

# Penerapan Algoritma Naive Bayes dan Forward Selection dalam Pengklasifikasian Status Gizi Stunting pada Puskesmas Pandanaran Semarang

**Junta Zeniarja<sup>1</sup>, Kiki Widia<sup>2</sup>, Ramadhan Rakhmat Sani<sup>3</sup>**

Fakultas Ilmu Komputer, Universitas Dian Nuswantoro Semarang

Jl.Imam Bonjol No.207 Semarang, telp. (024) 3569196

E-mail: <sup>1</sup>junta@dsn.dinus.ac.id, <sup>2</sup>kikiwidia47@gmail.com, <sup>3</sup>ramadhan\_rs@dsn.dinus.ac.id

Diterima: 12 Agustus 2019; Direvisi: 24 Februari 2020; Disetujui: 20 April 2020

## **Abstrak**

Masalah stunting pada balita tidak dapat diremehkan begitu saja, karena dapat berdampak pada kemampuan berbahasa, kognitif, motorik, dan berisiko juga pada kecacatan, terserang penyakit infeksi, hingga kematian. Meningkatnya kasus stunting pada balita ini memerlukan suatu upaya dalam penanganan dan pencegahan secara dini. Untuk memperoleh informasi tersebut diperlukan metode data mining dengan menerapkan Naive Bayes dan penggunaan fitur Forward Selection. Untuk mendapatkan hasil keputusan dari klasifikasi status gizi stunting digunakanlah algoritma NBC, sedangkan untuk meningkatkan nilai akurasi menggunakan forward selection dengan melakukan seleksi fitur yaitu menghapus sebagian atribut yang tidak sesuai di dalam tahapan klasifikasinya. Hasil akurasi klasifikasi status gizi stunting pada balita dengan algoritma NBC saja pada penelitian ini sebesar 83,33%, sedangkan untuk algoritma NBC dengan fitur Forward Selection mencapai 86,00%. Peningkatan hasil akurasi tampak baik ketika dilakukan penggabungan algoritma NBC dengan fitur Forward Selection.

**Kata kunci:** status gizi, stunting, naive bayes, fitur forward selection

## **Abstract**

The problem of stunting in children under five cannot underestimated, because it can have an impact on language, cognitive, motoric, and risk of disability, infection, and even death. The increasing cases of stunting in toddlers requires an effort in early treatment and prevention. To obtain this information a data mining method is require by applying Naive Bayes and using the Forward Selection feature. To get the results of the decision of the classification of nutritional status of stunting, NBC algorithm is used, while to increase its accuracy value using forward selection by selecting features, which is removing some of the attributes that are not appropriate in the classification stage. The accuracy of stunting nutritional status classification for toddlers with NBC algorithm only in this study amounted to 83.33%, while for the NBC algorithm with the Forward Selection feature reached 86.00%. Improved accuracy results look good when combining NBC algorithms with the Forward Selection feature.

**Keywords:** nutritional status, stunting, naive bayes, forward selection feature

## **1. PENDAHULUAN**

Pada periode 2015-2019 dalam program pembangunan sektor kesehatan diutamakan untuk empat poin yaitu untuk pengurangan angka kematian ibu dan anak, menurunnya prevalensi

balita pendek (*stunting*), penanggulangan penyakit menular dan penanggulangan penyakit tidak menular. Tujuan dari pembangunan nasional yang terkandung dalam inti Rencana Pembangunan jangka Menengah Tahun 2015-2019 sebagai upaya untuk meningkatkan status gizi masyarakat dan juga untuk penurunan prevalensi *stunting* atau balita pendek [1].

Indonesia memiliki kontribusi 90% dari masalah gizi di antara 36 negara di dunia. Pada tahun 2010 hasil Riset Kesehatan Dasar (RISKESDAS) menunjukkan prevalensi *stunting* pada balita di Indonesia sebesar 35,6% yang terdiri atas kriteria pendek dan sangat pendek sekitar 17,1% dan 18,5%. Prevalensi persentase tersebut lebih tinggi dari persentase prevalensi kekurangan gizi sekitar 17,9%, gizi buruk 13,3% dan gizi lebih 14%. Adapun provinsi Jawa Tengah sendiri memiliki persentase prevalensi 33,9% dari *stunting* yang meliputi kriteria sangat pendek sekitar 16,9% dan kriteria pendek 17% [2]. Sedangkan indikator Pemantauan Status Gizi berdasarkan PB/U atau TB/U untuk kasus *stunting* pada Dinas Kesehatan Kota Semarang sekitar 20,37% pada daerah kota Semarang [3].

