

PERBANDINGAN ALGORITMA KNN DAN NAÏVE BAYES UNTUK MEMPREDIKSI JUMLAH GIZI BURUK DI JAWA BARAT

Wisnu Arbyanza Tulus^{1*}, Wiwin Windihastuty²

^{1,2} Sistem Informasi, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

Email: ^{1*} wisnuarbyanza83@gmail.com, ² wiwin.windihastuty@budiluhur.ac.id

(* : corresponding author)

Abstrak- Stunting pada anak usia dini merupakan masalah serius yang menghambat tumbuh kembang serta proses belajar anak. Kondisi ini juga menurunkan efisiensi kinerja mereka, yang dapat berdampak jangka panjang terhadap kualitas sumber daya manusia di masa depan. Provinsi Jawa Barat masih menghadapi jumlah kasus gizi buruk yang cukup tinggi, dan belum ada prediksi yang akurat mengenai jumlah kasus gizi buruk untuk tahun-tahun mendatang. Penelitian ini bertujuan untuk menemukan model algoritma terbaik yang dapat memprediksi jumlah kasus gizi buruk di Provinsi Jawa Barat dengan membandingkan dua algoritma data mining, yaitu *K-Nearest Neighbor* (KNN) dan *Naive Bayes Classifier* (NBC). *Dataset* yang digunakan dalam penelitian ini berasal dari situs resmi penyedia data pemerintah Jawa Barat, yaitu *Open Data Jabar*. Data tersebut diolah menggunakan algoritma KNN dan NBC dengan metode *split data*, dan hasilnya dievaluasi menggunakan berbagai matriks evaluasi seperti *akurasi*, *presisi*, *recall*, dan *F1 Score*. Hasil penelitian menunjukkan bahwa algoritma *K-Nearest Neighbor* (KNN) lebih akurat dalam memprediksi jumlah kasus gizi buruk pada balita dibandingkan dengan algoritma *Naive Bayes Classifier* (NBC). Algoritma KNN memiliki nilai akurasi 0,50, presisi 0,55, recall 0,51, dan F1-Score 0,47, sedangkan NBC memiliki nilai akurasi 0,34, presisi 0,23, recall 0,35, dan F1-Score 0,28. Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa algoritma *K-Nearest Neighbor* terbukti lebih unggul dalam memprediksi jumlah kasus gizi buruk di Provinsi Jawa Barat dibandingkan dengan algoritma *Naive Bayes Classifier*. Temuan ini dapat membantu dalam pengambilan keputusan yang lebih tepat terkait penanganan dan pencegahan gizi buruk di daerah tersebut. Penelitian ini juga membuka peluang bagi penelitian selanjutnya untuk mengeksplorasi algoritma lain atau memperbaiki model yang ada untuk mendapatkan hasil prediksi yang lebih akurat dan andal. Dengan demikian, upaya penanggulangan gizi buruk dapat dilakukan dengan lebih efektif dan efisien, membawa dampak positif yang signifikan bagi kesehatan dan kesejahteraan anak-anak di Provinsi Jawa Barat.

Kata Kunci: Data mining, *K-Nearest Neighbor*, *Naive Bayes Classifier*, Gizi Buruk

