

# Penerapan Algoritma Naïve Bayes Classifier Dalam Klasifikasi Status Gizi Balita dengan Pengujian K-Fold Cross Validation

Nurainun\*, Elin Haerani, Fadhilah Syafria, Lola Oktavia

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

Email: <sup>1</sup>\*11950125156@students.uin-suska.ac.id, <sup>2</sup>elin.haerani@uin-suska.ac.id, <sup>3</sup>fadhilah.syafria@uin-suska.ac.id, <sup>4</sup>lola.oktavia@uin-suska.ac.id

Email Penulis Korespondensi: 11950125156@students.uin-suska.ac.id

**Abstrak**—Status gizi merupakan suatu kondisi terkait gizi yang bisa diukur dan merupakan hasil dari adanya keseimbangan kebutuhan gizi pada tubuh dengan asupan gizi dari makanan. Di Indonesia masalah gizi masih banyak ditemukan seperti gizi buruk, dan masalah gizi lainnya. Pada penelitian ini akan menggunakan algoritma Naïve Bayes Classifier dengan pengujian K-Fold Cross Validation. Data yang digunakan adalah data status gizi balita pada bulan Agustus 2022 di Puskesmas Rambah Samo I. Atribut pada penelitian ini diantaranya Jenis Kelamin, Berat Badan Lahir, Tinggi Badan Lahir, Usia Saat Ukur, Berat Badan, Tinggi Badan, ZS BB/U, BB/U, ZS TB/U, dan TB/U. Penentuan status gizi balita pada penelitian ini berdasarkan indeks BB/TB yang terdiri dari 6 kelas, yaitu gizi buruk, gizi kurang, gizi baik, risiko gizi lebih, gizi lebih, dan obesitas. Dari penelitian yang dilakukan didapatkan bahwa algoritma Naïve Bayes Classifier dengan K-Fold Cross Validation bisa mengklasifikasikan status gizi balita secara tepat. Dari pengolahan data dengan menggunakan 10-Fold Cross Validation pada algoritma Naïve Bayes Classifier yang dilakukan, diketahui bahwa nilai akurasi yang paling tinggi yaitu sebesar 82,94% pada iterasi ke-5, sedangkan nilai akurasi paling rendah yaitu sebesar 65,88% pada iterasi ke-6. Dengan rata-rata nilai akurasi secara keseluruhan yaitu sebesar 75,47%. Sedangkan untuk rata-rata nilai presisi yang didapatkan yaitu sebesar 81,36% dan rata-rata nilai recall sebesar 75,47%.

**Kata Kunci:** Status Gizi Balita; Klasifikasi; Algoritma Naive Bayes Classifier

**Abstract**—Nutritional status is a condition related to nutrition that can be measured and is the result of a balance between nutritional needs in the body and nutritional intake from food. In Indonesia, there are still many nutritional problems such as malnutrition and other nutritional problems. This research will use the Naïve Bayes Classifier algorithm with K-Fold Cross Validation testing. The data used is data on the nutritional status of toddlers in August 2022 at the Rambah Samo I Health Center. Attributes in this study include Gender, Birth Weight, Birth Height, Age at Measurement, Weight, Height, ZS BB/U, BB/U, ZS TB/U, and TB/U. Determination of the nutritional status of toddlers in this study was based on the BB/TB index which consisted of 6 classes, namely severely wasted, wasted, normal, possible risk of overweight, overweight, and obese. From the research conducted, it was found that the Naïve Bayes Classifier algorithm with K-Fold Cross Validation can correctly classify the nutritional status of toddlers. From data processing using 10-Fold Cross Validation on the Naïve Bayes Classifier algorithm, it is known that the highest accuracy value is 82.94% in the 5th iteration, while the lowest accuracy value is 65.88% in 6th iteration. With an average overall accuracy value of 75.47%. Meanwhile, the average precision value obtained is 81.36% and the average recall value is 75.47%.

**Keywords:** Nutritional Status Of Children Under Five; Classification; Naïve Bayes Classifier Algorithm

## 1. PENDAHULUAN

Status gizi merupakan suatu kondisi terkait gizi yang bisa diukur dan merupakan hasil dari adanya keseimbangan kebutuhan gizi pada tubuh dengan asupan gizi dari makanan[1]. Status gizi balita merupakan suatu hal yang penting karena berpengaruh akan pertumbuhan dan perkembangannya. Gizi yang seimbang memiliki beberapa manfaat yang penting bagi tubuh, diantaranya: tumbuh sehat, perkembangan otak sempurna, daya tahan tubuh kuat, kesehatan gigi dan mulut terjaga, lebih konsentrasi, suasana hati lebih stabil dan lebih berenergi[2]. Akan tetapi di Indonesia masalah gizi masih banyak ditemukan. Berdasarkan dari beberapa penelitian menunjukkan bahwa di Indonesia masalah gizi cenderung meningkat, berbeda dengan di negara Malaysia atau Thailand[3]. Masalah kesehatan yang cukup banyak diderita di Indonesia khususnya wilayah bagian timur adalah gizi buruk[4]. Selain gizi buruk, anemia, stunting, anak yang terlalu kurus, dan kelebihan berat badan juga menjadi masalah gizi lainnya yang terdapat di Indonesia.

Pada tahun 2018 data dari Global Nutrition Report menunjukkan bahwa secara umum untuk di seluruh dunia terdapat 22,2% balita stunting (kekerdilan), sebanyak 7,5% balita kurus, serta terdapat sekitar 5,6% balita gemuk. Sedangkan menurut Budi dalam webinar Hari Puncak Pekan Menyusui Sedunia 2021 diketahui bahwa secara khusus di Indonesia pada tahun 2018 terdapat sebanyak 30,8% balita kerdil, sebanyak 10,2% balita kurus dan sebanyak 8% balita gemuk berdasarkan data riset Kesehatan dasar. Sehingga dapat diketahui bahwa di Indonesia masalah gizi pada balita masih cukup tinggi[5]. Selain itu permasalahan gizi (gizi kurang) mengakibatkan dampak negative bagi balita, diantaranya pertumbuhan menjadi terhambat, sistem imunitas terganggu yang menyebabkan balita menjadi mudah terserang penyakit, dapat mengakibatkan pertumbuhan otak menjadi kurang optimal, hingga terjadinya perubahan pada perilaku anak seperti anak menjadi tidak tenang, mudah menangis, dan juga berperilaku apatis[6]. Itulah kenapa status gizi balita harus dijaga agar dapat mengurangi permasalahan gizi di Indonesia serta dapat terhindar dari dampak negatif permasalahan gizi tersebut.

