

## PENERAPAN NAÏVE BAYES UNTUK MENGETAHUI STATUS GIZI BALITA

Muhammad Ardiansyah Sembiring<sup>1</sup>, Mustika Fitri Larasati Sibuea<sup>2</sup>, Ika Rahmanda Sitorus<sup>3</sup>

Sekolah Tinggi Manajemen Informatika dan Komputer Royal

e-mail: <sup>1</sup>adinmantap88@gmail.com

**Abstract:** *The increasing population accompanied by the increasing development of technological advances is one of the efforts to maintain the level of health by improving people's nutrition. Balanced nutrition can increase body resistance, increase intelligence and promote normal growth. In daily practice, nutritional status is obtained through anthropometric measurements at the posyandu. Generally, people use the BB/U index or body weight compared to age to determine nutritional status, so it takes a long time and calculations can cause errors. This research aims to make it easier to determine the nutritional status of toddlers using machine learning with the naive Bayes classification algorithm with 4 data tests. Where the ratio test that produces the best accuracy value in 4 accuracy measurements with ratios of 90:10, 80:20, 70:30 and 60:40 includes the first ratio of 90:10, the second ratio of 80:20, and the third ratio of 70:30 in the data training and testing data. The research results show an accuracy of 100.0%. So the Naive Bayes Classification is considered good for testing the nutritional status of toddlers.*

**Keywords:** *Machine Learning, Naïve Bayes, Malnutrition.*

**Abstrak:** Meningkatnya populasi penduduk yang diiringi dengan semakin berkembangnya kemajuan teknologi merupakan salah satu upaya untuk menjaga derajat kesehatan dengan perbaikan gizi masyarakat, gizi yang seimbang dapat meningkatkan ketahanan tubuh, dapat meningkatkan kecerdasan dan menjadikan pertumbuhan yang normal. Dalam praktik sehari-hari status gizi didapatkan melalui pengukuran antropometri di posyandu. Umumnya masyarakat menggunakan indeks BB/U atau berat badan dibanding usia untuk menentukan status gizi sehingga membutuhkan waktu yang lama dan perhitungan yang bisa terjadi kesalahan. Penelitian ini bertujuan untuk memudahkan penentuan status gizi balita menggunakan *machine learning* dengan algoritma naive bayes classification dengan 4 kali pengujian data. Dimana pengujian rasio yang menghasilkan nilai akurasi terbaik dalam 4 pengujian akurasi dengan rasio 90:10, 80:20, 70:30 dan 60:40 diantaranya rasio pertama 90:10, rasio kedua 80:20, dan rasio ketiga 70:30 pada Data training dan data testing. Hasil penelitian menunjukkan akurasi sebesar 100.0%. Maka Naive bayes Classification dikategorikan baik untuk pengujian status gizi balita.

**Kata kunci:** Machine Learning, Naïve Bayes, Gizi Buruk.

### PENDAHULUAN

Teknologi membuat segala sesuatu yang dilakukan menjadi lebih mudah. Manusia selalu berusaha untuk menciptakan sesuatu yang dapat mempermudah aktivitasnya, hal inilah yang mendorong perkembangan teknologi yang telah banyak menghasilkan alat

sebagai piranti untuk mempermudah kegiatan manusia, bahkan menggantikan peran manusia dalam suatu fungsi tertentu. Teknologi memegang peran penting di era globalisasi pada saat ini, dimana teknologi telah menjadi bagian yang tidak dapat dipisahkan dalam kehidupan sehari-hari. Perkembangan teknologi saat ini telah merambah ke

segala aspek kehidupan, salah satunya yaitu dibidang kesehatan khususnya pada status gizi.

Status gizi adalah ekspresi dari keadaan keseimbangan dalam bentuk variabel tertentu atau perwujudan nutrire dalam bentuk variabel tertentu, status gizi optimal adalah keseimbangan antara asupan dan kebutuhan zat gizi (Hasrul et al., 2020). Berdasarkan hasil Riset Kesehatan Dasar (Kemenkes, 2018) didapatkan status gizi anak 5- 12 tahun menurut indeks massa tubuh/umur di Indonesia, yaitu prevalensi kurus adalah 9,3, terdiri dari 2,5% sangat kurus dan 6,8% kurus. Masalah gemuk pada anak di Indonesia juga masih tinggi dengan prevalensi 20,6% terdiri dari gemuk 11,1% dan sangat gemuk (obesitas) 9,5%. Sedangkan prevalensi pendek yaitu 23,6% terdiri dari 6,7 sangat pendek dan 16,9% pendek (Farida et al., 2022).