COMPARISON OF KNN AND NAÏVE BAYES ALGORITHM FOR PREDICTING THE AMOUNT OF BAD NUTRITION IN WEST JAVA

Abstract- Stunting in early childhood is a serious problem that hinders children's growth and development and learning processes. This condition also reduces the efficiency of their performance, which can have a long-term impact on the quality of human resources in the future. West Java Province is still facing a fairly high number of malnutrition cases, and there are no accurate predictions regarding the number of malnutrition cases for the coming years. This research aims to find the best algorithm model that can predict the number of malnutrition cases in West Java Province by comparing two data mining algorithms, namely *K-Nearest Neighbor* (KNN) and *Naive Bayes Classifier* (NBC). The dataset used in this research comes from the official website of the West Java government data provider, namely *Open Data Jabar*. The data is processed using the KNN and NBC algorithms with the *split data* method, and the results are evaluated using various evaluation matrices such as *accuracy*, *precision*, *recall* and *F1 Score*. The research results show that the *K-Nearest Neighbor* (KNN) algorithm is more accurate in predicting the number of cases of malnutrition in toddlers compared to the *Naive Bayes Classifier* (NBC) algorithm. The KNN algorithm has an accuracy value of 0.50, precision 0.55, recall 0.51, and F1-Score 0.47, while NBC has an accuracy value of 0.34, precision 0.23, recall 0.35, and F1-Score 0.28. Based on the results of this research, it can be concluded that the *K-Nearest Neighbor* algorithm has proven to be superior in predicting the number of malnutrition cases in West Java Province compared to the *Naive Bayes Classifier* algorithm. It is hoped that these findings will help in making more appropriate decisions regarding the handling and prevention of malnutrition in the area. Apart from that, this research also opens up opportunities for further research to explore other algorithms or improve existing models to get more accurate and reliable prediction results. In this way, efforts to overcome malnutrition can be carried out more effectively and efficiently, bringing a significant positive impact on the health and welfare of children in West Java Province.

Keywords: Data mining, *K-Nearest Neighbor*, *Naive Bayes Classifier*, *Malnutrition*

1. PENDAHULUAN

Keberhasilan tumbuh kembang anak usia dini merupakan komponen penting dalam kemajuan suatu negara. Diakui bahwa gangguan tumbuh kembang balita membutuhkan perhatian khusus dari pemerintah dan pengetahuan orang tua[1]. WHO melaporkan bahwa pada tahun 2020, 149,2 juta anak di bawah usia 5 tahun mengalami stunting, yang berarti balita memiliki tinggi badan rendah dibandingkan teman yang seusianya; 45,4 juta anak mengalami kekurangan berat badan, yang berarti balita memiliki berat badan rendah dibandingkan usia teman

yang seusianya; dan 38,9 juta anak mengalami kelebihan berat badan atau obesitas[2]. Dengan hampir dua pertiga anak-anak di dunia yang mengalami stunting dan tiga perempat anak-anak terlantar di seluruh dunia, negara-negara berpendapatan menengah ke bawah, juga dikenal sebagai LMIC, menyumbang proporsi tertinggi anak-anak yang kekurangan gizi pada tahun 2020. Namun, meskipun setengah dari populasi global yang berusia dibawah 5 tahun.

Dalam bidang nutrisi, kesehatan, dan kedokteran, "gizi buruk" adalah istilah khusus yang biasanya digunakan dalam bidang gizi dan kesehatan. Istilah ini didasarkan pada catatan berat badan menurut usia (BB/U), yang merupakan singkatan dari kata kurang berat badan (gizi kurang) dan gizi buruk yang parah[3]. Salah satu jenis proses yang berkelanjutan adalah gizi buruk. Malnutrisi adalah kesalahan dalam penyediaan nutrisi, baik kekurangan atau kelebihan[4]. Tingkat gizi buruk di Jawa Barat masih sangat tinggi dan memerlukan perhatian serius, meskipun telah banyak upaya yang dilakukan oleh pemerintah dan berbagai organisasi untuk mengatasi masalah ini. Berbagai faktor yang saling berkaitan bertanggung jawab atas tingkat gizi yang rendah di Jawa Barat. Status sosial ekonomi, ketidaktahuan ibu tentang pemberian gizi yang baik untuk anak mereka, dan Berat Badan Lahir Rendah (BBLR) adalah beberapa faktor yang dapat menyebabkan gizi buruk[5].

Banyak orang tua tidak tahu bagaimana memberikan makanan yang seimbang dan cukup kepada anak-anak untuk pertumbuhan dan perkembangan anak. Akses yang terbatas terhadap layanan kesehatan adalah faktor tambahan yang menyebabkan gizi buruk di Jawa Barat. Anak-anak dengan gizi buruk tidak mendapatkan perawatan yang memadai karena fasilitas kesehatan kurang dan sulit dijangkau di daerah terpencil. Selain itu, distribusi makanan sehat ke seluruh wilayah Jawa Barat terhambat oleh kekurangan infrastruktur. Berdasarkan data Dinas Kesehatan Provinsi Jawa Barat pada Tahun 2019 sampai 2023 banyak balita yang mengalami Balita Gizi Kurang, Balita Kurus dan Balita Pendek.

