



PERBANDINGAN METODE K-NEAREST NEIGHBOR DAN MODIFIED K-NEAREST NEIGHBOR UNTUK KLASIFIKASI KELUARGA BERESIKO STUNTING

Dandi Irwayunda Pratama¹, Fitri Insani², Febi Yanto³, Iis Afrianty⁴

^{1,2,3,4}Universitas Islam Negeri Sultan Syarif Kasim Riau

(Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau)

(Jl. HR. Soebrantas KM. 15 No. 155, Simpang Baru, Kota Pekanbaru, 28293, Indonesia)

email: 12050113429@students.uin-suska.ac.id, fitri.insani@uin-suska.ac.id,

3febi.yanto@uin-suska.ac.id, 4iis.afrianty@uin-suska.ac.id

Abstrak

Stunting disebabkan oleh kekurangan gizi kronis, yang menghambat pertumbuhan terhambat pada anak dan dapat memengaruhi kesehatan jangka panjang. Penelitian ini bertujuan untuk mengklasifikasikan keluarga beresiko *stunting* menggunakan metode *K-Nearest Neighbor* (*K-NN*) dan *Modified K-Nearest Neigbor* (*MK-NN*). Perbandingan keduanya dilakukan dengan tujuan memberikan gambaran lebih jelas mengenai metode mana yang lebih cocok dalam membantu dalam memilih algoritma yang memberikan hasil yang optimal. Data yang digunakan terdiri dari 23607 data keluarga dan 20 parameter, diperoleh dari Balai Penyuluhan KB (Kampung Berencana) di Kecamatan Tuah Madani. Hasil menunjukkan bahwa *MK-NN* memberikan performa lebih konsisten pada berbagai nilai *k* dengan akurasi mencapai 99.28% terutama pada rasio 80:20 dan 70:30. Sebaliknya, *K-NN* mencapai akurasi maksimum 99.36% tetapi mengalami fluktuasi pada nilai *k* tertentu. *MK-NN* juga unggul dalam metrik precision, recall dan f1-score menunjukkan mampu menghadapi data yang kompleks. Dapat disimpulkan bahwa *MK-NN* lebih efektif dan stabil dibandingkan *K-NN*. Penelitian ini menyarankan penggunaan data ekonomi seperti pendapatan dan pekerjaan orang tua pada studi mendatang untuk memberikan hasil klasifikasi yang lebih menyeluruh dan akurat dalam mendukung kebijakan *stunting*.

Kata kunci: *stunting, klasifikasi, manhattan distance, k-nearest neighbor, Modified K-Nearest Neigbor*

Abstract

Stunting is caused by chronic malnutrition, inhibits children's growth in children and can impact long-term health. This study aims to classify families at risk of stunting using the K-Nearest Neighbor (K-NN) and Modified K-Nearest Neighbor (MK-NN). The comparison aims to provide a clearer understanding of which methods is more suitable for selecting an algorithm that delivers optimal results. The data used consists of 23607 family records and 20 parameters, The data were obtained from the Family Planning Counseling Center in Tuah Madani District. The results show that the MK-NN method delivers more consistent performance across various k values, achieving an accuracy of 99.28%, particularly with 80:20 and 70:30 data ratios. In contrast, K-NN reached a maximum accuracy of 99.36% but experienced fluctuations at certain k values. MK-NN also excelled in precision, recall and F1-score metrics, demonstrating its reliability in handling complex data. It can be concluded that MK-NN is more effective and stable than K-NN. This study recommends including economic data, such as family income and parental employment, in future research to produce more comprehensive and accurate classification results in support of stunting policies.

Keywords: *stunting, classification, manhattan distance, K-nearest neighbor, Modified K-Nearest Neigbor*

1. PENDAHULUAN

Di Indonesia, tingkat kesejahteraan masih belum merata, yang berdampak pada kondisi kesehatan. Gizi yang baik berarti status gizi seseorang dapat terpenuhi dengan baik, sehingga mendukung tumbuh kembang yang optimal[1]. Berdasarkan studi kasus gizi di Indonesia tahun 2021, prevalensi *stunting* di Indonesia mencapai 24,4 persen. Angka ini menunjukkan penurunan sebesar 3,3 persen dibandingkan tahun 2019, yang sebelumnya 27,7 persen. [2].

Stunting merupakan kondisi dimana anak memiliki tinggi badan yang kurang sesuai dengan usia, yang dimulai dari faktor lingkungan yang menyebabkan berat badan lahir rendah terus berlanjut dengan pertumbuhan yang terhambat pada 2 tahun pertama kehidupannya[3]. Menurut definisi Organisasi Kesehatan Dunia (WHO), *stunting* terjadi sejak dini dalam kehidupan, terutama dalam 1000 hari pertama sejak awal kehamilan hingga usia dua tahun[4].

Faktor penyebab *stunting* mencakup kekurangan gizi dan juga melibatkan pendidikan[5]. Selain itu, *stunting* juga berkaitan dengan ketidaksetaraan status sosial dan ekonomi keluarga[6]. Meskipun dampak *stunting* tidak langsung terlihat, kondisi ini dapat berpengaruh jangka panjang, seperti masalah dalam perkembangan postur tubuh pada dewas serta gangguan kesehatan reproduksi. Orang tua memegang peran penting dalam upaya pencegahan *stunting* pada anak-anak[7]. Faktor-faktor dalam keluarga, seperti pola pemenuhan gizi yang tidak seimbang, ketersediaan fasilitas kesehatan dasar dan pola pengasuhan yang kurang baik menjadi penyebab terjadinya *stunting*. Oleh karena itu, mengurangi jumlah keluarga beresiko *stunting* sangatlah krusial[8].