Status gizi dapat ditentukan melalui pemeriksaan laboratorium maupun secara antropometri. Akan tetapi pengukuran indeks antropometri sering terjadi kerancuan karena Pengolahan data status gizi balita yang dilakukan oleh puskesmas masih menggunakan perhitungan dan analisa secara manual sehingga hasil yang didapatkan dirasa kurang efektif dalam menentukan status gizi pada balita. Sama halnya dengan Pengolahan data status gizi balita yang dilakukan oleh puskesmas Kecamatan Air Batu. Pegawai Puskesmas masih menggunakan perhitungan dan analisa secara manual dengan perhitungan gizi balita yang banyak jumlahnya dan dihitung secara manual, artinya prosesnya sangat memakan waktu dan sering menimbulkan kesalahan dalam memasukan data membuat proses tidak efisien.

Oleh karena itu diperlukan suatu metode yang digabungkan dengan ilmu komputer untuk mendapatkan hasil yang akurat. Teknologi informasi di dunia kesehatan dapat membantu memudahkan permasalahan yang dihadapi. Teknologi informasi dapat digunakan untuk membantu pegawai puskesmas atau posyandu dalam menentukan klasifikasi

status gizi balita. Metode yang digunakan adalah Naïve Bayes yang merupakan sebuah metode pengelompokan statistik yang bisa dipakai untuk memprediksi probabilitas anggota suatu class (Evanko, 2010). Algoritma Naïve Bayes adalah algoritma yang memanfaatkan teori kemungkinan, yakni memprediksi kemungkinan di masa depan berdasarkan pengalaman pada masa sebelumnya ( et al., 2021).

Machine Learning (ML) adalah salah satu aplikasi dari Artificial Intilligent(AI) yang fokus kepada pengembangan sebuah sistem yang mampu belajar sendiri tanpa harus diprogram berulang kali(Chazar & Erawan, 2020). ML membutuhkan sebuah data (data traning) sebagai proses learningsebelum menghasilkan sebuah hasil. Jadi, secara sederhana(Alfianty & Mulyati, 2022)na dapat dijelaskan bahwa Machine Learning adalah pemograman komputer untuk mencapai kriteria/performa tertentu denganmenggunakan sekumpulan data traningatau pengalaman di masa lalu. Machine Learning (ML) dapat digunakan untuk mencari dan menemukan pola yang unik dari sekumpulan data(Rizal et al., 2022).

Tujuan penelitian ini adalah menerapkan proses machine learning dalam menentukan status gizi balita terdapat 5 parameter yang diperlukan sebagai inputan dalam perhitungan metode Naive Bayes, yaitu jenis kelamin, umur, berat badan, tinggi badan dan gaji orang tua. Sehingga akan menghasilkan sebuah informasi tentang gizi pada balita berdasarkan semua kriteria yang ada. Dengan adanya teknologi ini diharapkan nantinya proses penentuan status gizi bisa dilakukan dengan lebih cepat dan dapat mengurangi tingkat akurasi penentuan status gizi.

Beberapa penelitian terdahulu yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Pada penerapannya dalam penelitian di bidang Kesehatan, metode Naive Bayes juga telah banyak digunakan salah satunya oleh Ayu Novita Sari dengan judul “Klasifikasi Penyakit Jantung Menggunakan Metode Naive Bayes”(Sari & Alfionita, 2022). Pada penelitian tersebut Naive Bayes digunakan sebagai metode untuk mengklasifikasi resiko penyakit jantung dari segi umur, jenis kelamin, kolestrol dan kriteria lainnya.
2. Penelitian lainnya yang dilakukan oleh Handoko menunjukkan bahwa ini algoritma naïve bayes digunakan untuk melakukan penentuan penyakit selama masa kehamilan. Hasil dari pengujian yang dilakukan dengan ketepatan diagnosa yang diperoleh dari perbandingan antara hasil diagnosa sistem dengan diagnosa pakar adalah sebesar 77%.
3. Penelitian lainnya yang dilakukan oleh Wibisono dengan tujuan utama dalam penelitiannya adalah mengkaji penerapan metode Naive Bayes untuk menentukan email spam dan email ham.(Wibisono et al., 2020) Hasil pengujian aplikasi terhadap 5 email yang terdiri dari 2 email spam dan 3 email ham