Pada puskesmas Rambah Samo I ini pengukuran status gizi dilakukan secara manual di posyandu yang terdapat di setiap desa di Kecamatan Rambah Samo. Tenaga medis dari puskesmas akan mengukur berat badan dan tinggi badan setiap balita dan selanjutnya di data secara manual. Selanjutnya tenaga medis akan melihat dan melakukan perhitungan z skor (standar deviasi) pada tabel standar antropometri penilaian status gizi balita. Akan tetapi puskesmas sering kali memerlukan waktu yang relatif lama untuk menangani pasien dengan jumlah yang banyak serta dengan tenaga medis yang terbatas. Sehingga hal ini menyebabkan penilaian status gizi menjadi kurang optimal. Oleh karena itu peneliti akan menerapkan data mining untuk proses pengklasifikasian status gizi balita di puskesmas Rambah Samo I.

Data mining adalah suatu proses untuk mendapatkan pengetahuan dan informasi di dalam suatu database dengan menggabungkan beberapa teknik seperti machine learning, statistik matematika, dan artificial intelligence sehingga ditemukan suatu pola dari hal tersebut[7]. Klasifikasi dalam data mining adalah proses untuk mengetahui kelas atau kategori dari suatu objek diantara beberapa kelas atau kategori yang sudah didefinisikan sebelumnya[8]. Pada penelitian ini akan menggunakan algoritma Naïve Bayes Classifier. Algoritma Naïve Bayes Classifier (NBC) adalah suatu metode dalam klasifikasi yang dapat digunakan untuk memprediksi kemungkinan anggota dari suatu kelas[9]. Pada penelitian sebelumnya dengan judul “Perbandingan Metode Classification and Regression Trees (CART) dengan Naïve Bayes Classification (NBC) dalam Klasifikasi Status Gizi Balita di Kelurahan Pagesangan Barat” didapatkan hasil dari akurasi ketiga perbandingan data training dan data testing yang dilakukan memperlihatkan bahwa algoritma Naïve Bayes Classification (NBC) lebih baik dibandingkan dengan algoritma Classification and Regression Trees (CART). Hal ini menunjukkan bahwa Algoritma Naïve Bayes Classifier merupakan algoritma yang baik untuk digunakan. Algoritma NBC ini juga digunakan pada penelitian lainnya seperti pada kasus klasifikasi penyakit jantung[10], klasifikasi penyakit stroke[11], klasifikasi resiko penyakit diabetes mellitus[12], klasifikasi penyakit paru-paru[13], dan prediksi penyakit ginjal kronis[14].

Pada penelitian ini peneliti juga menambahkan penggunaan metode K-Fold Cross Validation. K-Fold Cross Validation berguna untuk mengevaluasi performa dari proses suatu algoritma dengan cara data sampel yang digunakan akan dibagi secara random dan kemudian akan dikelompokkan sebanyak nilai K yang digunakan[15]. Dan K-Fold Cross Validation termasuk jenis pengujian Cross Validation. Salah satu kelebihan yang didapatkan dari penggunaan K-Fold Cross Validation dalam pengujian adalah bisa mengetahui model dengan nilai akurasi paling baik dikarenakan sampel data dibagi secara random ke dalam K-partisi sehingga komposisi model yang paling baik dapat diketahui[16].

Pada penelitian ini akan menggunakan parameter yang umum digunakan untuk klasifikasi status gizi balita seperti umur, jenis kelamin, berat badan, dan tinggi badan. Akan tetapi selain itu peneliti juga akan menggunakan parameter berat badan lahir, dan tinggi badan lahir. Parameter berat badan lahir dan tinggi badan lahir digunakan karena adanya kaitan antara BB dan TB balita pada saat lahir dengan status gizi balita tersebut[17]. Pada penelitian tersebut diketahui bahwa balita dengan berat badan lahir rendah berisiko mengalami keterlambatan dalam perkembangan dan pertumbuhannya sehingga dapat mengalami kekurangan gizi, dibandingkan dengan balita yang tidak memiliki riwayat berat badan lahir rendah. Kemudian pada jurnal kesehatan lainnya juga memaparkan bahwa berat badan lahir dan panjang badan lahir termasuk salah satu faktor yang bisa meningkatkan stunting pada balita yang berarti juga mempengaruhi status gizinya[18][19]. Selain BB Lahir dan TB Lahir, pada penelitian ini juga akan menggunakan parameter BB/U, ZS BB/U, TB/U, ZS TB/U yang akan membantu proses klasifikasi untuk menghasilkan status gizi balita sesuai dengan indeks BB/TB yang terdiri dari 6 kelas, yaitu gizi buruk, gizi kurang, gizi baik, risiko gizi lebih, gizi lebih, dan obesitas.

Penelitian ini dilakukan dengan tujuan untuk menerapkan algoritma Naïve Bayes Classifier dan K-Fold Cross Validation dalam klasifikasi status gizi balita serta dapat menghasilkan model yang terbaik. Sehingga diharapkan dapat memberikan kontribusi dan manfaat sehingga pengukuran status gizi balita menjadi lebih efektif dan efisien.

## 2. METODOLOGI PENELITIAN

### 2.1 Klasifikasi

Suatu teknik dalam data mining yang bekerja dengan mengidentifikasi sifat suatu atribut dari kelompok data yang telah diketahui dan ditentukan sebelumnya disebut dengan klasifikasi. Salah satu ciri dari klasifikasi adalah memiliki kelas atau target. Teknik ini bekerja dengan cara mengklasifikasikan data yang ada pada suatu kelas atau target untuk mendapatkan suatu model atau aturan[20]. Model atau aturan yang sudah didapatkan tersebut akan digunakan untuk mengklasifikasikan data baru ke dalam kelas atau target tertentu pada masa mendatang.