Mengatasi resiko *stunting* pada keluarga adalah salah satu fokus pemerintah melalui bantuan dan pendampingan. Akan tetapi, sejumlah program pemerintah masih belum efektif dalam menjangkau target yang tepat[9]. Dengan banyaknya data yang didapatkan dari keluarga beresiko *stunting* maka digunakanlah metode *data mining* untuk mengolah data yang sudah ada dan kemudian nantinya dapat menetukan suatu keluarga beresiko *stunting* atau tidak. Kampung KB (Kampung Berkualitas) Tunas Harapan merupakan salah satu contoh yang berperan dalam menangani *stunting* di kecamatan Tuah Madani kota Pekanbaru.

Berbagai penelitian mengenai *stunting* telah dilakukan, dengan salah satu metode yang umum digunakan yaitu data mining. Teknik data mining merupakan proses pengolahan data yang terkandung dalam data. Metode ini dibagi menjadi beberapa kategori berdasarkan tujuannya, seperti deskripsi, estimasi, prediksi, klasifikasi, pengelompokan dan asosiasi[10]. Penelitian ini mengalipkasikan teknik klasifikasi yaitu sebuah teknik yang menetapkan sebuah objek ke dalam kelas yang sudah ditentukan[11]. Jenis algoritma untuk klasifikasi diantaranya adalah *K-Nearest Neighbor* dan *Modified K-Nearest Neighbor*.

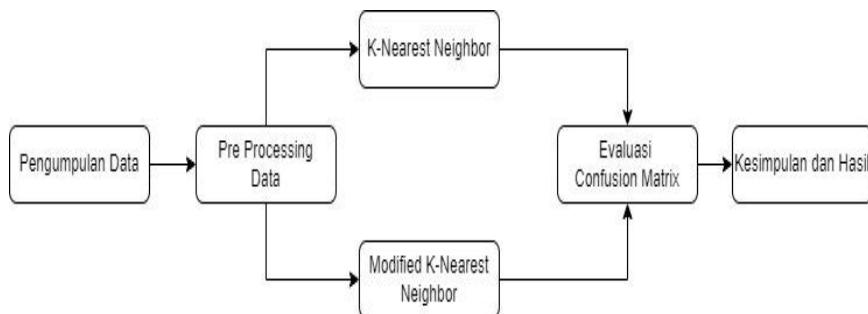
K-NN adalah metode yang banyak digunakan untuk melakukan klasifikasi dengan menentukan kelas suatu objek berdasarkan tetangga terdekat atau nilai k. Namun, *K-NN* memiliki beberapa keterbatasan dalam pemilihan nilai k yang menyebabkan akurasi lebih rendah. *MK-NN* merupakan pengembangan metode dari *K-NN* yang dirancang untuk mengatasi masalah akurasi pada *K-NN* dengan menambahkan proses validasi data dan proses penambahan bobot[12]. Menurut[13], meskipun *MK-NN* lebih unggul dari *K-NN* dalam beberapa aspek, algoritma *MK-NN* justru bisa memiliki akurasi rendah dibandingkan *K-NN* pada nilai k tertentu. Perbandingan keduanya dilakukan dengan tujuan memberikan gambaran lebih jelas mengenai metode mana yang lebih cocok dalam membantu dalam memilih algoritma yang memberikan hasil optimal dalam klasifikasi keluarga beresiko *stunting*.

Berdasarkan penelitian sebelumnya menggunakan metode NBC dan *K-NN* mendapatkan hasil dimana NBC sebesar 90,94% dan *K-NN* sebesar 96,10% yang mana *K-NN* lebih baik dibandingkan dengan NBC. Selanjutnya, penelitian menggunakan *MK-NN* yang menghasilkan akurasi sebesar 85,10% dengan nilai k=5.

Berdasarkan permasalahan yang sudah dijelaskan, penelitian ini bertujuan untuk mengklasifikasikan keluarga beresiko *stunting* dengan data yang sudah didapatkan menggunakan perbandingan metode *K-Nearest Neighbor* dan *Modified K-Nearest Neighbor*. Penelitian ini akan mengukur tingkat akurasinya dalam melakukan perbandingan keluarga beresiko *stunting*.

2. METODE PENELITIAN

Pada penelitian kali ini dibagi menjadi lima tahapan, yang dimulai dari pengumpulan data, pre processing, Klasifikasi K-Nearest Neighbor dan Modified K-Nearest Neighbor, Evaluasi Confusion Matrix dan diakhiri dengan kesimpulan. Alur penelitian dapat dilihat pada Gambar 1 yang berupa flowchart.



Gambar 1. Flowchart tahap penelitian

2.1. Pengumpulan Data

Pengumpulan data adalah suatu tahapan yang memiliki tujuan untuk mendapatkan informasi berdasarkan fakta yang mendukung terkait dengan penelitian[14]. Adapun data yang digunakan dalam penelitian ini merupakan data yang didapatkan dari balai Penyuluhan KB di Kecamatan Tuah Madadi berdasarkan pendataan keluarga tahun 2021. Data yang digunakan keluarga beresiko *stunting* yang berupa kumpulan data data tiap kelapa keluarga di Kecamatan Tuah Madani.