Malnutrisi yang kronis selama pertumbuhan dan perkembangan awal dalam hidup digambarkan oleh *stunting* dari z-skor tinggi badan untuk usia (TB/U) kurang dari -2 standar deviasi (SD) berdasarkan standar pertumbuhan WHO. Dalam *stunting* umum mempengaruhi 1 dari 4 balita [4]. Ada efek jangka pendek yang disebabkan oleh *stunting* yaitu penghambatan kemampuan untuk berbicara, perkembangan motorik dan kognitif, risiko kecacatan, penyakit infeksi, dan kematian. Ada juga dampak jangka panjang yang mengakibatkan risiko penyakit degeneratif, seperti penyakit jantung coroner, stroke, hipertensi dan diabetes mellitus. Selanjutnya bisa mengurangi produktivitas kerja di masa dewasa [5].

Metode klasifikasi merupakan metode yang digunakan pada penelitian ini dalam pengolahan data dimana metode ini sering digunakan dalam ilmu data mining untuk menentukan suatu rekaman data baru ke salah satu dari beberapa jenis yang sudah di definisikan sebelumnya [6]. Salah satu dari algoritma klasifikasi yaitu algoritma klasifikasi Naive Bayes. Algoritma klasifikasi *Naive Bayes* merupakan algoritma pengkategorian statistik yang bisa dipakai untuk memperkirakan probabilitas keanggotaan suatu target (label/output). Algoritma klasifikasi *Naive Bayes* mempunyai keakuratan dan komputasi yang sangat cepat ketika diaplikasikan ke dalam basis data yang jumlahnya sangat banyak. Algoritma *Naive Bayes* mampu meminimalkan tingkat kesalahan dibandingkan dengan semua algoritma klasifikasi lainnya [7]. Berdasarkan [8] data hasil validasi kinerja dari beberapa algoritma klasifikasi antara lain : *Decision TreeJ48*, *Zero-R*, *Lazy-IBK*, dan *Naive Bayes*; algoritma *Naive Bayes* memiliki kinerja yang paling baik dalam hal akurasi. Menunjukkan bahwa nilai dari *Correctly Classified instances* pada data set *ecoli* mencapai nilai prosentase tertinggi sebesar 85,12% dengan menggunakan model tes *cross-validation*. Begitu juga pada model tes *percentage-split*, algoritma *Naive Bayes* mencapai nilai prosentase tertinggi sebesar 82,46%. Algoritma *Naive Bayes* juga mempunyai kinerja yang lebih baik daripada algoritma *K-NN*. Algoritma *Naive Bayes* dipilih karena bisa meningkatkan akurasi hingga mencapai optimal dengan data training yang sedikit. Hasil akurasi *Naive Bayes* lebih unggul yaitu sebesar 70% sedangkan *K-NN* memiliki nilai akurasi hanya 40% [9].

Pada penelitian ini digunakan salah satu jenis dari metode seleksi fitur yaitu *Forward Selection* yang bertujuan untuk mengoptimalkan nilai akurasi dari algoritma klasifikasi *Naive Bayes* dengan cara menghapus atribut yang tidak sesuai. Riset ini dilaksanakan dengan tujuan untuk mengoptimalkan kinerja algoritma klasifikasi *Naive Bayes* itu sendiri pada proses klasifikasi status gizi *stunting* balita dengan menambahkan metode seleksi fitur yaitu *Forward Selection*.

## 2. METODE PENELITIAN

### 2.1 Algoritma Klasifikasi Naive Bayes

Naive Bayes merupakan salah satu bagian algoritma pengaplikasian dari teorema Bayes dengan memperkirakan probabilitas masa mendatang menurut data historis sebelumnya atau

pengalaman yang sudah ada. Hubungan antara Naive Bayes dengan klasifikasi, korelasi hipotesis dan bukti klasifikasi menunjukkan bahwa hipotesis dalam teorema Bayes menggambarkan label kelas yang menjadi sasaran pemetaan dalam suatu proses klasifikasi. Sedangkan bukti sendiri merupakan atribut-atribut yang menjadi masukan dalam sebuah proses klasifikasi. Apabila X merupakan nilai masukan dari fitur, dan Y adalah sebuah label class, Naive Bayes dituliskan dengan  $P(X|Y)$  yang artinya probabilitas label Y didapatkan setelah fitur-fitur X diamati.  $P(X|Y)$  juga disebut probabilitas akhir untuk Y, sedangkan  $P(Y)$  disebut probabilitas awal [10].

