Penerapan Algoritma KNN dan *Naive Bayes* untuk Klasifikasi *Stunting* pada Balita di Desa Pasirjengkol

Putri Meriyana[1], Adi Rizky Pratama[2], Euis Nurlaelasari[3], Ayu Ratna Juwita[4]
Universitas Buana Perjuangan Karawang
If21.putrimeriyana@mhs.ubpkarawang.ac.id[1], adi.rizky@ubpkarawang.ac.id[2],
euis.nurlaelasari@ubpkarawang.ac.id[3], ayurj@ubpkarawang.ac.id[4]

Abstrak: *Stunting* merupakan kondisi gangguan tumbuh anak balita yang terjadi karena kekurangan asupan gizi secara terus-menerus dalam waktu yang lama, sehingga menyebabkan anak mengalami gagal tumbuh secara fisik maupun perkembangan kemampuan berpikir. Penelitian ini menerapkan algoritma *K-Nearest Neighbor* (KNN) dan *Naïve Bayes* dalam klasifikasi status *stunting* balita di Desa Pasirjengkol berdasarkan data usia, jenis kelamin, dan tinggi badan. Dataset yang digunakan berjumlah 1.195 data yang dikumpulkan pada tahun 2023 dan 2024, namun setelah proses pembersihan data menjadi 1.192 data. Proses penelitian meliputi pengumpulan data, *pre-processing* data penghapusan data duplikat, *transformasi* data mencakup *label encoding*, dan normalisasi dengan *Min-Max Scaling*, kemudian pemilihan fitur, pelatihan model, dan evaluasi menggunakan *confusion matrix*. Hasil evaluasi menunjukkan algoritma KNN dengan k=1 memberikan akurasi tinggi sebesar 97.90%, sedangkan *Naïve Bayes* sebesar 54.39%. Berdasarkan hasil tersebut, KNN menunjukkan kinerja yang lebih baik dalam mengklasifikasikan status *stunting* pada balita di Desa Pasirjengkol.
Kata Kunci : Balita, Klasifikasi, *K-Nearest Neighbor*, *Naive Bayes*, *Stunting.*

Abstract: *Stunting is a condition of impaired growth in toddlers caused by prolonged inadequate nutritional intake, leading to physical growth failure as well as delays in cognitive development. This study applies the K-Nearest Neighbor (KNN) and Naïve Bayes algorithms to classify stunting status in toddlers in Pasirjengkol Village based on age, gender, and height data. The dataset consists of 1,195 records collected in 2023 and 2024, which was reduced to 1,192 after the data cleaning process. The research process includes data collection, preprocessing (removal of duplicate data), data transformation using label encoding, normalization using Min-Max Scaling, feature selection, model training, and evaluation using a confusion matrix. The evaluation results show that the KNN algorithm with k=1 achieved a high accuracy of 97.90%, while Naïve Bayes reached 54.39%. Based on these results, KNN demonstrated better performance in classifying stunting status among toddlers in Pasirjengkol Village.*
Keywords*: Classification, K-Nearest Neighbor, Naïve Bayes, Stunting, Toddlers.*

1. Pendahuluan

Kemajuan teknologi informasi telah menjadi pendorong utama *transformasi* di berbagai bidang kehidupan manusia, termasuk sektor kesehatan. Masalah kesehatan yang membutuhkan perhatian serius salah satunya *stunting*, gangguan tumbuh pada balita yang disebabkan oleh kekurangan gizi dalam jangka waktu yang lama. *Stunting* tidak hanya berdampak pada tinggi badan yang lebih pendek dari standar usia, tetapi juga dapat menghambat perkembangan mental dan kesehatan anak di masa depan (Pebrianti et al., 2024). Menurut standar *antropometri* Kementerian Kesehatan Republik Indonesia, seorang

anak dikategorikan *stunting* jika tinggi badannya di bawah -2 standar deviasi dari median *z-score* . Data Survei Status Gizi Indonesia (SSGI) menunjukkan persentase *stunting* secara nasional mengalami penurunan, dari 24,4% pada tahun 2021 menjadi 21,6% pada tahun 2022. Namun demikian, capaian tersebut belum mencapai target nasional yang di tetapkan sebesar 14% pada 2024.

Desa Pasirjengkol dipilih sebagai objek studi karena memiliki data pertumbuhan balita yang terdokumentasi dengan baik melalui Posyandu setempat, serta merupakan bagian dari upaya kontribusi daerah dalam penurunan angka *stunting*. Pemahaman dan identifikasi risiko *stunting* pada balita di wilayah ini penting dilakukan guna mendukung program penanganan yang lebih tepat sasaran. Oleh karena itu, diperlukan pendekatan berbasis teknologi untuk meningkatkan efektivitas proses identifikasi tersebut.

Pendekatan yang dapat dimanfaatkan untuk mengidentifikasi status *stunting* pada balita yaitu pemanfaatan algoritma pembelajaran mesin (*machine learning*). Metode ini bekerja dengan mengolah data kesehatan balita, seperti usia, jenis kelamin, dan tinggi badan, untuk mengelompokkan mereka ke dalam kategori tertentu. Beberapa penelitian terdahulu menunjukkan bahwa algoritma pembelajaran mesin dapat mengklasifikasikan status *stunting* dengan baik.