2.2. Pre Processing

Tahap *pre processing* merupakan tahap pertama dalam proses pengolahan data[15]. Langkah-langkah ini mencakup tindakan penting untuk segera dilakukan agar memastikan efisiensi waktu dan kerjaan. Proses *Knowledge Discovery* terdiri dari beberapa tahapan yang harus dijalani pada data agar dapat diproses lebih lanjut dalam *data mining*. *Pre processing* atau proses mempersiapkan data dari data mentah diubah menjadi format data yang siap digunakan. Pada tahap *pre processing* ini dilakukan proses *cleaning* data dan transformasi data.

2.3. Klasifikasi K-Nearest Neighbor

K-Nearest Neighbor (K-NN) adalah metode klasifikasi yang menentukan kategori objek baru berdasarkan tetangga terdekat,yang diatur oleh parameter K. metode ini merupakan algoritma *Supervised Learning*, di mana klasifikasi ditentukan oleh kategori mayoritas dalam diantara tetangga terdekat. Dalam metode ini, vektor fitur disimpan dan digunakan untuk klasifikasi pada data pelatihan. Pada fase klasifikasi, fitur yang relevan dihitung untuk data uji. Jarak antara vektor baru dan semua vektor data dihitung[16], logika *K-NN* adalah menjelajahi lingkungan sekitar, menganggap titik data pengujian serupa dengan titik data tersebut dan didapatkan keluaran[17]. Dihitung menggunakan *Manhattan distance*. *Manhattan distance* adalah metode untuk mencari jarak minimal dari dua titik[18] dengan rumus perhitungan sebagai berikut

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

Dimana:

d : Jarak antara x dan y

- i : Total banyak data latih
xi : Sampel data
yi : Data uji yang dilatih

2.4. Klasifikasi Modified K-Nearest Neighbor

Metode *MK-NN* tidak jauh berbeda dengan *K-NN* yang juga memiliki proses perhitungan jarak dan menghitung kedekatan antara data uji dan data latih yang besarnya bergantung pada parameter K. Pada metode *MK-NN* ada penambahan beberapa proses yaitu perhitungan nilai validitas dan perhitungan bobot[19]

$$Validitas_{(x)} = \frac{1}{k} \sum_{i=0}^n S(\text{lbl}(x), \text{lbl}(N_i(x))) \quad (2)$$

Dimana:

- i : Total banyak data latih
k : Banyak titik terdekat
 $\text{lbl}(x)$: Kelas x pada label x
 $\text{lbl}(N_i(x))$: Label kelas titik terdekat pada x

sebagai Langkah untuk mempermudah, ditetapkan fungsi “S” pada titik “x” dan data tetangga terdekat ke-i. Fungsi S dapat diperoleh dengan Persamaan berikut.

$$S(a, b) = \begin{cases} 1 & \text{jika } a = b \\ 0 & \text{jika } a \neq b \end{cases} \quad (3)$$

Dimana:

- s : Kemiripan tiap data latih
a : Kelas a pada data latih
b : Kelas lainnya selain kelas a pada data latih

Menghitung persamaan bobot dengan persamaan berikut.

$$W^{(i)} = Validitas^{(x)} x \frac{1}{d+a} \quad (4)$$

Dimana:

- w(i) : Weight Voting Validity
(i) : Nilai validitas
d : Nilai jarak
a : 0,5 untuk menghindari nilai 0

2.5. Evaluasi Confusion Matrix

Confusion matrix merupakan salah satu metrik yang digunakan oleh para peneliti untuk menilai efektivitas Teknik pembelajaran mesin, khususnya dalam penelitian *Supervised Learning* [20]. Melalui *Confusion Matrix* dapat menghitung presisi, akurasi, recall, f1-score[21].

Tabel 1. Tabel Confusion Matrix

PREDICTION		
ACTUAL	TRUE	FALSE
TRUE (POSITIVE)	TP	FN
FALSE (NEGATIVE)	FP	TN

Sebagai representasi dari hasil proses klasifikasi, terdapat empat istilah yang akan digunakan untuk mengukur kinerja *Confusion Matrix*. Jumlah data negatif yang diprediksi benar adalah angka *true negative* (TN) sedangkan *false positive* (FB) adalah data negatif namun diidentifikasi sebagai. *True positive* (TP) adalah data positif yang diprediksi benar, sedangkan *false negative* (FN) adalah data positif namun diidentifikasi sebagai negatif[22]. Keterangan dan penjelasan tentang *Confusion Matrix* adalah sebagai berikut:

$$1. \text{ Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \quad (5)$$

$$2. \text{ Presisi} = \frac{TP}{FP+TP} * 100\% \quad (6)$$

$$3. \text{ Recall} = \frac{TP}{FN+TP} * 100\% \quad (7)$$

$$4. \text{ F1-Score} = \frac{Recall*Presisi}{Recall+Presisi} * 2 \quad (8)$$

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Pengumpulan data dilakukan di Balai Penyuluhan KB di Kecamatan Tuah Madani. 7 dari 20 total parameter ditampilkan pada Tabel 2.