Pada saat proses pembelajaran berlangsung, harus dilaksanakan pelatihan probabilitas akhir  $P(Y|X)$  terhadap pola data untuk setiap kombinasi X dan Y berdasarkan nilai yang didapat dari data training. Dengan membentuk pola tersebut suatu data uji X' dapat dikategorikan dengan mencari nilai Y' dengan memaksimalkan nilai  $P(X'|Y')$  yang diperoleh.

Rumus Naive Bayes untuk klasifikasi seperti pada persamaan (1) berikut :

$$P(X|Y) = \frac{P(Y)\pi_{i=1}^q P(X_i|Y)}{P(X)} \tag{1}$$

Keterangan:

$P(Y|X)$  : probabilitas data vektor X pada class Y

$P(Y)$  : probabilitas awal class Y

$\prod_{i=1}^q P(X_i|Y)$  : probabilitas independen dari class Y dari seluruh fitur vektor X

$P(X)$  : nilai tetap yang digunakan dalam perhitungan prediksi yang nantinya menjadi hasil prediksi terbesar pada class yang akan dipilih sebagai hasil prediksi

Sedangkan probabilitas Independen  $\prod_{i=1}^q P(X_i|Y)$  ini merupakan pengaruh semua fitur dari data terhadap class Y, yang diformulasikan pada persamaan (2) berikut:

$$P(X|Y = y) = \pi_{i=1}^q P(X_i|Y = y) \tag{2}$$

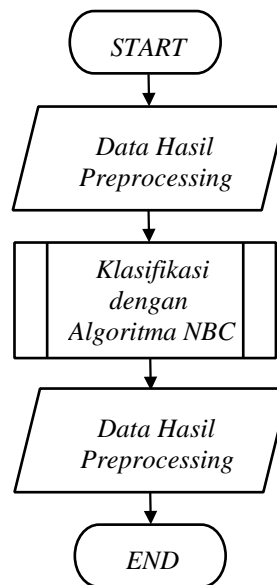
Setiap set fitur  $X = [X_1, X_2, X_3, X_4, \dots, X_q]$  terdiri atas q atribut.

Formula yang dijelaskan diatas merupakan perhitungan prediksi dengan untuk fitur dengan tipe kategori. Namun untuk atribut bertipe angka memiliki *special treatment* sebelum dimasukkan ke dalam Naive Bayes dan berikut adalah caranya:

1. Menukar nilai fitur kontinu dan melakukan diskretisasi pada setiap atribut tersebut dengan nilai interval diskret dimana dalam pendekatan ini melakukan transformasi atribut kontinu ke dalam atribut ordinal.
2. Mengasumsikan bentuk tertentu dari disribusi probabilitas untuk atribut kontinu dan memperkirakan parameter distribusi dengan data training.

Kinerja algoritma Naive bayes digambarkan pada gamabr 1. Algoritma ini beroperasi atas dasar teori probabilitas yang melihat seluruh atribut dari data sebagai bukti dalam probabilitas. Hal tersebut memberikan karakteristik pada algortima Naive Bayes yaitu sebagai berikut:

- a. Algoritma Naive Bayes yang solid (*robust*) terhadap data yang terisolasi yang biasanya merupakan data dengan karakteristik berbeda (*outlier*) dan juga dapat menangani atribut yang salah dengan mengabaikan data training selama proses pembentukan pola atau prediksi yang dilakukan.
- b. Solid dalam menangani atribut yang tak sesuai.
- c. Atribut yang berkolerasi dapat menurunkan kinerja algoritma klasifikasi Naive Bayes yang disebabkan oleh asumsi independensi atribut telah hilang.



Gambar 1. Kinerja algoritma *klasifikasi Naive Bayes*

## 2.2 Forward Selection

*Feature selection* biasa disebut dengan seleksi fitur adalah bagian dari proses penting yang digunakan pada *data preprocessing* di dalam data mining untuk mengoptimalkan kinerja serta mempercepat proses dari sebuah algoritma. Teknik tersebut dapat memilih subset dari fitur dengan jumlah yang cukup besar yaitu mengarah pada pengurangan atau penghilangan fitur-fitur yang kurang berpengaruh atau tidak relevan untuk klasifikasi [11].

