

PENERAPAN ALGORITMA K-NEAREST NEIGHBOR (KNN) UNTUK KLASIFIKASI STATUS GIZI BALITA DI KECAMATAN RUMBAI TIMUR

Marshanda¹, Nurliana Nasution^{1*}

¹Teknik Informatika, Universitas Lancang Kuning, Pekanbaru Riau

*email : nurliana@unilak.ac.id

Abstract: *This study aims to classify the nutritional status of toddlers based on anthropometric data using the K-Nearest Neighbor (KNN) algorithm. Data were obtained from 20 Integrated Health Posts (Posyandu) in Rumbai Timur District, including Lembah Sari Village and Limbungan Village with a total of 1,000 toddler data. After cleaning and preprocessing, 782 data were obtained ready for use. The preprocessing stages include data cleaning and transformation, outlier removal, minority class handling, and data normalization. Next, data balancing was carried out using the Synthetic Minority Oversampling Technique (SMOTE) to address class imbalance. The data was divided into 80% training data and 20% test data, then the K parameter was tested from 1 to 15 using 5-fold cross-validation. The results showed that the value of $K = 1$ provided the best performance with a macro recall of 0.8827 and an accuracy of 86.26%. These results indicate that the combination of the KNN algorithm with the SMOTE method and Min-Max normalization is effective in improving classification performance on imbalanced data and producing accurate and balanced predictions of toddler nutritional status between classes.*

Keywords: *k-nearest neighbor; toddler nutritional status; SMOTE; min-max scaling; classification; anthropometric data*

Abstrak: Penelitian ini bertujuan untuk mengklasifikasikan status gizi balita berdasarkan data antropometri menggunakan algoritma *K-Nearest Neighbor* (KNN). Data diperoleh dari 20 Posyandu di Kecamatan Rumbai Timur, meliputi Kelurahan Lembah Sari dan Kelurahan Limbungan dengan total 1.000 data balita. Setelah melalui proses *cleaning* dan *preprocessing*, diperoleh 782 data yang siap digunakan. Tahapan pra-pemrosesan meliputi pembersihan dan transformasi data, penghapusan outlier, penanganan kelas minoritas, serta normalisasi data. Selanjutnya dilakukan penyeimbangan data menggunakan *Synthetic Minority Oversampling Technique* (SMOTE) untuk mengatasi ketidakseimbangan kelas. Data dibagi menjadi 80% data latih dan 20% data uji, kemudian dilakukan pengujian parameter K dari 1 hingga 15 menggunakan *5-fold cross-validation*. Hasil penelitian menunjukkan bahwa nilai $K = 1$ memberikan performa terbaik dengan *recall macro* sebesar 0,8827 dan akurasi 86,26%. Hasil ini menunjukkan bahwa kombinasi algoritma KNN dengan metode SMOTE dan normalisasi *Min-Max* efektif dalam meningkatkan kinerja klasifikasi pada data tidak seimbang serta menghasilkan prediksi status gizi balita yang akurat dan seimbang antar kelas.

Kata kunci: *k-nearest neighbor; status gizi balita; SMOTE; min-max scaling; klasifikasi; data antropometri*

PENDAHULUAN

Perkembangan teknologi komputer saat ini telah mendorong terciptanya berbagai sistem yang dirancang untuk memenuhi kebutuhan manusia dan memudahkan aktivitas sehari-hari. Hal ini terutama diterapkan pada aktivitas-aktivitas yang kompleks dan membutuhkan tingkat ketelitian yang tinggi, seperti dalam hal pemantauan kebutuhan gizi. Jika kualitas dan kuantitas makanan tidak terpenuhi dengan baik, maka dapat terjadi kekurangan gizi yang berdampak pada terganggunya proses pertumbuhan dan perkembangan pada balita [1].

Prediksi status gizi yang tepat untuk anak sangat dibutuhkan guna mendapatkan penanganan dan memulihkan kesehatan. Dalam prediksi status gizi ini butuh keakuratan dalam mendiagnosa, hal itu dimaksudkan untuk mendapatkan hasil status gizi anak sesuai dengan kondisinya [2]. Apabila balita mengalami permasalahan terhadap gizi maka dikhawatirkan akan dapat mempengaruhi perkembangan psikologis dan fisiknya, hingga dapat menyebabkan kematian [3].

