

Implementasi Algoritma K-Nearest Neighbors (KNN) untuk Prediksi Gizi Buruk

Dian Hasna Ramadhani^{1*}
Jumadi²
Gitarja Sandi³

^{1,2,3}Teknik Informatika, UIN Sunan Gunung Djati Bandung, Jalan A.H Nasution No. 105, Cipadung,
Cibiru, Kota Bandung, Jawa Barat 40614, Indonesia
¹1207050026@uinsgd.ci.id, ²jumadi@uinsgd.ac.id, ³sandi@uinsgd.ac.id

***Penulis Korespondensi:**
Dian Hasna Ramadhani
1207050026@uinsgd.ac.id

Abstrak

Gizi buruk menjadi masalah serius di negara-negara berkembang, disebabkan oleh kekurangan asupan makanan yang mengandung zat-zat esensial seperti protein dan energi. Implementasi algoritma machine learning dapat memberikan diagnosis yang akurat terhadap kondisi kesehatan gizi buruk pada balita, memfasilitasi deteksi dini dan intervensi yang tepat. Tujuan penelitian ini untuk mengetahui kinerja algoritma K-Nearest Neighbors (KNN) dalam memprediksi gizi buruk berdasarkan ciri klinis yang dimiliki oleh balita. Data yang digunakan adalah ciri-ciri klinis dari gizi buruk yang bersumber dari seorang nutrisionis. Dari hasil penelitian, nilai akurasi yang paling optimal dalam memprediksi gizi buruk adalah 87%. Dengan dataset yang ada dapat dibuktikan bahwa algoritma K-Nearest Neighbors (KNN) mampu mengklasifikasikan gizi buruk kedalam 2 kondisi, yaitu marasmus dan kwashiorkor.

Kata Kunci: gizi buruk; K-Nearest Neighbors; kwashiorkor; marasmus; pembelajaran mesin

Abstract

Malnutrition is a serious problem in developing countries, caused by a lack of food intake containing essential substances such as protein and energy. The implementation of machine learning algorithms can provide an accurate diagnosis of malnutrition health conditions in toddlers, facilitating early detection and appropriate interventions. The purpose of this study is to determine the performance of the K-Nearest Neighbors (KNN) algorithm in predicting malnutrition based on clinical characteristics possessed by toddlers. The data used are clinical characteristics of malnutrition sourced from a nutritionist. From the research results, the most optimal accuracy value in predicting malnutrition is 87%. With the existing dataset, it can be proven that the K-Nearest Neighbors (KNN) algorithm is able to classify malnutrition into 2 conditions, namely marasmus and kwashiorkor.

Keywords: K-Nearest Neighbors; kwashiorkor; machine learning; malnutrition; marasmus.

1. Pendahuluan

Menurut data Survei Status Gizi Indonesia (SSGI) tahun 2022, terdapat 7,7% balita yang mengalami wasting atau gizi buruk, angka ini meningkat 0,6% dari tahun 2021 [1]. Kelompok usia yang paling berisiko mengalami gizi buruk adalah balita. Kebutuhan nutrisi balita sangat penting untuk perkembangan otak mereka [2].

Gizi buruk merupakan masalah kesehatan pada anak-anak dan balita yang timbul akibat kekurangan asupan makanan yang mengandung nutrisi penting seperti protein dan energi. Gizi buruk meningkatkan risiko infeksi, menghambat pertumbuhan dan perkembangan anak, serta memperburuk kesehatan anak saat dewasa [3].

Terjadinya gizi buruk di Indonesia merupakan akibat dari konsumsi makanan yang tidak merata sehingga mengganggu perkembangan fisik dan psikologis anak. Hal ini menyebabkan menurunnya kesehatan sumber daya manusia (SDM) secara umum. Gizi buruk saat ini menjadi isu nasional yang perlu dikaji secara berkala karena menjadi penyebab utama kematian di negara-negara berkembang ketika kebutuhan gizi tidak terpenuhi [1].

Gizi buruk berkembang secara bertahap. Pada fase awal, asupan protein tidak perlu diberikan untuk memenuhi kebutuhan energi mereka; sebagai gantinya, asupan kalori yang harus diberikan. Asupan protein dapat dimulai setelah jumlah kalori yang dibutuhkan tercapai. Pemberian protein dapat dimulai dengan dosis yang sederhana dan meningkatkannya secara bertahap [4]. Terdapat beberapa jenis gizi buruk, diantaranya kwashiorkor, marasmus, dan marasmus-kwashiorkor [5].

Jenis penyakit gizi buruk yang dikenal sebagai kwashiorkor ditandai dengan pembengkakan di seluruh tubuh, terutama di bagian belakang kaki. Tanda dan gejala tambahan termasuk wajah yang membengkak dan bulat, perut yang membesar, otot-otot yang menyusut, mata yang terkulai, dan rambut yang tipis atau berwarna kemerahan. Marasmus adalah jenis malnutrisi di mana anak-anak mengalami dua fase malnutrisi parah yang berbeda yang disebabkan oleh asupan makanan yang tidak mencukupi. Kwashiorkor-Marasmus menggabungkan gejala-gejala kwashiorkor dan marasmus [6].