Seperti pada penelitian oleh Azizah & fatah 2024 menggunakan algoritma *K-Nearest Neighbor* (KNN) untuk klasfikasi *stunting* pada anak berdasarkan usia, jenis kelamin, dan tinggi badan, dimana *variabel* usia memiliki pengaruh terbesar terhadap klasifikasi (Azizah & Fatah, 2024). Penelitian oleh Pebrianti dkk 2023, menunjukkan algoritma *K-Nearest Neighbor* (KNN) mampu menghasilkan klasifikasi yang baik, dengan akurasi mencapai 92% berdasarkan 503 data (Pebrianti et al., 2024). Sementara itu penelitian oleh Titimeidara dan Hadikurniawati 2021 menggunakan algoritma *Naive Bayes* menunjukkan hasil akurasi 88% dari 300 data (Titimeidara & Hadikurniawati, 2021.). Adapun penelitian oleh Noer Azzahra dkk 2024 menggunakan algoritma *Naive Bayes* menunjukkan hasil akurasi 93.2% (Noer Azzahra et al., 2024).

Penelitian ini menyajikan pendekatan baru dengan memanfaatkan data riil dari Posyandu di Desa Pasirjengkol. Tidak hanya menerapkan satu algoritma, penelitian ini juga membandingkan dua metode klasifikasi, yaitu *K-Nearest Neighbor* (KNN) dan *Naïve Bayes.* Selain itu, untuk meningkatkan akurasi model KNN, dilakukan pencarian nilai *k* yang optimal menggunakan teknik *cross-validation*, sehingga diperoleh hasil klasifikasi yang lebih akurat.

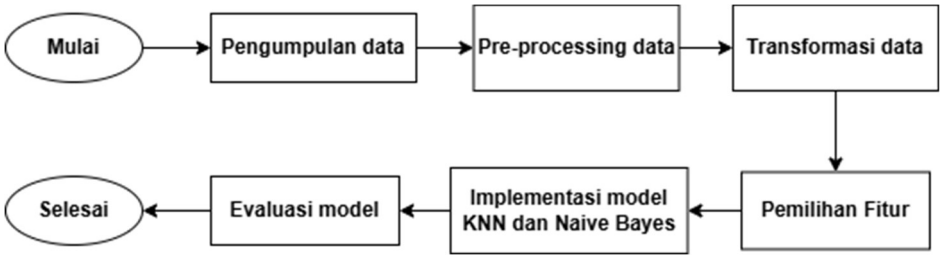
Tujuan dari penelitian ini adalah untuk mengklasifikasikan status stunting pada balita ke dalam tiga kategori, yaitu normal, *stunting*, dan *severely stunting*, serta membandingkan kinerja algoritma *K-Nearest Neighbor* (KNN) dan *Naïve Bayes* guna menentukan metode klasifikasi yang paling optimal.

2. Metode Penelitian

A. Tahapan Penelitian

Langkah pertama penelitian ini dimulai dengan pengumpulan data sebagai tahapan awal, kemudian dilanjutkan dengan tahap *pre-processing* data, *transformasi* data, dan pemilihan fitur. Selanjutnya dilakukan *implementasi* algoritma *K-Nearest Neighbor* (KNN) dan *Naïve Bayes* untuk klasifikasi, kemudian dilanjutkan dengan evaluasi model

menggunakan *confusion matrix*. Gambar 1 menampilkan alur tahapan metode penelitian ini.



Gambar 1. Tahapan metode penelitian

B. Pengumpulan Data

Data dikumpulkan melalui Posyandu yang berada di wilayah Desa Pasirjengkol, Kabupaten Karawang dengan total sebanyak 1195 data. Data yang dikumpulkan merupakan data pertumbuhan balita dan mencakup lima atribut , yaitu nama, usia, jenis kelamin, tinggi badan dan status. Data balita yang digunakan pada penelitian ini ditunjukkan pada Tabel 1.

Tabel 1. Dataset Penelitian

No	Nama	Usia Bulan	Jenis Kelamin	Tinggi Badan	Status
1	Mizan	24	Laki-Laki	73	*Severely* Stunting
2	Rafa	24	Laki-Laki	78	*Stunting*
3	Raizi	23	Laki-Laki	86	Normal
4	Elvano	23	Laki-Laki	82	Normal
5	Nadira	36	Perempuan	85.4	*Stunting*
….	….	….	….	….	….
1195	Hamdan	28	Laki-Laki	78.3	*Severely Stunting*

C. *Pre-processing*

Preprocessing data merupakan tahap untuk membersihkan dan mempersiapkan data sebelum memasuki proses klasifikasi (Rahayu et al., 2022). Tahap ini bertujuan untuk meningkatkan kualitas data agar dapat digunakan sebaik mungkin dalam pelatihan model atau analisis lanjutan. Salah satu tahapan utama dalam *preprocessing Data* yaitu menghapus data duplikat.

D. *Transformasi* Data

Transformasi data mengubah data ke dalam format yang lebih tepat guna mendukung proses analisis dan pemodelan. Dalam penelitian ini, *transformasi* data dilakukan dengan menggunakan *Label Encoding* untuk mengubah *variabel* kategorikal menjadi nilai numerik, dan normalisasi data untuk mengubah nilai atribut numerik ke dalam rentang tertentu.

