PREDIKSI STATUS GIZI BALITA DI KABUPATEN BANTAENG MENGGUNAKAN K-NEAREST NEIGHBOUR (KNN)

Hafsah. HS[1], Ayu Bella Fauziah[2], Resky Akmal[3], Dedy Qadry Puryadi Putra[4]
[1234]Universitas Prof. Dr. H. M. Arifin Sallatang
[1]hafsah.hsulaeman@gmail.com,[2] AyuBellaFauziah@gmail.com,[3]reskyakmal04@gmail.com,
[4]Puryadi.putra01@gmail.com

ABSTRAK

Masalah gizi balita masih menjadi perhatian penting dalam pembangunan kesehatan masyarakat, termasuk di Kabupaten Bantaeng. Status gizi yang tidak seimbang dapat menimbulkan risiko terhadap pertumbuhan fisik maupun perkembangan kognitif balita. Upaya intervensi lapangan perlu didukung oleh pemanfaatan teknologi, salah satunya melalui algoritma *K-Nearest Neighbour* (KNN) untuk memprediksi status gizi balita. Penelitian ini bertujuan untuk membangun model klasifikasi status gizi balita dengan KNN serta mengevaluasi performanya berdasarkan variasi nilai K dan proporsi pembagian data latih dan data uji. Data penelitian bersumber dari Dinas Kesehatan Kabupaten Bantaeng tahun 2024 sebanyak 1.200 record dengan atribut usia, jenis kelamin, tinggi badan, berat badan, dan status gizi. Tahapan penelitian meliputi *preprocessing* (pembersihan data, penanganan *missing value*, normalisasi *Z-Score* dan *min-max*), pembagian data dengan tiga komposisi (90:10, 70:30, dan 50:50), pemodelan dengan KNN menggunakan nilai K=3, K=5, dan K=7, serta evaluasi menggunakan akurasi, presisi, *recall*, dan *F1-score*. Hasil pengujian menunjukkan bahwa KNN mampu menghasilkan akurasi tinggi dengan kisaran 92%–94%. Nilai K=3 memberikan hasil terbaik, khususnya pada komposisi 70:30 dan 50:50 yang mencapai akurasi 94%. Dengan demikian, algoritma KNN terbukti efektif dalam memprediksi status gizi balita dan berpotensi mendukung tenaga kesehatan dalam melakukan deteksi dini masalah gizi.

Kata kunci: gizi balita, klasifikasi, *K-Nearest Neighbour*, prediksi, data mining

1. PENDAHULUAN

Gizi merupakan komponen penting yang terkandung dalam makanan yang meliputi karbohidrat, protein, vitamin, mineral, lemak dan air yang diperlukan oleh tubuh untuk pertumbuhan, perkembangan, dan pemeliharaan yang dimanfaatkan secara langsung oleh tubuh untuk memperbaiki jaringan tubuh.[1] Gizi sangat diperlukan oleh setiap orang khususnya pada masa balita karena gizi berfungsi untuk mempertinggi derajat kesehatan. Status gizi anak balita adalah cerminan ukuran terpenuhinya kebutuhan gizi anak balita yang didapatkan dari asupan dan penggunaan zat gizi oleh tubuh. [2]

Gizi yang baik sangat penting untuk mendukung pertumbuhan dan perkembangan optimal, terutama pada masa balita dan anak-anak. Asupan nutrisi yang seimbang membantu membangun sistem kekebalan tubuh yang kuat, mencegah berbagai penyakit dan meningkatkan kemampuan belajar serta konsentrasi. Status gizi adalah keadaan tubuh sebagai akibat dari pemakaian, penyerapan, dan penggunaan makanan [3]. Status gizi yang optimal memastikan tubuh mendapatkan semua nutrisi yang dibutuhkan untuk berbagai fungsi vital, seperti pertumbuhan dan perkembangan, memperbaiki jaringan, dan menjaga sistem kekebalan tubuh yang kuat.[4]

Status gizi balita merupakan salah satu indikator penting dalam pembangunan kesehatan masyarakat. Balita yang mengalami masalah gizi berisiko mengalami gangguan pertumbuhan fisik maupun perkembangan kognitif. Hal ini menjadi perhatian serius pemerintah pusat maupun daerah, termasuk di Kabupaten Bantaeng [5]

Berdasarkan laporan Dinas Kesehatan Bantaeng pada Juli 2025, angka stunting di Kabupaten Bantaeng tercatat sebesar 5,39%. Meski angka ini menunjukkan penurunan berkat Program Aksi Stop Stunting (ASS) yang telah dijalankan sejak 2022, permasalahan gizi balita tetap menjadi isu yang membutuhkan penanganan berkelanjutan. Pemerintah daerah melalui Tenaga Pendamping Gizi Desa (TPGD) terus melakukan pendampingan kepada ibu hamil, bayi, dan balita, serta memberikan edukasi mengenai pentingnya gizi seimbang.

