

Metode Machine Learning untuk Klasifikasi Data Gizi Balita dengan Algoritma Naïve Bayes, KNN dan Decision Tree

Ramadhani

Fakultas Teknik, Program Studi Teknik Informatika
Universitas Islam Syeck Yusuf
Email: ramadhanigb19@gmail.com

Ramadhanu

Fakultas Teknik, Program Studi Teknik Informatika
Universitas Islam Syeck Yusuf
Email: rhamadhanu2019@gmail.com

ABSTRAK

Stunting pada balita merupakan masalah kesehatan yang serius, *Stunting* adalah istilah yang digunakan untuk menggambarkan keterlambatan perkembangan fisik anak sejak konsepsi atau pembentukan hingga usia 2 tahun, sehingga mengakibatkan tinggi badan lebih rendah dari usia kronologisnya. *Stunting* pada balita dapat disebabkan oleh kondisi sosial ekonomi, gizi ibu saat hamil, penyakit bayi, dan asupan gizi bayi yang tidak memadai. Penyakit menular adalah penyebab paling langsung dan umum dari kegagalan pertumbuhan pada anak kecil, dan diperlukan strategi yang efektif untuk mengurangi faktor risiko keterlambatan perkembangan pada anak di bawah usia lima tahun. Metode untuk mengatasi permasalahan tersebut adalah metode klasifikasi *Machine Learning (ML)* yang menggunakan algoritma Naive Bayes, KNN dan Decision Tree untuk mengklasifikasikan data gizi anak kecil, sehingga membantu mengatasi keterlambatan perkembangan, intervensi dini. Hasil penelitian ini adalah kinerja algoritma Naive Bayes buruk presisi tertinggi terdapat pada kategori gizi buruk yaitu 38% dan recall terdapat dua kategori yang tidak bisa diidentifikasi. *Algoritma* KNN ada satu kategori berisiko gizi lebih yang tidak bisa diidentifikasi presisi dan *recall*, KNN lebih tinggi dari naive bayes yaitu 40%. Decision Tree terlihat normal dan memiliki akurasi 48%, dengan *recall* dan presisi lebih baik dibandingkan Naive Bayes dan KNN.

Kata kunci: klasifikasi, Naive bayes, gizi balita, *stunting*, knn, *machine learning*

ABSTRACT

Stunting in toddlers is a serious health problem, Stunting is a term used to describe the delay in physical development of children from conception or formation to the age of 2 years, resulting in height lower than their chronological age. Stunting in toddlers can be caused by socioeconomic conditions, maternal nutrition during pregnancy, infant diseases, and inadequate infant nutritional intake. Infectious diseases are the most direct and common cause of growth failure in young children, and effective strategies are needed to reduce risk factors for developmental delays in children under the age of five. The method to overcome this problem is a machine learning (ML) classification method that uses Naive Bayes, KNN and Decision Tree algorithms to classify nutritional data of young children, thus helping to overcome developmental delays, early intervention. The result of this study is the highest precision poor naïve bayes algorithm performance found in the malnutrition category at 38% and recall there are two categories that cannot be identified. The KNN algorithm has one category of nutritional risk that cannot be identified precision and recall, KNN is higher than naïve bayes at 40%. The Decision Tree looks normal and has 48% accuracy, with better recall and precision than Naive Bayes and KNN.

Keywords: *classification, Naive bayes, toddler nutrition, stunting, knn, machine learning*

1. PENDAHULUAN

Stunting pada anak balita merupakan masalah kesehatan yang serius, terutama di banyak negara berkembang termasuk Indonesia [1]. *Stunting* adalah istilah yang digunakan untuk menggambarkan keterlambatan perkembangan fisik dan pertumbuhan anak sejak konsepsi atau pembentukan janin hingga anak berusia 2 tahun, sehingga mengakibatkan tinggi badan lebih rendah dari usia kronologisnya. Hal ini dapat berdampak negatif pada kesehatan, pendidikan, dan produktivitas anak di masa depan [2]. Pada tahun 2022, angka *stunting* di Indonesia meningkat sebesar 24,4% pada tahun 2021 dan turun menjadi 21,6% pada tahun 2022 [3]. Hal ini akan menjadi langkah yang baik dalam menurunkan angka *stunting* di Indonesia pada tahun-tahun mendatang. Keterlambatan perkembangan merupakan indikator terbaik kesehatan anak kecil selama masa kanak-kanaknya. Kondisi ini secara akurat mencerminkan kesenjangan sosial dan kondisi lingkungan. Untuk mengatasi permasalahan *stunting* ini, diperlukan strategi yang efektif untuk mengurangi faktor risiko terjadinya *stunting* pada usia dibawah lima tahun. Berbagai faktor dapat berkontribusi terhadap masalah gizi jangka panjang pada anak kecil yang mengalami keterlambatan perkembangan. *Stunting* pada balita dapat disebabkan oleh kondisi sosial ekonomi, gizi ibu saat hamil, penyakit bayi, dan asupan gizi bayi yang tidak memadai. Penyakit menular adalah penyebab paling langsung dan umum dari gangguan pertumbuhan pada anak kecil [4][5].