Sedangkan penelitian yang akan peneliti lakukan yaitu menerapkan metode naïve bayes pada data set gizi buruk dengan menggunakan beberapa skenario pengujian data testing. Adapun pengujian data testing yang akan dilakukan yaitu dengan split data training dan testing 90:10, 80:20, 70:30, dan 60:40. Tujuan dari skenario pengujian ini adalah untuk mendapatkan performance naïve bayes terbaik yang dilihat dari performance akurasi yang dihasilkan dari masing-masing pengujian.

## METODE

Metode riset memuat hal – hal penting terkait pelaksanaan penelitian dan tahapan – tahapan yang akan ditempuh dalam penelitian ini untuk mencapai

suatu hasil atau *output* yang diinginkan. Berikut ini hal – hal yang mencakup ke dalam metode riset antara lain :

### 1. Peralatan Riset

Peralatan yang akan digunakan dalam penelitian ini adalah berupa (1) Komputer dengan spesifikasi yang mendukung, (2) *software* yang akan digunakan dalam proses pengolahan berupa *software Jupyter notebook* dalam *Anaconda Navigator* dimana bahasa pemrograman yang akan digunakan adalah bahasa pemrograman *Python*, *Visual Studio Code* sebagai *teks editor* dan menggunakan HTML sebagai bahasa *markup* yang digunakan untuk membangun *deploy* aplikasinya.

### 2. Variabel yang Terdapat dalam Riset

Dalam *machine learning*, penelitian ini dikategorikan sebagai penelitian *supervised learning* yaitu sebuah teknik model dimana data yang akan diproses memiliki label/*target/class* dengan tujuan untuk mengetahui hubungan kausalitas antara variabel bebas (variabel independen) dan variabel yang menjadi target/labelnya (variabel y). Oleh sebab itu, variabel yang terdapat dalam penelitian ini terdiri dari variabel independen atau variabel x meliputi *Jenis Kelamin, Umur, Berat Badan, Tinggi Badan, dan Gaji Orang Tua*. Serta yang menjadi variabel dependennya adalah Status Gizi pada Balita.

Tahapan yang akan ditempuh dalam penelitian ini dimulai dari tahapan pertama yang akan dilakukan adalah *Data Preprocessing* dimana dalam proses tersebut terdapat *data cleaning* untuk melakukan *handling missing* pada data, *data selection* untuk menyeleksi data berdasarkan kolom yang akan diteliti dan *data transformation*. Tahapan kedua adalah melakukan *data visualization* yaitu untuk menyajikan representasi dari sebaran data yang digunakan. Tahapan selanjutnya adalah melakukan pemisahan data set menjadi 2 bagian yaitu *data training* dan *data testing*. Setelah itu, memasuki tahapan proses klasifikasi dengan mencari rasio terbaik.

## HASIL DAN PEMBAHASAN

Berdasarkan pengumpulan data yang telah dilakukan data riset yang akan digunakan adalah dataset untuk menentukan status gizi yang berasal dari Puskesmas Air Batu pada tahun 2022. Data tersebut berjumlah sebanyak 21 baris data. Terdapat 5 kolom (*Umur, Jenis Kelamin, Berat*

*Badan, Tinggi Badan, Gaji Orang Tua*) yang akan digunakan untuk pengujian dan 1 kolom (*Status Gizi*) sebagai kolom targer/label. Sampel dataset yang akan digambarkan dalam penelitian ini menggunakan 5 data teratas seperti yang terdapat dalam Tabel (1) berikut ini :

**Tabel 1.** Sampel Dataset

Umur(bulan)	JK	BB(kg)	TB(cm)	Gaji(ribu)	Status gizi
3	0	12	8	4	2
8	0	11	14	7	0
13	1	9	13	3	0
5	0	8	6	3	0
7	0	3	4	6	1

### Pembagian Data

Pembagian data dalam penelitian ini menggunakan *data training* dan *data testing* dengan rasio sebesar 90:10 Dimana *data training* memiliki kisaran data sebanyak 90 dari dataset serta *data testing* sebanyak 10 dari dataset. Algoritma yang menggunakan rasio 90:10 memiliki performa yang lebih baik dibandingkan dengan algoritma lainnya dimana tingkat akurasi dapat diperoleh sebesar 99,40%. Adapun rasio 80:20, 70:30, dan 60:40 yang digunakan dalam penelitian ini sehingga rasio pembagian data lebih bervariasi dimana tujuan yang diharapkan adalah agar memperoleh rasio terbaik manakah yang memiliki hasil klasifikasi terbaik.

