

Article history

Received Jan 8, 2024

Accepted Nov 30, 2024

Publish Nov 30, 2024

PENERAPAN ALGORITMA K-NEAREST NEIGHBOR UNTUK DETEKSI DINI STATUS GIZI PASIEN DEWASA

Dian Wijayanti^{1),2)}, Arief Hermawan¹⁾, Donny Avianto¹⁾

¹Magister Teknologi Informasi, Universitas Teknologi Yogyakarta

²Rumah Sakit Umum Daerah Sleman

email: dianwijayanti077@gmail.com, ariefdbuty@gmail.com, donny@uty.ac.id

Abstract

Assessing the nutritional status of adult patients is essential to gain a comprehensive understanding of their condition and assist healthcare workers in planning appropriate treatment. However, manual assessment is time-consuming and labor-intensive, especially when the number of patients exceeds the number of available healthcare workers. This can hinder the timely and accurate delivery of nutritional care. The K-Nearest Neighbor (KNN) algorithm is a commonly used method for nutritional status classification, particularly in toddlers, pregnant women, or for obesity classification in adults. The use of KNN for early detection of adult nutritional status remains rarely explored. This study applies the KNN algorithm to classify the nutritional status of adult patients using data from the Alamanda 1 ward and the ICU ward at Sleman Regional General Hospital, collected from January 2 to October 18, 2023. The dataset includes patient height, weight, and nutritional status. The algorithm was implemented using RapidMiner with odd k-values less than 20, and data splits of 90:10, 70:30, and 50:50 for training and testing. Results show that the optimal k-values for the highest accuracy were $k = 1$ and $k = 3$ using the 70:30 data split, both achieving an accuracy of 96.77%. The highest sensitivity, 97.61%, was also achieved at $k = 3$ with the same data split. The KNN algorithm demonstrates strong potential to be developed into an early detection system for assessing the nutritional status of adult patients in hospitals, supporting faster and more accurate nutritional care services.

Keywords: K-Nearest Neighbor algorithm, nutritional status classification, adult patients, rapidminer.

Abstrak

Penilaian status gizi pasien dewasa penting untuk memahami kondisi gizi secara menyeluruh dan membantu tenaga kesehatan merencanakan perawatan yang sesuai. Namun, proses penilaian manual memerlukan waktu dan tenaga besar, terutama saat jumlah pasien tidak sebanding dengan jumlah tenaga kesehatan. Hal ini dapat menghambat pemberian perawatan gizi yang cepat dan tepat. Algoritma K-Nearest Neighbor (KNN) merupakan metode yang sering digunakan dalam klasifikasi status gizi, terutama pada balita, ibu hamil, atau klasifikasi obesitas pada orang dewasa. Penggunaan KNN untuk deteksi dini status gizi pasien dewasa masih jarang dibahas. Penelitian ini menerapkan algoritma KNN untuk mengklasifikasikan status gizi pasien dewasa menggunakan data dari ruang Alamanda 1 dan ruang ICU RSUD Sleman, yang dikumpulkan dari 2 Januari hingga 18 Oktober 2023. Dataset mencakup tinggi badan, berat badan, dan status gizi pasien. Implementasi dilakukan menggunakan RapidMiner dengan nilai k ganjil kurang dari 20, serta pembagian data pelatihan dan pengujian dalam rasio 90:10, 70:30, dan 50:50. Hasil menunjukkan bahwa nilai k optimal untuk akurasi tertinggi adalah $k = 1$ dan $k = 3$ pada rasio data 70:30, dengan akurasi sebesar 96,77%. Sensitivitas tertinggi sebesar 97,61% juga dicapai pada $k = 3$ dengan rasio 70:30. Algoritma KNN menunjukkan potensi tinggi untuk dikembangkan menjadi sistem deteksi dini status gizi pasien dewasa di rumah sakit, guna mendukung pelayanan gizi yang lebih cepat dan akurat.

Kata Kunci: algoritma K-Nearest Neighbor, klasifikasi status gizi, pasien dewasa, rapidminer.

1. PENDAHULUAN

Status gizi pasien merupakan faktor penting yang mempengaruhi respons terhadap pengobatan dan tingkat pemulihan pasien. Penilaian status gizi dapat memberikan gambaran komprehensif mengenai kondisi nutrisi pasien, termasuk berat badan dan tinggi badan, Indeks Massa Tubuh (IMT), lingkaran lengan atas, kadar albumin serum, dan kadar hemoglobin.

Penilaian status gizi juga dapat membantu petugas kesehatan memantau kemajuan pasien dan menyesuaikan rencana perawatan sesuai kebutuhan. Dengan demikian, penilaian status gizi merupakan komponen penting dalam perawatan kesehatan yang berkualitas dan dapat berkontribusi pada peningkatan kesehatan pasien. Penilaian status gizi secara manual merupakan proses yang memakan waktu dan tenaga, terutama ketika menangani sejumlah besar pasien. Hal ini dapat menjadi kendala dalam pemberian perawatan gizi yang tepat dan efektif, terutama jika jumlah pasien tidak sebanding dengan jumlah tenaga kesehatan yang melayani.