Anak balita perlu mengonsumsi zat gizi yang cukup dalam jumlah dan kualitas karena masih dalam proses tumbuh kembang dan belajar [6], Untuk mencapai pertumbuhan yang baik diperlukan pangan yang berkualitas agar anak tidak tumbuh lambat. *Stunting* menurunkan produktivitas kerja sehingga menghambat pertumbuhan ekonomi dan memperburuk kemiskinan di negara tersebut. Keterlambatan perkembangan dapat menyebabkan kerusakan permanen pada perkembangan kognitif akibat terganggunya proses pematangan saraf otak dan perubahan struktur dan fungsi otak [7][8]. Kondisi ini dapat mengganggu kemampuan berpikir dan belajar anak. Terdapat kebutuhan untuk memilah data gizi anak-anak secara teratur untuk mengurangi tingkat keterlambatan perkembangan pada anak-anak. Berdasarkan permasalahan tersebut, kami menggunakan metode machine learning (ML), sebuah solusi yang dapat membantu mengklasifikasikan data gizi balita secara teratur dan efisien [9][10], sehingga dapat membantu upaya penanggulangan *stunting* [11], klasifikasi data gizi balita dengan machine learning dapat membantu mengidentifikasi balita yang berisiko *stunting* dan memungkinkannya intervensi dini untuk mencegah *stunting*.

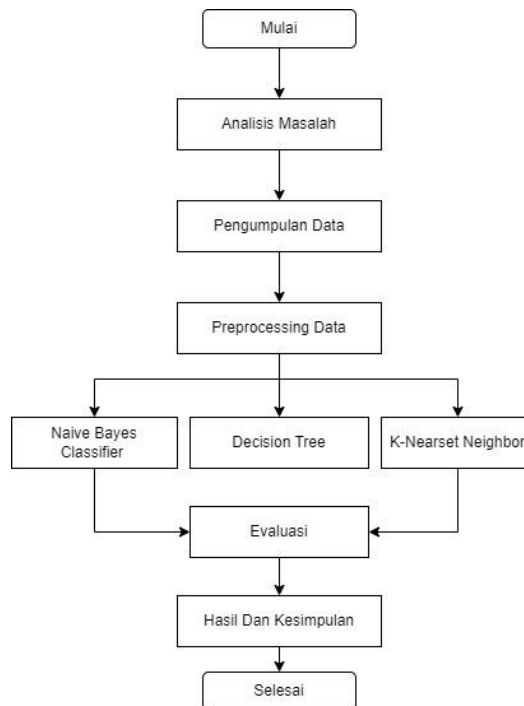
Dalam penelitian ini, kami akan mengklasifikasikan data gizi balita menggunakan beberapa variabel antara lain NIK, berat badan (Kg), tinggi badan (cm), jenis kelamin, kategori. NIK akan menjadi Merupakan identitas ID balita, Variabel kategori bergantung pada parameter variabel nilai data. Kami menggunakan pembelajaran mesin untuk mengklasifikasikan data nutrisi anak usia dini. Penelitian ini bertujuan untuk menguji kembali kinerja algoritma K-nearest neighbour decision tree dan algoritma Naive Bayes dalam mengklasifikasikan data gizi balita.

Penelitian yang dilakukan oleh [12] memperoleh hasil klasifikasi algoritma Naive Bayes, dengan akurasi sebesar 80,60% dan K tetangga terdekat sebesar 91,79%. Hasil perbandingan prediksi Naive Bayes menunjukkan status gizi balita “baik” sebesar 81,98%, sedangkan status gizi balita tetangga terdekat K adalah “baik” sebesar 100% [13] menyimpulkan bahwa algoritma KNN lebih baik dibandingkan algoritma NBC karena akurasi KNN yang diperoleh untuk nilai KNN dan NBC sebesar 96,10%, sedangkan NBC sebesar 90,94% [14] dalam penelitiannya menggunakan 200 *record* dan 5 atribut yaitu jenis kelamin, usia, berat badan, tinggi badan, dan pola makan dan minum. Yang di olah menggunakan tools dengan algoritma C.45. hasil yang didapat yaitu 96.00%.