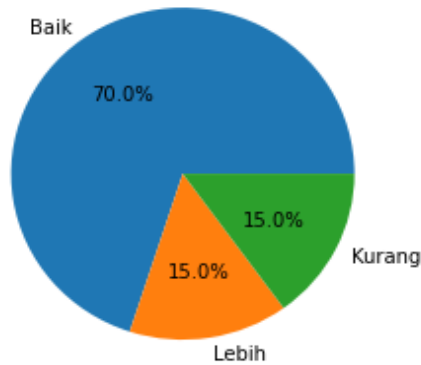
Pada pembahasan dalam penelitian ini akan dilakukan persentase data agar data dapat lebih mudah untuk diamati serta dapat mengetahui hubungan antara tiap variabel *x* terhadap variabel *y*. Selanjutnya, data akan diuji dengan 4 bentuk pembagian data dengan rasio yang berbeda – beda Cara ini bertujuan untuk

menentukan rasio manakah yang paling baik dengan menghasilkan nilai akurasi yang paling akurat pada klasifikasi status gizi buruk pada balita. Adapun tahapan pembahasan yang dilakukan yaitu :

### Persentase Data

Persentase data juga memiliki tujuan untuk memudahkan seseorang dalam mengamati dan memahami sebuah data. Menggunakan berbagai model yang bisa digunakan dalam menyajikan sebuah data kedalam bentuk visualnya. *PIE Chart* menggunakan *library* yaitu “*matplotlib.pyplot*”. Dalam penelitian tersebut persentase data menggunakan *PIE Chart* juga bertujuan untuk menampilkan total persentase yang harus mencapai 100%, yang mana setiap potongan pie akan menampilkan ukuran tertentu. Data pada status gizi balita yang digunakan untuk mengklasifikasikan data pada status gizi balita yang telah divisualisasikan terhadap sumbu *x* nya. Visualisasi data disajikan dalam bentuk antar variabel *x* dengan variabel *y* dan penyajian *pie chart* tiap variabel

menggunakan warna yang berbeda agar lebih mudah dipahami dan mudah diingat. Visualisasinya data tersebut sebagai berikut ini :



**Gambar 2.** Pie chart terhadap persentase pada status gizi balita

Berdasarkan chart diatas dataset yang dikumpulkan oleh puskesmas memiliki porsi kelas target yaitu baik 70%, sedangkan kurang dan lebih memiliki target yang balance.

#### Pengujian Data dengan Naïve Bayes

Pada penelitian ini menggunakan metode naïve bayes dimana akan dilakukan perbandingan dari pembagian data untuk diuji sebanyak 4 bentuk diantaranya pembagian rasio pertama yaitu 90:10, kedua 80:20, ketiga 70:30 dan rasio keempat yaitu 60:40. Pengujian data dengan rasio yang bervariasi untuk mencari serta menentukan model regresi manakah yang paling sering muncul dalam menghasilkan nilai klasifikasi yang terbaik. Sehingga, nantinya akan diketahui rasio mana yang terbaik dalam melakukan klasifikasi terhadap dataset untuk mencari hasil status gizi pada balita. Berikut pembahasan penelitian untuk menentukan rasio mana yang terbaik diuraikan kedalam rincian poin sebagai berikut :

#### Pengujian Akurasi Pada Rasio

Pada pengujian data dilakukan perbandingan dari pembagian data untuk diuji sebanyak 4 bentuk diantaranya pembagian rasio pertama yaitu 90:10,

kedua 80:20, ketiga 70:30 dan rasio keempat yaitu 60:40. Pengujian data dengan rasio yang bervariasi untuk mencari serta menentukan model regresi manakah yang paling sering muncul dalam menghasilkan nilai klasifikasi yang terbaik. Dengan melakukan pembagian pada data training dan data testing dari dataset, diperoleh hasil seperti tabel dibawah ini:

**Tabel 2.** Pengujian Akurasi

RASIO	Hasil Akurasi
90:10	1.0
80:20	1.0
70:30	1.0
60:40	0.75
Jumlah	3.75
Rata-rata	0.9375

Berdasarkan tabel diatas, dari pengujian data yang dilakukan sebanyak 4 bentuk diantaranya rasio pertama 90:10 dengan hasil 1.0, rasio kedua 80:20 dengan hasil 1.0, rasio ketiga 70:30 dengan hasil 1.0, dan rasio keempat 60:40 adalah 0.75. Maka dapat diketahui pengujian akurasi data menghasilkan 3 hasil rasio yang sama pada rasio 90:10, 80:20, dan 70:30.

#### KESIMPULAN

Berdasarkan hasil pengolahan data menggunakan metode naïve bayes menggunakan *software jupyter notebook*, *visual studio code* dan *python* sebagai bahasa pemrogramannya serta pengujiannya menggunakan 4 rasio pengukuran nilai akurasi yaitu 90:10 dengan nilai akurasi 1.0, 80:20 dengan nilai akurasi 1.0, 70:30 memiliki nilai akurasi 1.0 dan 60:40 dengan nilai akurasi 0.75 maka diperoleh rasio terbaik menghasilkan 3 nilai akurasi paling bagus diantaranya yaitu pada nilai akurasi 1.0 pada pengujian rasio 90:10, 80:20 dan 70:30. Pada pengujian 4 rasio menghasilkan performance dari naïve

bayes secara akumulatif rata-rata dengan nilai 0.9375. Dengan demikian pada pengujian rasio 80:20 digunakan sebagai metode naïve bayes yang akan di implementasikan dalam pembuatan *deploy* aplikasi untuk klasifikasi hasil pada status gizi Balita.

#### DAFTAR PUSTAKA

- Alfianty, N. H., & Mulyati, S. (2022). Penerapan Naïve Bayes untuk Klasifikasi Data Penyakit Pada Anak. *Automata*. <https://journal.uui.ac.id/AUTOMAT-A/article/view/21914>
- Chazar, C., & Erawan, B. (2020). Machine Learning Diagnosis Kanker Payudara Menggunakan Algoritma Support Vector Machine. *INFORMASI (Jurnal Informatika Dan Sistem Informasi)*, 12(1), 67–80. <https://doi.org/10.37424/informasi.v12i1.48>
- Darnila, E., Maryana, M., & Azmi, M. (2021). Aplikasi Klasifikasi Status Gizi Balita Menggunakan Metode Naïve Bayes Berbasis Android. *METHOMIKA Jurnal Manajemen Informatika Dan Komputerisasi Akuntansi*, 5(2), 135–141. <https://doi.org/10.46880/jmika.vol5n02.pp135-141>
- Evanko, D. (2010). Optical imaging of the native brain. *Nature Methods*, 7(1), 34. <https://doi.org/10.1038/nmeth.f.284>
- Farida, N., Farida, I. N., & Sahertian, J. (2022). Implementasi Metode Naïve Bayes Classification Dalam Menentukan Gizi Balita Menggunakan Indeks Antropometri. 107–112.
- Hasrul, H., Hamzah, H., & Hafid, A. (2020). Pengaruh Pola Asuh Terhadap Status Gizi Anak. *Jurnal Ilmiah Kesehatan Sandi Husada*, 12(2), 792–797. <https://doi.org/10.35816/jiskh.v12i2.403>
- Rizal, R., Martanto, M., & Arie Wijaya, Y. (2022). Analisa Dataset Software Defined Network Intrusion Menggunakan Algoritma Deep Learning H2O. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 6(2), 747–757. <https://doi.org/10.36040/jati.v6i2.5724>
- Sari, A. N., & Alfionita, S. (2022). Klasifikasi Penyakit Jantung Menggunakan Metode Naïve Bayes. *AMRI (Analisa Metode Rekayasa Informatika)*, 1(1), 22–26. <https://doi.org/10.12487/AMRI.v1i1.xxxxx>
- Wibisono, A. D., Dadi Rizkiono, S., & Wantoro, A. (2020). Filtering Spam Email Menggunakan Metode Naive Bayes. *TELEFORTECH: Journal of Telematics and Information Technology*, 1(1). <https://doi.org/10.33365/tft.v1i1.685>