Stunting pada anak usia dini menghambat tumbuh kembang serta proses belajar pada anak, stunting juga menurunkan efisiensi kinerja anak. Hal tersebut membuat penulis tertarik untuk mengetahui kasus gizi buruk di Jawa Barat. Untuk memperoleh model terbaik untuk melakukan prediksi data dengan akurat, penulis menggunakan

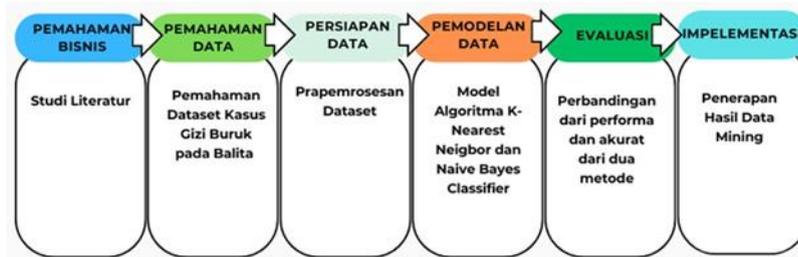
2 (dua) algoritma kemudian melakukan perbandingan kedua algoritma tersebut dengan mengukurnya menggunakan evaluasi metrik *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Algoritma yang penulis gunakan adalah Algoritma *K-Nearest Neighbor* dan Algoritma *Naïve Bayes Classifier*.

Berdasarkan penelitian sebelumnya, Pada penelitian yang berjudul "Metode *K-Nearest Neighbor* Dan *Naive Bayes* Dalam Menentukan Status Gizi Balita" Menggunakan Algoritma *K-Nearest Neighbor* & *Naive Bayes*. Penelitian ini membahas telah berhasil membuat sebuah sistem informasi yang dapat digunakan oleh user dengan cara melakukan proses input kriteria kondisi untuk dilakukan proses klasifikasi status gizi. Tingkat akurasi metode *Naive Bayes* lebih unggul dengan persentase nilai sebesar 87,5% sedangkan metode *K-Nearest Neighbor* dengan $k = 3$ mempunyai persentase akurasi sebesar 71,25%[6].

Penelitian ini bertujuan untuk membandingkan algoritma *K-Nearest Neighbor* dan *Naive Bayes Classifier* untuk mengetahui nilai akurasi masing-masing, menggunakan data dari Dinas Kesehatan Provinsi Jawa Barat untuk melakukan pengklasifikasian status gizi balita. untuk membantu Dinas Kesehatan Provinsi Jawa Barat memantau status gizi balita dan meningkatkan pengetahuan.

2. METODE PENELITIAN

Kerangka penelitian menunjukkan tahapan-tahapan penelitian menggunakan metodologi CRISP-DM, yang merupakan standar industri *cross-industry* untuk data *mining*, terdiri dari enam fase[7]. Kemudian dilakukan *preprocessing* terhadap data yang sudah dikumpulkan yang terdiri dari pemilihan atribut, pembersihan dataset dan seleksi data. Selanjutnya perbandingan masing-masing algoritma dilakukan menggunakan algoritma *K-Nearest Neighbor* dan *Naive Bayes Classifier* serta tool *Google Colab*. Seluruh tahapan penelitian terdapat pada Gambar 1.



Gambar 1. Kerangka Penelitian implementasi Algoritma *K-Nearest Neighbor* dan *Naive Bayes Classifier*

2.1 Pemahaman Bisnis

Tahapan pertama yang dilakukan adalah memahami situasi dan topik serta tujuan penelitian yang ingin dicapai. Penelitian ini bertujuan untuk mencari model algoritma klasifikasi terbaik untuk melakukan prediksi gizi buruk pada balita di Provinsi Jawa Barat.