Salah satu cara memenuhi kebutuhan gizi tersebut adalah dengan cara mengkonsumsi gizi yang cukup dan sesuai untuk tubuh. Selain itu orang tua juga perlu mengetahui tingkat kesehatan si balita yang dapat dilihat dari status gizinya. Penilaian status gizi balita dapat ditentukan melalui pengukuran tubuh manusia yang dikenal dengan istilah antropometri [6]. Standar acuan status gizi balita adalah Berat Badan menurut Umur (BB/U) yang menggambarkan berat badan relative anak dengan umur, Berat Badan menurut Tinggi Badan (BB/TB) yang menggambarkan apakah berat badan anak sesuai dengan pertumbuhan tinggi badannya dan Tinggi Badan menurut Umur (TB/U) menggambarkan pertumbuhan tinggi badan anak berdasarkan umurnya.[7]. Namun, upaya intervensi lapangan juga perlu didukung oleh pemanfaatan teknologi dalam hal pengolahan data kesehatan. Salah satu pendekatan yang dapat dilakukan adalah menggunakan algoritma *K-Nearest Neighbors (KNN)* untuk memprediksi status gizi balita [8].

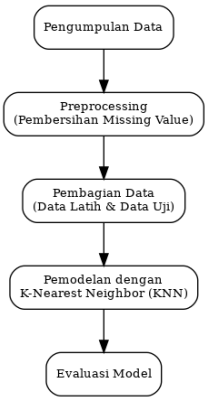
Selain relevansi dengan program pemerintah dalam upaya penanggulangan stunting, penelitian ini juga merupakan pengembangan dari penelitian sebelumnya. Penelitian sebelumnya menggunakan *algoritma K-Nearest Neighbor (KNN)* untuk mengklasifikasikan status gizi balita ke dalam empat kategori (buruk, kurang, normal, lebih). Eksperimen dilakukan dengan satu variasi pembagian data, yaitu 90:10 untuk training dan testing, serta hanya menggunakan K = 5 sebagai parameter KNN. Selain itu, dataset yang digunakan relatif terbatas, yaitu 170 data balita**.** Hasil penelitian menunjukkan bahwa normalisasi *Z-Score* memberikan kinerja terbaik dengan akurasi 95%, presisi 98,08%**,** dan recall 93,75%**,** lebih tinggi dibandingkan tanpa normalisasi (80%) maupun normalisasi *min-max* (85%) [9].

Kebaruan penelitian ini terletak pada perluasan dari penelitian sebelumnya yang hanya menggunakan satu variasi komposisi data (90:10) dan satu nilai parameter K (K=5) dengan jumlah data yang terbatas. Dalam penelitian ini, dilakukan pengujian dengan tiga variasi komposisi data training–testing, yaitu 90:10, 70:30, dan 50:50, serta menggunakan tiga variasi nilai K (K=3, K=5, K=7) untuk mengetahui sensitivitas performa algoritma KNN terhadap perubahan parameter. Selain itu, jumlah data yang digunakan lebih banyak sehingga hasil penelitian lebih representatif dan reliabel. Penelitian ini juga fokus pada dua teknik normalisasi terbaik dari penelitian sebelumnya, yaitu *Z-Score* dan *min-max*, untuk membandingkan kinerja keduanya pada dataset yang lebih besar dan konfigurasi data yang lebih bervariasi. Dengan demikian, penelitian ini berkontribusi dalam memberikan analisis yang lebih komprehensif mengenai penerapan KNN pada klasifikasi status gizi balita.

Oleh karena itu, maka perlu menerapkan metode klasifikasi untuk menentukan status gizi anak karena metode ini memungkinkan kita untuk mengorganisir data gizi anak-anak ke dalam kategori yang lebih mudah dipahami dan digunakan oleh para professional kesehatan, peneliti dan pembuat kebijakan, hal ini membantu dalam menyajikan informasi yang lebih terstruktur dan mudah diinterpretasikan [10]. Pada sistem ini pengguna akan menginputkan data balita yang telah diketahui. Alasan menggunakan metode KNN karena metode ini merupakan salah satu pendekatan yang digunakan dalam pengklasifikasian secara mudah dan efisien.