### 2.2 Algoritma Naïve Bayes Classifier

Pada teknik klasifikasi ada berbagai algoritma yang bisa dipakai, salah satunya yaitu algoritma Naïve Bayes Classifier. Algoritma Naïve Bayes Classifier (NBC) adalah suatu metode dalam klasifikasi yang dapat digunakan untuk memprediksi kemungkinan anggota dari suatu kelas[9]. Berikut persamaan dari naïve bayes :

$$P(A|B) = \frac{P(A)}{P(B)} \times P(B|A) \quad (1)$$

Keterangan:

A : Data sampel yang tidak diketahui label kelasnya.

B : Kelas hasil klasifikasi

$P(A|B)$  : Probabilitas terjadinya A jika B diketahui

$P(B|A)$  : Probabilitas terjadinya B jika A diketahui

$P(A)$  : Probabilitas terjadinya A

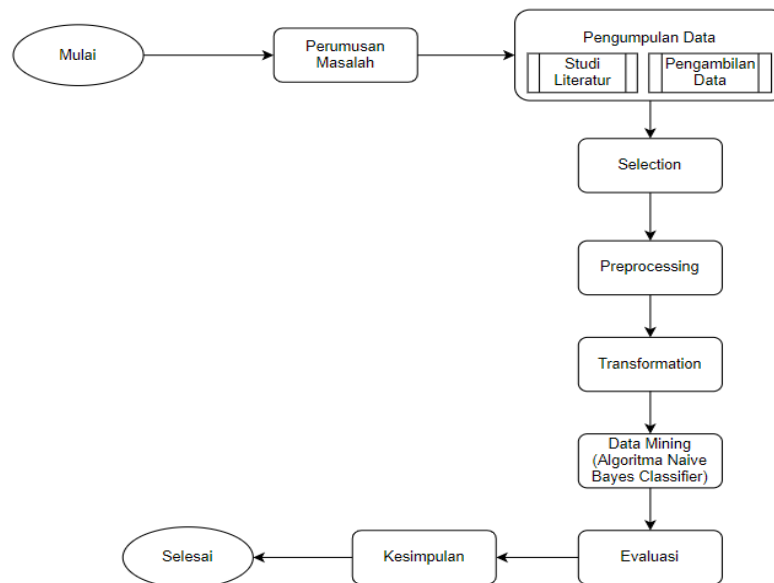
$P(B)$  : Probabilitas terjadinya B

### 2.3 K-Fold Cross Validation

Suatu metode dalam data mining yang dipakai untuk mendapatkan nilai akurasi paling baik ketika data dibagi menjadi data uji dan data latih disebut cross validation. K-Fold Cross Validation berguna untuk mengevaluasi performa dari proses suatu algoritma dengan cara data sampel yang digunakan akan dibagi secara random dan kemudian akan dikelompokkan sebanyak nilai K yang digunakan[15]. Pada metode K-Fold Cross Validation, dataset yang digunakan akan dibagi menjadi beberapa partisi secara random. Setelah dibagi ke dalam beberapa partisi maka data tersebut akan diolah sebanyak K kali percobaan dengan setiap K kali percobaan, data testing yang digunakan adalah data partisi ke-K dan sisa partisi yang lain digunakan sebagai data training. Begitu seterusnya hingga pengolahan berakhir sesuai dengan jumlah K dalam k-fold yang digunakan pada penelitian.

### 2.4 Tahapan Penelitian

Adapun alur dari penelitian yang dilakukan ditunjukkan pada gambar 1 berikut :



Gambar 1. Tahapan Penelitian

Pada gambar 1 tersebut menunjukkan tahapan dari penelitian yang akan dilakukan secara keseluruhan. Mulai dari tahap merumuskan masalah, kemudian mengumpulkan data yang dibutuhkan, hingga kemudian masuk ke tahap KDD yaitu selection, preprocessing, transformation, data mining, dan evaluasi, sehingga kemudian didapatkan kesimpulan dari penelitian yang sudah dilakukan.

### 2.5 Perumusan Masalah

Tahap awal dalam melakukan penelitian adalah rumusan masalah. Pada tahap ini peneliti akan mencari masalah yang ingin diteliti dari berbagai sumber baik itu dari lingkungan sekitar, buku, dll. Kemudian dari masalah yang ditemukan, peneliti akan mempelajari masalah tersebut, dan menemukan solusi untuk mengatasi hal tersebut. Adapun rumusan masalah dalam penelitian ini adalah “Bagaimana menerapkan algoritma Naïve Bayes Classifier dalam klasifikasi status gizi balita pada puskesmas Rambah Samo I” serta “Pengujian K-Fold Cross Validation mana yang menghasilkan akurasi tertinggi”.

### 2.6 Pengumpulan Data

Setelah perumusan masalah selanjutnya akan dilakukan tahap pengumpulan data dengan dua cara yaitu :

#### a. Studi Pustaka

Pada suatu penelitian, tahap awal dalam pengumpulan data adalah studi pustaka (studi literatur). Studi pustaka dilakukan dengan menemukan informasi atau data yang dapat digunakan dan juga mendukung penelitian yang dilakukan dengan menelusurinya dari berbagai sumber seperti buku, karya ilmiah maupun jurnal.

**b. Pengambilan Data**

Selanjutnya akan dilakukan pengambilan data yang digunakan pada penelitian. Data yang akan digunakan berupa data primer yang diambil dari status gizi balita pada bulan Agustus 2022 di Puskesmas Rambah Samo I yang berjumlah sebanyak 1700 data.

**2.7 Data Selection**

Data pada penelitian selanjutnya akan melalui proses KDD atau Knowledge Discovery Database. Tahapan pertama dari proses KDD adalah data selection. Pada tahapan ini terdapat pemilihan atau penseleksian data dari sekumpulan data operasional.