Salah satu langkah penting untuk meningkatkan kualitas kesehatan pada anak adalah dengan melakukan pemantauan status gizi secara rutin dan sistematis di fasilitas layanan seperti Posyandu. Di Posyandu, masyarakat bisa mendapatkan pelayanan seperti imunisasi dan pemeriksaan tumbuh kembang anak. Selain itu, Puskesmas juga berfungsi dalam membina masyarakat dalam meningkatkan kemampuan hidup sehat, pengumpulan serta pendataan, melaksanakan komunikasi serta informasi kepada masyarakat dalam bidang kesehatan. [4].

Posyandu yang berada di Kelurahan Lembah Sari dan Kelurahan Limbungan saat ini masih aktif dalam melakukan kegiatan pemantauan status gizi pada balita. Di setiap Posyandu tersebut, proses klasifikasi status gizi balita yang sedang berjalan saat ini masih dilakukan secara konvensional oleh tenaga kerja. Proses ini melibatkan pencatatan data antropometri balita, seperti usia, jenis kelamin, berat badan dan tinggi badan yang selanjutnya akan ditulis dalam buku besar. Namun, proses pencatatan kesehatan balita tersebut memiliki beberapa permasalahan, yaitu pencatatan yang dilakukan secara manual dapat menyebabkan kesalahan dalam memasukkan atau menghitung data, yang dapat berpengaruh pada hasil klasifikasi. Kemudian, metode yang digunakan masih bersifat konvensional, sehingga proses analisis dapat memakan waktu dan menjadi kurang efisien. Oleh karena itu, diperlukan suatu sistem klasifikasi berbasis teknologi yang mampu mengolah data balita secara terstruktur sehingga proses penentuan status gizi dapat dilakukan dengan lebih cepat, akurat, dan efisien.

Salah satu metode yang digunakan dalam proses klasifikasi adalah algoritma *K-Nearest Neighbor* (KNN). Metode ini dipilih karena memiliki konsep sederhana dan mampu mengklasifikasikan data berdasarkan kedekatan jarak antara data uji dan data latih. Dalam penelitian ini, KNN digunakan untuk mengklasifikasikan status gizi balita berdasarkan variabel jenis kelamin, usia, berat badan, dan tinggi badan sehingga diharapkan dapat menghasilkan klasifikasi status gizi secara lebih akurat.

Seiring dengan perkembangan teknologi informasi, berbagai penelitian telah dilakukan untuk mengklasifikasikan status gizi balita menggunakan metode data mining. Penelitian oleh Dewi Fitrianiingsih, Martaleli Bettiza, dan Alena Uperiati menerapkan algoritma *K-Nearest Neighbor* (KNN) untuk mengklasifikasikan status gizi balita berdasarkan pengukuran tubuh balita. Penelitian ini hanya menghasilkan akurasi

sebesar 73,53%, rendahnya performa algoritma KNN dikarenakan jumlah data yang digunakan relatif sedikit serta tidak adanya menerapkan teknik penyeimbangan data [5]. Kemudian, penelitian oleh Muzizah Annabaa' Aulia, Rito Goejantoro, dan Memi Nor Hayati juga menerapkan metode klasifikasi KNN. Penelitian ini menghasilkan akurasi yang cukup tinggi yaitu sebesar 90,57%, namun penelitian ini tidak menerapkan metode penyeimbangan data seperti SMOTE melainkan hanya menggunakan pra-pemrosesan data berupa randomisasi data dan standarisasi data menggunakan *Z-score*.

Selain itu, metode lain juga telah digunakan dalam penelitian sebelumnya. Penelitian oleh Harliana, Rizqi Darma Rusdian Yusron, dan Imam Machfud menggunakan metode *Naïve bayes* berbasis *Geographic Information System* (GIS) untuk klasifikasi dan monitoring status gizi balita [6]. Penelitian oleh Azkiyatun Nadroh, Deny Nugroho Triwibowo, dan R. Bagus Bambang Sumantri menggunakan metode *Support Vector Machine* (SVM) dengan optimasi parameter menggunakan *Grid Search* untuk meningkatkan akurasi klasifikasi status gizi balita berdasarkan data antropometri [7]. Penelitian oleh Harifa Hananti dan Kartika Sari melakukan perbandingan menggunakan metode *Support Vector Machine* (SVM) dan *Artificial Neural Network* (ANN) untuk pengklasifikasian dan mendapatkan metode terbaik pada gizi balita (gizi baik & gizi kurang) di Puskesmas Salisingan Sulawesi Barat [8]. Lalu, penelitian oleh Gina Purnama Insany, Indra Yustiana, dan Sri Rahmawati menerapkan metode KNN dan ANN pada klasifikasi status gizi balita berdasarkan indeks antropometri [9]. Kemudian, penelitian oleh Alter Lasarudin, Hilmansyah Gani, dan Misran Tomayahu melakukan perbandingan akurasi menggunakan metode *Naïve Bayes* dan C4.5 untuk mengklasifikasi status gizi pada bayi dan balita [10].