Berdasarkan uraian tersebut, *stunting* pada anak kecil diketahui dapat menghambat tumbuh kembang dan proses belajar anak, serta *stunting* juga menurunkan efisiensi kerja. Oleh karena itu, penelitian dengan judul “Metode Machine Learning Untuk Klasifikasi Data Gizi Balita Dengan Algoritma K-Nearest Neighbor dan Naive Bayes” penting untuk dikaji. Untuk mengetahui algoritma mana yang bekerja dengan baik, kami menggunakan metode machine learning yang berfokus pada pengujian ulang algoritma K-Nearest Neighbors, decision tree dan Naive Bayes untuk menentukan algoritma mana yang bekerja dengan baik, akurat, dan efisien.

2. METODOLOGI PENELITIAN

Pada penelitian ini akan memusatkan pada penggunaan metode machine learning, khususnya algoritma k-nearest neighbor, decision tree dan naive bayes, untuk mengklasifikasi data gizi balita. Dimulai dengan mencari masalah yang akan dibahas dari berbagai sumber. Data yang akan digunakan dalam penelitian ini berasal dari posyandu. Setelah itu, data yang telah dikumpulkan diproses dengan cleaning, transformasi. Selanjutnya dilakukan proses klasifikasi data terhadap data yang sudah di preprocessing lalu akan dilakukan pengujian algoritma K-Nearest Neighbor, Decision Tree dan Naive Bayes Classifier, hasilnya akan menentukan hasil penelitian. Tahapan penelitian dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1. Data Mining

Data mining adalah proses menganalisis dan merangkum data dari sudut pandang yang berbeda, memperoleh model, pola, dan aturan dari kumpulan data yang sangat besar, dan menggunakan berbagai metode untuk menghasilkan hasil yang dapat digunakan untuk meningkatkan pendapatan dan mengurangi biaya [15]. Selain itu, data mining diklasifikasikan berdasarkan tujuannya, seperti asosiasi, clustering, prediksi, estimasi, dan klasifikasi [16]. penelitian ini akan menggunakan K-Nearest Neighbor Classifier, Decision Tree dan Naive Bayes Classifier.

2.2. Klasifikasi

Klasifikasi data merupakan sebuah metode dalam data mining yang digunakan untuk mengkategorikan data berdasarkan atribut dan karakteristiknya [17]. Algoritma ini mempelajari pola dan hubungan antar data, kemudian menggunakan pengetahuan tersebut untuk memprediksi kategori data baru.

2.3. Analisis Masalah

Tahap awal dari penelitian ini adalah analisis masalah. Pada tahap ini kami mencari masalah yang terjadi, kemudian kami akan mempelajari masalah tersebut, dan menemukan solusi dari permasalahan yang terjadi. Masalah yang dihadapi adalah keterlambatan intervensi dini dalam mencegah stunting dan hasil akhir untuk mengetahui klasifikasi data gizi balita.

2.4. *Preprocessing Data*

Pra-pemrosesan adalah teknik yang melibatkan perubahan data mentah menjadi format yang dapat dimengerti. Pemrosesan awal data diperlukan untuk mengatasi berbagai masalah seperti ketidakakuratan data, redundansi, dan nilai data yang hilang [18]. Sebelum masuk kedalam proses klasifikasi data akan di cleaning dan transformasi terlebih dahulu, Data yang bertipe kategorikal akan di transformasi menjadi numerik seperti Jenis Kelamin, dan Kategori.

2.5. *Algoritma K-Nearest Neighbor*

Algoritma K-Nearest Neighbor (KNN) adalah salah satu algoritma machine learning yang termasuk kategori supervised learning. Supervised learning berarti algoritma ini belajar dari data berlabel untuk memprediksi label pada data baru. KNN sendiri berfokus pada klasifikasi, yaitu mengelompokkan data ke dalam kategori tertentu [19]. Dalam algoritma KNN, digunakan permasalahan berikut.

$$dic(x_1, x_2) = \sqrt{\sum_{i=0}^n (x_{1i} - x_{2i})^2} \quad (1)$$

Keterangan:

dic x_1, x_2 = adalah Jarak sampel
n = adalah jumlah dimensi data
 x_1 = adalah data sampel
 x_2 = adalah data input
i = Variabel data

2.6. *Algoritma Naive Bayes Classifier*

Naive Bayes merupakan algoritma klasifikasi statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas. Algoritma Naive Bayes terbukti memiliki akurasi dan kecepatan tinggi ketika diterapkan pada database big data [20]. Dalam algoritma Naive Bayes, digunakan permasalahan berikut [21].

$$P(C|X) = \frac{(P(X|C) \times P(C))}{P(X)} \quad (2)$$

Keterangan:

$P(C|X)$ = adalah probabilitas kelas C terjadi, dengan X sebagai fitur yang diamati
 $P(X|C)$ = adalah probabilitas fitur X muncul dalam kelas C.
 $P(C)$ = adalah probabilitas kelas C terjadi secara umum.
 $P(X)$ = adalah probabilitas fitur X muncul secara umum.