**2.8 PreProcessing/ Cleaning**

Pembersihan data merupakan proses yang dilakukan dalam tahap preprocessing yaitu proses untuk menghapus data yang sama (duplikat), memeriksa data yang inkonsistensi dan data yang memiliki kesalahan akan diperbaiki dan menghapus data yang tidak memiliki value pada salah satu atributnya.

**2.9 Transformation**

Data transformation digunakan untuk mengubah data kedalam tipe yang sesuai dalam data mining. Pada penelitian ini data yang digunakan adalah data dalam tipe numerical, sehingga data yang masih dalam tipe kategorial harus diubah terlebih dahulu. Seperti pada data penelitian ini, atribut jenis kelamin akan diubah kedalam bentuk biner yaitu 0 dan 1, dimana 0 menunjukkan jenis kelamin perempuan dan 1 menunjukkan jenis kelamin laki-laki.

**2.10 Data Mining**

Data mining merupakan proses untuk mendapatkan model, pola, dan aturan dari sekumpulan data besar dengan menggunakan berbagai teknik untuk memprediksi hasil. Pada penelitian ini akan menggunakan algoritma Naïve Bayes Classifier. Sebelum dilakukan klasifikasi data akan dibagi terlebih dahulu menggunakan k-fold cross validation dengan penggunaan k=10.

**2.11 Interpretation / Evaluation**

Tahap evaluasi ini dilakukan dengan cara mengecek nilai akurasi, presisi, dan recall. Akurasi adalah suatu proses yang dilakukan untuk mengevaluasi performa dengan melihat kedekatan nilai prediksi dengan nilai sebenarnya[16]. Nilai akurasi diperoleh dari hasil analisis terkait jumlah data yang dapat diklasifikasikan secara tepat. Adapun persamaan akurasi sebagai berikut.

$$Accuracy = \frac{N_{benar}}{N} \times 100\% \tag{2}$$

Recall adalah suatu proses yang dilakukan untuk mengevaluasi performa dengan melihat perbandingan antara jumlah informasi relevan yang ditemukan dengan jumlah total informasi relevan pada koleksi informasi[16]. Adapun persamaan recall sebagai berikut.

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

Precision atau presisi adalah suatu proses yang dilakukan untuk mengevaluasi performa dengan melihat perbandingan antara jumlah informasi relevan yang ditemukan dengan jumlah total informasi yang diambil oleh sistem baik yang relevan maupun tidak[16]. Adapun persamaan presisi sebagai berikut.

$$Precision = \frac{TP}{TP+FP} \tag{4}$$

**3. HASIL DAN PEMBAHASAN**

**3.1 Data**

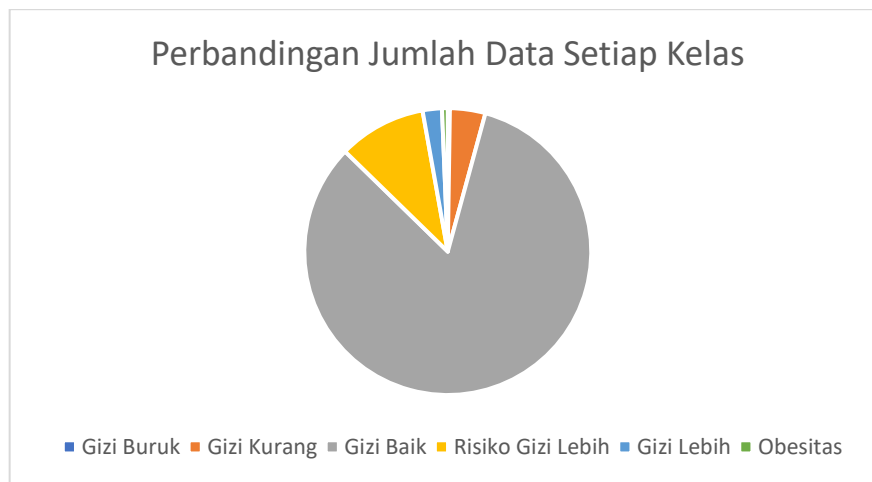
Pada penelitian ini data yang digunakan adalah data status gizi balita pada Puskesmas Rambah Samo I dengan jumlah data sebanyak 1700 record data. Atribut yang digunakan yaitu Jenis Kelamin, Berat Badan Lahir, Tinggi Badan Lahir, Usia Saat Ukur, Berat Badan, Tinggi Badan, ZS BB/U, BB/U, ZS TB/U, TB/U, dan atribut terakhir yaitu sebagai class dengan atribut BB/TB. Atribut dari data pada penelitian ini ditunjukkan pada tabel 1 berikut :

**Tabel 1.** Atribut Data Status Gizi Balita

No	Item	Penjelasan	Jumlah Data
1	Jenis Kelamin	Jenis Kelamin Balita yaitu Perempuan atau Laki-laki	1700
2	Berat Badan Lahir	Berat Badan Balita saat Lahir	1700
3	Tinggi Badan Lahir	Tinggi Badan Balita saat Lahir	1700

No	Item	Penjelasan	Jumlah Data
4	Usia Saat Ukur	Usia Balita pada saat pengukuran	1700
5	Berat Badan	Berat Badan Balita pada saat pengukuran	1700
6	Tinggi Badan	Tinggi Badan Balita pada saat pengukuran	1700
7	ZS BB/U	Nilai Z-Skor Berat Badan menurut Umur	1700
8	BB/U	Berat Badan menurut Umur yaitu Sangat Kurang, Kurang, Berat Badan Normal, Resiko Lebih	1700
9	ZS TB/U	Nilai Z-Skor Tinggi Badan menurut Umur	1700
10	TB/U	Tinggi Badan menurut Umur yaitu Sangat Pendek, Pendek, Normal, dan Tinggi	1700
11	BB/TB	Berat Badan menurut Tinggi Badan yaitu Gizi Buruk, Gizi Kurang, Gizi Baik, Risiko Gizi Lebih, Gizi Lebih, Obesitas	1700