1. Label *Encoding*

   Pada penelitian ini, dilakukan *Label Encoding* untuk kolom jenis kelamin dan status. Untuk kolom jenis kelamin, kategori laki-laki dikodekan sebagai 0, dan perempuan sebagai 1. Sementara itu, untuk kolom status, kategori normal di kodekan sebagai 0, *stunting* sebagai 1 dan *severely stunting* sebagai 2.

2. *Normalisasi* Data

   Normalisasi data merupakan proses penyesuaian nilai numerik agar berada dalam rentang tertentu, umumnya antara 0 hingga 1. Tujuannya adalah menyamakan skala data agar algoritma pembelajaran mesin beroperasi secara lebih optimal. Penelitian ini, menggunakan metode *Min-Max Scaling* untuk menyesuaikan nilai data sehingga lebih seragam.

3. Pembagian Data

   Setelah data diproses melalui tahapan *preprocessing* yang mencakup data *cleaning* dan *transformasi*, langkah selanjutnya adalah memisahkan data menjadi dua bagian utama: *training* dan *testing*. Tujuan pembagian ini untuk melatih model menggunakan sebagian data, sementara sebagian lainnya digunakan untuk menguji performa model.

E. Pemilihan Fitur

Pemilihan fitur bertujuan menyederhanakan data dan meningkatkan kinerja model. Korelasi antar fitur serta dengan *variabel* target dianalisis menggunakan *heatmap.* Melalui proses ini, fitur-fitur yang kurang berkontribusi atau memiliki hubungan yang terlalu kuat dengan fitur lain diputuskan untuk tidak digunakan, guna menghindari *overfitting* dan mendukung kinerja model yang lebih optimal.

F. *Implementasi* algoritma KNN dan *Naïve Bayes*

1. Algoritma KNN

   K-Nearest Neighbor (KNN) sebuah metode dalam *data mining* yang mengklasifikasikan data baru dengan membandingkannya dengan data yang sudah yang sudah memiliki label sebelumnya (Musthafa et al., 2023.). Proses klasifikasi dilakukan dengan mencari tetangga terdekat dari data yang akan diprediksi (Puspita Hidayanti, 2020). Langkah-langkah algoritma KNN sebagai berikut (Duwo Jiwo Saputro et al., 2023):

   a. Menyiapkan data balita sebagai data *training* dan data *testing.*
   b. Menentukan nilai k sebagai jumlah tetangga terdekat yang digunakan dalam proses klasifikasi.
   c. Menghitung jarak antara data *testing* dengan setiap data *training* menggunakan rumus *Euclidean Distance*.

$$D(x,y) = \sqrt{\sum_{k=1}^{n}\left(X_{training} - Y_{testing}\right)^2} \quad (1)$$

   d. Mengurutkan data berdasarkan nilai jarak dari yang terkecil ke yang terbesar.
   e. Menentukan klasifikasi data *testing* berdasarkan kelas mayoritas dari k tetangga terdekat.

   Keterangan pada rumus (1):
   n = Jumlah data *training*
   ^{X}*training* = Nilai atribut ke-k dari data *training*
   ^{Y}*testing* = Nilai atribut ke-k dari data *testing*
   k = Indeks fitur dari 1 sampai n.

2. Algoritma *Naïve Bayes*

Naïve Bayes algoritma klasifikasi yang menggunakan *teorema Bayes* untuk menentukan kategori suatu input, dengan menghitung kemungkinan berdasarkan data pelatihan yang ada (Khoiruddin et al., 2023.). Algoritma ini dikenal memiliki akurasi yang tinggi dan kinerja yang cepat, terutama digunakan pada kumpulan data dalam skala besar. (Moch. Rizky Yuliansyah et al., 2022). Langkah-langkah algoritma *Naive Bayes* sebagai berikut:

a. Menyiapkan data yang sudah memiliki label sebagai data *training* dan data *testing*.
b. Menghitung *probabilitas* awal (*a priori*) untuk setiap kelas berdasarkan frekuensi kelas dalam data *training*.
c. Menghitung *probabilitas* kondisional untuk setiap fitur dalam dataset, yaitu peluang munculnya fitur tertentu dalam kelas tertentu.
d. Menghitung *probabilitas posterior* untuk setiap kelas berdasarkan *Teorema Bayes*, dengan menggabungkan *probabilitas a priori* dan *probabilitas* kondisional, menggunakan rumus:

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)} \quad (2)$$

e. Menentukan kelas yang diprediksi untuk data *testing* berdasarkan kelas dengan *probabilitas posterior* tertinggi.

Keterangan pada rumus (2) :
A = data dengan kelas yang belum diketahui
B = data yang dijadikan sebagai hipotesis.
P(A|B) = kemungkinan hipotesis A mengacu pada kondisi B
P(A) = kemungkinan hipotesis A
P(B|A) = kemungkinan B berdasarkan kondisi pada A
P(B) = kemungkinan dari B

G. Evaluasi Model

Pada penelitian ini, evaluasi model dilakukan dengan memanfaatkan *Confusion Matrix*. *Confusion Matrix* merupakan metode evaluasi untuk menilai performa model klasifikasi dengan membandingkan hasil prediksi dengan nilai yang sebenarnya (Normawati & Prayogi, 2021). Berikut tabel yang menggambarkan struktur *confusion matrix*, yang terdiri dari empat komponen utama, seperti yang ditunjukkan pada Tabel 2.