2. METODE PENELITIAN

Gambar 2.1 menunjukkan tahapan penelitian secara sistematis, dimulai dari pengumpulan data hingga evaluasi model. Langkah-langkan ini dirancang untuk memastikan setiap proses dilakukan secara terstruktursehingga tujuan penelitian dapat tercapai dengan hasil yang valid dan bermanfaat



Gambar 2.1 Tahapan penelitian

2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini bersumber dari Dinas Kesehatan Kabupaten Bantaeng Tahun 2024. Data yang diperoleh meliputi variabel-variabel yang berhubungan dengan kondisi gizi balita seperti berat badan, tinggi badan, umur, serta status gizi yang ditetapkan berdasarkan standar kesehatan yang berlaku. Jumlah data yang digunakan dalam penelitian ini sebanyak 1200 *record.*

2.2 *Preprocessing*

Tahap *preprocessing* dilakukan untuk menyiapkan data agar siap diproses pada pemodelan. Pada penelitian ini, *preprocessing* meliputi pembersihan data dari duplikasi atau ketidakkonsitenan, penanganan *missing value,* normalisasi menggunakan metode *Z-score,* dan *min-max* agar variabel memiliki skala yang sebanding. Dengan tahapan ini, data menjadi lebih bersih, konsisten dan sesuai untuk digunakan dalam proses klasifikasi KNN.

2.3 Pembagian Data Latih dan Data Uji

Setelah melalui tahap *preprocessing,* dataset dibagi menjadi data latih (*training data*) dan data uji *(testing data)*. Data latih digunakan untuk membangun model klasifikasi menggunakan algoritma *K-Nearesr Neighbour (KNN,* sedangkan data uji dipakai untuk mengukur kinerja model terhadap data yang belum pernah dilatih sebelumnya. Dalam penelitian ini, tiga scenario pembagian data, yaitu:

Tabel 2.1 Pembagian Data

Variasi Pembagian Dataset	Data Latih	Data Uji
1	90%	10%
2	70%	30%
3	50%	50%

Penggunaan beberapa komposisi pembagian data ini, bertujuan untuk membandingkan pengaruh jumlah data latih dan data uji terhadap performa model KNN. Sehingga dapat diketahui komposisi mana yang menghasilkan tingkat akurasi terbaik.

2.4 Pemodelan *K-Nearest Neighbour (KNN)*

Pada tahap ini, peneliti menggunakan model KNN untuk melakukan klasifikasi status gizi balita. KNN bekerja dengan prinsip mencari sejumlah tetangga terdekat dari suatu data uji berdasarkan jarak kemiripan (*Euclidien distance)* kemudian menentukan kelas dari data uji tersebut berdasarkan mayoritas kelas tetangga terdekatnya. Dalam penelitian ini, KNN diuji dengan tiga nilai parameter K, yaitu K=3, K=5, K=7 untuk melihat pengaruh jumlah tetangga terhadap performa klasifikasi. Pemilihan nilai K ini didasarkan pada pertimbangan agar model tidak terlalu sensitive terhadap *noise* (K yang terlalu kecil) dan tidak kehilangan akurasi akibat oengaruh tetangga yang terlalu banyak (K yang terlalu besar). K adalah jumlah data tetangga terdekat, tidak ada standar baku dalam menentukan jumlah K.

Model yang dipilih dalam penelitian ini adalah KNN dengan menggunakan 2 bentuk pemodelan yaitu data yang dinormalisasi menggunakan *Z-Score* dan data yang dinormalisasi menggunakan *min-max*

Selanjutnya adalah menghitung jarak dari data uji dengan setiao baris data pelatihan. Perhitungannya menggunakan formula jarak. Formula jarak yang digunakan dalam penelitian ini adalah jarak *euclidien distance* dengan formula sebagai berikut:

$$d\,(x,y) = \sqrt{\sum_{i=1}^{m} (x_1 - y_1)^2} \qquad (1)$$

Dimana:

$d\,(x,y)$ = Jarak antara data pengujian baru dan data pelatihan

x_1 = Data latih

y_1 = Data uji

2.5 Evaluasi Model

Evaluasi yang dilakukan dalam penelitian ini adalah *Confussion matriks* untuk menilai kinerja model KNN dalam mengklasifikasikan status gizi balita. *Confussion matriks* digunakan untuk melihat jumlah prediksi yang benar maupun salah, yang kemudian menjadi dasar perhitungan matriks evaluasi seperti: Akurasi: Mengukur tingkat ketepatan keseluruhan prediksi

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \qquad (1)$$

Presisi: Menunjukkan ketepatan model dalam memprediksi kelas positif

$$Presisi = \frac{TP}{TP + FP} \quad (2)$$

Recall: Mengukur kemampuan model dalam menentukan seluruh data positif

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

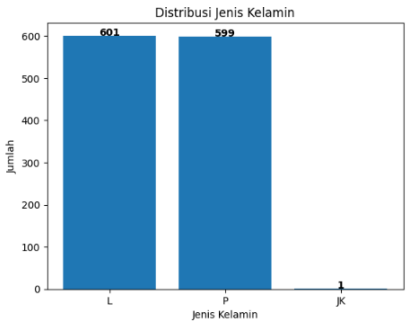
F1-Score: rata-rata harmonis antara *precission* dan *recall,* digunakan saat data tidak seimbang