Pada tabel 1 tersebut berisi seluruh atribut yang ada pada data status gizi balita beserta dengan penjelasan dari masing-masing atributnya, mulai dari atribut jenis kelamin hingga atribut BB/TB yang berfungsi sebagai kelas atau target. Adapun data yang digunakan pada penelitian masih belum seimbang antara setiap kelas targetnya dimana kelas gizi baik lebih banyak dibandingkan dengan kelas lainnya, perbandingan jumlah data antara masing-masing kelas ditunjukkan pada gambar 2 berikut :



Gambar 2. Perbandingan Jumlah Data Setiap Kelas

Pada gambar 2 tersebut menunjukkan bahwa perbandingan jumlah data antara setiap kelasnya tidak seimbang. Dari 6 kelas target tersebut lebih didominasi oleh kelas gizi baik, sedangkan kelas lainnya yaitu gizi buruk, gizi kurang, risiko gizi lebih, gizi lebih dan obesitas jumlahnya jauh lebih kecil dibandingkan gizi baik.

### 3.2 Data Selection

Setelah data pada penelitian didapatkan selanjutnya akan dilakukan proses Knowledge Discovery Database (KDD). Data tersebut akan melalui tahap data selection yang merupakan tahapan pertama dari Knowledge Discovery Database (KDD). Data tersebut akan dipilih atau diseleksi dari sekumpulan data yang ada.

### 3.3 PreProcessing/ Cleaning

Pada penelitian ini menggunakan data yang belum siap untuk diolah. Sehingga perlu dilakukan preprocessing terlebih dahulu yaitu dengan menghilangkan missing value, menghilangkan data duplikat dan juga memperbaiki data yang memiliki kesalahan.

### 3.4 Transformation

Setelah dilakukan preprocessing data, selanjutnya data yang tipenya belum sesuai akan ditransformasi terlebih dahulu. Dari data penelitian tersebut terdapat beberapa atribut yang belum sesuai diantaranya jenis kelamin, usia saat ukur, BB/U, TB/U, dan kelas targetnya yaitu BB/TB. Untuk atribut jenis kelamin akan diubah kedalam bentuk biner yaitu 0 dan 1, dimana 0 menunjukkan jenis kelamin perempuan dan 1 menunjukkan jenis kelamin laki-laki. Untuk atribut usia saat ukur yang awalnya berbentuk x tahun x bulan x hari diubah menjadi kedalam bentuk x bulan, seperti tabel 2 berikut :

Tabel 2. Transformasi Atribut Usia

Sebelum Transformasi	Sesudah Transformasi
0 Tahun – 1 Bulan – 9 Hari	1
0 Tahun – 1 Bulan – 29 Hari	1

Sebelum Transformasi	Sesudah Transformasi
1 Tahun – 2 Bulan – 22 Hari	14
2 Tahun – 3 Bulan – 3 Hari	27

Pada tabel 2 tersebut diketahui bahwa untuk atribut usia balita dihitung dalam bulan penuh, seperti untuk usia 0 Tahun – 1 Bulan – 29 Hari dihitung menjadi 1 bulan[21]. Kemudian untuk atribut BB/U, TB/U, dan kelas targetnya yaitu BB/TB akan ditransformasikan dengan menggunakan label encoding seperti tabel 3 berikut :

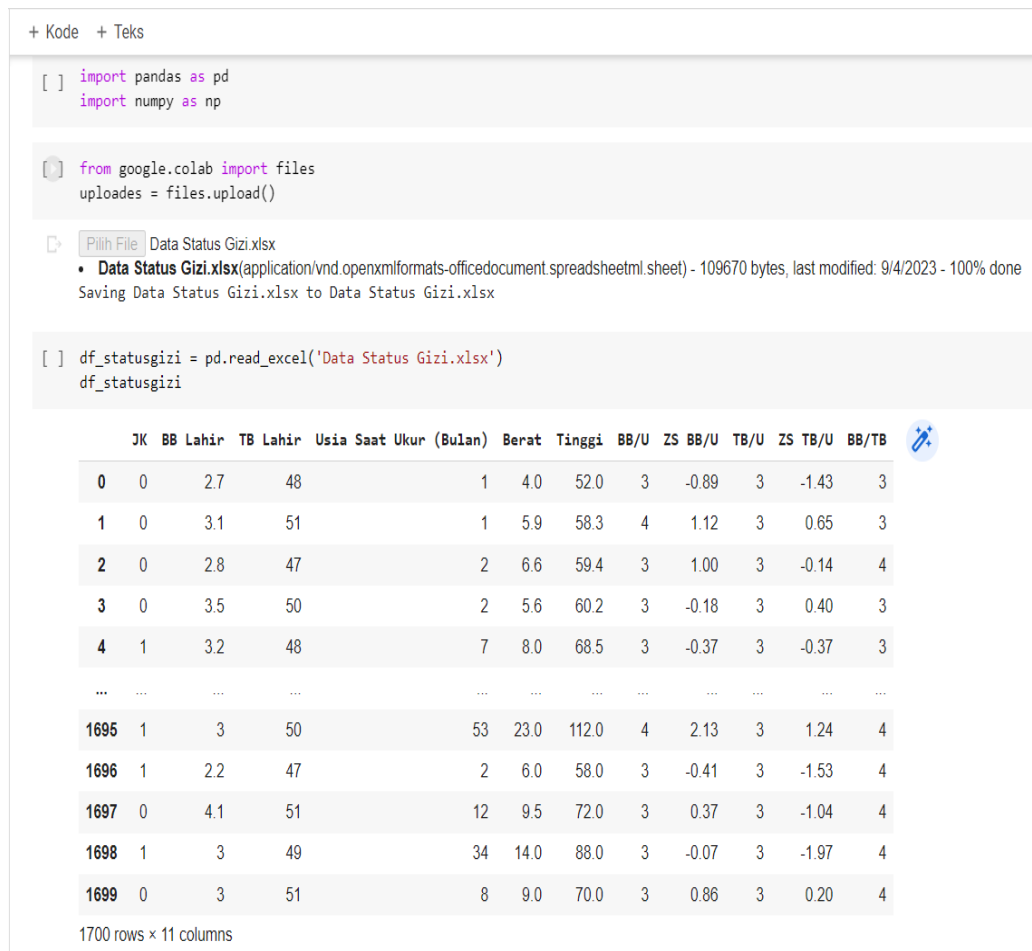
Tabel 3. Transformasi Atribut BB/TB

BB/TB Sebelum Transformasi	BB/TB Setelah Transformasi
Gizi Buruk	1
Gizi Kurang	2
Gizi Baik	3
Risiko Gizi Lebih	4
Gizi Lebih	5
Obesitas	6