Tabel 2. *Confusion Matrix Multiclass*

		Kelas Prediksi		
		Normal (1)	*Stunting* (2)	Severely *Stunting* (3)
aktual	Normal (1)	TP (1)	FN	FN
	Stunting (2)	FN	TP (2)	FN
	Severely Stunting(3)	FN	FN	TP (3)

Keterangan:
TP (*True Positive*) = Kondisi dimana prediksi yang dihasilkan sesuai dengan nilai aktual

FN (*False Negative*) = kondisi dimana suatu prediksi menunjukkan hasil negatif, namun nilai aktual yang sebenarnya adalah positif.

Accuracy rasio prediksi benar terhadap keseluruhan data (Ridhovan et al., 2022).

$$Accuracy = \frac{\text{TP}}{(\text{TP+FN})}\ x\ 100\% \quad (3)$$

Precision mengukur tingkat ketepatan prediksi model terhadap suatu kelas (Hikmayanti Handayani et al., 2023).

$$Precision = \frac{\text{TP}}{(\text{TP+FP})}\ x\ 100\% \quad (4)$$

Recall mengukur kemampuan model dalam mendeteksi data positif secara benar dari seluruh data yang memang termasuk kategori positif (Christopher et al., 2022).

$$Recall = \frac{\text{TP}}{(\text{TP+FN})}\ x\ 100\% \quad (5)$$

F1-Score merupakan metrik yang mengombinasikan nilai *precision* dan *recall* untuk menghasilkan satu ukuran yang seimbang (Maulidah et al., 2020).

$$F1 - Score\ = 2\ \text{x}\ \frac{Precision\ x\ Recall}{Precision+Recall}\ x\ 100\% \quad (6)$$

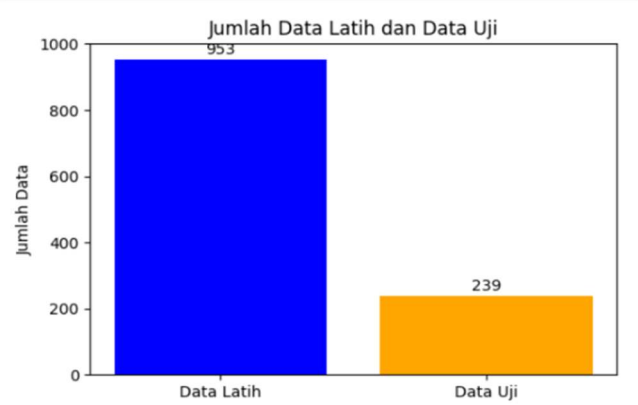
3. Hasil dan Pembahasan

Dataset pada penelitian ini awalnya terdiri dari 1.195 data balita, setelah dilakukan proses *data cleaning* ditemukan tiga data duplikat sehingga jumlah data yang digunakan berkurang menjadi 1.192 data. Selanjutnya dilakukan *transformasi* data agar setiap fitur siap digunakan dalam tahap klasifikasi. Tahap ini, *variabel* jenis kelamin dikonversi menjadi data numerik, yaitu 0 untuk laki-laki dan 1 untuk perempuan. Sementara itu, *variabel* Status *stunting* dikodekan menjadi tiga kategori 0 untuk normal, 1 untuk *stunting*, dan 2 untuk *severely stunting*. Setelah itu dilakukan normalisasi data untuk menyamakan skala antar fitur numerik pada *variabel* usia (bulan) dan tinggi badan menggunakan metode *Min-Max Scaling*. Tujuan dari normalisasi ini adalah untuk meningkatkan kinerja model klasifikasi dengan mengurangi perbedaan skala antar atribut. Hasil dari proses *preprocessing* dan *transformasi* data ditampilkan Tabel 8.

Tabel 8. Hasil preprocessing dan *transformasi* data

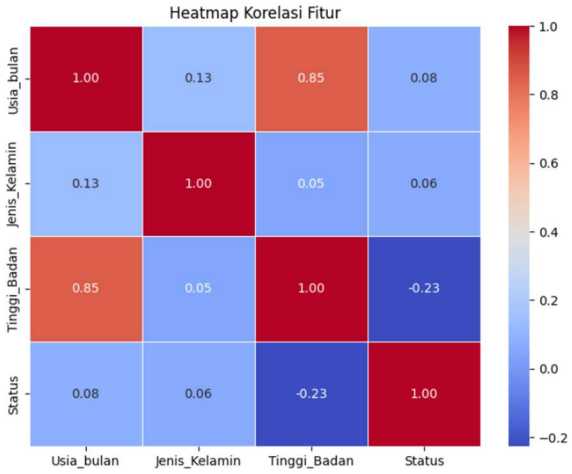
No	Nama	Usia Bulan	Jenis Kelamin	Tinggi Badan	Status
1	Mizan	0.25	0	0.23875	2
2	Rafa	0.25	0	0.32525	1
3	Raizi	0.2291	0	0.46366	0
4	Elvano	0.2083	0	0.39446	0
5	Nadira	0.5	1	0.43528	1
….	….	…	…	…	…
1192	Hamdan	0.333	0	0.33044	2

Setelah proses *transformasi* selesai, dataset dibagi menjadi dua bagian, data latih 80% dengan jumlah 953 data dan data uji 20% dengan jumlah 239 data . Pembagian ini bertujuan untuk melatih dan mengevaluasi performa model klasifikasi. Gambar 2 menampilkan grafik yang menunjukkan jumlah data latih dan data uji setelah proses pembagian dataset.