Tabel 2. Data Keluarga beresiko stunting

No.	Kode Keluarga	NIK	Nama Kepala Keluarga	Anak < 2 tahun	Anak < 5 tahun	Pasangan Usia Subur	Pus Hamil	...	Keluarga berpotensi stunting
1	147...	320...	A...Muna...	x	x	v	x	...	x
2	147...	140...	Aa...Musl...	v	v	v	x	...	v
3	147...	140...	A...Sup...	x	x	v	x	...	x
4	147...	149...	Ab...	x	v	v	x	...	v
...
23607	147...	147...	Zu.r Di..	x	v	v	x	...	v

Data yang didapatkan berjumlah 20 parameter yang berhubungan dengan faktor terjadinya stunting pada keluarga. Berikut seluruh atribut yang terdapat pada parameter bisa dilihat pada Tabel 3.

Tabel 3. Seluruh atribut parameter *stunting*

No.	Atribut
1.	Nomor data
2.	Kode Keluarga
3.	Nomor Induk Kepala Keluarga
4.	Nama Kepala Keluarga
5.	Baduta
6.	Balita
7.	Pus
8.	Pus hamil
9.	Tidak ada anggota keluarga yang memiliki sumber penghasilan untuk kebutuhan pokok
10.	Jenis lantai tanah
11.	Tidak setiap anggota makan makanan beragam
12.	Keluarga pra sejahtera
13.	Keluarga tidak memiliki sumber air minum yang layak
14.	Keluarga tidak memiliki jamban yang layak
15.	Keluarga tidak memiliki rumah layak huni
16.	Pendidikan terakhir ibu dibawah sltp
17.	Terlalu muda
18.	Terlalu tua
19.	Terlalu dekat
20.	Terlalu banyak

3.2 Pre Processing Data

Pada tahapan selanjutnya adalah menyiapkan data untuk digunakan dalam menganalisis *K-NN* dan *MK-NN*. *Cleaning* data, transformasi data dilakukan pada fase *pre processing*. Performa keakuratan antara metode *K-NN* dan *MK-NN* dapat ditingkatkan apabila melakukan *pre processing* data dengan benar

a. Cleaning Data

Pada tahap ini, menghapus parameter yang tidak diperlukan dalam proses klasifikasi. Parameter yang dihapus adalah kode keluarga, NIK dan nama kepala keluarga karena termasuk kategorikal untuk klasifikasi dengan model *K-NN* dan *MK-NN* menggunakan numerik sehingga kategorikal harus dihapus karena akan mempengaruhi hasil klasifikasi. Hasil cleaning data dapat dilihat pada Tabel 4.

Tabel 4. Hasil setelah *Cleaning data*

No.	Anak < 2 tahun	Anak < 5 tahun	Pasangan Subur	Pasangan Subur Hamil	Tidak ada sumber penghasilan	Lantai rumah jenis tanah	Tidak ada yang makan lebih dari 2 jenis	...	Keluarga berpotensi stunting
1	x	x	v	x	v	x	x	...	x
2	x	v	v	x	v	x	x	...	v
3	x	x	v	x	v	x	x	...	x
4	x	x	v	x	x	x	x	...	v
...
23607	x	v	v	v	v	x	x	...	v

b. Transformasi Data

Metode *K-NN* dan *MK-NN* menginput tugas ke kelas dengan jarak terkecil. Karena itu, simbol-simbol yang mewakili nilai kelas dan parameter diubah menjadi bentuk angka sehingga saat proses klasifikasi *K-NN* dan *KM-NN* dapat mengidentifikasi data dengan benar. Simbol x diubah menjadi angka 0 dan simbol v diubah menjadi angka 1, dan pada label diubah menjadi 1 dan 2 untuk membedakan nya dengan kelas. Hasil transformasi dapat dilihat pada Tabel 5.

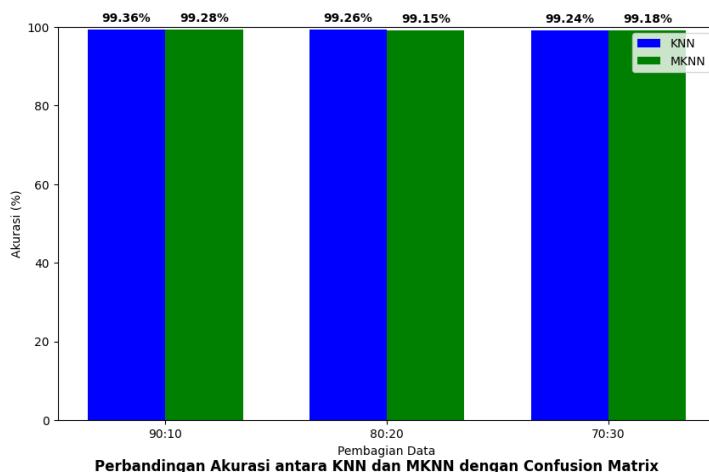
Tabel 5. Tabel setelah Transformasi data

No.	Anak < 2 tahun	Anak < 5 tahun	Pasangan Usia Subur	Pasangan Usia Subur Hamil	Tidak ada sumber penghasilan	Lantai rumah jenis tanah	Tidak ada yang makan lebih dari 2 jenis	...	Keluarga berpotensi stunting
1	0	0	1	0	1	0	0	...	2
2	0	1	1	0	1	0	0	...	1
3	0	0	1	0	1	0	0	...	2
4	0	0	1	0	0	0	0	...	1
...
23607	0	1	1	1	1	0	0	...	1

Ada 23607 data pada data set dengan 16 parameter yang akan dianalisis. Kelas 1 terdiri dari 17500 data yang mewakili keluarga beresiko *stunting*, sedangkan kelas 2 terdiri dari 6170 data mewakili keluarga tidak beresiko *stunting*.