Pada tabel 3 tersebut menunjukkan transformasi dari atribut BB/TB yang sebelumnya tipe datanya kategorial diubah menjadi tipe data numerical dengan menggunakan label encoding. Sehingga atribut BB/TB tersebut menjadi dalam bentuk angka 1-6. Adapun label encoding digunakan karena terdapat unsur ordinalitas pada atribut tersebut[22].

### 3.5 Klasifikasi dengan Algoritma Naïve Bayes Classifier

Pada penelitian ini terdapat data sebanyak 1700 data yang diklasifikasi menggunakan algoritma NBC dan pengujian K-Fold Cross Validation dengan bahasa pemrograman python di google colab. Berikut pada gambar 3 ditunjukkan proses input data pada google colab :



Gambar 3. Proses Input Data di Google Colab

Pada gambar 3 tersebut menunjukkan proses ketika memasukkan data yang digunakan, dan juga ditunjukkan isi dari data tersebut. Kemudian dilakukan pengolahan data di google colab dengan menggunakan algoritma NBC dan pengujian K-Fold Cross Validation yang ditunjukkan pada gambar 4 berikut :

```

+ Kode + Teks

[ ] from sklearn.naive_bayes import GaussianNB
    from sklearn.model_selection import KFold, cross_val_score

    k_folds = KFold(n_splits=10)

[ ] classifier = GaussianNB()

[ ] akurasi_10fold = cross_val_score(classifier, x, y, scoring='accuracy', cv=k_folds)
    print("Akurasi 10 Fold = {}".format(akurasi_10fold))
    print("Rata-rata akurasi 10 Fold = {}".format(akurasi_10fold.mean()*100))

Akurasi 10 Fold = [0.71176471 0.74117647 0.68823529 0.78823529 0.82941176 0.65882353
0.74117647 0.81176471 0.76470588 0.81176471]
Rata-rata akurasi 10 Fold = 75.47058823529412

▶ presisi_10fold = cross_val_score(classifier, x, y, scoring='precision_weighted', cv=k_folds)
    print("Presisi 10 Fold = {}".format(presisi_10fold))
    print("Rata-rata presisi 10 Fold = {}".format(presisi_10fold.mean()*100))

☐ Presisi 10 Fold = [0.72867428 0.78718636 0.72081275 0.84437811 0.88168948 0.73069204
0.82411667 0.90247293 0.85575923 0.85976967]
Rata-rata presisi 10 Fold = 81.35551531177863

▶ recall_10fold = cross_val_score(classifier, x, y, scoring='recall_weighted', cv=k_folds)
    print("Recall 10 Fold = {}".format(recall_10fold))
    print("Rata-rata recall 10 Fold = {}".format(recall_10fold.mean()*100))

Recall 10 Fold = [0.71176471 0.74117647 0.68823529 0.78823529 0.82941176 0.65882353
0.74117647 0.81176471 0.76470588 0.81176471]
Rata-rata recall 10 Fold = 75.47058823529412
    
```

Gambar 4. Proses Pengolahan Data di Google Colab

Pada gambar 4 tersebut menunjukkan proses membuat model NBC, kemudian menggunakannya pada data dan melakukan evaluasi berupa menilai akurasi, presisi, dan recall. Evaluasi ini bertujuan untuk menguji performa dari algoritma Naïve Bayes Classifier dalam klasifikasi status gizi balita, dan menentukan model terbaik, yang kemudian didapatkan hasil pengujian seperti tabel 4 berikut :

Tabel 4. Hasil Pengujian

Fold ke-	Akurasi	Presisi	Recall
1	71.18%	72.86%	71.18%
2	74.12%	78.72%	74.12%
3	68.82%	72.08%	68.82%
4	78.82%	84.44%	78.82%
5	82.94%	88.17%	82.94%
6	65.88%	73.07%	65.88%
7	74.12%	82.41%	74.12%
8	81.18%	90.25%	81.18%
9	76.47%	85.58%	76.47%
10	81.18%	85.98%	81.18%

Pada tabel 4 tersebut diketahui bahwa berdasarkan pengolahan data dengan penggunaan 10-Fold Cross Validation pada algoritma Naïve Bayes Classifier yang dilakukan, didapatkan nilai akurasi yang paling tinggi yaitu sebesar 82,94% pada iterasi ke-5, sedangkan nilai akurasi paling rendah yaitu sebesar 65,88% pada iterasi ke-6. Dengan rata-rata nilai akurasi secara keseluruhan yaitu sebesar 75,47%. Sedangkan untuk rata-rata nilai presisi yang didapatkan yaitu sebesar 81,36% dan rata-rata nilai recall sebesar 75,47%.