Menganalisis status gizi anak dapat dilakukan salah satunya dengan mengukurnya menggunakan data antropometri [1]. Ciri klinis gizi buruk juga dapat digunakan untuk mendiagnosa kondisi tersebut, namun untuk mendapatkan hasil diagnosa dan pengobatan yang terbaik, dibutuhkan seorang pakar yang memiliki keahlian di bidang gizi. Untuk mengatasi masalah gizi ini, *machine learning* dapat diterapkan dalam membantu tenaga medis memprediksi gizi buruk dengan akurat. Hasilnya, proses analisis dan pengambilan keputusan akan menghasilkan hasil yang lebih tepat dan efektif, sehingga dapat dilakukan penanganan dini [4].

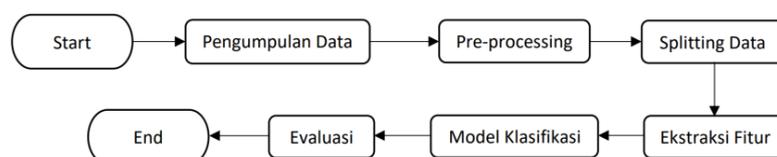
Telah banyak penelitian mengenai deteksi atau diagnosis penyakit menggunakan machine learning seperti penelitian yang dilakukan oleh Ferliandini and Risnanto (2023) menggunakan metode K-Nearest Neighbor (KNN) dalam memprediksi status gizi balita dengan tingkat akurasi 91,94%. [7] Penelitian lain dilakukan oleh Lonang et al., (2023) menggunakan metode Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbors (KNN), dan Support Vector Machine (SVM) dalam mendeteksi stunting, dari metode yang digunakan, metode K-Nearest Neighbors (KNN) memiliki tingkat akurasi tertinggi yaitu 94,85%. [8] Serta pada penelitian Viransyah and Sugiarto (2023) menggunakan metode Naïve Bayes dalam mendiagnosa penyakit gizi buruk menghasilkan tingkat akurasi 80%. [9]

Berdasarkan penelitian [7][8] metode K-Nearest Neighbors (KNN) memiliki akurasi yang baik dalam mendiagnosis penyakit. Dalam penelitian Amien et al., (2023) disebutkan bahwa KNN memiliki kelebihan, yaitu proses pelatihan yang sangat cepat, sederhana, dan efektif terutama pada dataset yang besar. [6] Pada penelitian sebelumnya yang menggunakan algoritma K-Nearest Neighbors banyak yang menggunakan data antropometri, sedangkan pada penelitian ini data yang digunakan adalah data ciri klinis.

Tujuan penelitian ini adalah untuk melakukan deteksi gizi buruk marasmus dan kwashiorkor dilihat dari ciri klinis atau gejala menggunakan algoritma K-Nearest Neighbors (KNN) serta mengetahui kinerja algoritma K-Nearest Neighbors (KNN) terhadap dataset yang digunakan.

2. Metode Penelitian

Berdasarkan masalah diatas, penelitian ini akan dilakukan berdasarkan tahapan-tahapan yang digambarkan pada Gambar 1 berikut.



Gambar 1. Alur Penelitian

Dari gambar diatas, tahap pertama adalah pengumpulan data, proses pengumpulan data dilakukan dengan wawancara bersama petugas dari lembaga. Selain itu, data juga diambil dari sumber-sumber terpercaya di internet dan telah divalidasi oleh ahli gizi untuk memastikan akurasi dan relevansinya. Ciri klinis kondisi gizi buruk pada balita menjadi fokus utama, dan data diperoleh dengan mengumpulkan informasi mengenai ciri-ciri tersebut. Dari hasil wawancara diperoleh 50 data ciri klinis. Setelah memperoleh data, dilakukan *preprocessing* pada data sehingga model machine learning dapat menanganinya dengan lebih efisien. Tahapan *preprocessing* yang dilakukan antara lain menghapus tanda baca, tokenisasi, *stop word*, *stemming*, dan augmentasi data. Tahap selanjutnya yaitu *splitting data* dimana data dibagi menjadi dua subset yaitu data latih dan data uji. Data latih digunakan untuk melatih model algoritma dan data uji digunakan untuk melihat kinerja model yang telah dibuat [10]. Menggunakan library Scikit-Learn pada Python untuk secara otomatis membagi dataset menjadi data uji dan data latih.