$$F1 - Score = \frac{2\ X\ presisi\ X\ recall}{presisi\ +\ recall} \quad (4)$$

3. HASIL DAN PEMBAHASAN

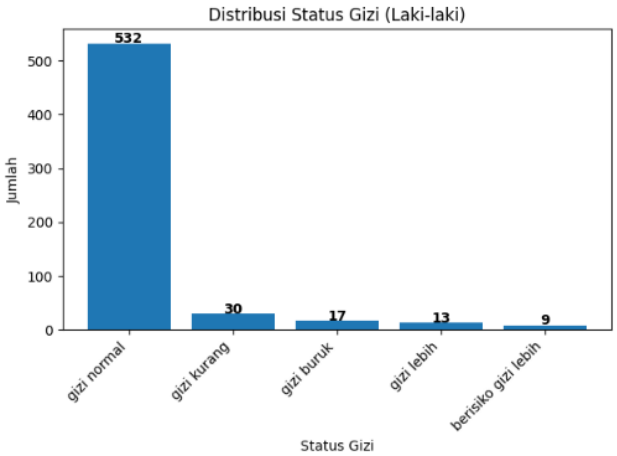
3.1 *Data Preparation*

Berdasarkan data yang diperoleh setelah melalui tahap pembersihan data dan penanganan *missing value,* jumlah sampel yang digunakan dalam penelitian ini adalah sebanyak 1.200 *record* balita dengan parameter Usia, Jenis Kelamin, tinggi badan, berat badan, dan status gizi. Distribusi berdasarkan jenis kelamin menunjukkan bahwa 601 anak balita laki-laki dan 599 anak balita perempuan.



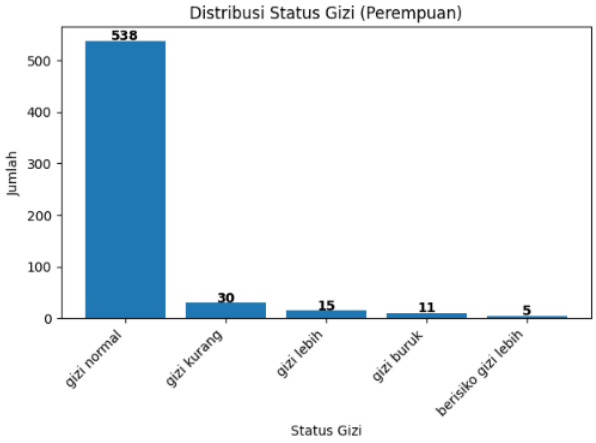
Gambar 3.1 distribusi jenis kelamin balita

Sebagian besar balita laki-laki memiliki status gizi normal yaitu 532, meskipun masih terdapat sebagian kecil yang mengalami masalah gizi, yaitu gizi kurang sebanyak 30, gizi buruk sebanyak 17, gizi lebih sebanyak 13, dan berisiko gizi lebih sebanyak 9.



Gambar 3.2 distribusi status gizi balita laki-laki

Mayoritas balita perempuan atau sebanyak 538 anak berada pada kategori gizi normal, sementara sebagian kecil lainnya mengalami masalah gizi, yaitu gizi kurang sebanyak 30, gizi buruk sebanyak 11, gizi lebih sebanyak 15, dan berisiko gizi lebih sebanyak 5.



Gambar 3.3 distribusi status gizi balita perempuan

Setelah dilakukan pembersihan data untuk memeriksa adanya duplikasi serta penanganan *missing value* dengan menghapus *record* yang tidak memenuhi kelengkapan parameter, data yang tersedia kemudian dinormalisasi agar setiap atribut berada pada skala yang sebanding. Teknik normalisasi yang dilakukan pada penelitian adalah *Z-score* dan normalisasi *min-max*.

Proses untuk normalisasi menggunakan *Z-score* dapat dilihat pada formula berikut ini:

$$x_{stand} = \frac{x - mean(x)}{stdev\,(x)} \quad (4)$$

Sebelum melakukan normalisasi *Z-Score,* terlebih dahulu menentukan nilai rata *(mean)* setiap parameter Tabel berikut adalah nilai *mean* dan standar deviasi setiap parameter:

tabel 3.1 nilai *mean* dan standar deviasi

Parameter	*mean*	Standar deviasi
Jenis Kelamin	0,499	0,50
Usia (bulan)	48,57	9,21
Berat Badan (kg)	14,77	3,14
Tinggi Badan (cm)	96,94	12,22