Gambar 2. Grafik pembagian dataset

Setelah data dibagi dan siap digunakan, dilakukan analisis korelasi guna memahami hubungan antara *variabel* dalam dataset. Tahap ini menggunakan teknik visualisasi berupa *heatmap* yang menampilkan matriks korelasi antar *variabel*. Visualisasi *heatmap* tersebut ditampilkan pada Gambar 3.



Gambar 3. *Heatmap* korelasi

Berdasarkan Gambar 3, *heatmap* korelasi menunjukkan atribut tinggi badan memiliki korelasi signifikan terhadap status (-0.23), sehingga dipilih sebagai fitur utama. Usia juga berkorelasi tinggi dengan tinggi badan (0.85), mencerminkan pertumbuhan alami balita. Meskipun usia dan jenis kelamin memiliki korelasi rendah terhadap status, keduanya tetap digunakan karena dapat meningkatkan akurasi prediksi. Oleh karena itu, penelitian ini menggunakan fitur tinggi badan, usia, dan jenis kelamin untuk klasifikasi status *stunting* pada balita.

Kemudian selanjutnya *implementasi* algoritma *K-Nearest Neighbor* (KNN) dan *Naïve Bayes* untuk mengklasifikasikan status *stunting* pada balita. *Implementasi* algoritma KNN menentukan kelas data berdasarkan mayoritas dari *k* tetangga terdekatnya. Sebelum digunakan, perlu ditentukan nilai *k* yang optimal agar model dapat bekerja secara maksimal. Untuk itu, dilakukan pengujian menggunakan teknik *cross-validation* guna memperoleh nilai *k* terbaik. Proses ini diawali dengan menentukan nilai *k* ganjil, yaitu dari k = 1 hingga k = 19, dengan tujuan menghindari kemungkinan hasil seri dalam proses pemilihan

mayoritas kelas. Setiap nilai *k* diuji menggunakan metode *k-fold cross-validation*, di mana data balita dibagi menjadi beberapa bagian (*fold*), dan secara bergantian digunakan sebagai data latih dan data uji. Model KNN dilatih dan dievaluasi pada setiap *fold*, kemudian dihitung rata-rata akurasi dari seluruh *fold* untuk mengetahui performa model pada masing-masing nilai *k*. Setelah seluruh nilai *k* diuji, dipilih nilai dengan rata-rata akurasi tertinggi sebagai *k* optimal. Hasil dari pengujian ini ditunjukkan pada Gambar 4.

```
Optimal K berdasarkan Cross Validation (K Ganjil): 1
    K   Akurasi
0   1  0.994660
1   3  0.989734
2   5  0.981937
3   7  0.978655
4   9  0.975371
5  11  0.973317
6  13  0.972495
7  15  0.970856
8  17  0.968807
9  19  0.969215
```

Gambar 4. Hasil nilai K menggunakan *cros- validation*

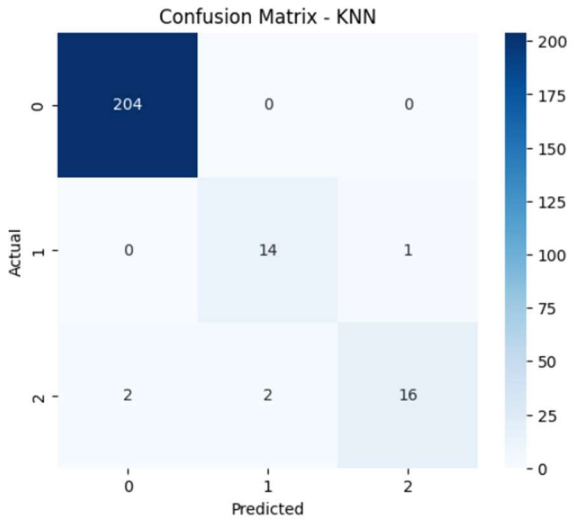
Berdasarkan Gambar 4, nilai *k* terbaik diperoleh pada *k* = 1 dengan tingkat akurasi tertinggi sebesar 0.994660. Hasil ini menunjukkan bahwa pemilihan *k* = 1 memberikan performa terbaik dalam klasifikasi status *stunting*. Sementara itu, *implementasi* algoritma *Naïve Bayes* tidak memerlukan penyesuaian *parameter*, sehingga dapat langsung diterapkan. Algoritma ini melakukan klasifikasi dengan menghitung *probabilitas* masing-masing kelas berdasarkan distribusi fitur, dengan asumsi bahwa setiap fitur bersifat *independen*. *Probabilitas* untuk masing-masing kelas dihitung dan dibandingkan, lalu kelas dengan *probabilitas* tertinggi dipilih sebagai hasil klasifikasi. Hasil perhitungan *probabilitas* untuk beberapa sampel data ditampilkan pada Tabel 9.