Tahap selanjutnya ekstraksi fitur yaitu Proses pengambilan fitur yang dapat menggambarkan informasi yang diperlukan dikenal sebagai ekstraksi fitur. Dalam penelitian ini, TF-IDF digunakan untuk proses ekstraksi fitur [11]. TF adalah frekuensi kata yang ditemukan dalam dokumen, sedangkan IDF adalah nilai invers dari dokumen yang mengandung kata tersebut.[12] Dengan menggunakan TF-IDF, proses pembobotan pada kata terdiri dari perhitungan nilai TF (frekuensi kata) dan IDF (frekuensi dokumen balik). Pembobotan TF memiliki beberapa formula, pada penelitian ini menggunakan raw term, yaitu dihitung frekuensi kemunculan term pada setiap dokumen, dan IDF (frekuensi dokumen balik) menghitung frekuensi sebuah term yang tersebar luas di seluruh dokumen. Berikut formula untuk menghitung IDF :

$$IDF_t = \log \log \left(\frac{N}{df_j} \right) + 1 \quad (1)$$

N adalah total jumlah dokumen dalam data latih, dan df_j adalah total dokumen yang didalamnya terdapat term.

TF-IDF membantu menghitung bobot dari setiap kata yang dihasilkan dari preprocessing. Pembobotan kata mengubah kata menjadi angka yang dapat menampilkan data secara keseluruhan. Hasil dari proses ini adalah matriks yang terdiri dari kolom dan baris, dengan data dalam bentuk baris dan fitur dalam bentuk kolom [11]. Untuk nilai TF-IDF, rumus umumnya adalah menggabungkan rumus perhitungan TF dengan rumus IDF seperti berikut:

$$w_{tj} = TF \times IDF_t \quad (2)$$

Di mana w_{tj} adalah bobot term (t) terhadap dokumen (j). Sedangkan TF adalah jumlah kemunculan term (t) dalam dokumen (d). IDF_j adalah hasil dari perhitungan seberapa luas term tersebar dalam kumpulan dokumen yang relevan.

Setelah melakukan ekstraksi fitur pada data, selanjutnya membuat model klasifikasi menggunakan algoritma K-Nearest Neighbor (KNN) Prinsip dasar K-Nearest Neighbor adalah mencari jarak terdekat antara data yang akan dievaluasi dengan k tetangga dalam dataset pelatihan. Ini merupakan salah satu algoritma pembelajaran mesin yang bersifat non parametrik, artinya tidak membuat asumsi tentang distribusi data [13].

Menggunakan dataset sebagai data latih, menghitung nilai k, memasukkan nilai data baru untuk pengujian, menghitung nilai Euclidean distance, mengurutkan data dari yang terkecil hingga terbesar, mengambil data sesuai dengan angka nilai k, dan mengambil prediksi yang keluar teratas sebagai hasil prediksi merupakan langkah-langkah yang dilakukan dalam implementasi algoritma KNN pada penelitian ini. Dataset yang diberi label x adalah data latih untuk model dan y adalah data uji atau data baru.[14] Pada penelitian ini metode perhitungan jarak menggunakan metode Euclidean Distance dengan rumus seperti persamaan (3).

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3)$$

Dalam persamaan diatas, x_i merupakan data latih, y_i merupakan data uji dan n adalah banyaknya data yang digunakan dalam penelitian. Jarak antara data latih dengan data masukan baru (pengujian) dapat dihitung dengan menggunakan hasil komputasi Euclidean Distance. Data tersebut kemudian diurutkan dari terkecil hingga terbesar, dan jarak ke tetangga terdekat dibatasi dengan menggunakan nilai k . Nilai mayoritas kemudian dipilih sebagai hasil ramalan dari jumlah maksimum data yang dikumpulkan dengan menggunakan nilai k [15].

Setelah proses klasifikasi selesai, langkah terakhir adalah evaluasi kinerja sistem. Ini dilakukan untuk mengevaluasi kinerja sistem yang akan dibangun. Penelitian ini menggunakan confusion matrix sebagai metode yang digunakan untuk menguji kinerja sistem klasifikasi, yang akan membandingkan data hasil klasifikasi dengan data sebenarnya. Dengan bantuan hasil confusion matrix, perhitungan akurasi, precision, dan recall juga dapat dilakukan [16].

Akurasi didefinisikan sebagai perilaku kedekatan antara hasil prediksi dan nilai sebenarnya (aktual). Presisi adalah tingkat ketepatan antara hasil yang dihasilkan oleh sistem dan informasi yang diminta. Tingkat keberhasilan sistem dalam mengumpulkan informasi dikenal sebagai recall. Selain itu, nilai F1 terdiri dari nilai recall dan nilai ketepatan yang telah dihitung sebelumnya [17]. Tabel 13 adalah konsep yang digunakan pada confusion matrix.