Oleh karena itu, penerapan teknologi dalam bidang kesehatan menghadirkan alternatif untuk deteksi dini status gizi pasien dewasa. Salah satu pendekatan yang relevan adalah penggunaan algoritma *K-Nearest Neighbor* (KNN) dalam klasifikasi status gizi pasien dewasa.

Algoritma *K-Nearest Neighbors* telah digunakan dalam klasifikasi status gizi balita, ibu hamil dan klasifikasi obesitas. Metode *Modified K-Nearest Neighbor* (M-KNN) telah digunakan dalam mengidentifikasi faktor risiko status gizi balita dengan akurasi sebesar 75% menggunakan 295 data status gizi balita pada nilai $k=4$ [1]. *K-Nearest Neighbor* memiliki kelebihan dalam mengatasi data yang memiliki kesalahan acak (*noise*). Penelitian menggunakan 412 data gizi balita, membandingkan metode *Naïve Bayes* dan KNN untuk klasifikasi status gizi balita diperoleh hasil akurasi 80,60% dengan *Naïve Bayes* dan 91,79% dengan *K-Nearest Neighbor* [2]. Tingkat akurasi sistem yang diuji dengan metode KNN untuk penentuan status gizi ibu hamil mencapai 86,7% dari 30 data sampel yang diuji. Hasil akurasi tertinggi diperoleh melalui pengujian k -optimal dengan nilai $k=3, 5, \text{ dan } 6$ [3]. Klasifikasi Status Gizi Balita Menggunakan Algoritma *Fuzzy K-Nearest Neighbor in Every Class* (FK-NNC) di Puskesmas Wonorejo menunjukkan bahwa nilai k optimal yang menghasilkan akurasi klasifikasi tertinggi adalah $k=8$ dengan akurasi 100% [4].

Pada penelitian Implementasi Algoritma *K-Nearest Neighbor* (KNN) untuk Klasifikasi Tingkat Obesitas, nilai k optimal yang menghasilkan akurasi tertinggi adalah $k=2$ dengan akurasi sebesar 78,98%. Nilai k yang lebih kecil menghasilkan akurasi yang lebih tinggi namun nilai k yang terlalu kecil dapat menyebabkan *overfitting*, sehingga sebaiknya dipilih nilai k yang optimal untuk menghasilkan akurasi yang tinggi tanpa menyebabkan *overfitting* [5].

Algoritma KNN telah digunakan untuk mengklasifikasikan sentimen publik mengenai vaksinasi COVID-19 menjadi kategori positif dan negatif. Dengan nilai $k=3$ diperoleh akurasi 85% [6]. Algoritma *K-Nearest Neighbor* (KNN) digunakan untuk mengklasifikasikan penyakit diabetes dengan menggunakan algoritma *Particle Swarm Optimization* (PSO) sebagai metode seleksi fitur. Sebelum dilakukan seleksi fitur menggunakan PSO, akurasi algoritma KNN adalah 75% dengan k -optimal 19. Namun, setelah dilakukan seleksi fitur menggunakan PSO, terjadi peningkatan akurasi menjadi 77,213% pada $k=19$ [6].

Klasifikasi kelayakan darah yang dapat didonorkan menggunakan metode K-NN (*K-Nearest Neighbor*) dengan optimasi algoritma *K-Means* menggunakan nilai $k=3$ untuk K-NN, nilai sensitivitas mencapai 100% dengan nilai spesifisitas sebesar 90,48% dan AUC (*Area Under the Curve*) sebesar 95%. Akurasi tertinggi yang dicapai adalah 98,91% dengan tingkat kesalahan sebesar 1,09% [7].

Pengujian kinerja algoritma *K-Nearest Neighbor* (KNN) dalam mengklasifikasikan kerentanan penyakit jantung menunjukkan bahwa KNN dengan nilai $k=5$ menghasilkan akurasi tertinggi sebesar 0,95 [8]. Komparasi Metode *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN), dan *Random Forest* (RF) untuk Prediksi Penyakit Gagal Jantung, metode KNN menghasilkan akurasi sebesar 93% dengan $k=20$ [9]. Optimasi Algoritma *K-Nearest Neighbor* (KNN) dengan Normalisasi dan Seleksi Fitur untuk Klasifikasi Penyakit Liver menghasilkan akurasi tertinggi pada normalisasi *min-max* dengan seleksi fitur *Information Gain* dan *Gain Ratio*, yaitu sebesar 78,2%, dengan $k=10$ [10].