Setelah menentukan nilai rata-rata dan standar deviasi setiap paramater, langkah selanjutnya adalah menghitung nilai normalisasi setiap parameter dari setiap *record* dengan menggunakan formula *Z-Score.* Sebagai contoh, jika terdapat data balita dengan karakteristik jenis kelamin perempuan (P = 1), usia 48 bulan, berat badan 15 kg, dan tinggi badan 67 cm maka prosesnya sebagai berikut

$$x_1 = \frac{1 - 0{,}499}{0{,}50} = 1{,}002$$

Proses diatas dilakukan pada setiap parameter untuk semua baris data latih. Tabel berikut adalah hasil normalisasi *Z-score* pada data latih pertama.

tabel 3.2 hasil normalisasi *Z-score*

Parameter	Nilai asli	*Z-score*
Jenis Kelamin	1	1,002
Usia (bulan)	48	–0,062
Berat Badan (kg)	15	0,073
Tinggi Badan (cm)	67	–2,450

Proses untuk normalisai menggunakan *min-max* formula yang digunakan adalah sebagai berikut:

$$x_{norm} = \frac{x' - \min(x)}{\max(x) - \min(x)} \quad (6)$$

Sebelum melakukan normalisasi *min-max,* terlebih dahulu menentukan minimal *(min)* dan maksimal *(max)* setiap parameter Tabel berikut adalah nilai *min* dan *max* setiap parameter:

tabel 3.3 nilai *min* dan *max*

Parameter	*min*	*max*
Jenis Kelamin	0	1
Usia (bulan)	4	60
Berat Badan (kg)	5	22.3
Tinggi Badan (cm)	49	119.2

Setelah menentukan nilai *min* dan *max* setiap parameter, langkah selanjutnya adalah menghitung nilai normalisasi setiap parameter dari setiap *record* sampai akhir dengan menggunakan formula *min-max.* Sebagai contoh, jika terdapat data balita dengan karakteristik jenis kelamin perempuan (P = 1), usia 48 bulan, berat badan 15 kg, dan tinggi badan 67 cm maka prosesnya sebagai berikut

$$x_{norm} = \frac{1-0}{1-1} = 1$$

Proses diatas dilakukan pada setiap parameter untuk semua baris data latih. Tabel berikut adalah hasil normalisasi *min-max* pada data latih pertama.

tabel 3.4 hasil normalisasi *min-max*

Parameter	Nilai asli	*Min-max*
Jenis Kelamin	1	1
Usia (bulan)	48	0.7857
Berat Badan (kg)	15	0.5780
Tinggi Badan (cm)	67	0.2564

3.2 Komposisi *Split* data

Data yang telah melewati tahap *preprocessing* selanjutnya dilakukan pembagian atau *split* data. Komposisi *split* data disajikan dalam tabel 3.1

tabel 3.1 Komposisi *Split* data

Komposisi %	Data Latih	Data Uji
90:10	1080	120
70:30	840	360
50:50	600	600

Pembagian data menjadi tiga variasi komposisi (90:10, 70:30, dan 50:50) dilakukan untuk mengevaluasi kinerja model pada kondisi yang berbeda. Komposisi 90:10 memberikan lebih banyak data latih sehingga model berpotensi belajar lebih optimal, meskipun data uji relatif sedikit. Komposisi 70:30 menawarkan keseimbangan antara data latih dan data uji sehingga kemampuan generalisasi model dapat diuji dengan lebih baik. Adapun komposisi 50:50 digunakan untuk menilai *robustnes* model ketika data latih lebih sedikit, sehingga dapat dilihat sejauh mana model tetap mampu bekerja secara efektif. Dengan variasi ini, penelitian ini dapat menilai stabilitas, konsistensi, dan performa model secara lebih komprehensif.

3.3 Pengujian Normalisasi *Z-score*

Pengujian dilakukan dengan menggunakan data yang telah dinormalisasi menggunakan metode *Z-score.* Proses normalisasi ini bertujuan untuk menyetarakan skala antar variabel sehingga tidak ada atribut dengan nilai yang lebih besar mendominasi perhitungan jarak pada algoritma *K-Nearest Neighbour (KNN)*. Dengan normalisasi *Z-Score,* setiap atribut memiliki nilai rata-rata *(mean)* nol dan standar deviasi satu, sehingga distribusi data menjadi lebih seimbang untuk proses klasifikasi.

Agoritma *K-Nearest Neighbour (KNN)* digunakan sebagai metode klasifikasi dengan tiga variasi nilai parameter K=3, K=5, dan K=7. Pemilihan variasi nilai K dimaksudkan untuk mengetahui pengaruh jumlah tetangga terdekat terhadap performa klasifikasi. Untuk memperoleh hasil yang lebih komprehensif