Tabel 9. Hasil Perhitungan *Probabilitas*

No	P (Normal)	P (*Stunting*)	P (*Severely Stunting*)	Kelas Terpilih
1	0.39	0.23	0.36	0
2	0.52	0.09	0.37	0
3	0.13	0.66	0.20	1
4	0.79	0.00	0.19	0
5	0.32	0.32	0.34	2
…	…	…	…	…
239	0.68	0.03	0.28	0

Pada tabel 9 , setiap baris menunjukkan hasil klasifikasi untuk satu sampel data. Kolom P (Normal), P (*Stunting*), dan P (*Severely Stunting*) menunjukkan *probabilitas* bahwa data tersebut termasuk ke dalam masing-masing kategori status *stunting*. Kolom Kelas Terpilih menampilkan kelas yang dipilih berdasarkan *probabilitas* tertinggi.

Selanjutnya langkah terakhir dalam tahap *implementasi* yaitu evaluasi model, yang dilakukan menggunakan *confusion matrix*. Evaluasi ini bertujuan mengukur kinerja klasifikasi dari algoritma KNN dan *Naïve Bayes*. Gambar 5 berikut menampilkan *confusion matrix* dari model KNN.



Gambar 5. *Confusion matrix* KNN

Berdasarkan *confusion matrix* pada gambar 5 :
Kelas 0 (Normal): Semua 204 sampel diklasifikasikan dengan benar.
Kelas 1 (*Stunting*): 14 dari 15 sampel diklasifikasikan dengan benar, 1 sampel salah sebagai kelas 2.
Kelas 2 (*Severely Stunting*): 16 dari 20 sampel diklasifikasikan dengan benar, 2 sampel salah sebagai kelas 0, 2 sampel lainnya diklasifikasikan sebagai 1.

Perhitungan manual akurasi sebagai berikut :

$$Accuracy\ KNN = \frac{TP_0 + TP_1 + TP_2}{Total\ sampel} = \frac{204 + 14 + 16}{239} = 0.9790\ atau\ 97.90\%$$

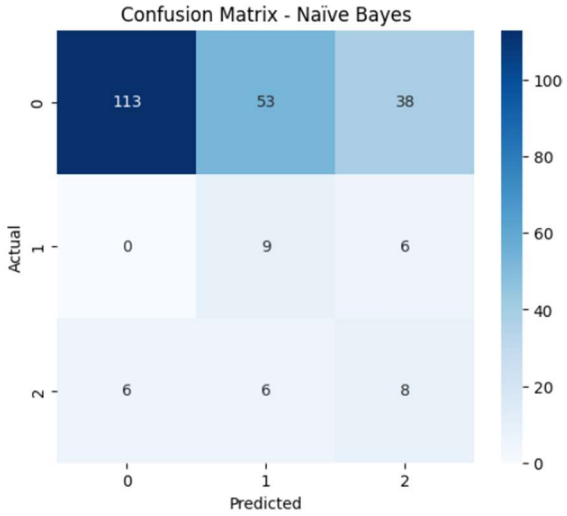
Setelah mendapatkan *confusion matrix* KNN pada gambar 5, selanjutnya melakukan perhitungan untuk menentukan nilai *precision*, *recall dan f1-score* pada model klasifikasi KNN. Hasil perhitungan dapat dilihat pada tabel 10 berikut:

Tabel 10. Evaluasi Kinerja Model KNN

	Precision	*Recall*	*F1-Score*	*Support*
Normal (0)	0.99	1.00	0.99	204
Stunting (1)	0.87	0.93	0.90	15
Severely Stunting (2)	0.94	0.80	0.86	20
Accuracy		97.90		239

Berdasarkan Tabel 10, model KNN menunjukkan hasil *precision* yang tinggi, yaitu 99% kelas Normal, serta 87% kelas *Stunting* dan 94% *Severely Stunting*. Ini menunjukkan bahwa model sangat tepat dalam mengklasifikasikan data, terutama pada kelas Normal. Dari sisi *recall*, model mencapai 100% pada kelas Normal, 93% pada kelas *Stunting* dan 80% pada kelas *Severely Stunting*. Ini menunjukkan meskipun model sangat baik dalam mengenali data pada kelas Normal, masih terdapat beberapa data pada kelas *Stunting* dan *Severely Stunting* yang tidak berhasil dikenali dengan benar. Nilai *F1-score* yang diperoleh masing-masing sebesar 99%, 90%, dan 96%. Secara keseluruhan, penerapan model KNN

menghasilkan akurasi sebesar 97,90%, yang menunjukkan bahwa model ini mampu melakukan klasifikasi data dengan cukup baik secara umum.



Gambar 6. *Confusion matrix Naive Bayes*

Berdasarkan *confusion matrix* pada gambar 6 :
Kelas 0 (Normal) : 113 dari 204 sampel diklasifikasikan dengan benar, 53 sampel diklasifikasikan sebagai 1, 38 sampel diklasifikasikan sebagai 2.
Kelas 1 (*Stunting*) : 9 dari 15 sampel diklasifikasikan dengan benar, 6 sampel diklasifikasi sebagai 2.
Kelas 2 (*Severely Stunting*) : 8 dari 20 sampel diklasifikasikan dengan benar, 6 sampel diklasifikasi sebagai kelas 1, 6 sampel lain diklasifikasikan sebagai 0.