Pada penelitian berjudul Sistem Deteksi Pelat Kendaraan dengan Menggunakan Metode *K-Nearest Neighbour* (KNN), nilai k yang optimal adalah 1 dengan akurasi 88,88% dalam jarak 40 cm. Semakin besar nilai k , maka semakin banyak data yang digunakan untuk memprediksi kelas dari data baru. Namun, jika nilai k terlalu besar, maka

model akan menjadi *overfitting*, yaitu model akan terlalu baik dalam memprediksi data training tetapi tidak baik dalam memprediksi data baru [11].

Pada penelitian yang berjudul Optimasi Nilai K pada Algoritma KNN untuk Klasifikasi Spam dan Ham Email nilai k optimal yang menghasilkan akurasi tertinggi adalah k=49 dengan akurasi sebesar 96.2%. Semakin besar nilai k, maka akurasi klasifikasi juga semakin tinggi. Namun nilai k yang terlalu besar dapat menyebabkan *overfitting*, sehingga sebaiknya dipilih nilai k yang optimal untuk menghasilkan akurasi yang tinggi tanpa menyebabkan *overfitting* [12].

Penelitian yang mengklasifikasikan penyakit pada daun jagung menggunakan algoritma *K-Nearest Neighbors* (KNN) dengan ekstraksi fitur *Gray-Level Co-occurrence Matrix* (GLCM) dan *Hue Saturation Value* (HSV), menunjukkan bahwa akurasi tertinggi diperoleh pada algoritma KNN-GLCM-HSV dengan nilai k = 3 dan jarak piksel = 1, yaitu sebesar 85% [13]. Klasifikasi Citra Penyakit Daun Padi Menggunakan KNN Berdasarkan Ekstraksi Fitur GLCM mendapatkan nilai k optimal untuk KNN adalah 11 dengan akurasi sebesar 65.83% [14].

Pada penelitian yang berjudul Analisis Perbandingan Algoritma SVM Dan KNN Untuk Klasifikasi Anime Bergenre Drama, algoritma KNN menggunakan nilai k=10 menghasilkan akurasi sebesar 84%. Algoritma KNN memiliki akurasi yang lebih tinggi dibandingkan dengan algoritma SVM karena algoritma KNN lebih sederhana dan mudah dipelajari [15].

Pada Prediksi Harga Saham PT Bank Central Asia Tbk Berdasarkan Data dari Bursa Efek Indonesia Menggunakan Metode *K-Nearest Neighbors* (KNN), nilai k=37 memberikan performa yang baik, dengan akurasi sebesar 61.79% dan kesalahan klasifikasi sebesar 38.21% [16].

Pada penelitian berjudul Perbandingan Metode *K-Nearest Neighbor* dan LBPH pada Klasifikasi Daun Herbal, metode *K-Nearest Neighbor* (KNN) dengan nilai k=9 menghasilkan akurasi tertinggi yaitu 97,5%. Nilai k yang lebih tinggi pada metode *K-Nearest Neighbor* (KNN) menghasilkan akurasi yang lebih tinggi karena lebih banyak data tetangga yang digunakan untuk klasifikasi [17].

Perbandingan Metode Klasifikasi KNN dan Naïve Bayes untuk Menentukan Tingkat Stres Siswa Sekolah Menengah Pertama, nilai k yang optimal untuk KNN adalah 3 dengan akurasi 86,61% [18]. Algoritma KNN merupakan algoritma klasifikasi yang dapat digunakan untuk memprediksi kelas data baru berdasarkan data yang sudah ada. Dalam

konteks penilaian status gizi, algoritma KNN dapat memanfaatkan data *antropometri* pasien, seperti tinggi badan dan berat badan untuk memprediksi status gizi pasien dewasa.

Penggunaan algoritma KNN memiliki beberapa keuntungan dalam penilaian status gizi pasien dewasa. Pertama, algoritma KNN mudah diimplementasikan dan tidak memerlukan pelatihan data yang ekstensif. Kedua, algoritma KNN dapat menangani data yang tidak lengkap atau *noisy*. Ketiga, algoritma KNN dapat memberikan hasil yang akurat dan *reliable*.

Dengan demikian, penerapan algoritma KNN dalam klasifikasi status gizi pasien dewasa dapat menjadi pendekatan yang efisien dan terukur untuk mendukung penilaian status gizi pasien dewasa di lingkungan rumah sakit.

Berdasarkan studi literatur, nilai k optimal untuk akurasi tertinggi sangat bervariasi pada masing-masing hasil penelitian. Penelitian ini berupaya menjawab pertanyaan: berapa nilai k optimal yang akan digunakan dalam algoritma *K-Nearest Neighbor* (K-NN) untuk mengklasifikasikan status gizi pasien dewasa supaya mendapatkan waktu tercepat, akurasi tertinggi, dan sensitivitas tertinggi?

2. METODE PENELITIAN

Rancangan Kegiatan

Tahapan dalam penelitian ini yaitu proses pengumpulan data, *preprocessing* yang bertujuan untuk membersihkan data dari *missing value*, setelah itu dilakukan pembagian data latih dan data uji, kemudian data akan dimodelkan menggunakan model klasifikasi *K-nearest neighbor*, dan dilakukan evaluasi.