Seleksi fitur juga dikenal dengan nama *feature selection*, *subset selection*, *attribute selection* atau *variable selection* dapat pula diartikan sebagai proses pemilihan atribut yang sesuai untuk dipakai pada proses klasifikasi atau clustering yang bertujuan untuk menghilangkan tingkat kompleksitas dari suatu algoritma klasifikasi, mengoptimalkan akurasi dari algoritma klasifikasi tersebut, dan dapat mengetahui atribut-atribut yang paling berpengaruh terhadap tingkat akurasi. Sedangkan model *Forward Selection* adalah pemodelan yang dimulai dari nol peubah (*empty model*), kemudian satu persatu peubah dimasukan sampai kriteria tertentu terpenuhi seperti proses pada gambar 2. Berikut ini adalah bagaimana cara model *Forward Selection* dilakukan [12] :

- Membangun model dengan meregresikan variabel respon Y dengan setiap variabel prediktor. Selanjutnya dipilih model yang mempunyai nilai R<sup>2</sup> tertinggi. Misal model tersebut adalah yang membuat prediktor X<sub>a</sub>, yaitu pada persamaan (3):

$$Y^{\wedge} = b_0 + b_a X_a. \quad (3)$$

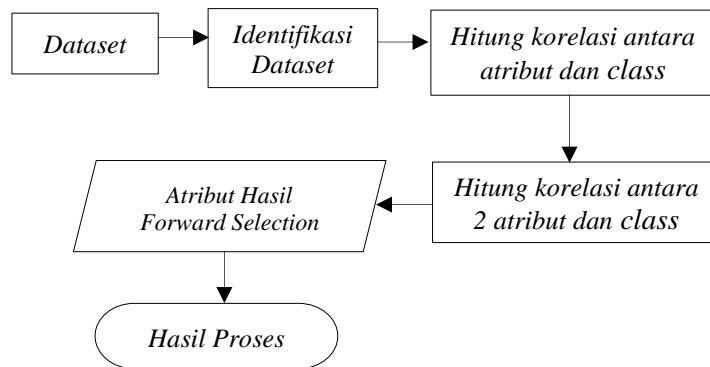
- Meregresikan variabel respon Y, dengan prediktor X<sub>a</sub>, ditambah dengan setiap prediktor selain X<sub>a</sub> dan prediktor lain. Kemudian dipilih model yang memiliki nilai R<sup>2</sup> paling tinggi, misal mengandung tambahan prediktor X<sub>b</sub>, yaitu model pada persamaan (4):

$$Y^{\wedge} = b_0 + b_a X_a + b_b X_b \quad (4)$$

- Prediktor terpilih X<sub>b</sub> berarti mempunyai F<sub>sequensial</sub> tertinggi. Formula F<sub>sequensial</sub> untuk X<sub>b</sub> adalah sebagai berikut pada persamaan (5):

$$F_{seq} = R(\beta_b | \beta_0, \beta_a) / MSE / db \quad (5)$$

- Nilai F<sub>sequensial</sub> untuk X<sub>b</sub> juga dapat diperoleh dengan cara mengkuadratkan nilai statistik uji T prediktor X<sub>b</sub>.
- Urutkan jarak tersebut dan tentukan tetangga mana yang terdekat berdasarkan jarak minimum ke-k.
- Menentukan kategori dari tetangga terdekat.
- Gunakan kategori mayoritas dari tetangga terdekat tersebut sebagai nilai prediksi data yang baru.



Gambar 2. Kinerja model *Forward Selection*

### 2.3 Confusion Matrix

*Confusion Matrix* merupakan sebuah metode yang sering diterapkan untuk mengetahui performa dari sebuah algoritma. Metode ini berbentuk diagram yang berisikan informasi aktual mengenai perkiraan klasifikasi dari sebuah sistem dengan memakai data dalam bentuk matriks sebagai hasilnya [13]. Tabel 1 dari confusion matrix digambarkan sebagai berikut :

Tabel 1. Diagram *Confusion Matrix* penelitian yang dilakukan

Classification (class)	Predicted Class	
	Yes	No
Yes	a (True Positive - TP)	b (False Negative - FN)
No	c (False Positive - FP)	d (True Negative - TN)

## 3. HASIL DAN PEMBAHASAN

Hasil dari penelitian ini berupa tingkat akurasi yang didapatkan dari pengujian yang telah dilakukan, dengan tujuan yaitu untuk menguji keakuratan dan kinerja dari algoritma klasifikasi *Naive Bayes* berbasis seleksi fitur *Forward Selection* dalam melakukan klasifikasi Status Gizi Stunting Balita. Metode *Forward Selection* digunakan dengan tujuan untuk menghilangkan atribut-atribut yang tidak berpengaruh dalam proses klasifikasi. Untuk menyeleksi atribut dapat dilakukan dengan cara mencari hubungan korelasi yang paling dekat dengan target sehingga dapat meningkatkan hasil akurasi dari klasifikasi menggunakan algoritma klasifikasi *Naive Bayes*.