2.2 Pemahaman Data

Langkah kedua jumlah balita berdasarkan kategori gizi buruk di Jawa Barat memberikan pondasi analitik untuk sebuah penelitian dengan membuat ringkasan dan mengidentifikasi masalah dalam data. Data yang diambil dari situs resmi Jawa Barat dari opendata.jawabarat.go.id sebanyak 405 data. Atribut yang digunakan dalam klasifikasi status gizi balita sebanyak 4 atribut yaitu: Nama Kabupaten/Kota, Kategori Gizi Buruk, Jumlah Balita, Tahun. Data yang akan diperoleh maka nantinya akan dilakukan *Pre-processing* data dan dilakukannya pemilihan atribut, pembersihan dataset, dan seleksi data. Untuk tampilan dataset asli dapat dilihat pada Tabel 1.

Tabel 1. Tabel Dataset Asli

NO	Kode Provinsi	Nama Provinsi	Kode Kabupaten/Kota	Nama Kabupaten/Kota	Kategori Gizi Buruk	Jumlah Balita	Satuan	Tahun
1	32	Jawa Barat	3201	Kabupaten Bogor	Balita Gizi Kurang	20999	Orang	2019
2	32	Jawa Barat	3201	Kabupaten Bogor	Balita Pendek	86706	Orang	2019
3	32	Jawa Barat	3201	Kabupaten Bogor	Balita Kurus	16179	Orang	2019
.....								
405	32	Jawa Barat	3279	Kota Banjar	Balita Kurus	707	Orang	2023

2.3 Persiapan data

Pada penelitian ini secara garis besar terdapat beberapa tahapan dalam pra-pemrosesan data antara lain:

a. Pemilihan Atribut

Pada penelitian ini, pemilihan atribut dalam *dataset* dilakukan berdasarkan masalah dan tujuan yang telah ditetapkan sebelumnya untuk menghasilkan model algoritma terbaik. Atribut yang digunakan dapat dilihat pada Tabel 2.

Tabel 2. Hasil Proses Pemilihan Atribut

No	Nama Kabupaten/Kota	Kategori Gizi Buruk	Jumlah Balita	Tahun
1	Kabupaten Bogor	Balita Gizi Kurang	20999	2019
2	Kabupaten Bogor	Balita Pendek	86706	2019
3	Kabupaten Bogor	Balita Kurus	16179	2019
.....				
405	Kota Banjar	Balita Kurus	707	2023

b. Pembersihan Dataset

Pada tahap ini menggunakan aplikasi *Microsoft Excel* untuk melakukan pembersihan data dengan memanfaatkan fitur *sorting* dan *filter*. Data yang dihapus adalah kode provinsi, nama provinsi, kode kabupaten kota dan satuan yang jumlahnya adalah 0, tujuannya adalah menghindari overfitting dan membuat akurasi dari pemodelan prediksi menjadi lebih optimal. Setelah dilakukan penghapusan data, maka dataset tersisa 125 baris dan 15 kabupaten/kota. Berikut tabel data Setelah Pembersihan di Tabel 3.

Tabel 3. Tabel Data Setelah Pembersihan

No	Nama Kabupaten/Kota	Kategori Gizi Buruk	Jumlah Balita	Tahun
1	Kabupaten Bandung	Balita Gizi Kurang	12706	2019
2	Kabupaten Bandung	Balita Pendek	15020	2019
3	Kabupaten Bandung	Balita Pendek	8151	2019
.....				
125	Kota Bandung	Balita Kurus	6794	2023

c. Seleksi Data

Pada tahap ini melakukan seleksi data, proses seleksi data menggunakan *Microsoft Excel* dengan fitur yaitu *Sorting* dan *Filter* pada atribut nama kabupaten/kota, kategori gizi buruk, jumlah balita, dan tahun. Pada proses seleksi data diseleksi lagi karena ada *missing value* di nama kabupaten/kota sehingga tersisa 84 baris dan 7 kabupaten/kota, menggunakan kriteria yang ditentukan pada penelitian ini yaitu:

a) Menentukan Jumlah Balita pada rata-rata pertahun

Pada tahap ini, menentukan proses seleksi data dari rata-rata pertahun. Proses ini dilakukan dengan cara menghitung rata-rata jumlah balita yang dikirimkan pertahun kemudian dibagi dengan jumlah periode selama 5 Tahun (2019-2023). Berdasarkan proses perhitungan pada tabel 4, maka rata-rata pertahun jumlah balita yang dikirimkan ke Jawa Barat adalah 14156,95 atau dibulatkan menjadi 14157 jumlah balita seperti pada Tabel 4.