3.3 Klasifikasi K-Nearest Neighbor dan Modified K-Nearest Neighbor

Tiga percobaan telah dilakukan dalam klasifikasi *K-Nearest Neighbor* dan *Modified K-Nearest Neighbor* dengan menggunakan skema untuk training dan juga data testing. Dapat dilihat pada Gambar.2 pembagian data yang memperoleh akurasi klasifikasi tertinggi 99.36% dan 99.28% menggunakan data training sebesar 10% dimana tingkat akurasi *K-NN* lebih baik dibandingkan dengan *MK-NN*. Meskipun nilai akurasi dan performa *K-NN* lebih baik dibandingkan *MK-NN* di beberapa nilai k yaitu k=5,7,9 namun *MK-NN* cukup stabil dalam beberapa nilai k lainnya. Percobaan juga dilakukan dengan menggunakan nilai k yang berbeda. Tabel 6 dan 7 memperlihatkan hasil perbandingan akurasi *K-NN* dan *MK-NN* dengan nilai k yang berbeda.



Gambar 2. Hasil perbandingan akurasi *K-NN* dan *MK-NN*

Tabel 6. Hasil pengujian nilai k K-NN

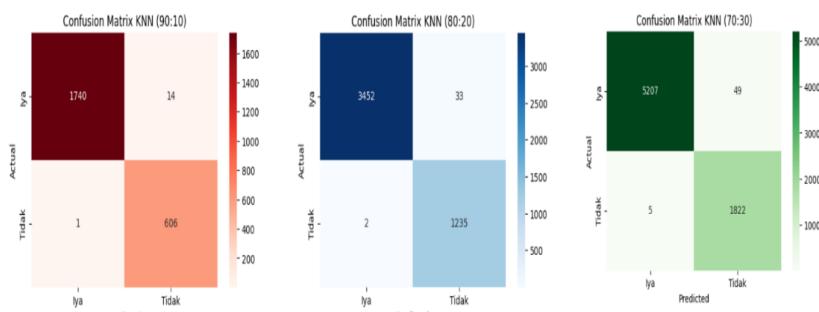
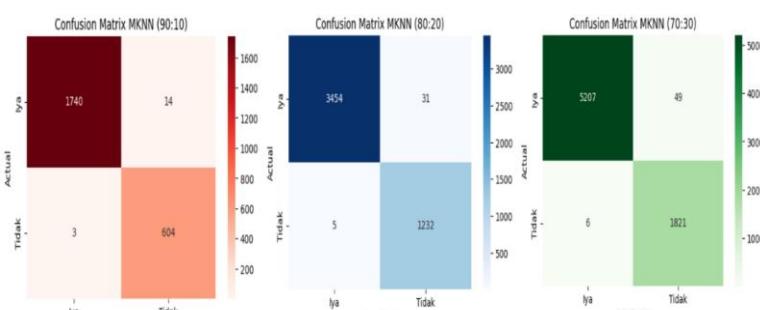
k	90:10	80:20	70:30
3	99.28%	99.15%	99.22%
5	99.36%	99.26%	99.24%
7	99.36%	99.26%	99.12%
9	99.36%	99.22%	99.11%
11	99.24%	97.17%	99.10%
13	99.24%	99.15%	99.05%
15	99.24%	99.13%	99.00%

Tabel 7. Hasil pengujian nilai k MK-NN

k	90:10	80:20	70:30
3	99.28%	99.26%	99.18%
5	99.28%	99.15%	99.18%
7	99.28%	99.15%	99.18%
9	99.28%	99.24%	99.22%
11	99.28%	99.24%	99.22%
13	99.28%	99.24%	99.22%
15	99.28%	99.24%	99.22%

3.4 Pengujian

Dalam mengevaluasi kinerja metode yang sudah dikembangkan, dilakukanlah pengujian pada 23607 data yang dibagi menjadi data latih dan data uji. Pembagian dilakukan dengan menggunakan tiga skenario yang berbeda. Skenario yang berbeda. Skenario yang pertama menggunakan data latih 90% dan data uji 10%. Skenario kedua menggunakan data latih 80% dan data uji 20%. Skenario ketiga menggunakan data latih 70% dan data uji 30%. Berikut pengujian confusion matrix dapat dilihat pada gambar 3 dan 4.

**Gambar 3.** Confusion Matrix seluruh skenario K-NN**Gambar 4.** Confusion Matrix seluruh skenario MK-NN

Tabel 8,9,10 memperlihatkan hasil dari pengujian menggunakan *confusion matrix* untuk setiap pembagian data. Hasil dapat diperoleh dengan cara menghitung menggunakan rumus yang telah dijelaskan sebelumnya dan juga melihat elemen-elemen yang terdapat dalam *confusion matrix*.