Ruang Lingkup

Penelitian ini menggunakan algoritma KNN dengan aplikasi *Rapidminer*. Penelitian ini berfokus mencari k optimal untuk mendapatkan waktu tercepat, akurasi tertinggi dan sensitivitas tertinggi.

Teknik Pengumpulan Data

Data dikumpulkan dari laporan bulanan asuhan gizi pasien dewasa di Bangsal Alamanda 1 dan Bangsal *Intensive Care Unit (ICU)* Rumah Sakit Umum Daerah Sleman. Data yang digunakan adalah data tanggal 2 Januari hingga 18 Oktober 2023. Data set yang digunakan adalah tinggi badan, berat badan dan status gizi pasien.

Teknik Analisis

Atribut diberikan pada data set yang tersedia. Tinggi badan dan berat badan pasien termasuk dalam kategori atribut reguler, sedangkan data status gizi termasuk dalam kategori atribut khusus. Label status gizi dibagi menjadi lima kriteria, yaitu kurang, normal, lebih, obesitas I, dan obesitas II. Data set dibagi menjadi 2, yaitu data latih dan data uji. Dilakukan 3 variasi pembagian data latih dan data uji seperti disajikan pada Tabel 1 berikut ini:

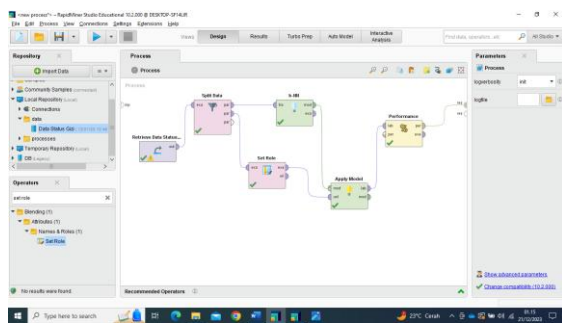
Tabel 1. Pembagian Data Set

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

Pada penelitian ini data set akan dilatih menggunakan menggunakan *tool RapidMiner Studio* dengan model klasifikasi *K-nearest neighbor*, sebanyak 30 kali, dengan k bilangan ganjil < 20, lalu akan dievaluasi waktu pemrosesan, tingkat akurasi, dan sensitivitasnya.

Model Klasifikasi

Gambar 1 adalah desain *RapidMiner K-Nearest Neighbor* (KNN) yang menggambarkan langkah-langkah penerapan algoritma KNN pada data.



Gambar 1. Desain RapidMiner

Pada tahap awal, data disiapkan dan dimuat ke dalam platform *RapidMiner*. Selanjutnya, dilakukan tahap *preprocessing* untuk membersihkan data dari *missing value*. Setelah itu dilakukan *split* data dan algoritma KNN dijalankan pada data set dengan k yang telah ditetapkan. Hasil dari klasifikasi data dilihat melalui *output* yang dihasilkan. Desain ini memberikan pandangan visual tentang implementasi KNN menggunakan *RapidMiner*, memfasilitasi pemahaman yang lebih baik

terhadap proses dan hasil yang dihasilkan dari algoritma ini.

Evaluasi

Evaluasi sensitivitas dilakukan untuk hasil akurasi tertinggi menggunakan *confusion matrix multiple classes classification*. *Confusion Matrix*, adalah cara tabel untuk memvisualisasikan kinerja model prediksi pada pembelajaran supervised learning. Setiap data dari masing-masing kelas dalam tabel confusion matrix menunjukkan jumlah prediksi yang dibuat guna untuk mengklasifikasikan kelas yang benar atau salah.

Tabel 2 menunjukkan contoh format *confusion matrix* dengan n kelas:

Tabel 2. *Confusion Matrix Multiple Classes* [19]

		Predicted Number			
		Class 1	Class 2	...	Class n
Actual Number	Class 1	X11	X12	...	X1n
	Class 2	X21	X22	...	X2n

	Class n	Xn1	Xn2	...	Xnn

Gambar 2 menunjukkan *confusion matrix* untuk klasifikasi *multiple classes* dengan tiga kelas (A, B, dan C).

		True Class		
		A	B	C
Predicted Class	A	TP _A	E _{BA}	E _{CA}
	B	E _{AB}	TP _B	E _{CB}
	C	E _{AC}	E _{BC}	TP _C

Gambar 2. *Confusion Matrix Tiga Kelas* [20]

Diagonal hijau mewakili prediksi yang benar dan kotak merah muda menunjukkan prediksi yang salah. Jika sampel positif dan diklasifikasikan sebagai positif, yaitu diklasifikasikan positif dengan benar sampel dihitung positif benar atau *true positive* (TP); jika diklasifikasikan sebagai negatif, maka dianggap sebagai kesalahan negatif palsu *false negative* (FN) atau kesalahan tipe II. Jika sampelnya negatif dan diklasifikasikan sebagai negatif, maka disebut sebagai *true negative* (TN); jika diklasifikasikan positif maka dihitung *false positive* (FP), atau kesalahan Tipe I.