Tabel 1. Konsep Confusion Matrix

Nilai Prediksi	Nilai Sebenarnya	
	Positif	Negatif
Positif	TP	FP
Negatif	FN	TN

Tabel confusion matrix di atas menunjukkan bahwa TP adalah data yang diproyeksi positif dan data yang sebenarnya positif; FP adalah data yang diproyeksi positif tetapi data yang sebenarnya negatif; TN adalah data yang diproyeksi negatif tetapi data yang sebenarnya negatif; dan FN adalah data yang diproyeksi negatif tetapi data yang sebenarnya positif [16]. Untuk menghitung akurasi, precision, dan recall menggunakan confusion matrix dapat menggunakan persamaan (4), (5), dan (6).

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (4)$$

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (6)$$

3. Hasil

Dataset yang digunakan pada penelitian ini adalah 50 data ciri klinis dari gizi buruk yang terdiri dari 27 ciri klinis kondisi marasmus dan 23 ciri klinis kondisi kwashiorkor. Wawancara dilakukan dengan nutrisionis dalam sebuah lembaga kesehatan yaitu puskesmas untuk mendapatkan pemahaman mendalam mengenai ciri klinis yang mungkin muncul dan data didapat dari sumber-sumber terpercaya di internet. Dataset yang digunakan seperti pada tabel 1 dan 2.

Tabel 2. Data Kondisi Gizi Buruk

Kondisi	Kode
Marasmus	P1
Kwashiorkor	P2

Pada Tabel 1 merupakan data penyakit yang nantinya akan diprediksi menggunakan data ciri klinis yang terdapat pada Tabel 2.

Tabel 3. Data Ciri Klinis

Ciri Klinis	Kode	Ciri Klinis	Kode
Rakhitis	G1	Koilonikia	G26
Infeksi perut dan intoleran laktosa	G2	Kuku berbentuk sendok	G27
Imunitas rendah	G3	Penurunan badan yang signifikan	G28
Mudah terkena penyakit infeksi	G4	Kehilangan otot	G29
Mudah menangis dan apatis (kesadaran menurun)	G5	Wajah yang bulat dan sembab (Moon Face)	G30
Pertumbuhan terhambat	G6	Penurunan selera makan	G31
Diare kronik	G7	Penurunan Kekebalan otot	G32
Susah buang air kecil	G8	Dermatosis	G33
Kulit tubuh longgar	G9	Kulit berwarna kemerahan	G34
Kepala tampak lebih besar	G10	Kulit bersisik	G35
Wajah terlihat tua	G11	Kulit terkelupas	G36
Kulit kering dan kendur	G12	Kuku rapuh	G37
Rambut kering dan rapuh	G13	Kuku mudah patah	G38
Perut cekung pada bayi	G14	Warna rambut kuning kemerahan	G39
Lesu dan lemah	G15	Gangguan tumbuh kembang	G40
Tensi rendah	G16	Sering lemas	G41
Hipotensi	G17	Pembengkakan	G42
Perut kembung	G18	Tidak memiliki nafsu makan	G43
Dehidrasi	G19	Ruam yang meluas	G44
Sering menangis	G20	Penyembuhan luka yang lambat	G45
Iritabilitas	G21	Nadi lambat	G46
Oedema	G22	Bradikardi	G47
Mata Kering	G23	Suhu rendah	G48
Bitot spot di mata	G24	Kulit Kering	G49
Hipotermia	G25	Bengkak yang ditekan menimbulkan lubang	G50

Tabel 2 diatas merupakan data yang digunakan dalam memprediksi kondisi gizi buruk, setiap kondisi memiliki gejala seperti pada Tabel 3.

Tabel 4. Ciri Klinis Dari Setiap Kondisi

Kondisi	Ciri Klinis
Marasmus	G1, G2, G3, G4, G5, G6, G7, G8, G9, G10, G11, G12, G13, G14, G15, G16, G17, G23, G24, G25, G26, G27, G28, G29, G46, G47, G48,
Kwashiorkor	G18, G19, G20, G21, G22, G30, G31, G32, G33, G34, G35, G36, G37, G38, G39, G40, G41, G42, G43, G44, G45, G49, G50

Tabel 3 merupakan pembagian ciri klinis berdasarkan kondisi yang dialami balita, untuk memprediksi kondisi gizi buruk sesuai ciri pada data tabel 3.

Preprocessing atau pengolahan data adalah proses membersihkan, menstandarisasi, dan menyiapkan data sehingga model machine learning dapat menanganinya dengan lebih efisien [10]. Tahapan preprocessing yang dilakukan antara lain menghapus tanda baca, tokenisasi, stopword, stemming, dan augmentasi data. Pada Tabel 4 merupakan potongan dataset sebelum dan sesudah dilakukan preprocessing.