### 3.1 Algoritma Klasifikasi Naive Bayes

Algoritma klasifikasi *Naive Bayes* dalam penelitian ini digunakan sebagai metode dalam melakukan klasifikasi Status Gizi Stunting Pada Balita. Dataset yang digunakan merupakan dataset valid yang berjumlah 300 *record* dengan jumlah atribut awal sebanyak 6 atribut dan 1 label. Untuk mengetahui kinerja algoritma klasifikasi *Naive Bayes* dalam melakukan klasifikasi Status Gizi Stunting Pada Balita maka dilakukan pengujian dengan menggunakan *confusion matrix* sehingga diperoleh hasil perhitungan seperti tabel di bawah ini:

Tabel 2. Diagram *Confusion Matrix* menggunakan Algoritma klasifikasi *Naive Bayes*

	<i>True Stunting</i>	<i>True Normal</i>
<i>Stunting Prediction</i>	249	41
<i>Normal Prediction</i>	9	1

Berdasarkan tabel 2 di atas hasil dari tabel confusion matrix dapat dilakukan perhitungan nilai akurasi pada persamaan (6) berikut:

$$\begin{aligned}
 \text{Akurasi} &= \frac{249+1}{249+1+9+41} * 100\% \\
 &= 0.833333333 * 100\% \\
 &= 83.33\%
 \end{aligned}
 \tag{6}$$

### 3.2 Algoritma klasifikasi *Naive Bayes* berbasis *Forward Selection*

Terdapat 300 *record* data yang terdiri dari 6 atribut dan 1 label yang akan digunakan untuk melakukan perhitungan *Forward Selection* dengan menghitung hubungan korelasi yang dihasilkan dari setiap atribut. Kemudian dilakukan inisialisasi dari ke 6 atribut tersebut yaitu atribut pertama menjadi X1 sampai dengan atribut terakhir menjadi X6.

Dari semua atribut mulai X1 sampai X6 dilakukan perhitungan korelasinya dengan variabel target (*label/class*) yang diinisialisasikan dengan Y, dimana nilai dari label atau class itu sendiri diasumsikan yaitu NORMAL = 1 dan STUNTING = 2. Hubungan dari koefisien korelasi pada regresi dihitung dengan rumus korelasi regresi sederhana seperti persamaan (7) berikut ini:

$$r = \frac{n\sum X_i Y_i - (\sum X_i)(\sum Y_i)}{\sqrt{\{n\sum X_i^2 - (\sum X_i)^2\} \{n\sum Y_i^2 - (\sum Y_i)^2\}}}
 \tag{7}$$

Perhitungan menggunakan rumus korelasi regresi sederhana di terapkan pada atribut X1 sampai dengan X6, tujuannya untuk mencari nilai R tertinggi.

Setelah memperoleh nilai koefisien korelasi yang pertama, selanjutnya dilakukan perhitungan koefisien korelasi antara atribut terpilih, Y dan atribut-atribut lain selain atribut terpilih itu sendiri menggunakan regresi linear berganda dengan rumus seperti pada persamaan (8) berikut:

$$r_{y.x1.x2} = \sqrt{\frac{r_{yx1}^2 + r_{yx2}^2 - 2r_{yx1}r_{yx2}r_{x1x2}}{1 - r_{x1x2}^2}}
 \tag{8}$$

Kemudian perhitungan dilanjutkan menggunakan rumus korelasi regresi berganda hingga didapatkan atribut-atribut yang mempunyai hubungan korelasi yang kuat dan berpengaruh. Dari atribut atribut yang telah diseleksi akan dilakukan klasifikasi menggunakan algoritma klasifikasi *Naive Bayes* untuk menentukan klasifikasi Status Gizi Stunting Pada Balita dengan hasil keputusan berupa probabilitas dari klasifikasi itu sendiri.