2.7. *Algoritma Decision Tree*

Decision Tree merupakan algoritma klasifikasi dan regresi yang berbentuk seperti pohon. Ia bekerja dengan memecah data menjadi sub-sub kelompok berdasarkan fitur atau karakteristiknya,

decision tree menggunakan serangkaian aturan yang ditampilkan pada cabang-cabang pohon untuk sampai pada klasifikasi atau prediksi akhir [22]. Berikut persamaan decision tree.

$$Entropy(S) = \sum_{i=0}^n - p_i \cdot \log_2 p_i \tag{3}$$

Keterangan:

S = Himpunan kasus

n = Jumlah partisi S

pi = Proporsi Si terhadap S

2.8. Evaluasi

Tahap evaluasi dilakukan dengan cara mengecek nilai akurasi, presisi, recall, akurasi adalah ukuran seberapa sering algoritma menghasilkan prediksi yang benar, ini dihitung sebagai rasio antara jumlah prediksi yang benar dan total prediksi [23]. Berikut persamaan akurasi.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

Recall adalah ukuran seberapa sering algoritma mengidentifikasi semua contoh positif. Ini dihitung sebagai rasio antara jumlah prediksi positif yang benar dan total contoh positif [24]. Berikut persamaan recall.

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

Presisi adalah proses yang dilakukan untuk mengevaluasi performa dengan melihat ukuran seberapa sering prediksi positif algoritma benar [25]. Ini dihitung sebagai rasio antara jumlah prediksi positif yang benar dan total prediksi positif. Berikut persamaan precision

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

3. HASIL DAN PEMBAHASAN

3.1. Preprocessing Data

Pada tahap ini data yang sudah dikumpulkan akan di proses melalui tahap awal yaitu cleaning data, cleaning data akan membantu untuk mengetahui apakah ada ada yang missing, biasanya jika pada dataset format excel atau csv tidak boleh ada (Space) karena akan menimbulkan perbedaan nantinya. Tabel 1 memperlihatkan tampilan data awal yang belum di transformasi.

Tabel 1. Data Blita

No	NIK	BB (Kg)	TB (CM)	Jenis Kelamin	Kategori
0	3603141210180012	15	105	L	Berisiko Gizi Lebih
1	3603140205780001	15	108	L	gizi buruk
2	360 3145412180004	14	103	L	gizi kurang
3	360314550690007	17	103	P	gizi kurang
.....
495	8890123456	11	82	0	Berisiko Gizi Lebih
496	9901234567	9	73	0	Gizi Lebih

497	1001234567	11	78	0	Gizi Lebih
498	1112345678	11	78	0	Gizi Lebih
499	1223456789	11	83	0	Berisiko Gizi Lebih

Setelah di transformasi, data akan menjadi seperti ini bisa dilihat pada tabel 2, kami melakukan transformasi data pada Kategori menjadi numerik yang bertipe obesitas "6", berisiko gizi lebih "5", gizi lebih "4", gizi baik "3", gizi kurang "2", gizi buruk "1". Dan jenis kelamin L "1", P "0". Dan kami juga menghapus variabel NIK, karena NIK bukan bagian dari parameter yang menentukan kategori. Lalu dataset di split menjadi data testing 30% dan data training 70% untuk mencapai tahap berikutnya.

Tabel 2. Transformasi Data

No	BB (Kg)	TB (CM)	Jenis Kelamin	Kategori
0	15	105	1	0
1	15	108	1	1
2	14	103	1	2
3	17	103	0	2
.....
495	11	82	0	5
496	9	73	0	4
497	11	78	0	4
498	11	78	0	4
499	11	83	0	5