Tabel 5. Potongan Dataset Setelah Preprocessing

Sebelum	Sesudah
Mudah terkena penyakit infeksi	mudah kena sakit infeksi
Diare kronik	diare kronik
Susah buang air kecil	susah buang air
Warna rambut kuning kemerahan	warna rambut kuning merah
Kulit kering	kulit kering

Setelah melakukan preprocessing data dan data siap dipakai, selanjutnya adalah pembagian data menjadi dua atau lebih subset terpisah dikenal sebagai *splitting data*, dan biasanya dilakukan untuk melatih dan mengevaluasi model pembelajaran mesin. Umumnya dataset dibagi menjadi 2 subset, yaitu data latih dan data uji. Pada penelitian ini, pembagian data dibagi menjadi 4 skenario yaitu komposisi 90:10, 80:20, 70:30, dan 60:40.

Tabel 6. Potongan Data Latih

Ciri Klinis	Kondisi
kulit kelupas	kwashiorkor
diare kronik	marasmus
tensi rendah	marasmus
kulit kering kendur	marasmus
iritabilitas	kwashiorkor
suhu rendah	marasmus
wajah bulat moon face	kwashiorkor
sembuh luka lambat	kwashiorkor

Setelah melakukan pembagian data, selanjutnya proses ekstraksi fitur dengan menggunakan TF-IDF. Hasil dari proses ini adalah matriks yang terdiri dari kolom dan baris, dengan data dalam bentuk baris dan fitur dalam bentuk kolom [11]. Untuk menghitung TF-IDF data latih, setiap term data dihitung pada seluruh data latih, hasil dari TF-IDF ini akan menjadi acuan dalam menghitung TF-IDF data baru. Berikut hasil dari TF-IDF data baru tensi rendah, suhu rendah, dan iritabilitas terhadap data latih.

Tabel 7. Hasil TF-IDF pada Data Baru

Term	TF	IDF	TF-IDF
tensi	1	1,9030	1,9030
rendah	2	1,6020	3,2040
suhu	1	1,9030	1,9030
iritabilitas	1	1,9030	1,9030

Tabel 6 merupakan hasil perhitungan pada proses TF-IDF, data baru dihitung dengan acuan data latih pada Tabel 5 menggunakan persamaan (1) dan (2). Nilai TF-IDF ini nantinya akan digunakan sebagai inputan yang masuk ke model untuk memprediksi kondisi gizi buruk.

Untuk prediksi kondisi dilakukan menggunakan algoritma K-Nearest Neighbors. K-Nearest Neighbor (KNN) mengukur kemiripan data untuk mengukur jarak. Untuk menghitung jarak antar data yang digunakan pada metode KNN ini menggunakan jarak Euclidean [16]. Dari perhitungan TF-IDF pada Tabel 6 dihitung jarak Euclidean menggunakan persamaan (3). Setelah dihitung, diambil jarak terdekat sebanyak nilai K. Nilai K sangat berpengaruh terhadap hasil prediksi [16]. Pada penelitian ini nilai K yang digunakan ada 5.

Tabel 8. Jarak Euclidean Antara Data Latih Dengan Data Baru

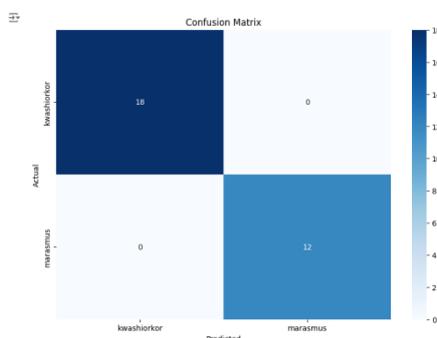
Dokumen	Jarak Euclidean	Kondisi
Doc3	2,9892	Marasmus
Doc6	2,9892	Marasmus
Doc5	4,1843	Kwashiorkor
Doc1	4,5957	Kwashiorkor
Doc2	4,5957	Marasmus

Berdasarkan tabel 7 merupakan hasil perhitungan jarak Euclidean antara data latih dan data baru yang sudah diurutkan dan dipilih 5 jarak paling dekat, dari 5 dokumen terdekat kondisi marasmus terdapat pada 3 dokumen dan kwashiorkor terdapat pada 2 dokumen. Hasil prediksi dapat dilihat dari label yang lebih banyak, sehingga data baru yaitu tensi rendah, suhu rendah, dan iritabilitas termasuk dalam klasifikasi kondisi marasmus.

Setelah proses klasifikasi selesai, langkah terakhir adalah evaluasi kinerja sistem. Ini dilakukan untuk mengevaluasi kinerja sistem yang akan dibangun. Penelitian ini menggunakan confusion matrix sebagai metode yang digunakan untuk mengukur kinerja sistem klasifikasi, yang akan membandingkan data hasil klasifikasi dengan data sebenarnya. Dengan bantuan hasil confusion matrix, perhitungan akurasi, precision, dan recall juga dapat dilakukan [16].

4. Pembahasan
Evaluasi Komposisi 90:10

Dataset yang digunakan sejumlah 300 data hasil augmentasi, dibagi menjadi 2 subset yaitu data latih dan data uji. Komposisi yang digunakan yaitu 90% data latih dan 10% data uji. Pada gambar 1 merupakan hasil evaluasi atau pengujian dari komposisi ini.