Tabel 8. Pengujian *Confusion Matrix* dengan rasio 90:10

K-NN (k)	Accuracy	Precision	Recall	F1-Score	MK-NN (k)	Accuracy	Precision	Recall	F1-score
3	99.28%	97.73%	99.51%	98.61%	3	99.28%	97.73%	99.51%	98.61%
5	99.36%	97.74%	99.84%	98.78%	5	99.28%	97.73%	99.51%	98.61%
7	99.36%	97.74%	99.84%	98.78%	7	99.28%	97.73%	99.51%	98.61%
9	99.36%	97.74%	99.84%	98.78%	9	99.28%	97.73%	99.51%	98.61%
11	99.24%	97.27%	99.84%	98.54%	11	99.28%	97.73%	99.51%	98.61%
13	99.24%	97.27%	99.84%	98.54%	13	99.28%	97.73%	99.51%	98.61%
15	99.24%	97.27%	99.84%	98.54%	15	99.28%	97.73%	99.51%	98.61%

K-NN memiliki sedikit keunggulan pada beberapa nilai k (terutama pada nilai k=5,7,9) dengan akurasi maksimum yaitu 99.36%. Sementara MK-NN memiliki akurasi dibawah K-NN di beberapa nilai k jika dibandingkan, namun nilai akurasi tetap konsisten pada angka 99.28%. Untuk rasio ini K-NN sedikit lebih baik dalam hal akurasi, namun perbedaannya kecil sehingga tidak signifikan dibandingkan MK-NN yang memiliki stabilitas di semua nilai k.

Tabel 9. Pengujian *Confusion Matrix* dengan rasio 80:20

K-NN (k)	Accuracy	Precision	Recall	F1-Score	MK-NN (k)	Accuracy	Precision	Recall	F1-score
3	99.15%	97.01%	99.84%	98.41%	3	99.26%	97.55%	99.68%	98.60%
5	99.26%	97.40%	99.84%	98.60%	5	99.15%	97.01%	99.84%	98.41%
7	99.26%	97.40%	99.84%	98.60%	7	99.15%	97.01%	99.84%	98.41%
9	99.22%	97.24%	99.84%	98.52%	9	99.24%	97.47%	99.68%	98.56%
11	97.17%	97.09%	99.84%	98.45%	11	99.24%	97.47%	99.68%	98.56%
13	99.15%	97.01%	99.84%	98.41%	13	99.24%	97.47%	99.68%	98.56%
15	99.13%	96.94%	99.84%	98.37%	15	99.24%	97.47%	99.68%	98.56%

MK-NN memiliki keunggulan dengan nilai tertinggi mencapai 99.26% dibandingkan K-NN yang nilainya selalu naik turun dan juga tidak seimbang. MK-NN juga lebih stabil di berbagai nilai k, terutama pada k=3 dan k=11 hal ini ditunjukkan dengan kemampuan menjaga keseimbangan precision dan recall

Tabel 10. Pengujian *Confusion Matrix* dengan rasio 70:30

K-NN (k)	Accuracy	Precision	Recall	F1-Score	MK-NN (k)	Accuracy	Precision	Recall	F1-score
3	99.22%	97.38%	99.67%	98.51%	3	99.18%	97.02%	99.89%	98.44%
5	99.24%	97.33%	99.78%	98.54%	5	99.18%	97.02%	99.89%	98.44%
7	99.12%	96.92%	99.78%	98.33%	7	99.18%	97.02%	99.89%	98.44%
9	99.11%	96.87%	99.78%	98.30%	9	99.22%	97.38%	99.67%	98.51%
11	99.10%	96.81%	99.78%	98.27%	11	99.22%	97.38%	99.67%	98.51%
13	99.05%	96.66%	99.78%	98.20%	13	99.22%	97.38%	99.67%	98.51%
15	99.00%	96.46%	99.78%	98.09%	15	99.22%	97.38%	99.67%	98.51%

Nilai akurasi pada MK-NN unggul tipis dengan kestabilan di angka 99.22% dibandingkan dengan K-NN yang sedikit menurun pada K=7,9,11,13,15. MK-NN juga unggul dalam nilai F1-Score, terutama pada k=9 hingga k=15 yang menunjukkan keunggulan MK-NN dalam menangani data yang berukuran lebih besar. MK-NN lebih stabil dan akurat terutama jika rasio data uji meningkat.

3.5 Pembahasan

Penelitian ini membandingkan metode *K-NN* dan *MK-NN* untuk klasifikasi keluarga beresiko stunting. Hasil menunjukkan bahwa *MK-NN* unggul dalam stabilitas akurasi dengan nilai konsisten 99.28%, sementara *K-NN* mencapai akurasi maksimum 99.36% namun menunjukkan fluktuasi pada nilai parameter k tertentu. Stabilitas *MK-NN* didukung oleh proses validasi data tambahan dan pemberian bobot yang membantu *MK-NN* menghasilkan angka yang lebih stabil, terutama kondisi data yang memiliki variasi nilai k yang tinggi. Stabilitas ini yang menjadi kelebihan utama *MK-NN* dibandingkan dengan *K-NN* meskipun proses ini menambah waktu komputasi.

Pembagian rasio data latih uji turut memberikan pengaruh signifikan terhadap performa kedua algoritma. Pada rasio 90:10 *K-NN* unggul dalam akurasi, tetapi pada rasio 80:20 dan 70:30 terjadi perbedaan yang semakin mengecil. Pada rasio data yang lebih besar untuk pengujian, *MK-NN* mampu menjaga kestabilan metrik evaluasi seperti presisi, recall dan f1-score yang menunjukkan kemampuan menangani data uji dalam jumlah lebih banyak. Secara rinci *K-NN* memiliki keunggulan dalam akurasi saat digunakan dataset kecil dengan rasio data latih yang lebih besar. Di sisi lain *MK-NN* dengan penambahan bobot serta validasi data dapat diandalkan dalam menangani dataset yang lebih besar dan kompleks.