Tabel 4. Perhitungan Seleksi Data

Tahun	Rata-rata per tahun
2019	14390,50
2020	14259,10
2021	14172,37
2022	14058,49
2023	13904,31
Total rata-rata	14156,95

2.4 Pemodelan

Pada tahapan ini dilakukan pemodelan terhadap data yang telah diproses sebelumnya. Data akan dipecah menjadi data latih dan data uji model algoritma untuk prediksi jumlah kasus kategori gizi buruk balita di Provinsi Jawa Barat menggunakan metode *Split Data* dan *Cross Validation*. Pada *Split* data data peneliti akan membagi data analisis dengan perbandingan 60:40, 70:30, 80:20. Kedua algoritma yang dibandingkan adalah algoritma *K-Nearest Neighbor (KNN)* dan *Naïve Bayes Classifier (NBC)*. Algoritma KNN adalah sebuah metode untuk melakukan klasifikasi berdasarkan kedekatan lokasi (jarak) suatu data dengan data yang lain[8]. Sedangkan, algoritma NBC adalah klasifikasi statistik yang bisa digunakan untuk memprediksi probabilitas keanggotaan suatu kelas dengan akurasi dan kecepatan tinggi saat diterapkan ke database dengan data besar[9].

2.5 Evaluasi

Evaluasi dilakukan untuk mengevaluasi akurasi dan efektifitas hasil modeling yang dilakukan. Dengan ini dilakukan dengan menggunakan *Split Data* dan *Cross Validation*. *Split data* adalah Kumpulan data dibagi menjadi set pelatihan, validasi, dan pengujian melalui proses untuk mempersiapkan pelatihan dan pengujian model pembelajaran mesin. Sedangkan *Cross Validation* adalah Salah satu metode untuk mengevaluasi atau memverifikasi keakuratan model yang dibangun pada *dataset* tertentu[10].

2.6 Implementasi

Tahap ini melakukan pembuatan laporan yang diartikan dalam penerapan data *mining* dari hasil *modeling*. Hasil *modeling* menemukan dan memberikan rekomendasi kepada pemerintah Jawa Barat dan layanan masyarakat Jawa Barat.

3. HASIL DAN PEMBAHASAN

3.1 Evaluasi Perbandingan Model untuk Prediksi Jumlah Kasus Gizi Buruk di Jawa Barat

Pada evaluasi model menggunakan metode split data (0,2;0,3;0,4) untuk membandingkan algoritma KNN dan NBC. Pengujian ini menggunakan bahasa pemrograman *Python* dengan bantuan *software Google Colab*. Berikut adalah hasil perbandingan prediksi menggunakan algoritma KNN dan NBC:

a. Menggunakan split data 0,2 (80:20)

Hasil dari pemodelan data dengan menggunakan metode split data untuk algoritma KNN dan NBC dengan split data 0,2 adalah seperti pada tabel 5 dimana algoritma KNN lebih unggul dibandingkan algoritma NBC karena memiliki nilai lebih besar dilihat dari matriks *Accuracy*, *Precision*, *Recall*, dan *F1-Score* seperti pada Tabel 5.

Tabel 5. Tabel Split Data 0,2

<i>Metrix</i>	KNN	NBC
<i>Accuracy</i>	0,47	0,35
<i>Precision</i>	0,47	0,25
<i>Recall</i>	0,47	0,30
<i>F1 Score</i>	0,46	0,26

b. Menggunakan split data 0,3 (70:30)

Hasil dari pemodelan data dengan menggunakan metode split data untuk algoritma KNN dan NBC dengan split data 0,3 adalah seperti pada tabel 6 dimana algoritma KNN lebih unggul dibandingkan algoritma NBC karena memiliki nilai lebih besar dilihat dari matriks *Accuracy*, *Precision*, *Recall*, dan *F1-Score* seperti pada Tabel 6.