3.2. Penerapan Algoritma Klasifikasi

Pada tahap ini dataset di proses dengan algoritma naïve bayes, decision tree dan knn. Kami memisahkan label Kategori untuk membagi data menjadi data train dan test karena kami ingin memisahkan fitur atau atribut dari label atau target yang ingin di prediksi, dengan Hasil dari proses pengolahan algoritma klasifikasi dengan python dengan pembagian data yaitu 70:30, untuk semua algoritma yang digunakan pada penelitian ini, pada gambar dibawah menunjukkan proses membuat model algoritma yang digunakan. Untuk lebih detailnya bisa dilihat pada gambar-gambar dibawah ini.

```
[8] x=df.drop('Kategori',axis=1)
     y=df['Kategori']

[9] x_train,x_test,y_train,y_test = train_test_split(x,y,test_size = 0.30, random_state = 0)
     ss = StandardScaler()

     x_train_std = ss.fit_transform(x_train)
     x_test_std = ss.transform(x_test)

nb = GaussianNB()
nb.fit(x_train, y_train)
y_pred_nb = nb.predict(x_test)

[14] print(metrics.classification_report(y_test, y_pred_nb))
```

	precision	recall	f1-score	support
1	0.38	0.45	0.41	40
2	0.36	0.20	0.26	60
3	0.17	0.08	0.11	24
4	0.00	0.00	0.00	10
5	0.00	0.00	0.00	14
6	0.04	1.00	0.07	2
accuracy			0.23	150
macro avg	0.16	0.29	0.14	150
weighted avg	0.27	0.23	0.23	150

Gambar 2. Algoritma Naïve Bayes

Pada gambar 2 menunjukkan hasil klasifikasi report naive bayes. Kinerja Naive Bayes pada data ini bisa dikatakan sangat buruk. Akurasi secara keseluruhan rendah, dan kinerja di setiap kategori kurang optimal. Performa kelas 1 sedang, namun presisi dan recall masih rendah. Kelas 2, 3, 4 dan 5 mempunyai kinerja yang sangat buruk. Kategori ini hampir tidak memiliki prediksi yang benar (precision dan recall 0). Kelas 6 menarik, dengan recall mendekati 1 (sempurna) tetapi presisi sangat rendah (hanya 0,04). Ini berarti Naive Bayes dapat memprediksi semua data sebagai kelas 6, dan oleh karena itu banyak data lain yang tidak seharusnya berada di kelas 6 akan diklasifikasikan sebagai kelas 6, meskipun data kelas 6 itu sendiri telah diprediksi dengan baik.

```
x=df.drop('Kategori',axis=1)
y=df['Kategori']

[19] x_train,x_test,y_train,y_test = train_test_split(x,y,test_size = 0.30, random_state = 0)
     ss = StandardScaler()

     x_train_std = ss.fit_transform(x_train)
     x_test_std = ss.transform(x_test)

knn = KNeighborsClassifier()
knn.fit(x_train,y_train)
y_pred_knn = knn.predict(x_test)

[23] print(metrics.classification_report(y_test, y_pred_knn))
```

	precision	recall	f1-score	support
1	0.39	0.42	0.40	40
2	0.51	0.60	0.55	60
3	0.23	0.21	0.22	24
4	0.20	0.10	0.13	10
5	0.00	0.00	0.00	14
6	0.50	0.50	0.50	2
accuracy			0.40	150
macro avg	0.30	0.31	0.30	150
weighted avg	0.36	0.40	0.38	150

Gambar 3. Algoritma KNN

Pada gambar 3 hasil klasifikasi report, Secara keseluruhan kinerja klasifikasi model ini dapat dikatakan sedang. Akurasi keseluruhannya adalah 40%, tetapi performanya berbeda-beda di setiap kategori. Kelas 2 memiliki performa terbaik (presisi 0,51, recall 0,60). Artinya model tersebut dapat memprediksi kategori 2

lebih baik dibandingkan kategori lainnya. Kelas 1, 3, 4, dan 5 memiliki performa yang lebih buruk (presisi dan recall lebih rendah). Model mengalami kesulitan mengidentifikasi data ini dengan benar.

```

0s dtc=DecisionTreeClassifier(criterion='entropy',random_state=42)
   dtc.fit(x_train,y_train)

DecisionTreeClassifier
DecisionTreeClassifier(criterion='entropy', random_state=42)

1s [40] y_pred_dtc = dtc.predict(x_test)
    print(metrics.classification_report(y_test, y_pred_dtc))
    
```

	precision	recall	f1-score	support
1	0.44	0.50	0.47	40
2	0.59	0.65	0.62	60
3	0.42	0.42	0.42	24
4	0.14	0.10	0.12	10
5	0.25	0.07	0.11	14
6	0.25	0.50	0.33	2
accuracy			0.48	150
macro avg	0.35	0.37	0.34	150
weighted avg	0.46	0.48	0.46	150

Gambar 4. Algoritma Decision Tree

Pada gambar 4 merupakan hasil dari penerapan algoritma decision tree, Berdasarkan report klasifikasi ini, kita dapat melihat bahwa model decision tree memiliki akurasi keseluruhan yang moderat. Namun, kinerjanya sangat berbeda di antara kelas yang berbeda. Kelas 2 dan 6 memiliki nilai ingatan dan presisi yang relatif lebih tinggi, menunjukkan bahwa model bekerja dengan baik dalam mengklasifikasikan kelas ini. Sebaliknya, kelas 4 dan 5 memiliki nilai ingatan dan presisi yang lebih rendah, menunjukkan bahwa model kesulitan membedakannya dari kelas lain.