Walaupun berbagai penelitian telah dilakukan, sejumlah penelitian terdahulu masih menunjukkan keterbatasan, di antaranya jumlah dataset yang relatif terbatas serta belum dilaksanakannya proses *preprocessing* data secara lengkap. Selain itu, permasalahan ketidakseimbangan data (*imbalanced data*) dan keberadaan outlier juga belum sepenuhnya ditangani, yang pada akhirnya dapat memengaruhi kinerja model klasifikasi sehingga menjadi kurang optimal.

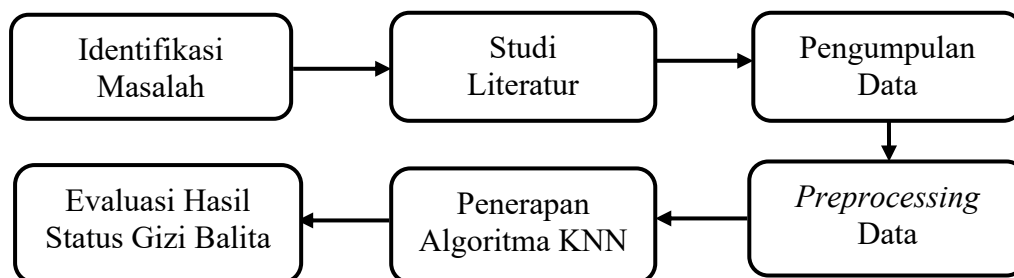
Oleh karena itu, penelitian ini memiliki kebaruan berupa penerapan algoritma *K-Nearest Neighbor* (KNN) menggunakan dataset balita sebanyak 1000 data yang diperoleh dari beberapa Posyandu di Kecamatan Rumbai Timur, disertai tahapan *preprocessing* data yang lebih menyeluruh. Proses *preprocessing* meliputi *cleaning* data, transformasi dan penghapusan outlier, normalisasi data, serta penanganan ketidakseimbangan kelas menggunakan metode *Synthetic Minority Over-sampling Technique* (SMOTE). Penelitian ini juga menangani kelas dengan jumlah data yang sangat terbatas, yaitu kategori gizi buruk yang hanya memiliki satu data, sehingga diharapkan mampu meningkatkan kualitas data sebelum proses klasifikasi dan menghasilkan akurasi yang lebih optimal dibandingkan penelitian sebelumnya.

Berdasarkan dari permasalahan di atas, penulis tertarik untuk melakukan penelitian dengan judul "Penerapan Algoritma *K-Nearest Neighbor* (KNN) Untuk Klasifikasi Status Gizi Balita di Kecamatan Rumbai Timur". Penelitian ini diharapkan dapat menghasilkan proses klasifikasi yang lebih akurat dan sistematis, serta dapat memberikan gambaran yang objektif bagi petugas kesehatan dalam melakukan evaluasi dan penanganan gizi pada balita secara lebih tepat.

METODE

Tahapan-Tahapan Penelitian

Tahapan penelitian menjelaskan urutan proses yang dilakukan secara terstruktur untuk mencapai hasil penelitian yang diharapkan. Adapun tahapan penelitian meliputi beberapa langkah yaitu sebagai berikut:



Gambar 1. Tahapan Penelitian

Tahapan penelitian dimulai dengan mengidentifikasi masalah di Kelurahan Lembah Sari dan Kelurahan Limbungan. Proses pengklasifikasian status gizi balita masih dilakukan secara manual oleh tenaga kerja, sehingga memerlukan waktu yang lama dan berisiko menimbulkan kesalahan. Selanjutnya dilakukan studi literatur untuk memperoleh pemahaman yang lebih mendalam terkait dengan konsep status gizi balita dan metode klasifikasi yang telah digunakan dalam penelitian sebelumnya.