Tabel 6. Tabel Split Data 0,3

<i>Metrix</i>	KNN	NBC
<i>Accuracy</i>	0,50	0,34
<i>Precision</i>	0,55	0,23
<i>Recall</i>	0,51	0,35
<i>F1 Score</i>	0,47	0,28

c. Menggunakan split data 0,4 (60:40)

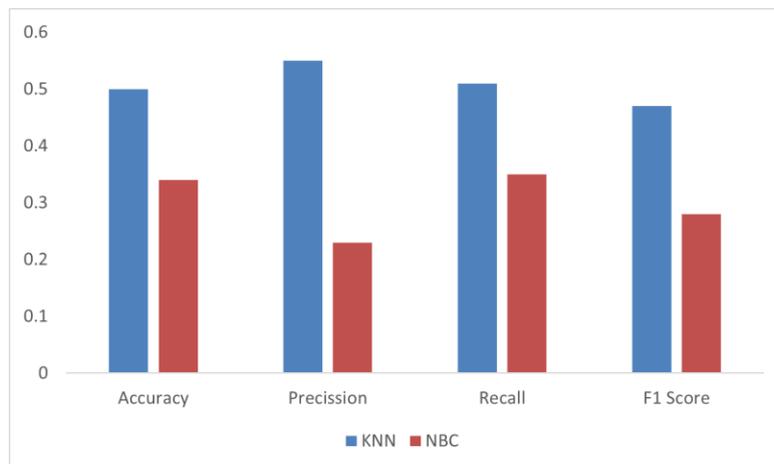
Hasil dari pemodelan data dengan menggunakan metode *split data* untuk algoritma KNN dan NBC dengan split data 0,4 adalah seperti pada tabel 7 dimana algoritma KNN lebih unggul dibandingkan algoritma NBC karena memiliki nilai lebih besar dilihat dari matriks *Accuracy*, *Precision*, *Recall*, dan *F1-Score* seperti pada Tabel 7.

Tabel 7. Tabel Split Data 0,4

<i>Metrix</i>	KNN	NBC
<i>Accuracy</i>	0,41	0,35
<i>Precision</i>	0,47	0,26
<i>Recall</i>	0,43	0,36
<i>F1 Score</i>	0,38	0,30

3.2 Penyajian Model Terbaik

Berdasarkan hasil pengujian performa model, didapatkan bahwa dengan metode split data 0.3 kedua algoritma menghasilkan performa terbaik, dengan algoritma KNN menjadi algoritma terbaik dibanding algoritma NBC. Visualisasi hasil performa terbaik algoritma KNN dan NBC dapat dilihat pada Gambar 2.


Gambar 2 Grafik Visualisasi hasil Algoritma KNN dan NBC

3.3 Perbandingan Data Aktual dan Data Prediksi

Pada tabel 8 merupakan perbandingan antara data aktual yang menggunakan *dataset* penelitian seperti Tabel 8, serta data prediksi menggunakan algoritma terbaik yaitu algoritma KNN untuk melakukan prediksi.

Tabel 8. Perbandingan Data Aktual dan Data Prediksi

Tahun	Data Aktual	Data Prediksi
2019	366.222	163.842
2020	378.450	160.572
2021	291.601	172.090
2022	282.826	196.538
2023	204.383	201.957
Jumlah	1.523.482	894.999
Rata-rata	304.696	179.000

3.4 Prediksi Sample di Tahun 2019 dan 2021

Menentukan prediksi *sample* berdasarkan mayoritas sample pada Tahun 2019 dan 2021 dari jarak terdekat.

- a. Untuk Data Uji: 'jumlah_balita = 12706', 'tahun = 2019':
 - a) Jarak Terdekat: 'BALITA KURUS', 'BALITAPENDEK', 'BALITA GIZI KURANG'
 - b) Prediksi: 'BALITA GIZI KURANG' (mayoritas)
- b. Untuk Data Uji: 'jumlah_balita = 6543', 'tahun = 2021':
 - a) Jarak Terdekat: 'BALITA GIZI KURANG', 'BALITAKURUS', 'BALITA PENDEK'
 - b) Prediksi: 'BALITA GIZI KURANG' (mayoritas)

3.5 Evaluasi

Membandingkan prediksi dengan sample aktual dan membentuk *confusion matrix*.