Rekomendasi untuk penelitian mendatang mencakup penambahan data ekonomi keluarga, seperti pendapatan dan pekerjaan orang tua perlu dievaluasi secara mendalam untuk hasil klasifikasi yang lebih menyeluruh, penambahan data tersebut dapat memberikan dampak signifikan terhadap akurasi model klasifikasi yang dihasilkan dan menjadi lebih relevan serta dapat memberikan data pendukung yang lebih kuat untuk perumusan kebijakan yang efektif dalam menangani *stunting*.

4. KESIMPULAN

Penelitian ini menunjukkan *K-NN* dan *MK-NN* mampu mengklasifikasikan keluarga beresiko *stunting* dengan baik. *MK-NN* secara konsisten unggul dalam menjaga akurasi tinggi, dengan stabilitas yang terlihat di seluruh rasio data (90:10, 80:20, 70:30) dan berbagai nilai k. Hasil menunjukkan bahwa *MK-NN* mencapai akurasi maksimum hingga 99,28% secara konsisten, sementara *K-NN* menunjukkan fluktuasi atau perubahan angka yang tidak stabil meski mencapai akurasi tertinggi pada nilai k tertentu. Selain akurasi, *MK-NN* menunjukkan keunggulan dalam menjaga keseimbangan metrik evaluasi lainnya meskipun rasio data uji meningkat.

Berbeda dengan *K-NN* yang mengalami penurunan saa nilai k meningkat dan data uji bertambah. Proses validasi dan penambahan bobot pada *MK-NN* menunjukkan kelebihannya dalam melakukan klasifikasi tidak hanya akurat tetapi juga dapat diandalkan dalam skenario klasifikasi yang kompleks. Menjadikan *MK-NN* adalah metode yang lebih baik dalam mendeteksi keluarga beresiko *stunting*.

Namun, penelitian ini memiliki keterbatasan dalam parameter yang digunakan. Data yang digunakan belum mencakup faktor ekonomi keluarga, yang merupakan salah satu faktor penting dalam mempengaruhi kejadian *stunting*. Oleh karena itu, saran untuk penelitian selanjutnya agar data yang digunakan melibatkan aspek ekonomi, seperti pendapatan keluarga, status pekerjaan orang tua. Faktor-faktor ekonomi ini memiliki dampak signifikan terhadapa pola konsumsi gizi dan akses terhadap layanan kesehatan. Dengan memasukkan data ekonomi, penelitian selanjutnya dapat memberikan gambaran yang lebih menyeluruh mengenai hubungan antara kondisi ekonomi keluarga dan resiko *stunting*. Penambahan parameter ini akan meningkatkan akurasi model klasifikasi dan memberikan data yang lebih kuat untuk perumusan kebijakan yang efektif dalam mengatasi *stunting*.

Daftar Pustaka

- [1] E. Marlena, S. Rahmatullah, S. Mintoro, and N. Ngajiyanto, "Implementasi Data Mining Untuk Memprediksi Status Gizi Anak Balita Pada Puskesmas Gedung Sari Menggunakan Polynomial Regression," *JTKSI (Jurnal Teknol. Komput. dan Sist. Informasi)*, vol. 6, no. 2, p. 193, 2023, doi: 10.56327/jtksi.v6i2.1490.
- [2] I. Y. Arulampalam Kunaraj, P. Chelvanathan, Ahmad AA Bakar, "Lailiyah, K. (2023). Peran Badan Kependudukan Dan Keluarga Berencana Nasional Dalam Percepatan Penurunan Stunting . Mendapo: Journal of Administrative Law, 4(1), 16-33.," *J. Eng. Res.*, vol. 4, 2023.
- [3] R. Roediger, D. Taylor Hendrixson, and M. J. Manary, "A roadmap to reduce stunting," *Am. J. Clin. Nutr.*, vol. 112, pp. 773S-776S, 2020, doi: 10.1093/ajcn/nqaa205.
- [4] A. Zikri, A. Nazir, S. Sanjaya, E. Haerani, and I. Afrianty, "The Random Forest algorithm for classifying stunting in toddlers based on anthropometric data," *Int. J. Multidiscip. Res. Growth Eval.*, vol. 5, no. 3, pp. 931–937, 2024, doi: 10.54660/ijmrge.2024.5.3.931-937.
- [5] A. Ramdhani, H. Handayani, and A. Setiawan, "Hubungan Pengetahuan Ibu Dengan Kejadian Stunting," *Semnas Lppm*, vol. ISBN: 978-, pp. 28–35, 2020.
- [6] S. Widia Pebrianti, R. Astuti, and F. M Basysyar, "Penerapan Algoritma K-Nearest Neighbor Dalam Klasifikasi Status Stunting Balita Di Desa Bojongemas," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 2, pp. 2479–2488, 2024, doi: 10.36040/jati.v8i2.8448.
- [7] E. E. P. Agustina and R. Dwijayanti, "Peran Orang Tua Dalam Pencegahan Stunting Melalui Program Sekolah Orang Tua Hebat Di Kelurahan Lakarsantri Ellyssia Eka Putri Agustina Rizky Dwijayanti infeksi ibu , kehamilan ibu saat remaja , gangguan jiwa ibu , jarak antar kehamilan yang pendek , anak," *An-Najat J. Ilmu Farm. dan Kesehat.*, vol. 1, no. 4, pp. 220–227, 2023.
- [8] S. Mustika and C. Khairunnisa, "Prevalensi Stunting pada Siswa SMP Negeri 7 Lhokseumawe Abstrak Pendahuluan malnutrisi zat gizi kronis atau penyakit infeksi kronis maupun berulang yang ditunjukkan Republik Indonesia (Kemenkes RI) hanya terbatas pada kelompok usia Balita . Penelitian," *Galen. J. Kedokt. dan Kesehat. Mhs. Malikussaleh*, vol. 1, no. 4, 2022.
- [9] A. Aziz, F. Insani, J. Jasril, and F. Syafria, "Implementasi Metode Learning Vector Quantization (LVQ) Untuk Klasifikasi Keluarga Beresiko Stunting," *Build. Informatics, Technol. Sci.*, vol. 5, no. 1, pp. 12–21, 2023, doi: 10.47065/bits.v5i1.3478.
- [10] A. O. P. Dewi, "Big Data di Perpustakaan dengan Memanfaatkan Data Mining," *Anuva J. Kaji. Budaya, Perpustakaan, dan Inf.*, vol. 4, no. 2, 2020, doi: 10.14710/anuva.4.2.223-230.
- [11] W. A. Istiqhfarani, I. Cholissodin, and F. A. Bachtiar, "Klasifikasi Penyakit Dental caries menggunakan Algoritme Modified K- Nearest Neighbor," *J. Pengemb. Teknol. Inf. dan*