Pengumpulan data dilakukan dengan cara mewawancarai petugas Posyandu yang dikunjungi serta mengumpulkan data-data balita dari setiap Posyandu. Data yang dikumpulkan terdiri dari 1000 dataset yang meliputi informasi seperti nama anak balita, jenis kelamin, tanggal/bulan/tahun lahir, usia, berat badan, tinggi badan, dan status gizi yang diketahui. Setelah itu, dilakukan *preprocessing* untuk meningkatkan kualitas dan integritas data sebelum digunakan dalam penerapan algoritma pembelajaran mesin. Proses *preprocessing* dalam penelitian ini mencakup pembersihan data (*cleaning data*), penanganan *missing values*, transformasi data, penghapusan outlier, penanganan kelas dengan jumlah data sedikit (gizi buruk), dan normalisasi data.

Kemudian, dilakukan penerapan algoritma *K-Nearest Neighbor* (KNN) yang bekerja dengan mencari sejumlah tetangga terdekat (K) dari data uji berdasarkan jarak *Euclidean*. Nilai K dipilih agar hasil klasifikasi tidak *overfitting* maupun terlalu umum. Evaluasi hasil dilakukan untuk menilai sejauh mana kinerja model K-NN dalam mengklasifikasikan status gizi balita. Pengukuran dilakukan menggunakan metrik evaluasi yaitu akurasi, presisi, *recall*, dan *F1-score* untuk melihat tingkat ketepatan model. Selain itu, *confusion matrix* digunakan untuk menganalisis seberapa baik model mengenali setiap kategori gizi dan mengidentifikasi kesalahan klasifikasi yang terjadi.

Metode / Pemodelan yang Digunakan

K-Nearest Neighbor (KNN) merupakan algoritma *supervised learning* yang digunakan untuk mengklasifikasikan status gizi balita berdasarkan atribut numerik seperti berat badan, tinggi badan, dan usia. Algoritma ini bekerja dengan menghitung jarak antara data uji dan data latih menggunakan *Euclidean Distance*, lalu menentukan

kelas berdasarkan mayoritas tetangga terdekat. KNN dipilih karena sederhana, akurat dan efektif untuk data berukuran kecil tanpa memerlukan proses pelatihan yang kompleks, serta komputasinya yang sederhana terhadap data yang banyak. Selain itu, KNN juga akan menghasilkan akurasi yang tepat dan tinggi ketika nilai K yang dipilih tepat, hal ini disebabkan karena KNN akan menghitung jarak terpendek dari *sample* uji ke *sample* latih yang tidak memperhitungkan kemungkinan distribusi dari masing-masing kelas [11].

HASIL DAN PEMBAHASAN

Dataset yang digunakan dalam penelitian ini diperoleh dari 20 Posyandu di Kecamatan Rumbai Timur, yang mencakup Kelurahan Lembah Sari dan Kelurahan Limbungan, dengan total 1.000 data balita. Dataset ini berisi informasi antropometri berupa umur (bulan), berat badan (kg), tinggi badan (cm), jenis kelamin, serta status gizi. Data-data tersebut dikumpulkan dari catatan hasil penimbangan balita di setiap Posyandu yang kemudian diolah ke dalam format elektronik (*Excel*) untuk kepentingan analisis.

No	Nama_Balita	Jenis_Kelamin	Tanggal_Lahir	Umur_(bulan)	Berat_Badan_(kg)	Tinggi_Badan_(cm)	Status_Gizi_(BB/TB)
1	Risky Raja Zaid	Laki-Laki	09/05/2021	48	12.2	93	Baik
2	Inayah Putri	Perempuan	29/07/2023	21	9.9	88	Baik
3	Cut Zara	Perempuan	08/04/2023	25	11	91	Baik
4	Anindya Putri	Perempuan	23/01/2022	41	12.7	88	Baik
5	Ibra. R	Laki-Laki	03/09/2023	20	10	75	Baik
6	Khairi Kinan	Laki-Laki	22/11/2021	43	10.7	88	Baik
7	Arsyla Oktavia. H	Perempuan	02/10/2024	8	8.91	73.2	Baik
8	Adnan Aresta	Laki-Laki	21/07/2022	33	13.1	90.5	Baik
9	Aidan	Laki-Laki	31/01/2022	41	8.8	77	Baik
10	M. Soyan	Laki-Laki	19/07/2021	47	10.9	81	Baik
11	Tiara Aisyah Ramadhani	Perempuan	15/07/2024	10	9.9	82	Baik
12	Reyfanza Alex Andor	Laki-Laki	05/09/2022	33	10.23	77.5	Baik
13	Hafizah Aira	Perempuan	26/02/2022	40	9.6	80	Baik
14	Kautszara Elsafm	Perempuan	28/08/2023	20	9.1	76.6	Baik
15	Riri Lismayenti	Perempuan	02/02/2025	3			
16	Khayira Azizah	Perempuan	02/04/2023	25			
17	Malik Abqory	Laki-Laki	28/04/2022	38	9.7	77	Baik
18	M. Bilal Artharaska	Laki-Laki	22/01/2025	5	6.46	65.5	Baik
19	M. Rizky Ramadhan	Laki-Laki	03/12/2020	53	16.4	107.1	Baik