Tabel 3.6 Pengujian Normalisasi *Z-score*

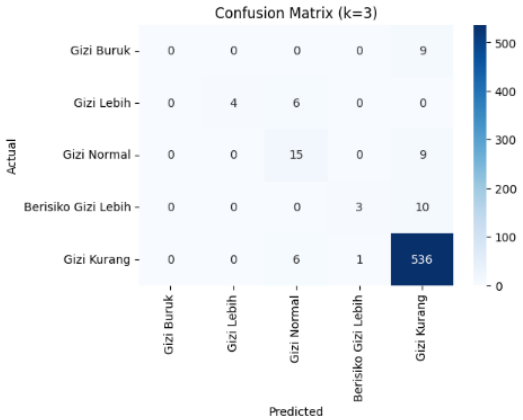
Komposisi Data	K	Akurasi (%)
90:10	3	93%
90:10	5	92%
90:10	7	93%

70:30	3	93%
70:30	5	93%
70:30	7	93%
50:50	3	94%
50:50	5	93%
50:50	7	93%

Hasil pengujian *model K-Nearest Neighbor (KNN)* dengan variasi nilai K dan komposisi data latih serta data uji menunjukkan bahwa tingkat akurasi yang diperoleh relatif stabil pada setiap skenario. Pada komposisi data 90:10, akurasi yang dicapai berada pada rentang 92% hingga 93%, dengan akurasi tertinggi yaitu 93% pada K=3 dan K=7. Pada komposisi 70:30, performa model lebih konsisten dengan akurasi 93% pada semua variasi nilai K, baik K=3, K=5, maupun K=7. Sementara itu, pada komposisi 50:50, akurasi yang diperoleh justru sedikit lebih tinggi, yaitu 94% pada K=3, sedangkan pada K=5 dan K=7 akurasi tetap stabil pada 93%.

Secara umum, hasil ini menunjukkan bahwa perubahan nilai K tidak memberikan perbedaan yang signifikan terhadap kinerja model. Nilai K yang relatif kecil (K=3) cenderung memberikan hasil yang lebih baik karena keputusan klasifikasi lebih dipengaruhi oleh tetangga terdekat yang paling relevan dengan data uji. Konsistensi akurasi pada berbagai proporsi pembagian data juga menunjukkan bahwa model KNN memiliki ketahanan terhadap variasi jumlah data latih dan data uji. Bahkan pada proporsi data uji yang lebih besar (50%), model tetap mampu mempertahankan akurasi tinggi, yang mengindikasikan bahwa pola data sudah cukup representatif untuk dilatih dan diuji.

Confusion matrix ditampilkan pada hasil pengujian dengan komposisi data 50:50 dan nilai K=3. Skenario ini dipilih karena memberikan akurasi tertinggi yaitu 94%, sehingga dapat merepresentasikan performa model secara optimal dibandingkan skenario lainnya.



Gambar 3.4 *Confussion matriks* K=3 Komposisi 50:50 normalisasi *Z-Score*

Selain akurasi, evaluasi performa model *K-Nearest Neighbor (KNN)* juga dilakukan dengan melihat nilai presisi, recall, dan F1-score pada masing-masing kelas. Presisi digunakan untuk mengukur sejauh mana prediksi yang dihasilkan benar sesuai dengan label aktual, recall menunjukkan sejauh mana model mampu mengenali data pada setiap kelas secara benar, sedangkan F1-score merupakan rata-rata harmonis dari presisi dan recall yang memberikan gambaran keseimbangan antara keduanya. Nilai presisi, recall, dan F1-score ini disajikan dalam tabel berikut untuk memberikan gambaran yang lebih detail mengenai kinerja model pada setiap kategori klasifikasi.

Tabel 3.7 *Classification Report* K=3 Komposisi 50:50 normalisasi *Z-Score*

Classification Report:	*precision*	*recall*	*f1-score*	*support*
Berisiko Gizi Lebih	0.00	0.00	0.00	9
Gizi Buruk	1.00	0.40	0.57	10
Gizi Kurang	0.56	0.62	0.59	24
Gizi Lebih	0.75	0.23	0.35	13
Gizi Normal	0.95	0.99	0.97	543

3.4 Pengujian Normalisasi *min-max*

Pengujian yang dinormalisasi menggunakan metode *Min-Max* dengan menerapkan algoritma K-Nearest Neighbor (KNN). Variasi nilai K yang digunakan adalah K=3, K=5, dan K=7, dengan tiga skenario pembagian data latih dan data uji, yaitu 90:10, 70:30, dan 50:50**.** Pengujian ini bertujuan untuk mengetahui

pengaruh variasi nilai K serta komposisi data terhadap tingkat akurasi model setelah dilakukan normalisasi Min-Max.