Tabel 3. Hasil *Confusion Matrix* menggunakan Algoritma *NBC* berbasis *Forward Selection*

	<i>True Stunting</i>	<i>True Normal</i>
<i>Stunting Prediction</i>	258	42
<i>Normal Prediction</i>	0	0

Berdasarkan tabel 3 di atas hasil dari tabel *confusion matrix* dapat dilakukan perhitungan nilai akurasi sebagai berikut:

$$\begin{aligned}
 \text{Akurasi} &= \frac{258+0}{258+0+0+42} * 100\% \\
 &= 0.86 * 100\% \\
 &= 86\%
 \end{aligned}
 \tag{9}$$

### 3.3 Evaluasi dan Validasi

Tahap validasi dan evaluasi ini bertujuan untuk menguji algoritma yang telah diterapkan dengan menggunakan metode *k-fold cross validation* yaitu dengan membentuk k subset dari dataset yang ada. Dalam penelitian ini dilakukan pengujian dengan nilai k sebanyak 10 *fold*, dimana tujuannya adalah untuk mengetahui performa akurasi dari algoritma yang digunakan yaitu algoritma klasifikasi *Naive Bayes* berbasis *Forward Selection* yang diterapkan pada klasifikasi Status Gizi Balita apabila diuji dengan menggunakan data *training* dan *testing* yang berbeda. Pengujian menggunakan metode *K-Fold Cross Validation* dengan nilai K adalah 10 *fold* merupakan pemilihan *fold* terbaik dalam uji validitas. Metode *10-Fold Cross Validation* akan bekerja dengan melakukan pengujian yang diulang sebanyak *K-fold*, dalam kasus ini *K-fold* bernilai 10. Hasil pengukuran dari pengujian yang diulang 10 kali ini yaitu berupa nilai akurasi rata-rata dari 10 kali pengujian.

Tahapan awal dalam metode validasi *10-fold cross validation* yaitu membagi dataset, dimana dalam penelitian ini jumlah data 300 *record*, kemudian data tersebut dibagi menjadi 10 subset data (bagian). Maka setiap 1 subset masing- masing berisi 30 data untuk setiap iterasinya. *Fold* pertama berisikan data kombinasi dari 9 subset berbeda yang telah digabungkan dan berperan sebagai data *training*. Sedangkan 1 subset tersisa digunakan untuk data *testing*, kemudian proses *training* dan *testing* data dilakukan berulang hingga *fold ke-10*. Gambaran pembagian data *training* dan data *testing* dengan menggunakan metode *10-fold cross validation* adalah sebagai berikut:

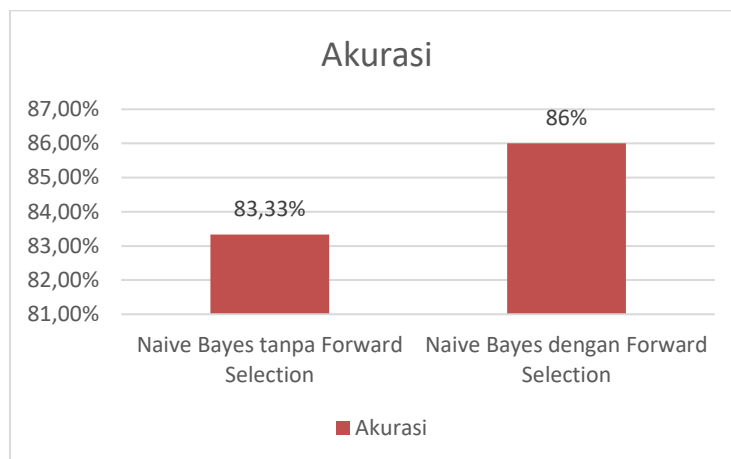
Tabel 4. Pembagian Data *Training* dan *Testing*