Seperti yang ditunjukkan, pada Gambar 2, TP_A adalah jumlah sampel positif sebenarnya di kelas A, yaitu jumlah sampel yang diklasifikasikan

dengan benar dari kelas A, dan E_{AB} sebagai sampel dari kelas A yang salah diklasifikasikan sebagai kelas B, yaitu sampel yang salah diklasifikasikan. Jadi, negatif palsu pada kelas A (FN_A) adalah jumlah E_{AB} dan E_{AC} ($FN_A = E_{AB} + E_{AC}$) yang menunjukkan jumlah semua sampel kelas A yang salah diklasifikasikan sebagai kelas B atau C.

Sederhananya, FN kelas mana pun yang terletak di suatu kolom dapat dihitung dengan menambahkan kesalahan pada kelas/kolom tersebut. Sedangkan yang *false positive* (FP) untuk setiap kelas prediksi yang terletak pada suatu baris mewakili jumlah semua kesalahan pada baris tersebut. Misalnya, positif palsu di kelas A (FP_A) dihitung sebagai berikut, $FP_A = E_{BA} + E_{CA}$ [20].

Sensitivitas, atau *true positive rate* (TPR), atau *hit rate*, atau *recall*, dari suatu klasifikasi adalah sampel positif yang diklasifikasikan dengan benar ke jumlah total sampel positif, dan diperkirakan berdasarkan rumus:

$$TPR = \frac{TP}{TP + FN}$$

Keterangan:

TPR : *True Positive Rate* (Sensitivitas)

TP : *True Positive*

FN : *False Negative*

Setelah diketahui nilai TPR masing-masing kelas, kemudian dihitung nilai rata-rata sensitivitas dari semua kelas dengan menjumlahkan dan dibagi jumlah kelas, menggunakan rumus:

$$All\ TPR = \frac{TPR\ (Class\ 1 + Class\ 2 + \dots + Class\ n)}{n}$$

Keterangan:

All TPR : Rata-rata sensitivitas

TPR : *True Positive Rate*

n : jumlah kelas

3. HASIL DAN PEMBAHASAN

Hasil Penelitian

Data set yang digunakan adalah data tinggi badan, berat badan dan status gizi pasien dewasa. Data diperoleh dari data primer pasien di bangsal 1 Alamanda dan ICU RSUD Sleman pada tanggal 2 Januari sampai dengan 18 Oktober 2023. Jumlah data set sebanyak 312 dengan 2

atribut reguler dan 1 atribut khusus. Pada saat *preprocessing* data dilakukan pengecekan *missing value* dan data set ini tidak ada *missing value*.

Variabel tinggi badan dan berat badan diberi atribut reguler, sedangkan variabel status gizi diberi atribut khusus, yaitu label. Distribusi klasifikasi status gizi disajikan dalam Tabel 3.

Tabel 3. Distribusi Klasifikasi Status Gizi

No.	Klasifikasi Status Gizi	Jumlah
1.	Kurang	45
2.	Normal	118
3.	Lebih	61
4.	Obes I	73
5.	Obes II	15
Jumlah		312

Setelah dilakukan *preprocessing* kemudian dilakukan proses *split* data. Skema *split* data disajikan dalam Tabel 4.

Tabel 4. Skema *Split* Data

Persentase (Data Latih:Data Uji)	Data Latih	Data Uji
90:10	281	31
70:30	218	94
50:50	156	156

Proses pengujian data menggunakan *tool RapidMiner Studio* dengan model klasifikasi *K-nearest neighbor*. Tabel 5 merupakan hasil pengujian pada k bilangan ganjil < 20 dengan *split* data 90:10, 70:30, dan 50:50.

Tabel 5. Hasil Pengujian

<i>Split</i> Data	k	Akurasi (%)	<i>ExecutionTime</i> (detik)
90:10	1	96.67	0
90:10	3	96.67	0
90:10	5	93.33	0
90:10	7	93.33	0
90:10	9	90.00	0
90:10	11	86.67	0
90:10	13	90.00	0
90:10	15	90.00	0
90:10	17	90.00	0
90:10	19	93.33	0
70:30	1	96.77	0
70:30	3	96.77	0
70:30	5	93.55	0

Split Data	k	Akurasi (%)	ExecutionTime (detik)
70:30	7	93.55	0
70:30	9	92.47	0
70:30	11	87.10	0
70:30	13	83.87	0
70:30	15	80.65	0
70:30	17	81.72	0
70:30	19	81.72	0
50:50	1	90.38	0
50:50	3	87.82	0
50:50	5	87.82	0
50:50	7	85.26	0
50:50	9	83.97	0
50:50	11	80.77	0
50:50	13	73.08	0
50:50	15	71.15	0
50:50	17	68.59	0
50:50	19	67.95	0

Dari hasil pengujian diperoleh hasil akurasi tertinggi pada k=1 dan k=3 yaitu sebesar 96.77% dengan split data 70:30. Selanjutnya dilakukan evaluasi sensitivitas menggunakan Confusion Matrix seperti disajikan pada Tabel 6 dan Tabel 7.