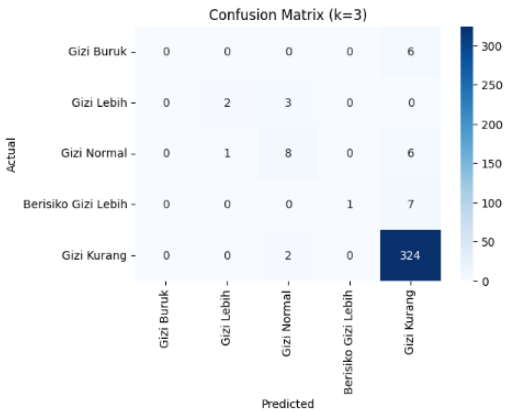
Tabel 3.8 Pengujian Normalisasi *min-max*

Komposisi Data	K	Akurasi (%)
90:10	3	93%
90:10	5	93%
90:10	7	93%
70:30	3	94%
70:30	5	93%
70:30	7	93%
50:50	3	94%
50:50	5	92%
50:50	7	92%

Hasil pengujian model *K-Nearest Neighbor (KNN)* yang dilakukan pada data yang telah dinormalisasi menggunakan metode *Min-Max* menunjukkan tingkat akurasi yang relatif tinggi pada setiap skenario. Pada komposisi data 90:10, akurasi yang diperoleh konsisten sebesar 93% untuk semua variasi nilai K, baik K=3, K=5, maupun K=7. Selanjutnya, pada komposisi 70:30, akurasi tertinggi dicapai pada K=3 dengan 94%**,** sedangkan pada K=5 dan K=7 akurasi sedikit lebih rendah namun tetap stabil di angka 93%. Sementara itu, pada komposisi 50:50**,** akurasi tertinggi kembali diperoleh pada K=3 dengan 94%**,** sedangkan pada K=5 dan K=7 akurasi menurun menjadi 92%.

Secara keseluruhan, hasil ini memperlihatkan bahwa nilai K=3 cenderung memberikan akurasi yang lebih tinggi dibandingkan nilai K yang lebih besar. Konsistensi akurasi pada komposisi data 90:10 dan 70:30 menunjukkan kestabilan model, sementara sedikit penurunan akurasi pada komposisi 50:50 dengan K=5 dan K=7 mengindikasikan bahwa peningkatan nilai K tidak selalu meningkatkan performa model. Dengan demikian, pemilihan nilai K yang lebih kecil, khususnya K=3, lebih optimal dalam klasifikasi pada data penelitian ini.

Confusion matrix pada penelitian ini ditampilkan dengan menggunakan hasil pengujian pada komposisi data 70:30 dengan nilai K=3. Pemilihan skenario ini didasarkan pada pertimbangan bahwa komposisi 70:30 memberikan keseimbangan antara jumlah data latih dan data uji, sehingga hasil yang diperoleh lebih representatif. Selain itu, pada skenario ini diperoleh tingkat akurasi tertinggi yaitu 94%, sehingga confusion matrix yang ditampilkan dapat menggambarkan performa model secara optimal. Sebagai perbandingan, pada komposisi 50:50 dengan K=3 juga diperoleh akurasi yang sama, namun pembagian data tersebut memberikan proporsi data latih yang relatif lebih sedikit sehingga berpotensi memengaruhi kestabilan model. Dengan demikian, penggunaan skenario 70:30 dengan K=3 dianggap paling sesuai untuk dianalisis lebih lanjut melalui *confusion matrix.*



Gambar 3.5 *Confussion matriks* K=3 Komposisi 70:30 normalisasi *min-max*

Selain akurasi, kinerja model dievaluasi menggunakan presisi, recall, dan F1-score pada masing-masing kelas. Tabel berikut menyajikan hasil pengukuran tersebut untuk memberikan gambaran yang lebih detail mengenai performa model.

Tabel 3.9 *Classification Report* K=3 Komposisi 70:30 normalisasi *min-max*

Classification Report:	*precision*	*recall*	*f1-score*	*support*
Berisiko Gizi Lebih	0.00	0.00	0.00	6
Gizi Buruk	0.67	0.40	0.50	5

Gizi Kurang	0.62	0.53	0.57	15
Gizi Lebih	1.00	0.12	0.22	8
Gizi Normal	0.94	0.99	0.97	326

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Berdasarkan hasil pengujian yang dilakukan, dapat disimpulkan bahwa algoritma *K-Nearest Neighbor (KNN)* mampu menghasilkan tingkat akurasi yang tinggi dan relatif stabil pada berbagai komposisi data latih dan uji. Nilai akurasi yang diperoleh berada pada kisaran 92% hingga 94%, dengan hasil terbaik dicapai pada nilai K=3, khususnya pada komposisi data 70:30 dan 50:50 yang sama-sama mencapai akurasi 94%. Temuan ini menunjukkan bahwa nilai K yang lebih kecil cenderung lebih optimal dibandingkan dengan nilai K yang lebih besar, serta proporsi pembagian data tidak memberikan pengaruh yang signifikan terhadap kinerja model. Bahkan pada komposisi 50:50, di mana jumlah data latih lebih sedikit, model tetap mampu mempertahankan performa yang baik.