Gambar 2. Dataset Asli dari Kader Posyandu

Sebelum dilakukannya *preprocessing* data, dataset yang telah disimpan dalam format *Excel* (.xlsx) di *Google Drive* diimpor terlebih dahulu ke dalam *platform Google Colab*. Selanjutnya proses pra-pemrosesan data, yang dimana pada tahap awal dilakukan pembersihan data untuk memastikan dataset hanya berisi data yang valid dan layak untuk diolah. Sebanyak 200 data dihapus karena mengandung nilai yang tidak wajar seperti outlier ekstrem pada tinggi dan berat badan, data duplikat, serta atribut yang tidak lengkap. Setelah proses ini, jumlah data berkurang dari 1.000 menjadi 800 data.

Missing values sebelum dihapus:		Missing values setelah dihapus:	
No	0	No	0
Nama_Balita	0	Nama_Balita	0
Jenis_Kelamin	0	Jenis_Kelamin	0
Tanggal_Lahir	7	Tanggal_Lahir	0
Umur_(bulan)	7	Umur_(bulan)	0
Berat_Badan_(kg)	158	Berat_Badan_(kg)	0
Tinggi_Badan_(cm)	160	Tinggi_Badan_(cm)	0
Status_Gizi_(BB/TB)	200	Status_Gizi_(BB/TB)	0
		dtype: int64	
dtype: int64		Jumlah data setelah cleaning: 800	

Gambar 3. Hasil *Cleaning* Data dan Penanganan *Missing Values*

Kemudian, dilakukan pemeriksaan lanjutan untuk memastikan konsistensi data, yaitu melakukan transformasi dan penyesuaian format data, untuk variabel Jenis Kelamin yang semula memiliki kategori “Laki-Laki” dan “Perempuan” dikonversi menjadi data biner, yaitu 0 untuk “Laki-Laki” dan 1 untuk “Perempuan”. Sementara itu, variabel Status_Gizi_(BB/TB) yang awalnya berupa label kategori “Baik”, “Buruk”, “Kurang”, dan “Lebih” dikodekan menjadi angka 0, 1, 2, dan 3 sesuai dengan pemetaan yang telah ditentukan.

```

Mapping Jenis Kelamin: {'Laki-Laki': 0, 'Perempuan': 1}
Mapping Status Gizi: {'Baik': 0, 'Buruk': 1, 'Kurang': 2, 'Lebih': 3}
Jumlah data asli: 800
Jumlah data setelah konversi numerik: 800
Jumlah baris yang jadi NaN: 0
  Jenis_Kelamin  Jenis_Kelamin_Encoded  Status_Gizi_(BB/TB)  Status_Gizi_Encoded
0      Laki-Laki                      0                Baik          0
1      Perempuan                      1                Baik          0
2      Perempuan                      1                Baik          0
3      Perempuan                      1                Baik          0
4      Laki-Laki                      0                Baik          0
..      ...                          ...                ...          ...
79     Laki-Laki                      0                Baik          0
80     Perempuan                      1                Baik          0
81     Perempuan                      1                Baik          0
82     Perempuan                      1                Kurang          2
83     Perempuan                      1                Baik          0
[70 rows x 4 columns]

```

Gambar 4. Tranformasi Data

Lalu dilanjutkan dengan penghapusan outlier, deteksi outlier ini dilakukan menggunakan metode *Interquartile Range* (IQR) dengan menghitung batas rentang normal berdasarkan kuartil 1 (Q1) dan kuartil 3 (Q3) hingga tersisa 783 data yang valid.

```

Jumlah data sebelum hapus outlier: 800
Baris yang terhapus sebagai outlier: 17
Jumlah data setelah hapus outlier: 783