Fold	Data Training		Data Testing		Akurasi
	Subset	Data	Subset	Data	
1	S <sub>2</sub> ,S <sub>3</sub> ,S <sub>4</sub> ,S <sub>5</sub> ,S <sub>6</sub> ,S <sub>7</sub> ,S <sub>8</sub> ,S <sub>9</sub> ,S <sub>10</sub>	270	S <sub>1</sub>	30	84,44%
2	S <sub>1</sub> ,S <sub>3</sub> ,S <sub>4</sub> ,S <sub>5</sub> ,S <sub>6</sub> ,S <sub>7</sub> ,S <sub>8</sub> ,S <sub>9</sub> ,S <sub>10</sub>	270	S <sub>2</sub>	30	84,44%
3	S <sub>1</sub> ,S <sub>2</sub> ,S <sub>4</sub> ,S <sub>5</sub> ,S <sub>6</sub> ,S <sub>7</sub> ,S <sub>8</sub> ,S <sub>9</sub> ,S <sub>10</sub>	270	S <sub>3</sub>	30	84,44%
4	S <sub>1</sub> ,S <sub>2</sub> ,S <sub>3</sub> ,S <sub>5</sub> ,S <sub>6</sub> ,S <sub>7</sub> ,S <sub>8</sub> ,S <sub>9</sub> ,S <sub>10</sub>	270	S <sub>4</sub>	30	87,41%
5	S <sub>1</sub> ,S <sub>2</sub> ,S <sub>3</sub> ,S <sub>4</sub> ,S <sub>6</sub> ,S <sub>7</sub> ,S <sub>8</sub> ,S <sub>9</sub> ,S <sub>10</sub>	270	S <sub>5</sub>	30	84,44%
6	S <sub>1</sub> ,S <sub>2</sub> ,S <sub>3</sub> ,S <sub>4</sub> ,S <sub>5</sub> ,S <sub>7</sub> ,S <sub>8</sub> ,S <sub>9</sub> ,S <sub>10</sub>	270	S <sub>6</sub>	30	84,44%
7	S <sub>1</sub> ,S <sub>2</sub> ,S <sub>3</sub> ,S <sub>4</sub> ,S <sub>5</sub> ,S <sub>6</sub> ,S <sub>8</sub> ,S <sub>9</sub> ,S <sub>10</sub>	270	S <sub>7</sub>	30	85,19%
8	S <sub>1</sub> ,S <sub>2</sub> ,S <sub>3</sub> ,S <sub>4</sub> ,S <sub>5</sub> ,S <sub>6</sub> ,S <sub>7</sub> ,S <sub>9</sub> ,S <sub>10</sub>	270	S <sub>8</sub>	30	84,44%
9	S <sub>1</sub> ,S <sub>2</sub> ,S <sub>3</sub> ,S <sub>4</sub> ,S <sub>5</sub> ,S <sub>6</sub> ,S <sub>7</sub> ,S <sub>8</sub> ,S <sub>10</sub>	270	S <sub>9</sub>	30	85,19%
10	S <sub>1</sub> ,S <sub>2</sub> ,S <sub>3</sub> ,S <sub>4</sub> ,S <sub>5</sub> ,S <sub>6</sub> ,S <sub>7</sub> ,S <sub>8</sub> ,S <sub>9</sub>	270	S <sub>10</sub>	30	95,56%
Rata-rata Nilai Akurasi					86,00%

Berdasarkan tabel 4 sebelumnya terlihat bahwa penggabungan algoritma klasifikasi *Naive Bayes* dan metode *Forward Selection* tergolong baik untuk digunakan dalam proses klasifikasi Status Gizi Stunting Balita karena menghasilkan nilai akurasi rata-rata sebesar 86,00% yang diuji menggunakan validasi *10-fold cross validation* pada 270 data *training* dan 30 data *testing*.

### 3.4 Analisis Pembahasan

Dari hasil yang diperoleh terlihat pada gambar 3, hasil pengujian dengan algoritma klasifikasi *Naive Bayes* saja, yang dilakukan pada seluruh data pada dataset berjumlah 300 *record* akurasinya adalah 83,33%, Sedangkan hasil akurasi dari percobaan menggunakan algoritma klasifikasi *Naive Bayes* dengan metode *Forward Selection* adalah 86,00%. Artinya pengujian menggunakan metode *Forward Selection* dapat meningkatkan hasil akurasi.



Gambar 3. Chart Perbandingan Hasil Akurasi

## 4. KESIMPULAN

Pada penelitian ini didapatkan beberapa kesimpulan berdasarkan hasil penggunaan algoritma klasifikasi *Naive Bayes* dengan metode *Forward Selection* pada klasifikasi status gizi stunting balita. Penggunaan metode *forward selection* sebagai seleksi fitur terbukti meningkatkan nilai akurasi dari algoritma *NBC*, dimana terjadi pengurangan atribut dari yang berjumlah 6 menjadi 4 atribut. Hasil pengujian algoritma *NBC* tanpa menggunakan seleksi fitur adalah sebesar 83,33%, sedangkan untuk kinerja algoritma *NBC* dengan metode *forward selection* mengalami peningkatan sebesar 2,67% menjadi 86,00%. Telihat bahwa kinerja algoritma *NBC* dengan metode *forward selection* pada kasus klasifikasi status gizi stunting balita lebih unggul dibandingkan hanya menerapkan algoritma *NBC* saja.

