13] A. Ridwan, "Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 15–21, 2020.

[14] T. Imandasari, E. Irawan, A. P. Windarto, and A. Wanto, "Algoritma Naive Bayes Dalam Klasifikasi Lokasi Pembangunan Sumber Air," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 750, 2019.

[15] M. Safaat, A. Sahari, and D. Lusiyanti, "Implementasi Metode K-Nearest Neighbor Untuk Mengklasifikasi Jenis Penyakit Katarak," *J. Ilm. Mat. Dan Terap.*, vol. 17, no. 1, pp. 92–99, 2020.

[16] K. Eliyen, H. Tolle, and M. A. Muslim, "K-Nearest Neighbor Untuk Klasifikasi Penilaian Pada Virtual Patient Case," *J. Arus Elektro Indones.*, vol. 3, no. 1, pp. 15–18, 2017.

[17] A. Bode, "K-Nearest Neighbor Dengan Feature Selection Menggunakan Backward Elimination Untuk Prediksi Harga Komoditi Kopi Arabika," *Ilk. J. Ilm.*, vol. 9, no. 2, pp. 188–195, 2017.

[18] F. Yunita, "Sistem Klasifikasi Penyakit Diabetes Mellitus Menggunakan Metode K-Nearest Neighbor (K-NN)," *Bappeda*, vol. 2, pp. 223–230, 2016.

[19] M. Sharma, S. Kumar Singh, P. Agrawal, and V. Madaan, "Classification of Clinical Dataset of Cervical Cancer using KNN," *Indian J. Sci. Technol.*, vol. 9, no. 28, 2016.

[20] D. A. Nasution, H. H. Khotimah, and N. Chamidah, "Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN," *Comput. Eng. Sci. Syst. J.*, vol. 4, no. 1, p. 78, 2019.

[21] N. L. Ratniasih, "Optimasi Data Mining Menggunakan Algoritma Naïve Bayes Dan C4.5 Untuk Klasifikasi Kelulusan Mahasiswa," *J. Teknol. Inf. dan Komput.*, vol. 5, no. 1, pp. 28–34, 2019.

[22] F. Liantoni, "Klasifikasi Daun Dengan Perbaikan Fitur Citra Menggunakan Metode K-Nearest Neighbor," *J. Ultim.*, vol. 7, no. 2, pp. 98–104, 2016.

[23] F. Kurnia, S. Kom, J. Kurniawan, and I. F. St, "Klasifikasi Keluarga Miskin Menggunakan Metode K- Nearest Neighbor Berbasis Euclidean Distance," no. November, pp. 230–239, 2019.

[24] Y. Pratama, A. Roberto Tampubolon, L. Diantri Sianturi, R. Diana Manalu, and D. Frietz Pangaribuan, "Implementation of Sentiment Analysis on Twitter Using Naïve Bayes Algorithm to Know the People Responses to Debate of DKI Jakarta Governor Election," *J. Phys. Conf. Ser.*, vol. 1175, no. 1, 2019.

[25] M. M. Baharuddin, H. Azis, and T. Hasanuddin, "Analisis Performa Metode K-Nearest Neighbor Untuk Identifikasi Jenis Kaca," *Ilk. J. Ilm.*, vol. 11, no. 3, pp. 269–274, 2019.