Tabel 3. Perbandingan Precision

<i>Kategori</i>	<i>Precision</i>		
	<i>Naïve Bayes</i>	<i>Decision Tree</i>	<i>KNN</i>
<i>Gizi buruk</i>	38%	44%	39%
<i>Gizi kurang</i>	36%	59%	51%
<i>Gizi baik</i>	17%	42%	23%
<i>Gizi lebih</i>	0%	14%	20%
<i>Berisiko gizi lebih</i>	0%	25%	0%
<i>Obesitas</i>	4%	25%	50%

Model Decision Tree memiliki kinerja terbaik dalam klasifikasi status gizi secara keseluruhan, dengan presisi tinggi untuk kategori gizi buruk, gizi baik. Hal ini menunjukkan bahwa model Decision Tree mampu membedakan kategori-kategori tersebut dengan cukup akurat. Model KNN memiliki presisi tinggi untuk kategori gizi lebih, menunjukkan kemampuannya dalam mengidentifikasi orang yang berisiko mengalami obesitas. Model Naïve Bayes memiliki presisi yang terendah dibandingkan dengan dua model lainnya. Hal ini menunjukkan bahwa model Naïve Bayes kurang akurat dalam klasifikasi status gizi.

Tabel 4. Perbandingan Recall

<i>Kategori</i>	<i>Recall</i>		
	<i>Naïve Bayes</i>	<i>Decision Tree</i>	<i>KNN</i>
<i>Gizi buruk</i>	45%	50%	42%
<i>Gizi kurang</i>	20%	65%	60%
<i>Gizi baik</i>	8%	42%	21%
<i>Gizi lebih</i>	0%	10%	10%
<i>Berisiko gizi lebih</i>	0%	7%	0%
<i>Obesitas</i>	100%	50%	50%

Model naive bayes ini cukup akurat dalam mengklasifikasikan kategori gizi buruk 45% dan obesitas 100%. Namun performanya kurang baik dalam mengklasifikasikan kategori gizi lainnya. Model decision tree memiliki performa terbaik dalam mengklasifikasikan kategori gizi kurang. Ini menunjukkan bahwa model Decision Tree mampu mengidentifikasi ciri-ciri individu yang mengalami gizi kurang dengan baik. Model KNN memiliki kinerja terbaik dalam mengklasifikasikan individu yang mengalami gizi kurang recall 42%. Meski tidak sebaik model Decision Tree dan Naïve Bayes, model KNN masih cukup andal dalam mengidentifikasi individu dengan kondisi gizi sangat serius tersebut.

Tabel 5. Perbandingan Akurasi

<i>Akurasi</i>		
<i>Naïve Bayes</i>	<i>Decision Tree</i>	<i>KNN</i>
23%	48%	40%

Di lihat dari tabel 5 decision tree unggul dalam akurasi, ini juga terlihat dari nilai presisi dan recall yang terlihat lebih baik dari dua algoritma lainnya Naive bayes dan KNN.

4. KESIMPULAN

Berdasarkan rumusan masalah yang telah kami jelaskan diatas dapat disimpulkan bahwa Pengujian kembali tiga algoritma yaitu naive bayes, decision tree dan knn. Kami bisa melihat algoritma naive bayes buruk dalam klasifikasi data gizi balita, presisi tertinggi terdapat pada kategori gizi buruk yaitu 38% dan recall terdapat dua kategori yang tidak bisa diidentifikasi, walaupun kategori obesitas mencapai 100%. Model KNN juga ada satu kategori berisiko gizi lebih yang tidak bisa diidentifikasi presisi dan recall, akurasi model KNN lebih tinggi dari naive bayes yaitu 40%. Model decision tree terlihat normal dengan akurasi 48%, dengan recall dan presisi lebih baik dibanding naive bayes dan knn.

Penggunaan metode klasifikasi telah memberikan kontribusi penting dalam meningkatkan intervensi dini mencegah stunting, namun, ada ruang untuk meningkatkan, khususnya mengatasi tantangan yang terjadi pada kurangnya variabel pada dataset data gizi balita dan penyesuaian parameter model.