- a. Data Uji: 'jumlah_balita = 12706', 'tahun = 2019', 'Aktual_Kategori_Gizi_Buruk = BALITA GIZI KURANG', 'Prediksi: 'BALITA GIZI KURANG'
- b. Data Uji: 'jumlah_balita = 6543', 'tahun = 2021', 'Aktual_Kategori_Gizi_Buruk = BALITA GIZI KURANG', 'Prediksi: 'BALITA GIZI KURANG'

3.6 Evaluasi Matriks dari Algoritma *K-Nearest Neighbor*

Berdasarkan pada Tabel 9 dan hasil perhitungan evaluasi metrik dimana hasil dari model prediksi dalam mengidentifikasi status gizi, khususnya untuk kategori "Gizi Kurang". Model ini berhasil dengan tepat mengenali dua kasus di mana seseorang memang mengalami gizi kurang (*True Positive* = 2). Tidak ada kesalahan di mana model mengira seseorang yang sebenarnya gizi kurang sebagai "Kurus/Pendek" (*False Negative* = 0). Di samping itu, model juga tidak pernah salah mengidentifikasi seseorang sebagai gizi kurang padahal kenyataannya tidak (*False Positive* = 0). Akan tetapi, karena tidak ada data yang diklasifikasikan sebagai "Kurus/Pendek" dalam Tabel 9, tidak ada hasil yang tercatat untuk kategori tersebut (*True Negative* = 0). Secara keseluruhan, model ini sangat baik dalam mendeteksi kasus gizi kurang, meskipun belum ada evaluasi untuk kategori "Kurus/Pendek".

Tabel 9. Tabel Evaluasi Matriks

	<i>Predicted Positive</i> (Gizi Kurang)	<i>Predicted Negative</i> (Kurus/Pendek)
<i>Actual Positive</i> (Gizi Kurang)	TP = 2	FN = 0
<i>Actual Negative</i> (Gizi Kurang)	FP = 0	TN = 0

3.7 Perhitungan Evaluasi Matriks

Untuk perhitungan evaluasi matriks dapat dilihat pada Persamaan 1, 2, 3, dan 4.

$$1. Accuracy : \quad = \frac{TP+TN}{TP+TN+FP+FN} = \frac{2+0}{2+0+0} = 1,0 \quad (1)$$

$$2. Precision: \quad = \frac{TP}{TP+FP} = \frac{2}{2+0} = 1,0 \quad (2)$$

$$3. Recall: \quad = \frac{TP}{TP+FN} = \frac{2}{2+0} = 1,0 \quad (3)$$

$$4. F1-Score: \quad = 2x \frac{precision \times recall}{precision+recall} = 2x \frac{1,0 \times 1,0}{1,0+1,0} = 1,0 \quad (4)$$

3.8 Pengujian

Berdasarkan pada Tabel 10 merupakan Tahapan evaluation atau pengujian pada klasifikasi dilakukan dengan membandingkan nilai *accuracy*, *precision*, *recall* dan *F1 Score* pada algoritma KNN dan NBC dengan split data 0,2-0,4 dilakukan dengan menggunakan bahasa pemrograman *Python* dan *tools* Google Colab seperti pada Tabel 10.

Tabel 10. Hasil Evaluasi Matriks

Split Data	Metrix	KNN	NBC
0.2	<i>Accuracy</i>	0.47	0.35
	<i>Precision</i>	0.47	0.25
	<i>Recall</i>	0.47	0.30
	<i>F1 Score</i>	0.46	0.26
0.3	<i>Accuracy</i>	0.50	0.34
	<i>Precision</i>	0.55	0.23
	<i>Recall</i>	0.51	0.35
	<i>F1 Score</i>	0.47	0.28
0.4	<i>Accuracy</i>	0.41	0.35
	<i>Precision</i>	0.47	0.26
	<i>Recall</i>	0.43	0.36
	<i>F1 Score</i>	0.38	0.30

4. KESIMPULAN

Penelitian ini menggunakan 2 (dua) algoritma klasifikasi yaitu *K-Nearest Neighbor* dan *Naïve Bayes Classifier* untuk memprediksi jumlah balita gizi buruk di Jawa Barat pada tahun 2019 sampai 2023 dengan metode split data menggunakan *tools* Google Colab. Berdasarkan dari hasil penelitian ini dapat disimpulkan bahwa algoritma terbukti KNN terbukti lebih akurat dalam memprediksi jumlah kasus gizi buruk balita dibandingkan algoritma NBC. Dimana algoritma KNN memiliki nilai *Accuracy* 0,50, *Precision* 0,55, *Recall* 0,51, dan *F1-Score* 0,47 sedangkan NBC memiliki nilai *Accuracy* 0,34, *Precision* 0,23, *Recall* 0,35, dan *F1-Score* 0,28.