Confusion Matrix pada penelitian ini adalah *Confusion Matrix Multiple Classes* yang terdiri dari 5 kelas klasifikasi yaitu normal, kurang, obes I, lebih, dan obes II. Kotak warna hijau menunjukkan jumlah prediksi yang benar dan kotak warna kuning menunjukkan jumlah prediksi yang salah. Jumlah prediksi diperoleh dari hasil pengujian menggunakan *tool RapidMiner Studio*. Sensitivitas dihitung pada masing-masing kelas, kemudian dihitung rata-ratanya.

Tabel 6. Confusion Matrix k=1 pada Split Data 70:30

		Actual Number				
		Normal	Kurang	Obes I	Lebih	Obes II
Predicted Number	Normal	35	0	0	0	0
	Kurang	0	14	0	0	0
	Obes I	0	0	19	0	0
	Lebih	0	0	3	18	0
	Obes II	0	0	0	0	4
Sensitivity		100%	100%	86,36 %	100 %	100 %
Rata-rata		97,27%				

Tabel 7. Confusion Matrix k=3 pada Split Data 70:30

		Actual Number				
		Normal	Kurang	Obes I	Lebih	Obes II
Predicted Number	Normal	34	0	0	0	0
	Kurang	0	14	0	0	0
	Obes I	0	0	20	0	0
	Lebih	1	0	2	18	0
	Obes II	0	0	0	0	4
Sensitivity		97,15 %	100%	90,91 %	100 %	100 %
Rata-rata		97,61%				

Hasil evaluasi sensitivitas disajikan pada Tabel 6.

Tabel 8. Evaluasi Sensitivitas

Split Data	k	Sensitivitas
70:30	1	97,27%
70:30	3	97,61%

Sensitivitas tertinggi diperoleh pada *split* data 70:30 dengan k=3.

Pembahasan

K-Nearest Neighbors (KNN) adalah algoritma pembelajaran mesin yang beroperasi berdasarkan prinsip "kesamaan tetangga." Jika suatu data memiliki tetangga-tetangga yang memiliki label tertentu, maka kemungkinan besar data tersebut juga akan memiliki label tersebut. Parameter utama dalam KNN adalah nilai k, yang menentukan jumlah tetangga terdekat yang akan diambil untuk membuat prediksi. Nilai k yang dipilih dapat memengaruhi tingkat kompleksitas dan performa model. KNN menggunakan metrik jarak, seperti *Euclidean distance* atau *Manhattan distance*, untuk mengukur seberapa dekat atau jauh antara dua data. Metrik ini menentukan cara mengukur "jarak" antara data dalam ruang fitur. Untuk memprediksi label suatu data, KNN mencari k tetangga terdekat berdasarkan metrik jarak. Prediksi dilakukan dengan mayoritas suara (klasifikasi) atau nilai rata-rata (regresi) dari tetangga-tetangga tersebut. Pemilihan nilai k berpengaruh pada kompleksitas model. Nilai k kecil dapat menyebabkan *overfitting*, sementara k besar dapat menyebabkan *underfitting* [21]. Hasil uji coba menunjukkan bahwa akurasi tertinggi diperoleh pada nilai k=1 dan k=3 yakni 96.77% dengan perbandingan data latihan:data uji sebesar 70:30. Pada penelitian ini, rasio data latihan

dan data uji 70:30 menghasilkan akurasi dan sensitivitas yang tertinggi. Evaluasi kinerja model pembelajaran mesin mengungkapkan bahwa kemampuan prediksi sangat dipengaruhi oleh proporsi data latih dan data uji. Perbandingan 70:30 menghasilkan kinerja terbaik berdasarkan parameter evaluasi [22]. Analisis lebih lanjut mengungkapkan bahwa sensitivitas tertinggi dicapai pada rasio 70:30 dengan nilai $k=3$. Uji sensitivitas dilakukan pada variasi nilai k yang berbeda yakni $k=1$ dan $k=3$ yang menggunakan perbandingan *split* data *training* dan *test* yang sama, yaitu 70:30. Hasil pengujian sensitivitas menunjukkan bahwa model KNN dengan $k=3$ menghasilkan sensitivitas sedikit lebih tinggi, yaitu sebesar 97,61%, dibandingkan model KNN dengan $k=1$ yang menghasilkan sensitivitas 97,27%. Hal ini mengindikasikan bahwa kemampuan model KNN dengan $k=3$ untuk mendeteksi sampel positif secara akurat (*true positive*) lebih baik dibandingkan model KNN dengan $k=1$. Dengan kata lain, model KNN dengan $k=3$ cenderung menghasilkan jumlah sampel positif yang terlewatkan atau salah diklasifikasi sebagai negatif (*false negative*) lebih sedikit. Secara keseluruhan, dalam ranah diagnostik model KNN dengan $k=3$ memiliki performa yang lebih unggul meskipun perbedaan sensitivitasnya relatif kecil.