DAFTAR PUSTAKA

- [1] V. Fitriyana, Lutfi Hakim, Dian Candra Rini Novitasari, and Ahmad Hanif Asyhar, "Analisis Sentimen Ulasan Aplikasi Jamsostek Mobile Menggunakan Metode Support Vector Machine," *Jurnal Buana Informatika*, vol. 14, no. 01, 2023, doi: 10.24002/jbi.v14i01.6909.
- [2] H. S. Mediani, A. Setyawati, S. Hendrawati, I. Nurhidayah, and N. F. Firdianty, "Pengaruh Faktor Maternal terhadap Insidensi Stunting pada Anak Balita di Negara Berkembang: Narrative Review," *Jurnal Obsesi : Jurnal Pendidikan Anak Usia Dini*, vol. 7, no. 2, pp. 1868–1886, Mar. 2023, doi: 10.31004/obsesi.v7i2.4160.
- [3] Rokom, "Prevalensi Stunting di Indonesia Turun ke 21,6% dari 24,4%," <https://sehatnegeriku.kemkes.go.id/baca/rilis-media/20230125/3142280/prevalensi-stunting-di-indonesia-turun-ke-216-dari-244/>.
- [4] A. D. Laksono, N. E. W. Sukoco, T. Rachmawati, and R. D. Wulandari, "Factors Related to Stunting Incidence in Toddlers with Working Mothers in Indonesia," *Int J Environ Res Public Health*, vol. 19, no. 17, Sep. 2022, doi: 10.3390/ijerph191710654.
- [5] A. J. Kowalski et al., "The Effects of Multiple Micronutrient Fortified Beverage and Responsive Caregiving Interventions on Early Childhood Development, Hemoglobin, and Ferritin among Infants in Rural Guatemala," *Nutrients*, vol. 15, no. 9, May 2023, doi: 10.3390/nu15092062.
- [6] M. Y. Titimeidara and W. Hadikurniawati, "Monica Yoshe Titimeidara Implementasi Metode Naive Bayes Implementasi Metode Naive Bayes Classifier Untuk Klasifikasi Status Gizi Stunting Pada Balita."