Gambar 2. Confusion Matrix Komposisi 90:10

Pada gambar 2 merupakan akurasi, presisi, dan recall dari model.

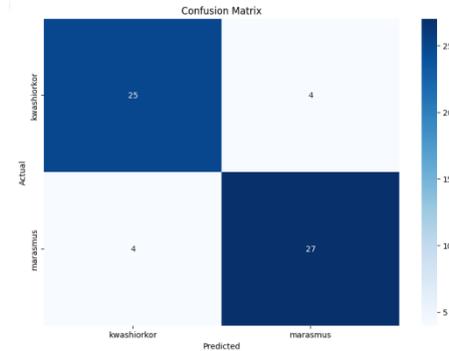
	precision	recall	f1-score	support
kwashiorkor	1.00	1.00	1.00	18
marasmus	1.00	1.00	1.00	12
accuracy			1.00	30
macro avg	1.00	1.00	1.00	30
weighted avg	1.00	1.00	1.00	30

Gambar 3. Akurasi, Presisi, dan Recall Komposisi 90:10

Dari gambar diatas terlihat bahwa hasil akurasi, presisi, dan recall semuanya 100% yang artinya model dapat memprediksi sempurna data uji, namun pada saat memprediksi data baru, hasil prediksinya tidak akurat, keadaan ini dapat disebut overfitting.[14]

Evaluasi Komposisi 80:20

Pada gambar 3 merupakan hasil evaluasi atau pengujian dengan komposisi 80% data latih dan 20% data uji.



Gambar 4. Confusion Matrix Komposisi 80:20

Pada gambar 3 terlihat bahwa model melakukan kesalahan prediksi terhadap 4 data pada masing-masing kondisi (FP dan FN). Pada gambar 4 merupakan hasil akurasi, presisi, dan recall.

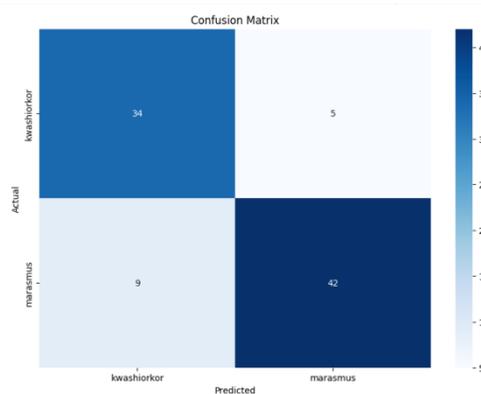
	precision	recall	f1-score	support
kwashiorkor	0.86	0.86	0.86	29
marasmus	0.87	0.87	0.87	31
accuracy			0.87	60
macro avg	0.87	0.87	0.87	60
weighted avg	0.87	0.87	0.87	60

Gambar 5. Akurasi, Presisi, dan Recall Komposisi 80:20

Akurasi model K-Nearest Neighbors (KNN) dengan komposisi 80:20 adalah 87%.

Evaluasi Komposisi 70:30

Pada gambar 5 merupakan hasil evaluasi dengan komposisi 70% data latih dan 30% data uji.



Gambar 6. Confusion Matrix Komposisi 70:30

Pada gambar 5 terlihat bahwa nilai FP adalah 5 dan nilai FN adalah 9. Pada gambar 6 merupakan hasil akurasi, presisi, dan recall pada komposisi ini.

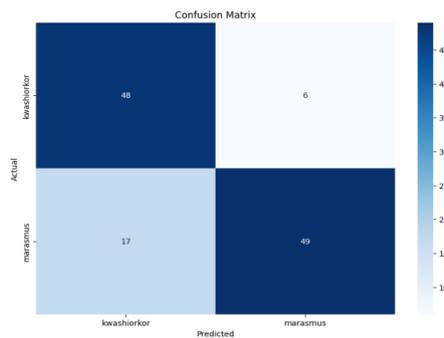
	precision	recall	f1-score	support
kwashiorkor	0.79	0.87	0.83	39
marasmus	0.89	0.82	0.86	51
accuracy			0.84	90
macro avg	0.84	0.85	0.84	90
weighted avg	0.85	0.84	0.85	90

Gambar 7. Akurasi, Presisi, dan Recall Komposisi 70:30

Pada gambar 6 terlihat bahwa nilai akurasi pada komposisi ini adalah 84%.

Evaluasi dengan Komposisi 60:40

Pada gambar 7 merupakan hasil evaluasi dengan komposisi 60% data latih dan 40% data uji.