Perhitungan manual akurasi *Naive Bayes* sebagai berikut :

$$Accuracy\ B = \frac{TP_0 + TP_1 + TP_2}{Total\ sampel} = \frac{113 + 9 + 8}{239} = 0.54.39\ atau\ 54.39\%$$

Setelah mengetahui hasil *confusion matrix Naive Bayes* pada gambar 6, dilakukan perhitungan untuk menentukan nilai *precision*, *recall dan f1-score* pada model klasifikasi *Naïve Bayes*. Hasil perhitungan dapat dilihat pada Tabel 11 berikut:

Tabel 11. Evaluasi Kinerja Model *Naïve Bayes*

	Precision	*Recall*	*F1-Score*	*Support*
Normal (0)	0.94	0.55	0.69	204
Stunting (1)	0.13	0.60	0.21	20
Severely Stunting (2)	0.15	0.40	0.22	15
Accuracy		54.39		239

Berdasarkan Tabel 11, model *Naïve Bayes* menghasilkan *precision* tertinggi pada kelas Normal sebesar 94%, sementara kelas *Stunting* dan *Severely Stunting* masing-masing sebesar 13% dan 15%. Hal ini menunjukkan bahwa model cenderung hanya mampu mengklasifikasikan data dengan benar pada kelas Normal, namun kurang akurat dalam mengidentifikasi dua kelas lainnya. Untuk *recall*, model mencatat hasil sebesar 55% kelas

Normal, 60% *Stunting*, dan 40% *Severely Stunting*. Meskipun *recall* pada kelas *Stunting* relatif lebih tinggi, nilai *precision* yang rendah menunjukkan banyak kesalahan dalam prediksi kelas tersebut. Nilai *F1-score* yang diperoleh masing-masing sebesar 69%, 21%, dan 22%. Secara keseluruhan, model *Naïve Bayes* hanya mencapai akurasi sebesar 54.39%, yang menunjukkan bahwa kinerjanya dalam mengklasifikasikan data masih tergolong rendah. Berdasarkan hasil evaluasi performa model, diketahui bahwa KNN memperoleh akurasi sebesar 97,90%, sementara *Naïve Bayes* hanya mencapai akurasi sebesar 54.39%.

Hasil penelitian ini menunjukkan bahwa algoritma *K-Nearest Neighbor* (KNN) memiliki performa yang sangat baik dalam mengklasifikasikan status *stunting* balita, dengan akurasi mencapai 97,90%. Peningkatan akurasi pada penelitian ini kemungkinan besar disebabkan oleh penggunaan data yang lebih besar 1.192 data, penerapan teknik *cross-validation* untuk menentukan nilai k terbaik, serta proses *preprocessing* yang lebih menyeluruh.

Sebaliknya, algoritma *Naïve Bayes* dalam penelitian ini hanya menghasilkan akurasi sebesar 54,39%, yang berbeda *signifikan* dengan beberapa studi sebelumnya. Perbedaan ini dapat dijelaskan oleh perbedaan *karakteristik* data. Penelitian sebelumnya umumnya menggunakan klasifikasi dua kelas *stunting* dan tidak *stunting*, sedangkan penelitian ini menggunakan tiga kelas normal, *stunting*, *severely stunting* yang lebih *kompleks*, terutama bagi algoritma seperti *Naïve Bayes* yang mengasumsikan *independensi* antar fitur.

4. Kesimpulan dan Saran

Penelitian ini menerapkan algoritma KNN dan *Naïve Bayes* untuk mengklasifikasikan *stunting* pada balita di Desa Pasirjengkol berdasarkan data *antropometri,* yaitu usia, jenis kelamin, dan tinggi badan. Kedua algoritma tersebut melalui tahapan prapemrosesan data, *transformasi* data, pembagian data menjadi data latih dan data uji, pemilihan fitur serta pelatihan dan pengujian model, dan keduanya dapat diterapkan dengan baik pada dataset yang digunakan.

Perbandingan akurasi dan kinerja antara kedua algoritma menunjukkan bahwa KNN dengan K=1 lebih unggul memperoleh akurasi 97.90% sementara *Naïve Bayes* hanya mencapai 54.39%. Perbedaan akurasi tersebut karena perbedaan cara kerja kedua algoritma. *Naïve Bayes* bekerja dengan asumsi bahwa setiap fitur dalam dataset *independen*, sehingga kurang optimal dalam mengidentifikasi hubungan antar fitur yang mungkin saling berkaitan. Di sisi lain, algoritma KNN tidak bergantung pada asumsi tersebut, melainkan melakukan klasifikasi berdasarkan kemiripan antar data, sehingga lebih fleksibel dalam mengenali pola-pola kompleks dalam data. Dengan pendekatan tersebut, KNN menunjukkan kinerja yang lebih baik dalam mengklasifikasikan status *stunting* pada balita di Desa Pasirjengkol.