- [7] B. Tan, Y. Wang, X. Zhang, and X. Sun, "Recent Studies on Protective Effects of Walnuts against Neuroinflammation," *Nutrients*, vol. 14, no. 20. MDPI, Oct. 01, 2022. doi: 10.3390/nu14204360.
- [8] M. S. Noor et al., "Analysis of Socioeconomic, Utilization of Maternal Health Services, and Toddler's Characteristics as Stunting Risk Factors," *Nutrients*, vol. 14, no. 20, Oct. 2022, doi: 10.3390/nu14204373.
- [9] A. Robi Padri, "HCI DAN MEDIA SOSIAL: STUDI KASUS ANALISIS SENTIMEN PILPRES 2024 DI TWITTER MENGGUNAKAN NAIVE BAYES CLASSIFIER," *Jurnal SIMETRIS*, vol. 14, no. 2, 2023.
- [10] [N. A. Widiastuti, M. Azhar, and H. Mulyo, "IMPLEMENTASI ALGORITMA K-NEAREST NEIGHBOR UNTUK KLASIFIKASI JURUSAN PADA PESERTA DIDIK BARU," *Jurnal SIMETRIS*, vol. 14, no. 2, 2023.
- [11] H. Ali, E. Hashmi, S. Yayilgan Yildirim, and S. Shaikh, "Analyzing Amazon Products Sentiment: A Comparative Study of Machine and Deep Learning, and Transformer-Based Techniques," *Electronics (Basel)*, vol. 13, no. 7, p. 1305, Mar. 2024, doi: 10.3390/electronics13071305.
- [12] R. Setiawan and A. Triayudi, "Klasifikasi Status Gizi Balita Menggunakan Naïve Bayes dan K-Nearest Neighbor Berbasis Web," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 2, p. 777, Apr. 2022, doi: 10.30865/mib.v6i2.3566.
- [13] J. Homepage, S. Kenia, P. Loka, and A. Marsal, "MALCOM: Indonesian Journal of Machine Learning and Computer Science Comparison Algorithm of K-Nearest Neighbor and Naïve Bayes Classifier for Classifying Nutritional Status in Toddlers Perbandingan Algoritma K-Nearest Neighbor dan Naïve Bayes Classifier Untuk Klasifikasi Status Gizi Pada Balita," vol. 3, pp. 8–14, 2023.
- [14] P. C. Algoritma, P. Klasifikasi Status Gizi Balita di Posyandu Desa Sukalilah Cibatu Kabupaten Garut Jawa Barat Sri Lestari, R. Amanda Amalia, and S. Lestari, "Penerapan Algoritma C.45 Pada Klasifikasi Status Gizi Balita di Posyandu Desa Sukalilah Cibatu Kabupaten Garut Jawa Barat," *Jurnal Sains dan Teknologi*, vol. 5, no. 1, pp. 177–182, 2023, doi: 10.55338/saintek.v5i1.1375.
- [15] A. I. Sang, E. Sutoyo, and I. Darmawan, "ANALISIS DATA MINING UNTUK KLASIFIKASI DATA KUALITAS UDARA DKI JAKARTA MENGGUNAKAN ALGORITMA DECISION TREE DAN SUPPORT VECTOR MACHINE DATA MINING ANALYSIS FOR CLASSIFICATION OF AIR QUALITY DATA DKI JAKARTA USING DECISION TREE ALGORITHM AND SUPPORT VECTOR MACHINER ALGORITHM."
- [16] N. Nurainun, E. Haerani, F. Syafria, and L. Oktavia, "Penerapan Algoritma Naïve Bayes Classifier Dalam Klasifikasi Status Gizi Balita dengan Pengujian K-Fold Cross Validation," *Journal of Computer System and Informatics (JoSYC)*, vol. 4, no. 3, pp. 578–586, May 2023, doi: 10.47065/josyc.v4i3.3414.
- [17] Y. A. Singgalen, "JURNAL MEDIA INFORMATIKA BUDIDARMA Penerapan Metode CRISP-DM dalam Klasifikasi Data Ulasan Pengunjung Destinasi Danau Toba Menggunakan Algoritma Naïve Bayes Classifier (NBC) dan Decision Tree (DT)," 2023, doi: 10.30865/mib.v7i3.6461.
- [18] S. Analisis, A. Satuselat, D. Wardhani, R. Astuti, and D. D. Saputra, "Optimasi Feature Selection Text Mining: Stemming dan Stopword," *INNOVATIVE: Journal Of Social Science Research*, vol. 4, pp. 7537–7548, 2024.
- [19] M. Akmal Hakim, P. Gunawan, R. Pratama, and F. Kurniawam, "Implementasi Algoritma K-Nearest Neighbors untuk Menganalisis Pendapat Pakar AI tentang Kemajuan Kecerdasan Buatan," 2024. [Online]. Available: <https://journal-computing.org/index.php/journal-cisa/index>
- [20] N. Nurainun, E. Haerani, F. Syafria, and L. Oktavia, "Penerapan Algoritma Naïve Bayes Classifier Dalam Klasifikasi Status Gizi Balita dengan Pengujian K-Fold Cross Validation,"

- Journal of Computer System and Informatics (JoSYC), vol. 4, no. 3, pp. 578–586, May 2023, doi: 10.47065/josyc.v4i3.3414.
- [21] J. Homepage, S. Kenia, P. Loka, and A. Marsal, “MALCOM: Indonesian Journal of Machine Learning and Computer Science Comparison Algorithm of K-Nearest Neighbor and Naïve Bayes Classifier for Classifying Nutritional Status in Toddlers Perbandingan Algoritma K-Nearest Neighbor dan Naïve Bayes Classifier Untuk Klasifikasi Status Gizi Pada Balita,” vol. 3, pp. 8–14, 2023.
- [22] M. R. Qisthiano, P. A. Prayesy, and I. Ruswita, “Penerapan Algoritma Decision Tree dalam Klasifikasi Data Prediksi Kelulusan Mahasiswa,” *G-Tech: Jurnal Teknologi Terapan*, vol. 7, no. 1, pp. 21–28, Jan. 2023, doi: 10.33379/gtech.v7i1.1850.
- [23] M. Mastur Alfitri and D. Rusda, “Evaluasi Performa Algoritma Naïve Bayes Dalam Mengklasifikasi Penerima Bantuan Pangan Non Tunai,” vol. 7, no. 3, pp. 1433–1445, 2023, doi: 10.30865/mib.v7i3.6151.
- [24] K. Saputra, “Perbandingan Kinerja Fungsi Kernel Algoritma Support Vector Machine Pada Klasifikasi Penyakit Padi,” *IJCCS*, vol. x, No.x, pp. 1–5.
- [25] G. M. C. Batubara, A. Desiani, and A. Amran, “Klasifikasi Jamur Beracun Menggunakan Algoritma Naïve Bayes dan K-Nearest Neighbors,” *Jurnal Ilmu Komputer dan Informatika*, vol. 3, no. 1, pp. 33–42, Jun. 2023, doi: 10.54082/jiki.68.