Sensitivitas merepresentasikan kemampuan model dalam mendeteksi kondisi positif secara akurat. Ukuran ini digunakan untuk menilai kinerja karena dapat mengetahui sejauh mana model dapat memprediksi kasus positif dengan tepat [19]. Di bidang kesehatan sensitivitas merupakan hal yang penting. Semakin besar persentase sensitivitas, semakin baik kemampuan model untuk memprediksi kasus positif dengan akurat. Masing-masing proses pengujian memerlukan durasi 0 detik, hal ini menunjukkan kemampuan algoritma KNN untuk memprediksi status gizi pasien secara *real-time*. Kecepatan pelayanan merupakan hal yang penting di dunia kesehatan. Deteksi dini status gizi pasien dewasa menggunakan algoritma KNN dapat mempercepat proses pengambilan keputusan dalam pelayanan pasien.

4. PENUTUP

Kesimpulan

Algoritma *K-Nearest Neighbors* (KNN) dapat digunakan untuk klasifikasi status gizi pasien

dewasa dengan memanfaatkan data *antropometri* tinggi badan dan berat badan. Nilai k yang optimal untuk mendapatkan akurasi klasifikasi tertinggi adalah $k=3$ dengan *split* data 70:30, akurasinya mencapai 96.77% dan sensitivitas 97.61%. Algoritma KNN dengan nilai k optimal dapat digunakan untuk mengklasifikasi status gizi pasien dewasa secara akurat dan dapat dikembangkan menjadi sistem deteksi dini status gizi pasien di rumah sakit.

Saran

Jumlah data sample yang digunakan saat ini masih relatif kecil yaitu 312 data. Penambahan jumlah data set akan meningkatkan akurasi model karena akan lebih mewakili populasi yang sebenarnya. Selain KNN, dapat diuji pula model klasifikasi lain seperti *Naive Bayes*, *Decision Tree*, *Random Forest* untuk dibandingkan kinerjanya. Model mana yang paling akurat untuk kasus ini.

5. REFERENSI

- [1] M. Musnaimah, A. Alifatin, and N. Hayatin, "Expert System to Identify Risk Factors of Toddler's Nutrition Status with Case Based Reasoning," *J. Peremp. dan Anak*, vol. 3, no. 1, p. 27, 2020, doi: 10.22219/jpa.v3i1.11810.
- [2] R. Setiawan and A. Triayudi, "Klasifikasi Status Gizi Balita Menggunakan Naïve Bayes dan K-Nearest Neighbor Berbasis Web," *J. Media Inform. Budidarma*, vol. 6, no. 2, p. 777, 2022, doi: 10.30865/mib.v6i2.3566.
- [3] A. Y. Budiman, "Penerapan K-Nearest Neighbor Classification dalam penentuan status Gizi Ibu Hamil berdasarkan tingkat konsumsi makanan secara harian," *Repository.Uinjkt.Ac.Id*, [Online]. Available: [https://repository.uinjkt.ac.id/dspace/handle/123456789/67817%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/67817/1/ADRIANTO YUSUF BUDIMAN-FST.pdf](https://repository.uinjkt.ac.id/dspace/handle/123456789/67817%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/67817/1/ADRIANTO_YUSUF_BUDIMAN-FST.pdf)
- [4] R. U. Izzah, S. Wahyuningsih, and F. D. T. Amijaya, "Classification of nutritional status of toddlers using fuzzy k-nearest neighbor in every class (FK-NNC)," *J. Phys. Conf. Ser.*, vol. 1277, no. 1, 2019, doi: 10.1088/1742-6596/1277/1/012050.
- [5] A. M. S. I. Dewi and I. B. G. Dwidasmara, "Implementation Of The K-Nearest Neighbor (KNN) Algorithm For