Pada penelitian ini juga dilakukan beberapa percobaan klasifikasi dengan algoritma Naïve Bayes Classifier dan pengujian K-Fold Cross Validation. Pada setiap percobaan menggunakan K-fold yang sama yaitu 10-Fold akan tetapi terdiri dari jumlah data yang berbeda yaitu 500 data, 1000 data, 1300 data, 1500 data, dan 1700 data. Kinerja klasifikasi dari setiap percobaan ditunjukkan pada tabel 5 sebagai berikut :

Tabel 5. Hasil Kinerja Klasifikasi Setiap Percobaan Dengan Jumlah Data yang Berbeda

Percobaan	Jumlah Data	Akurasi	Presisi	Recall
1	500 Data	57.8%	65.87%	57.8%
2	1000 Data	67%	72.87%	67%
3	1300 Data	71.62%	78.98%	71.62%
4	1500 Data	73.47%	80.33%	73.47%

Percobaan	Jumlah Data	Akurasi	Presisi	Recall
5	1700 Data	75.48%	81.36%	75.48%

Pada tabel 5 tersebut menunjukkan bahwa berdasarkan percobaan yang dilakukan didapatkan bahwa jumlah data mempengaruhi akurasi, presisi dan juga recall yang didapatkan. Semakin banyak jumlah datanya maka akan semakin baik nilai akurasi, presisi dan recall yang didapatkan. Pada percobaan tersebut nilai akurasi, presisi dan recall yang paling baik adalah percobaan dengan jumlah data yang paling banyak yaitu 1700 data.

Kemudian pada penelitian ini juga dilakukan beberapa percobaan terhadap jumlah data yang sama tetapi dengan perbandingan jumlah data setiap kelas yang berbeda. Dilakukan percobaan terhadap 500 data dengan jumlah data untuk setiap kelasnya masing-masing terdiri dari 4:68:212:168:37:11 menghasilkan akurasi 57,8% presisi 65,87% dan recall sebesar 57,8%. Kemudian dipercobaan selanjutnya dengan jumlah data yang sama tetapi jumlah data setiap kelas yang berbeda yaitu digunakan jumlah data setiap kelas masing-masing terdiri dari 4:40:296:130:20:10 menghasilkan nilai akurasi 65% presisi 72,56% dan recall 65%. Kemudian dipercobaan selanjutnya dengan jumlah data yang sama tetapi jumlah data setiap kelas yang berbeda yaitu digunakan jumlah data setiap kelas masing-masing terdiri dari 4:32:340:100:16:8 menghasilkan nilai akurasi 66,4% presisi 72,18% dan recall 66,4%. Adapun perbandingan dari setiap percobaan ditunjukkan pada tabel 6 sebagai berikut :

**Tabel 6.** Perbandingan Hasil Kinerja Klasifikasi dengan Jumlah Data Yang Sama

Jumlah Data Setiap Kelas	Perbandingan Jumlah Data Setiap Kelas	Akurasi	Presisi	Recall
4:68:212:168:37:11	1:17:53:42:9,25:2,75	57.8%	65.87%	57.8%
4:40:296:130:20:10	1:10:74:32,5:5:2,5	65%	72.56%	65%
4:32:340:100:16:8	1:8:85:25:4:2	66.4%	72.18%	66.4%

Pada tabel 6 tersebut menunjukkan bahwa berdasarkan percobaan yang dilakukan didapatkan bahwa jumlah data antara setiap kelas target berpengaruh pada nilai akurasi, presisi dan juga recall yang didapatkan. Dengan jumlah data yang sama, tapi dengan jumlah data setiap kelas target yang berbeda maka akan menghasilkan nilai akurasi, presisi, dan recall yang berbeda. Semakin seimbang jumlah data antara setiap kelas target maka nilai akurasi, presisi dan recall yang didapatkan akan semakin baik. Pada percobaan dengan jumlah data yang sama yaitu 500 data, didapatkan bahwa nilai akurasi, presisi, dan recall yang paling baik adalah percobaan dengan jumlah data antara setiap kelasnya adalah 4:32:340:100:16:8.

#### 4. KESIMPULAN

Hasil dari penelitian yang dilakukan memperlihatkan bahwa klasifikasi status gizi balita dengan menggunakan algoritma Naïve Bayes Classifier dan pengujian K-Fold Cross Validation bisa mengklasifikasikan status gizi balita secara tepat. Kemudian pada pengujian dengan 10-Fold Cross Validation didapatkan bahwa nilai akurasi yang paling tinggi yaitu sebesar 82,94% pada iterasi ke-5, sedangkan nilai akurasi paling rendah yaitu sebesar 65,88% pada iterasi ke-6. Dengan rata-rata nilai akurasi secara keseluruhan yaitu sebesar 75,47%. Sedangkan untuk rata-rata nilai presisi yang didapatkan yaitu sebesar 81,36% dan rata-rata nilai recall sebesar 75,47%. Dari percobaan yang dilakukan diketahui juga bahwa kinerja klasifikasi dipengaruhi oleh beberapa faktor lain diantaranya yaitu jumlah data yang digunakan serta keseimbangan data tersebut. Pada faktor jumlah data yang digunakan, diketahui bahwa semakin banyak jumlah datanya maka akan semakin baik nilai akurasi, presisi dan recall yang didapatkan. Selain itu untuk faktor keseimbangan data, diketahui bahwa semakin seimbang jumlah data antara setiap kelas target maka akan semakin baik nilai akurasi, presisi dan recall yang didapatkan. Sehingga untuk penelitian berikutnya diharapkan dapat menggunakan algoritma yang berbeda serta dengan jumlah data yang lebih banyak dan data yang digunakan juga lebih seimbang.

#### REFERENCES

- [1] W. Hadikristanto and T. D. Pungkas, "Klasifikasi Status Gizi Orang Dewasa Menggunakan Algoritma Naive Bayes (Studi Kasus Klinik Bhakti Mulia Cikarang)," *SIGMA - Jurnal Teknologi Pelita Bangsa*, vol. 9, 2019.
- [2] M. Platinum, "Pentingnya Memperhatikan Tercukupinya Kebutuhan Gizi pada Anak," <https://morinagaplatinum.com/id/milestone/pentingnya-memperhatikan-tercukupinya-kebutuhan-gizi-pada-ank>, Aug. 24, 2021.
- [3] A. M. Safitri, "6 Masalah Gizi yang Paling Sering Terjadi di Indonesia, dari Balita Hingga Dewasa," <https://helohehat.com/nutrisi/fakta-gizi/masalah-gizi-di-indonesia/>, 2023.
- [4] H. Yulian, "Inilah 5 Masalah Gizi yang Rentan Dialami Anak Indonesia," <https://www.momsmoney.id/news/inilah-5-masalah-gizi-yang-rentan-dialami-anak-indonesia>, 2022.
- [5] N. Azizah, "Masalah Gizi Anak di Indonesia Masih Tinggi," <https://www.republika.co.id/berita/qye6fm463/masalah-gizi-anak-di-indonesia-masih-tinggi>, 2021.
- [6] W. Mutika and D. Syamsul, "Analisis Permasalahan Status Gizi Kurang pada Balita di Puskesmas Teupah Selatan Kabupaten Simeuleu," *Jurnal Kesehatan Global*, vol. 1, no. 3, pp. 127–136, 2018.