## 5. SARAN

Ada beberapa saran yang perlu diperhatikan untuk dapat melakukan penelitian lebih lanjut agar mendapatkan hasil yang lebih baik lagi. Diantaranya perlu dilakukan uji perbandingan hasil akurasi menggunakan algoritma klasifikasi selain algoritma klasifikasi *Naive Bayes* maupun metode seleksi fitur selain *forward selection* yang mungkin dapat menghasilkan nilai akurasi yang lebih baik untuk penelitian selanjutnya. Perlu adanya pengembangan atau penambahan terhadap *dataset* agar klasifikasi status gizi stunting pada balita agar menghasilkan tingkat akurasi yang lebih tinggi.

## DAFTAR PUSTAKA

- [1] R. Wahyudi and Sufriani, "Pertumbuhan Dan Perkembangan Balita Stunting," *JIM FKep*, vol. IV, no. 1, pp. 56–62, 2018.



- 
- [2] E. Kusumawati, S. Rahardjo, and H. P. Sari, "Model Pengendalian Faktor Risiko Stunting pada Anak Bawah Tiga Tahun," *Kesmas Natl. Public Heal. J.*, vol. 9, no. 3, p. 249, 2015.
- [3] A. C. Mustikaningrum, H. W. Subagio, and A. Margawati, "Determinan kejadian stunting pada bayi usia 6 bulan di Kota Semarang," *J. GIZI Indones.*, vol. 4, no. 2, p. 82, 2017.
- [4] S. Hendro, A. Patria, P. Gastritis, D. Inggris, H. Riyanto, and D. Indonesia, "Faktor yang Berhubungan dengan Kejadian Gastritis," *Media Gizi Indones.*, vol. 9, no. 3, pp. 149–157, 2012.
- [5] B. R. Oktiva and M. Adriani, "Perbedaan Kadar Zinc Rambut pada Anak Stunting dan Non Stunting Usia 12-24 Bulan di Kelurahan Tambak Wedi Kenjeran, Surabaya," *Amerta Nutr.*, vol. 1, no. 2, p. 133, 2017.
- [6] A. B. da Silva, M. Arjmand, U. Sundararaj, and R. E. S. Bretas, "Novel composites of copper nanowire/PVDF with superior dielectric properties," *Polymer (Guildf.)*, vol. 55, no. 1, pp. 226–234, Jan. 2014.
- [7] M. F. Nugroho and S. Wibowo, "Fitur Seleksi Forward Selection Untuk Menentukan Atribut Yang Berpengaruh Pada Klasifikasi Kelulusan Mahasiswa Fakultas Ilmu Komputer UNAKI Semarang Menggunakan Algoritma Naive Bayes," *J. Inform. Upgris*, vol. 3, no. 1, pp. 63–70, 2017.
- [8] S. Fitri, "Perbandingan Kinerja Algoritma Klasifikasi Naive Bayesian, Lazy-Ibk, Zero-R, Dan Decision Tree- J48," *Data Manaj. dan Teknol. Inf.*, vol. 15, no. 1, p. 33, 2014.
- [9] R. N. Devita, H. W. Herwanto, and A. P. Wibawa, "Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa indonesia," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, p. 427, 2018.
- [10] K. A. E. T. Luthfi, "Algoritma Data Mining Yogyakarta," pp. 149–176, 2009.
- [11] S. Tabakhi, P. Moradi, and F. Akhlaghian, "An unsupervised feature selection algorithm based on ant colony optimization," *Eng. Appl. Artif. Intell.*, vol. 32, pp. 112–123, 2014.
- [12] W. Supriyanti, Kusrini, and A. Amborowati, "Perbandingan Kinerja Algoritma c4.5 Dan Naive Bayes Untuk Ketepatan Pemilihan Konsentrasi Mahasiswa," *J. Inf. Politek. Indonusa*, vol. 1, no. 3, pp. 61–67, 2016.
- [13] F. Gorinescu, *Data Mining. Concepts, Models and Techniques*. 2011.
-