Gambar 8. Confusion Matrix Komposisi 60:40

Pada gambar 7 terlihat bahwa nilai FP adalah 6 dan nilai FN adalah 17. Pada gambar 8 merupakan hasil akurasi, presisi, dan recall pada komposisi ini.

	precision	recall	f1-score	support
kwashiorkor	0.74	0.89	0.81	54
marasmus	0.89	0.74	0.81	66
accuracy			0.81	120
macro avg	0.81	0.82	0.81	120
weighted avg	0.82	0.81	0.81	120

Gambar 9. Akurasi, Presisi, dan Recall Komposisi 60:40

Pada gambar 6 terlihat bahwa nilai akurasi pada komposisi ini adalah 81%.

Pada tabel 13 merupakan hasil dari pengujian terhadap skenario yang dilakukan, yaitu membagi dataset menjadi beberapa komposisi.

Tabel 9. Hasil Akurasi Pada Setiap Komposisi

Splitting Data Training : Testing	Akurasi
90 : 10	100%
80 : 20	87%
70 : 30	84%
60 : 40	80%

Pada empat hasil pengujian diatas dapat disimpulkan bahwa komposisi 90:10 menjadi komposisi yang memiliki nilai akurasi paling tinggi, namun pada saat diimplementasikan untuk memprediksi data baru hasil prediksi seringkali salah, sehingga dari empat pengujian diatas yang

optimal saat diimplementasikan untuk memprediksi data baru adalah komposisi 80:20 dengan akurasi 87%.

Hasil penelitian ini sejalan dengan penelitian terdahulu yang menunjukkan keberhasilan penggunaan algoritma KNN dalam mendeteksi atau mendiagnosa kondisi menggunakan machine learning seperti penelitian yang dilakukan oleh Wulandari & Avianto (2024) melakukan penelitian menggunakan K-Nearest Neighbor (KNN) untuk diagnosis kelainan stunting pada balita. Dari penelitian ini dihasilkan tingkat akurasi 96% yang terdiri dari 119 prediksi benar dan 6 prediksi salah.[18] Selain itu, penelitian lain dilakukan oleh Safitri et.al (2023) menggunakan metode K-Nearest Neighbor (KNN) untuk menentukan status gizi dan memonitoring perkembangan balita. Tingkat akurasi dari metode K-Nearest Neighbor (K-NN) dengan nilai $k=3$ dan $k=5$ sebesar 87,5%.[19] Hasil penelitian ini menguatkan temuan bahwa KNN adalah alat yang andal untuk klasifikasi kondisi kesehatan, termasuk prediksi gizi buruk.

5. Penutup

Pada penelitian mengenai prediksi gizi buruk menggunakan K-Nearest Neighbors (KNN) . Sebelum memulai klasifikasi, dilakukan preprocessing terhadap data yang digunakan. Setelah itu dilakukan pembagian data menjadi data latih dan data set, lalu dilakukan ekstraksi fitur menggunakan TF-IDF sehingga dapat membuat model klasifikasi dan yang terakhir pengujian.

Dengan dataset yang digunakan sebanyak 50 data asli yang telah melalui proses augmentasi menjadi 300 data, model algoritma KNN mampu mengklasifikasikan kondisi menjadi 2 kelas yaitu kwashiorkor dan marasmus. Model ini dapat digunakan untuk alat pengambil keputusan oleh tenaga medis. Berdasarkan hasil pengujian dengan menggunakan 4 komposisi dataset yang berbeda, dapat disimpulkan bahwa komposisi 80:20 dengan akurasi 87% lebih optimal dalam memprediksi kondisi gizi buruk dibandingkan komposisi 90:10. Model ini memiliki potensi besar sebagai alat bantu bagi tenaga medis dalam pengambilan keputusan, memungkinkan diagnosis lebih cepat dan tepat. Saran untuk penelitian selanjutnya adalah memperbanyak dataset sehingga algoritma K-Nearest Neighbors (KNN) dapat lebih optimal dalam memprediksi dan mendapatkan nilai akurasi yang lebih tinggi.