```

Gambar 5. Hasil Penghapusan Outlier

Pada tahap berikutnya, dilakukan penanganan terhadap kelas dengan jumlah data sangat sedikit, yaitu kategori Gizi Buruk yang hanya memiliki satu data, sehingga

dikeluarkan karena dianggap tidak representatif. Dengan demikian, diperoleh 782 data bersih yang siap digunakan untuk tahap pemodelan.

```
Distribusi sebelum hapus kelas minoritas:
Status_Gizi_(BB/TB)
Baik      739
Kurang    27
Lebih     16
Buruk     1
Name: count, dtype: int64

Distribusi sesudah:
Status_Gizi_(BB/TB)
Baik      739
Kurang    27
Lebih     16
Name: count, dtype: int64

Jumlah data: 782
```

Gambar 6. Hasil Penanganan Kelas Dengan Jumlah Data Sedikit

Setelah itu, dilakukan normalisasi data menggunakan metode *Min-Max Scaler*, yang memetakan seluruh nilai fitur numerik ke dalam rentang 0-1. Atribut yang dinormalisasikan meliputi Umur (bulan), Berat Badan (kg), dan Tinggi Badan (cm), yang masing-masing memiliki rentang nilai berbeda. Langkah ini penting karena algoritma KNN bergantung pada perhitungan jarak *Euclidean Distance*, sehingga setiap fitur memiliki pengaruh yang seimbang dalam proses klasifikasi.

```
Data setelah normalisasi (5 baris pertama):
No  Nama_Balita  Jenis_Kelamin  Tanggal_Lahir  Umur_(bulan)  Berat_Badan_(kg)
0  1  Risky Raja Zaid  Laki-Laki      2021-05-09     0.839286        0.564103
1  2  Inayah Putri   Perempuan     2023-07-29     0.357143        0.416667
2  3  Cut Zara      Perempuan     2023-04-08     0.428571        0.487179
3  4  Anindya Putri Perempuan     2022-01-23     0.714286        0.596154
4  5  Ibra. R       Laki-Laki     2023-09-03     0.339286        0.423077

Tinggi_Badan_(cm)  Status_Gizi_(BB/TB)  Jenis_Kelamin_Encoded  Status_Gizi_Encoded
0  0.719640           Baik                  0                      0
1  0.644678           Baik                  1                      0
2  0.689655           Baik                  1                      0
3  0.644678           Baik                  1                      0
4  0.449775           Baik                  0                      0
```

Gambar 7. Hasil Normalisasi Data

Berdasarkan hasil analisis distribusi kelas, diketahui bahwa kelas Gizi Baik mendominasi dataset, sedangkan Gizi Kurang dan Gizi Lebih memiliki jumlah data yang jauh lebih sedikit. Ketidakseimbangan ini berpotensi menimbulkan bias pada model KNN karena data mayoritas akan lebih sering dijadikan acuan dalam proses pembelajaran. Untuk mengatasi hal tersebut, diterapkan metode *Synthetic Minority Over-sampling Technique* (SMOTE) yang berfungsi memperbanyak data pada kelas minoritas dengan cara menghasilkan sampel sintetis berdasarkan tetangga terdekat dari setiap data minoritas. Dengan penerapan SMOTE, distribusi antar kelas menjadi lebih proporsional dan seimbang yaitu masing-masing data tiap kelas menjadi 739 data.

Selanjutnya, dilakukan pembagian data menjadi dua bagian yaitu data latih (*training*) dan data uji (*testing*). Pada penelitian ini diterapkan rasio 80% untuk data latih dan 20% untuk data uji. Total data yang didapatkan setelah melakukan SMOTE adalah 2217 data, yang dimana jumlah data train sebanyak 1773 data dan data test sebanyak 444 data. Untuk distribusi antar kelas, data di train untuk masing-masing kelas terdiri dari 591 data, dan untuk masing-masing kelas di test yaitu sebanyak 148 data.

```

Jumlah data total setelah SMOTE: 2217
Jumlah data train: 1773
Jumlah data test : 444

Distribusi kelas di train:
Status_Gizi_Encoded
2    591
3    591
0    591
Name: count, dtype: int64

Distribusi kelas di test:
Status_Gizi_Encoded
3    148
2    148
0    148
Name: count, dtype: int64
    
```

Gambar 8. Hasil *Split Data* (Pembagian Data)

Setelah semua tahap pra-pemrosesan selesai, langkah selanjutnya adalah melakukan pemodelan menggunakan algoritma *K-Nearest Neighbor* (KNN). Nilai parameter K diuji mulai dari 1 hingga 15 untuk menentukan nilai optimal yang menghasilkan kinerja terbaik. Proses ini dilakukan menggunakan metode *5-fold cross-validation*, yang membagi dataset menjadi lima subset untuk meminimalkan bias hasil evaluasi dan memastikan stabilitas performa model. Dua metrik utama yang digunakan adalah akurasi dan *recall macro*.