4.2 Saran

Hasil penelitian menunjukkan bahwa algoritma *K-Nearest Neighbor (KNN)* mampu memprediksi status gizi balita dengan akurasi tinggi dan stabil, sehingga dapat membantu tenaga kesehatan di posyandu atau puskesmas mengidentifikasi balita berisiko gizi kurang atau gizi lebih secara cepat dan tepat. Berdasarkan temuan ini, disarankan untuk menggunakan nilai *K = 3*, namun penelitian berikutnya dapat mengeksplorasi variasi nilai K yang lebih luas dan menerapkan metode validasi silang *(cross-validation)* agar hasil lebih konsisten dan dapat digeneralisasi. Penelitian lanjutan juga dapat mempertimbangkan penerapan algoritma lain seperti *Deep Learning*, untuk membandingkan performa dan melihat apakah teknik lain dapat meningkatkan akurasi serta robustnes model pada dataset gizi balita.

DAFTAR PUSTAKA

[1] S. Anraeni, E. R. Melani, and H. Herman, "Ripeness Identification of Chayote Fruits using HSI and LBP Feature Extraction with KNN Classification," *Ilk. J. Ilm.*, vol. 14, no. 2, pp. 150–159, 2022, doi: 10.33096/ilkom.v14i2.1153.150-159.

[2] D. Fitrianingsih, M. Bettiza, and A. Uperiati, "Klasifikasi Status Gizi Pada Pertumbuhan Balita Menggunakan K-Nearest Neighbor (Knn)," *Student Online J.*, vol. 2, no. 1, pp. 106–111, 2021.

[3] L. Nurcholifah, D. Hartanti, and S. Sundari, "Klasifikasi Penentuan Gizi Balita dengan Algoritma K-Nearest Neighbor (Studi Kasus: Puskesmas Bringin)," *Smart Comp Jurnalnya Orang Pint. Komput.*, vol. 14, no. 2, pp. 296–307, 2025, doi: 10.30591/smartcomp.v14i2.7244.

[4] A. Yudhana, I. Riadi, and M. R. Djou, "Determining eligible villages for mobile services using k- NN algorithm," *Ilk. J. Ilm.*, vol. 15, no. 1, pp. 11–20, 2023.

[5] P. A. K. K. D. M. Putri, A. A. O. Lely, and L. G. Evayanti, "Hubungan antara Status Gizi dengan Perkembangan Kognitif pada Anak Usia 6-24 Bulan," *AMJ (Aesculapius Med. Journal)*, vol. 1, no. 1, pp. 1–7, 2021, [Online]. Available: https://www.ejournal.warmadewa.ac.id/index.php/amj/article/view/4003

[6] Dewi Marfuah, Siti Sarifah, Siti Khusnul Khotimah, and Dhinda Kusuma Hatifah, "Pengukuran Antropometri dan Penentuan Status Gizi Balita di Posyandu Balita Bina Sejahtera Kadipiro Banjarsari Surakarta," *ALKHIDMAH J. Pengabdi. dan Kemitraan Masy.*, vol. 2, no. 3, pp. 138–149, 2024, doi: 10.59246/alkhidmah.v2i3.983.

[7] A. L. M. Emma Anastya Puriastuti1*, Fresthy Astrika Yunita2, Kanthi Suratih3, Cahyaning Setyo Hutomo4, "Skrining status gizi balita melalui pengukuran berat badan menurut umur di posyandu kelurahan mojo kota surakarta," *J. Pengabdi. Masy. Nusant.*, vol. 3, pp. 1–7, 2024.

[8] D. A. Ferliandini and S. Risnanto, "Prosiding Seminar Sosial Politik, Bisnis, Akuntansi dan Teknik (SoBAT) ke-5 Bandung, 28 Oktober 2023 APLIKASI PREDIKSI STATUS GIZI BALITA BERBASIS WEB MENGGUNAKAN METODE K-NEAREST NEIGHBOR," pp. 622–631, 2023.

[9] N. Azmi, H. Hazriani, Y. Yuyun, and H. HS, "Klasifikasi Status Gizi Balita Menggunakan Algoritma K-Nearest Neighbor (KNN)," *Pros. SISFOTEK*, pp. 2–7, 2023, [Online]. Available: http://www.seminar.iaii.or.id/index.php/SISFOTEK/article/view/396%0Ahttp://www.seminar.iaii.or.id/index.php/SISFOTEK/article/download/396/328

[10] T. Prasetiya, I. Ali, C. L. Rohmat, and O. Nurdiawan, "Klasifikasi Status Stunting Balita Di Desa Slangit Menggunakan Metode K-Nearest Neighbor," *INFORMATICS Educ. Prof. J. Informatics*, vol. 5, no. 1, p. 93, 2020, doi: 10.51211/itbi.v5i1.1431.