UCAPAN TERIMA KASIH

Mengucapkan terima kasih yang sebesar-besarnya kepada Universitas Budi Luhur khususnya Fakultas Teknologi Informasi dan Program Studi Sistem Informasi, yang telah memberikan dukungan dan fasilitas kesempatan untuk melaksanakan penelitian ini.

DAFTAR PUSTAKA

- [1] P. Apriyani, A. R. Dikananda, and I. Ali, “Penerapan Algoritma K-Means dalam Klusterisasi Kasus Stunting Balita Desa Tegalwangi,” *Hello World J. Ilmu Komput.*, vol. 2, no. 1, pp. 20–33, 2023, doi: 10.56211/helloworld.v2i1.230.
- [2] Febriana Sulistya Pratiwi., “Pengaruh Latar Belakang Pendidikan Ibu Terhadap Pengetahuan Ibu Tentang Menu Berbasis Pangan Lokal Untuk Pencegahan Kejadian Stunting,” vol. I, no. 8.5.2017, pp. 2003–2005, 2022, [Online]. Available: <https://dataindonesia.id/sektor-riil/detail/angka-konsumsi-ikan-ri-naik-jadi-5648-kgkapita-pada-2022>
- [3] A. Fitria and Ilman Z. Yadi, “Pemanfaatan Business Intelligence Untuk Visualisasi Data Dan Pemetaan Kasus Gizi Buruk Dan Gizi Kurang Menggunakan Tableau,” *J. Mantik*, vol. 6, no. 3, pp. 3436–3445, 2022.
- [4] D. Simanjuntak and A. Sindar, “Sistem Pakar Deteksi Gizi Buruk Balita Dengan Metode Naïve Bayes Classifier,” *J. Inkofer **, vol. 1, no. 2, pp. 2581–2920, 2019.
- [5] R. Revinovita, “Hubungan Lama Penggunaan Intrauterine Device Dengan Kadar Hemoglobin Pada Akseptor Kb Iud Di Wilayah Kerja Puskesmas Bangko Tahun 2020,” *Din. Kesehat. J. Kebidanan Dan Keperawatan*, vol. 11, no. 1, pp. 264–271, 2020, doi: 10.33859/dksm.v11i1.609.
- [6] J. Pratama, F. Fauziah, and I. D. Sholihati, “Metode K-Nearest Neighbor dan Naive Bayes Dalam Menentukan Status Gizi Balita,” *Brahmana: Jurnal Penerapan Kecerdasan Buatan*, vol. 4, no. 2, pp. 214–221, 2023.
- [7] S. E. Damayanti and S. Kuswayati, “Analisis Dan Implementasi Framework CRISP-DM (Cross Industry Standard Process for Data Mining) Untuk Clustering Perguruan Tinggi Swasta,” *J. STT Bandung*, vol. 6, 2020.
- [8] Moch. Rizky Yuliansyah, M. B, and A. Franz, “Perbandingan Metode K-Nearest Neighbors dan Naïve Bayes Classifier Pada Klasifikasi Status Gizi Balita di Puskesmas Muara Jawa Kota Samarinda,” *Adopsi Teknol. dan Sist. Inf.*, vol. 1, no. 1, pp. 08–20, 2022, doi: 10.30872/atasi.v1i1.25.
- [9] S. K. P. Loka and A. Marsal, “Perbandingan Algoritma K-Nearest Neighbor dan Naïve Bayes Classifier untuk Klasifikasi Status Gizi Pada Balita,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 1, pp. 8–14, 2023, doi: 10.57152/malcom.v3i1.474.
- [10] C. H. Yutika, A. Adiwijaya, and S. Al Faraby, “Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naïve Bayes,” *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 422, 2021, doi: 10.30865/mib.v5i2.2845.