K	Accuracy_cv	Recall_macro_cv	
0	1	0.882731	0.882733
1	3	0.863790	0.863808
2	5	0.852508	0.852516
3	7	0.842136	0.842146
4	9	0.838526	0.838536
5	11	0.829060	0.829074
6	13	0.820043	0.820056
7	15	0.815987	0.815999

K terbaik berdasarkan Recall Macro = 1 (Recall = 0.882733)

Gambar 9. Hasil Uji Parameter K pada Algoritma KNN dengan *Cross-Validation*

Berdasarkan hasil evaluasi akurasi dan *recall* yang konsisten pada seluruh *fold*, terlihat bahwa model KNN dengan nilai K = 1 memiliki performa yang stabil, sehingga parameter tersebut digunakan pada tahap berikutnya, yaitu evaluasi model menggunakan data uji.

Setelah ditetapkan bahwa parameter terbaik adalah K = 1, model KNN kembali dilatih menggunakan keseluruhan data latihan (*training set*) dan selanjutnya diuji pada data uji (*testing set*). Evaluasi performa model dilakukan dengan sejumlah metrik, meliputi akurasi, *precision*, *recall*, *f1-score*, serta analisis tambahan menggunakan *confusion matrix* untuk melihat detail prediksi tiap kelas.

```

Classification Report:
              precision    recall  f1-score   support

   Gizi Baik      0.86      0.73      0.79      148
   Gizi Kurang    0.82      0.92      0.87      148
   Gizi Lebih     0.91      0.94      0.92      148

 accuracy          0.86
 macro avg         0.86
 weighted avg      0.86

Confusion Matrix:
[[108  29  11]
 [  9 136   3]
 [  8   1 139]]

Akurasi pada data uji: 0.8626126126126126
    
```

Gambar 10. Hasil *Classification Report*, *Confusion Matrix*, Akurasi Pada Data Uji

Berdasarkan hasil evaluasi pada Gambar 9, penerapan algoritma *K-Nearest Neighbor* (KNN) dengan parameter $K = 1$ menunjukkan performa klasifikasi yang optimal. Model ini memperoleh akurasi sebesar 86% pada data uji, yang berarti sekitar 86 dari 100 prediksi berhasil diklasifikasikan dengan benar. Nilai *precision* dan *recall* pada setiap kelas juga menunjukkan keseimbangan yang baik, dengan *precision* berada pada kisaran 0,86–0,91 dan *recall* antara 0,73–0,94. Selain itu, nilai *F1-score* yang cukup tinggi dan konsisten pada ketiga kelas yaitu antara 0,79–0,92 semakin memperkuat bahwa model KNN memiliki kinerja yang stabil serta mampu mengklasifikasikan status gizi balita dengan baik.

Secara keseluruhan, penerapan SMOTE dan normalisasi *Min-Max* meningkatkan kemampuan model dalam mempelajari pola seluruh kelas secara seimbang. Dengan akurasi 86%, *recall macro* 0,882733, serta distribusi prediksi yang baik pada *confusion matrix*, dapat disimpulkan bahwa algoritma KNN mampu mengklasifikasikan status gizi balita di Kelurahan Lembah Sari dan Kelurahan Limbungan secara efektif, stabil, dan representatif pada semua kategori kelas.

SIMPULAN

Proses klasifikasi status gizi balita berdasarkan data antropometri berhasil dilaksanakan menggunakan algoritma *K-Nearest Neighbor* (KNN). Tahapan klasifikasi diawali dengan proses pra-pemrosesan data yang meliputi pembersihan data, penanganan *missing values*, transformasi data, penghapusan outlier, dan pengeluan kelas Gizi Buruk karena hanya memiliki satu sampel data. Selain itu, dilakukan normalisasi menggunakan *Min-Max* serta penyeimbangan kelas menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE) untuk mengatasi ketidakseimbangan data. Setelah data siap untuk digunakan, model KNN dijalankan dengan mempertimbangkan variabel umur, berat badan, tinggi badan, dan jenis kelamin untuk mengklasifikasikan status gizi balita ke dalam kategori Gizi Baik, Gizi Kurang, dan Gizi Lebih.