Berdasarkan temuan penelitian ini, terdapat beberapa rekomendasi untuk penelitian selanjutnya. Pertama, meskipun KNN memberikan hasil yang lebih baik, disarankan untuk mengeksplorasi algoritma lain seperti SVM atau *Random Forest* guna membandingkan kinerjanya. Kedua, teknik prapemrosesan data seperti normalisasi dapat ditingkatkan untuk memperoleh hasil model yang lebih optimal. Selain itu, pengujian menggunakan dataset yang lebih besar dan bervariasi juga sangat penting untuk memastikan kemampuan generalisasi model yang baik.

Daftar Pustaka

Azizah, S. N., & Fatah, Z. (2024). Implementasi Metode K-Nearest Neighbor (K-NN) Pada Klasifikasi Stunting Balita. *Gudang Jurnal Multidisiplin Ilmu*, *2*, 282–288. https://doi.org/10.59435/gjmi.v2i10.1000

Duwo Jiwo Saputro, A., Darmawan, A., & Nurina Sari, B. (2023). Klasifikasi persentase kemiskinan di Jawa Barat menggunakan data mining algoritma K-Nearest Neighbor (KNN). *Jurnal Mahasiswa Teknik Informatika*, *7*(4). https://doi.org/10.36040/jati.v7i4.7178

Hikmayanti Handayani, H., Ahmad Baihaqi, K., & Buana Perjuangan Karawang, U. (2023). Implementasi Algoritma Logistic Regression Untuk Klasifikasi Penyakit Stroke. *Syntax: Jurnal Informatika*, *12*(01).

Khoiruddin, Y., Fauzi, A., & Siregar, A. M. (2023). Analisis Sentimen Gojek Indonesia Pada Twitter Menggunakan Algoritme Naïve Bayes Dan Support Vector Machine. *Jurnal Ilmiah Komputer*, *19*. https://doi.org/10.35889/progresif.v19i1.1173

Maulidah, M., Gata, W., Aulianita, R., Agustyaningrum, C. I., Studi, P., Komputer, I., & Mandiri, N. (2020). Algoritma Klasifikasi Decision Tree Untuk Rekomendasi Buku Berdasarkan Kategori Buku. *JURNAL ILMIAH EKONOMI DAN BISNIS*, *13*(2), 89–96. https://doi.org/10.51903/e-bisnis.v13i2.251

Moch. Rizky Yuliansyah, B, M., & Franz, A. (2022). Perbandingan Metode K-Nearest Neighbors dan Naïve Bayes Classifier Pada Klasifikasi Status Gizi Balita di Puskesmas Muara Jawa Kota Samarinda. *Adopsi Teknologi Dan Sistem Informasi (ATASI)*, *1*(1), 08–20. https://doi.org/10.30872/atasi.v1i1.25

Musthafa, K., 1✉, R., Witanti, W., & Yuniarti, R. (2023). Penerapan Algoritma K-Nearest Neighbor (KNN) Dengan Fitur Relief-F Dalam Penentuan Status Stunting. *INNOVATIVE: Journal Of Social Science Research*, *3*, 3555–3568.

Noer Azzahra, F., Rohana, T., Ratna Juwita, A., Karawang, P., Jl HSRonggo Waluyo, K., Timur, T., & Barat, J. (2024). Penerapan Metode Naive Bayes Dalam Klasifikasi Spam SMS Menggunakan Fitur Teks Untuk Mengatasi Ancaman Pada Pengguna. *Journal of Information System Research (JOSH)*, *5*(3), 880. https://doi.org/10.47065/josh.v5i3.5070

Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *Jurnal Sains Komputer & Informatika (J-SAKTI*, *5*(2), 697–711. https://doi.org/10.30645/j-sakti.v5i2.369

Pebrianti, S. W., Astuti, R., & Basysyar, F. M. (2024). Penerapan algoritma K-Nearest Neighbor dalam klasifikasi status stunting balita di Desa Bojongemas. *Jurnal Mahasiswa Teknik Informatika*, *8*(2). https://doi.org/10.29408/jit.v3i2.2279

Puspita Hidayanti, W. (2020). Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Efektivitas Penjualan Vape (Rokok Elektrik) pada "Lombok Vape On." *Jurnal Informatika Dan Teknologi*, *3*(2). https://doi.org/10.29408/jit.v3i2.2279

Rahayu, I. P., Fauzi, A., & Indra, J. (2022). Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Naive Bayes Dan Support Vector Machine. *Jurnal Sistem Komputer Dan Informatika (JSON)*, *4*(2), 296. https://doi.org/10.30865/json.v4i2.5381

Ridhovan, A., Suharso, A., Fakultas,), Komputer, I., Karawang, S., Ronggo Waluyo, J. H., Timur, T., & Karawang, K. (2022). Penerapan metode Residual Network (ResNet)

dalam klasifikasi penyakit pada daun gandum. *JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, *07*, 58–65. https://doi.org/10.29100/jipi.v7i1.2410

Titimeidara, M. Y., & Hadikurniawati, W. (n.d.). Monica Yoshe Titimeidara Implementasi Metode Naive Bayes Implementasi Metode Naive Bayes Classifier Untuk Klasifikasi Status Gizi Stunting Pada Balita. In *Tri Lomba Juang* (Vol. 50241, Issue 1).