Referensi

- [1] Badan Kebijakan Pembangunan Kesehatan, "Buku Saku Hasil Survei Status Gizi Indonesia (SSGI) 2022," 2022.
- [2] Ufiah Ramlah, "Gangguan Kesehatan Pada Anak Usia Dini Akibat Kekurangan Gizi Dan Upaya Pencegahannya," *Ana' Bulava: Jurnal Pendidikan Anak*, vol. 2, no. 2, pp. 12–25, Dec. 2021, doi: 10.24239/abulava.Vol2.Iss2.40.
- [3] D. P. Lestari, "Upaya Pencegahan Risiko Gizi Buruk pada Balita: Literature Review," *Jurnal Ilmiah Universitas Batanghari Jambi*, vol. 22, no. 1, p. 532, Mar. 2022, doi: 10.33087/jiubj.v22i1.1828.
- [4] A. S. R. Sinaga and D. Simanjuntak, "Sistem Pakar Deteksi Gizi Buruk Balita Dengan Metode Naïve Bayes Classifier," *Jurnal Inkofar*, vol. 1, no. 2, Jan. 2020, doi: 10.46846/jurnalinkofar.v1i2.110.
- [5] dr. Kartika Mayasari, "Berbagai Penyakit Akibat Malnutrisi yang Perlu Diwaspadai," *Klikdokter*. Accessed: Sep. 04, 2024. [Online]. Available: <https://www.klikdokter.com/ibu-anak/kesehatan-anak/7-penyakit-yang-disebabkan-karena-malnutrisi>
- [6] V. Viransyah and B. Sugiarto, "Sistem Pakar Diagnosis Penyakit Gizi Buruk Pada Balita Menggunakan Metode Naïve Bayes Berbasis Website," *Digital Transformation Technology*, vol. 3, no. 2, pp. 569–576, Nov. 2023, doi: 10.47709/digitech.v3i2.3074.
- [7] G. Gimnastiar, Y. I. Syuhardi, and I. Vandini, "Expert System Aplikasi Monitoring dan Pengendalian Gizi Berbasis Android," *JRKT (Jurnal Rekayasa Komputasi Terapan)*, vol. 1, no. 01, Apr. 2021, doi: 10.30998/jrkt.v1i01.4003.
- [8] D. A. Ferliandini and S. Risnanto, "Aplikasi Prediksi Status Gizi Balita Berbasis Web Menggunakan Metode K-Nearest Neighbor," *Prosiding Seminar Sosial Politik, Bisnis, Akuntansi dan Teknik*, vol. 5, p. 622, Dec. 2023, doi: 10.32897/sobat.2023.5.0.3136.

- [9] S. Lonang, A. Yudhana, and M. K. Biddinika, "Analisis Komparatif Kinerja Algoritma Machine Learning untuk Deteksi Stunting," *Jurnal Media Informatika Budidarma*, vol. 7, no. 4, p. 2109, Oct. 2023, doi: 10.30865/mib.v7i4.6553.
- [10] F. Putra, H. F. Tahiyat, R. M. Ihsan, R. Rahmaddeni, and L. Efrizoni, "Penerapan Algoritma K-Nearest Neighbor Menggunakan Wrapper Sebagai Preprocessing untuk Penentuan Keterangan Berat Badan Manusia," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, pp. 273–281, Jan. 2024, doi: 10.57152/malcom.v4i1.1085.
- [11] A. Ridho Lubis, M. K. M. Nasution, O. Salim Sitompul, and E. Muisa Zamzami, "The effect of the TF-IDF algorithm in times series in forecasting word on social media," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 22, no. 2, p. 976, May 2021, doi: 10.11591/ijeecs.v22.i2.pp976-984.
- [12] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," *SMATIKA JURNAL*, vol. 10, no. 02, pp. 71–76, Dec. 2020, doi: 10.32664/smatika.v10i02.455.
- [13] S. Lonang and D. Normawati, "Klasifikasi Status Stunting Pada Balita Menggunakan K-Nearest Neighbor Dengan Feature Selection Backward Elimination," *Jurnal Media Informatika Budidarma*, vol. 6, no. 1, p. 49, Jan. 2022, doi: 10.30865/mib.v6i1.3312.
- [14] M. M. Bejani and M. Ghatee, "A systematic review on overfitting control in shallow and deep neural networks," *Artif Intell Rev*, vol. 54, no. 8, pp. 6391–6438, Dec. 2021, doi: 10.1007/s10462-021-09975-1.
- [15] A. Yudhana, S. Sunardi, and A. J. S. Hartanta, "Algoritma K-NN Dengan Euclidean Distance untuk Prediksi Hasil Penggajian Kayu Sengon," *Transmisi*, vol. 22, no. 4, pp. 123–129, Nov. 2020, doi: 10.14710/transmisi.22.4.123-129.
- [16] Y. D. Kirana and S. al Faraby, "Sentiment Analysis of Beauty Product Reviews Using the K-Nearest Neighbor (KNN) and TF-IDF Methods with Chi-Square Feature Selection," 2021. [Online]. Available: <https://api.semanticscholar.org/CorpusID:245750721>
- [17] A. Putri *et al.*, "Komparasi Algoritma K-NN, Naïve Bayes dan SVM untuk Prediksi Kelulusan Mahasiswa Tingkat Akhir," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 1, pp. 20–26, May 2023, doi: 10.57152/malcom.v3i1.610.
- [18] A. R. Wulandari and D. Avianto, "Sistem Pakar Diagnosa Kelainan Stunting Balita Menggunakan Metode KNN Berbasis Web," *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, vol. 5, no. 1, pp. 1064–1072, Jan. 2024, doi: 10.35870/jimik.v5i1.587.
- [19] F. Safitri, R. Taufiq Subagio, and L. Norhan, "Implementasi Metode K-Nearest Neighbor (K-NN) dan Forward Chaining untuk Monitoring Tumbuh Kembang Balita," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 5, pp. 3491–3496, Jan. 2024, doi: 10.36040/jati.v7i5.7464.