Untuk menentukan tingkat akurasi algoritma KNN dalam klasifikasi status gizi balita, penelitian ini melakukan evaluasi model menggunakan pengujian nilai K dari 1 hingga 15 dengan metode *5-fold cross-validation*. Setelah melalui proses penyeimbangan data menggunakan SMOTE dan normalisasi *Min-Max*, nilai $K = 1$ menghasilkan performa terbaik dengan akurasi sebesar 86,26% dan nilai *recall macro* sebesar 0,882733. Selain itu, nilai *precision*, *recall*, dan *f1-score* pada setiap kelas menunjukkan hasil yang relatif seimbang, sehingga model mampu mengenali kelas mayoritas maupun minoritas secara efektif. Dengan demikian, dapat disimpulkan bahwa algoritma KNN mampu memberikan hasil klasifikasi yang optimal terhadap status gizi balita di kedua kelurahan tersebut.

DAFTAR PUSTAKA

- [1] M. Yunus and N. K. A. Pratiwi, "Prediksi Status Gizi Balita Dengan Algoritma K-Nearest Neighbor (KNN) di Puskesmas Cakranegara," *JTIM J. Teknol. Inf. dan*

- Multimed.*, vol. 4, no. 4, pp. 221–231, 2023, doi: 10.35746/jtim.v4i4.328.
- [2] D. A. Ferliandini and S. Risnanto, “APLIKASI PREDIKSI STATUS GIZI BALITA BERBASIS WEB MENGGUNAKAN METODE K-NEAREST NEIGHBOR,” pp. 622–631, 2023.
- [3] N. Mutiara Shandhini Maylita, H. Zulfia Zahro’, and N. Vendyansyah, “Penerapan Metode K-Nearest Neighbor (Knn) Untuk Menentukan Status Gizi Balita (STUDI KASUS: POSYANDU ANANDA KELURAHAN LANGKAI, KOTA PALANGKA RAYA, KALIMANTAN TENGAH),” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 6, no. 2, pp. 953–956, 2022, doi: 10.36040/jati.v6i2.5406.
- [4] M. Annabaa’ Aulia, R. Goejantoro, and M. N. Hayati, “Penerapan Metode Klasifikasi K-Nearest Neighbor (Studi Kasus: Data Status Gizi Balita di Puskesmas Baqa Samarinda Seberang),” *Pros. Semin. Nas. Mat. Stat. dan Apl. 2023*, pp. 128–142, 2023.
- [5] D. Fitriarningsih, M. Bettiza, and A. Uperiati, “Klasifikasi Status Gizi Pada Pertumbuhan Balita Menggunakan K-Nearest Neighbor (Knn),” *Student Online J.*, vol. 2, no. 1, pp. 106–111, 2021.
- [6] R. Darma, R. Yusron, and I. Machfud, “Klasifikasi dan Monitoring Status Gizi Balita Melalui Penerapan Metode Naïve Bayes Classification Berbasis GIS,” *J. Ilm. Intech Inf. Technol. J. UMUS*, vol. 4, no. 02, pp. 161–168, 2022.
- [7] A. Nadroh, D. N. Triwibowo, and R. B. B. Sumantri, “KLASIFIKASI STATUS GIZI BALITA MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE DENGAN OPTIMASI GRID SEARCH CROSS-VALIDATION,” *METHOMIKA J. Manaj. Inform. Komputerisasi Akunt.*, vol. 8, no. 2, pp. 250–257, 2024.
- [8] H. Hananti and K. Sari, “Perbandingan Metode Support Vector Machine (SVM) dan Artificial Neural Network (ANN) pada Klasifikasi Gizi Balita,” *Semin. Nas. Off. Stat.*, pp. 1036–1043, 2021.
- [9] G. P. Insany, I. Yustiana, S. Rahmawati, T. Informatika, and U. N. Putra, “Jurnal Computer Science and Information Technology (CoSciTech),” *J. Comput. Sci. Inf. Technol.*, vol. 4, no. 2, pp. 385–393, 2023.
- [10] A. Lasarudin, H. Gani, and M. Tomayahu, “Perbandingan Metode Naïve Bayes dan C4 . 5 Klasifikasi Status Gizi Bayi Balita,” *SPECTA J. Technol.*, vol. 6, no. 3, pp. 273–283, 2022.
- [11] O. S. Bachri and R. M. H. Bhakti, “Penentuan Status Stunting Pada Anak Dengan Menggunakan Algoritma Knn,” *J. Ilm. Intech Inf. Technol. J. UMUS*, vol. 3, no. 2, pp. 130–137, 2021.