- Ilmu Komput.*, vol. 4, no. 5, 2020.
- [12] S. Amelia, “Penerapan Metode Modified K-Nearest Neighbor pada Pengklasifikasian Status Pembayaran Kredit Barang Elektronik dan Furniture,” *Stat. J. Theor. Stat. Its Appl.*, vol. 22, no. 1, pp. 95–104, 2022, doi: 10.29313/statistika.v22i1.345.
- [13] I. Muslim, K. Karo, A. Tsany, R. Dzaky, and M. A. Saputra, “Comparative Analysis of K-Nearest Neighbor and Modified K-Nearest Neighbor Algorithm for Financial Well-Being Data Classification,” *J. Comput.*, vol. 6, no. 3, pp. 25–34, 2021, doi: 10.34818/indojc.2021.6.3.593.
- [14] M. Nafisatur, “Metode Pengumpulan Data Penelitian,” *Metod. Pengumpulan Data Penelit.*, vol. 3, no. 5, pp. 5423–5443, 2024.
- [15] F. Alghifari and D. Juardi, “Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes,” *J. Ilm. Inform.*, vol. 9, no. 02, pp. 75–81, 2021, doi: 10.33884/jif.v9i02.3755.
- [16] E. Sulistio, Y. Maulita, and M. Simanjuntak, “Classification For Predicting Heart Disease Using The K Nearest Neighbor Method Sylvani General Hospital Binjai City,” vol. 3, no. 1, 2023.
- [17] F. Handayani and R. M. Taufiq, “Jurnal Computer Science and Information Technology (CoSciTech) Komparasi Algoritma Menggunakan Teknik SMOTE Dalam Melakukan Klasifikasi Penyakit,” vol. 5, no. 2, pp. 367–372, 2024.
- [18] H. S. Amalia, U. Athiyah, and A. W. Muhammad, “The Application of Modified K-Nearest Neighbor Algorithm for Classification of Groundwater Quality Based on Image Processing and pH, TDS, and Temperature Sensors,” *Regist. J. Ilm. Teknol. Sist. Inf.*, vol. 9, no. 1, pp. 42–54, 2023, doi: 10.26594/register.v9i1.2827.
- [19] I. N. Y. T. Giri, L. A. A. Rahning Putri, G. A. V. Mastrika Giri, I. G. N. Anom Cahyadi Putra, I. M. Widiartha, and I. W. Supriana, “Music Genre Classification Using Modified K-Nearest Neighbor (MK-NN),” *JELIKU (Jurnal Elektron. Ilmu Komput. Udayana)*, vol. 10, no. 3, p. 261, 2022, doi: 10.24843/jlk.2022.v10.i03.p02.
- [20] L. Sari, A. Romadloni, and R. Listyaningrum, “Penerapan Data Mining dalam Analisis Prediksi Kanker Paru Menggunakan Algoritma Random Forest,” *Infotekmesin*, vol. 14, no. 1, pp. 155–162, 2023, doi: 10.35970/infotekmesin.v14i1.1751.
- [21] W. A. Naseer, S. Sarwido, and B. B. Wahono, “Gradient Boosting Optimization with Pruning Technique for Prediction of Bmt Al-hikmah Permata Customer Data,” *Jinteks*, vol. 6, no. 3, pp. 719–727, 2024.
- [22] P. Romadloni, B. Adhi Kusuma, and W. Maulana Baihaqi, “Komparasi Metode Pembelajaran Mesin Untuk Implementasi Pengambilan Keputusan Dalam Menentukan Promosi Jabatan Karyawan,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 6, no. 2, pp. 622–628, 2022, doi: 10.36040/jati.v6i2.5238.



ZONAsi: Jurnal Sistem Informasi

Is licensed under a [Creative Commons Attribution International \(CC BY-SA 4.0\)](https://creativecommons.org/licenses/by-sa/4.0/)