- [7] D. P. Utomo and M. Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 4, no. 2, pp. 437–444, Apr. 2020, doi: 10.30865/mib.v4i2.2080.
- [8] A. Tarigan, Mustakim, E. Wahyudi, and J. Adhiva, "Klasifikasi Status Kesejahteraan Rumah Tangga di Kabupaten Siak Menggunakan Algoritma Naive Bayes Classifier," *Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI)*, 2019.
- [9] E. Darnila, M. Maryana, and M. Azmi, "Aplikasi Klasifikasi Status Gizi Balita Menggunakan Metode Naive Bayes Berbasis Android," *METHOMIKA: Jurnal Manajemen Informatika & Komputerisasi Akuntansi*, vol. 5, no. 2, 2021, doi: 10.46880/jmika.Vol5No2.pp135-141.
- [10] M. A. Bianto, Kusri, and Sudarmawan, "Perancangan Sistem Klasifikasi Penyakit Jantung Menggunakan Naive Bayes," *Citec Journal*, vol. 6, no. 1, 2019.
- [11] N. Y. Paramitha, A. Nuryaman, A. Faisal, E. Setiawan, and D. E. Nurvazly, "Klasifikasi Penyakit Stroke Menggunakan Metode Naive Bayes," *Jurnal Siger Matematika*, vol. 04, no. 01, 2023, [Online]. Available: <https://www.kaggle.com/datasets/zzettrkalkpakbal/full-filled->
- [12] N. D. Lika, "PENERAPAN ALGORITMA NBC UNTUK KLASIFIKASI TINGKAT RESIKO PENYAKIT DIABETES MELLITUS," *Jurnal Perencanaan, Sains, Teknologi, dan Komputer (JuPerSaTeK)*, vol. 3, no. 1, pp. 45–54, 2020.
- [13] M. Y. Haffandi, E. Haerani, F. Syafria, and L. Oktavia, "KLASIFIKASI PENYAKIT PARU-PARU DENGAN MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER," *Jurnal Teknik Informasi dan Komputer (Teknikom)*, vol. 5, no. 2, pp. 176–186, Dec. 2022, doi: 10.37600/tekikom.v5i2.649.
- [14] T. Arifin and D. Ariesta, "PREDIKSI PENYAKIT GINJAL KRONIS MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER BERBASIS PARTICLE SWARM OPTIMIZATION," *Jurnal Tekno Insentif*, vol. 13, no. 1, pp. 26–30, Apr. 2019, doi: 10.36787/jti.v13i1.97.
- [15] A. Hutapea, M. T. Furqon, and I. Indriati, "Penerapan Algoritme Modified K-Nearest Neighbour Pada Pengklasifikasian Penyakit Kejiwaan Skizofrenia," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 10, pp. 3957–3961, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [16] R. R. R. Arisandi, B. Warsito, and A. R. Hakim, "Aplikasi Naive Bayes Classifier (NBC) pada Klasifikasi Status Gizi Balita Stunting dengan Pengujian K-Fold Cross Validation," *JURNAL GAUSSIAN*, vol. 11, no. 1, pp. 130–139, 2022, [Online]. Available: <https://ejournal3.undip.ac.id/index.php/gaussian/>
- [17] E. Yanti, E. Apriyeni, D. C. Rahayuningrum, Ibrahim, and D. Ayu K, "STATUS GIZI BAYI (6-12 bulan) DITINJAU DARI BERAT BADAN LAHIR DI POSYANDU BOUGENVILE I WILAYAH KERJA PUSKESMAS ANDALAS," *Jurnal Kesehatan Medika Sainika*, vol. 13, no. 1, 2022, doi: 10.30633/jkms.v13i1.1388.
- [18] S. Sutrio and M. Lupiana, "Berat Badan dan Panjang Badan Lahir Meningkatkan Kejadian Stunting," *Jurnal Kesehatan Metro Sai Wawai*, vol. 12, no. 1, pp. 21–29, 2019.
- [19] Sukmawati, Hendrayati, Chaerunnimah, and Nurhumaira, "STATUS GIZI IBU SAAT HAMIL, BERAT BADAN LAHIR BAYI DENGAN STUNTING PADA BALITA," *Media Gizi Pangan*, vol. 25, 2018.
- [20] L. Abdullah, R. Tamin, and A. A. Qashlim, "KLASIFIKASI PENERIMAAN BEASISWA MENGGUNAKAN ALGORITMA NAIVE BAYES DI UNIVERSITAS AL ASYARIAH MANDAR KABUPATEN POLEWALI MANDAR," *Journal Pegguruang: Conference Series*, vol. 3, no. 1, p. 183, May 2021, doi: 10.35329/jp.v3i1.1399.
- [21] PERATURAN MENTERI KESEHATAN REPUBLIK INDONESIA NOMOR 2 TAHUN 2020 TENTANG STANDAR ANTROPOMETRI ANAK. 2020.
- [22] N. Hajar, N. Y. Setiawan, and F. A. Bachtiar, "Pengelompokan Mahasiswa untuk Pengajuan Bantuan Uang Kuliah Tunggal menggunakan Metode K-Means Clustering (Studi Kasus BEM FILKOM UB)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 6, no. 5, pp. 2353–2361, 2022, [Online]. Available: <http://j-ptiik.ub.ac.id>