- Classification Of Obesity Levels,” *JELIKU (Jurnal Elektron. Ilmu Komput. Udayana)*, vol. 9, no. 2, p. 277, 2020, doi: 10.24843/jlk.2020.v09.i02.p15.
- [6] S. Suprayogi, C. A. Sari, and E. H. Rachmawanto, “Sentiment Analyst on Twitter Using the K-Nearest Neighbors (KNN) Algorithm Against Covid-19 Vaccination,” *J. Appl. Intell. Syst.*, vol. 7, no. 2, pp. 135–145, 2022, doi: 10.33633/jais.v7i2.6734.
- [7] Y. Fredyatama and Z. Abidin, “CLASSIFICATION FEASIBILITY OF BLOOD THAT CAN BE DONORED USING K-NN WITH K-MEANS OPTIMIZATION KLASIFIKASI KELAYAKAN DARAH YANG DAPAT DIDONORKAN MENGGUNAKAN K-NN DENGAN OPTIMASI K-MEANS,” vol. 4, no. 1, pp. 41–57, 2023.
- [8] H. Saputro, F. Zalmi, and R. Syahputra, “Performance Testing of KNN and Logistic Regression Algorithms in Classifying Heart Disease Susceptibility,” vol. 04, no. 04, pp. 140–144, 2023.
- [9] S. Adi and A. Wintarti, “Komparasi Metode Support Vector Machine (Svm), K-Nearest Neighbors (Knn), Dan Random Forest (Rf) Untuk Prediksi Penyakit Gagal Jantung,” *MATHunesa J. Ilm. Mat.*, vol. 10, no. 2, pp. 258–268, 2022, doi: 10.26740/mathunesa.v10n2.p258-268.
- [10] S. Zulaikhah Hariyanti Rukmana, A. Aziz, and W. Harianto, “Optimasi Algoritma K-Nearest Neighbor (Knn) Dengan Normalisasi Dan Seleksi Fitur Untuk Klasifikasi Penyakit Liver,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 6, no. 2, pp. 439–445, 2022, doi: 10.36040/jati.v6i2.4722.
- [11] F. Sunusi, Z. Zainuddin, and S. Sahibu, “Sistem Deteksi Plat Kendaraan Dengan Menggunakan Metode K-Nearest Neighbour (Knn),” *J. Ris. Inform.*, vol. 1, no. 2, pp. 9–14, 2019, doi: 10.34288/jri.v1i2.18.
- [12] E. P. Laksono, A. d Basuki, and F. Abdurrachman Bachtiar, “Optimasi Nilai K pada Algoritma KNN untuk klasifikasi Spam dan Ham Email,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 2, pp. 377–383, 2020.
- [13] E. H. Rachmawanto and H. P. Hadi, “Optimasi E Kstraksi F Itur P Ada Knn,” vol. 22, no. 2, pp. 58–67, 2021.
- [14] R. A. Saputra, Suharyanto, S. Wasiyanti, D. F. Saefudin, A. Supriyatna, and A. Wibowo, “Rice Leaf Disease Image Classifications Using KNN Based on GLCM Feature Extraction,” *J. Phys. Conf. Ser.*, vol. 1641, no. 1, 2020, doi: 10.1088/1742-6596/1641/1/012080.
- [15] Vika Vitaloka Pramansah, “Analisis Perbandingan Algoritma SVM Dan KNN Untuk Klasifikasi Anime Bergenre Drama,” *J. Inform. dan Teknol. Komput. (J-ICOM)*, vol. 3, no. 1, pp. 49–55, 2022, doi: 10.33059/j-icom.v3i1.4950.
- [16] E. R. Tauran, “Prediksi Harga Saham Pt Bank Central Asia Tbk Berdasarkan Data Dari Bursa Efek Indonesia Menggunakan Metode K-Nearest Neighbors (Knn),” *TeIKa*, vol. 11, no. 2, pp. 123–129, 2021, doi: 10.36342/teika.v11i2.2609.
- [17] Isman, Andani Ahmad, and Abdul Latief, “Perbandingan Metode KNN Dan LBPH Pada Klasifikasi Daun Herbal,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 3, pp. 557–564, 2021, doi: 10.29207/resti.v5i3.3006.
- [18] Y. C. Tapingdan and D. Paseru, “Comparative Analysis of Classification Methods of KNN and Naïve Bayes to Determine Stress Level of Junior High School Students,” *Indones. J. Inf. Syst.*, vol. 2, no. 2, pp. 80–89, 2020, doi: 10.24002/ijis.v2i2.3035.
- [19] C. Manliguez, “Generalized Confusion Matrix for Multiple Classes Generalized Confusion Matrix for Multiple Classes The total numbers of false negative (TFN), false positive (TFP), and true negative (TTN) for each class i will be calculated based on the Generalized,” no. November, pp. 5–7, 2016, doi: 10.13140/RG.2.2.31150.51523.
- [20] A. Tharwat, “Classification assessment methods,” *Appl. Comput. Informatics*, vol. 17, no. 1, pp. 168–192, 2018, doi: 10.1016/j.aci.2018.08.003.
- [21] S. Zhang, “Challenges in KNN Classification,” *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 10, pp. 4663–4675, 2021, doi: 10.1109/TKDE.2021.3049250.
- [22] Q. H. Nguyen *et al.*, “Influence of data splitting on performance of machine learning models in prediction of shear strength of soil,” *Math. Probl. Eng.*, vol. 2021, 2021, doi: 10.1155/2021/4832864.