

## OPTIMASI BACKWARD ELIMINATION PADA KLASIFIKASI PENYAKIT ISPA MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER

Anita Mila Oktafani<sup>1</sup>, Iut Tri Utami<sup>2</sup>, Puspita Kartikasari<sup>3</sup>

<sup>1,2,3</sup>Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro

\*e-mail : [anitaoktafani@gmail.com](mailto:anitaoktafani@gmail.com)

DOI: 10.14710/j.gauss.14.1.23-30

### Article Info:

Received: 2024-06-22

Accepted: 2025-02-12

Available Online: 2025-02-14

### Keywords:

ISPA; Data Mining; Naïve Bayes Classifier; Variable Selection; Backward Elimination

**Abstract:** ISPA cases in Indonesia are still a problem that threatens the health of various age groups, especially toddlers. The process of handling ISPA needs to be carried out quickly and precisely, one of the process is by classifying ISPA. This research classifies ISPA using the *Naive Bayes Classifier* algorithm with the addition of variable selection *Backward Elimination*. The *Naive Bayes Classifier* algorithm has the advantage of testing category-type data with a fast calculation process and high accuracy. The *Naive Bayes Classifier* test obtains an accuracy of 79,41%. The addition of *Backward Elimination* to the *Naive Bayes Classifier* algorithm aims to select irrelevant variables and is able to increase the previously accuracy of 7,84%, so that obtains an accuracy of 87,25%. This point out that the performance of the selection variable *Backward Elimination* is effective in optimizing the performance of the *Naive Bayes Classifier* in classifying ISPA.

## 1. PENDAHULUAN

ISPA merupakan infeksi pada saluran pernapasan yang disertai gejala seperti demam, batuk, pilek atau hidung tersumbat dan sakit tenggorokan (Kemenkes RI, 2018). ISPA disebabkan oleh 90% virus dan dapat menular melalui udara (Tandi *et al.*, 2018). Virus ISPA dapat menular terutama kepada individu dengan kelainan sistem kekebalan tubuh, orang lanjut usia, dan anak-anak. Kasus ISPA di Indonesia masih menjadi masalah yang mengancam kesehatan berbagai kalangan terutama balita. Permasalahan ini juga didukung dengan faktor-faktor penyebab ISPA yang secara umum masih banyak ditemui. Menurut Panduan Interim *World Health Organization* (2007), ISPA terjadi karena beberapa faktor intrinsik dan faktor ekstrinsik.

Salah satu penemuan kasus ISPA berada di Puskesmas Tanjungrejo, Kecamatan Jekulo, Kabupaten Kudus. Banyaknya kasus ISPA tiap tahunnya di Puskesmas Tanjungrejo tidak diimbangi dengan pendataan digital yang dapat memerlukan waktu lebih banyak untuk melakukan penanganan. Penelitian ini akan dilakukan klasifikasi ISPA untuk mengetahui apakah pasien penderita ISPA mengalami gejala dalam klasifikasi ISPA ringan, pneumonia atau gangguan pernapasan selain ISPA dan pneumonia. Penelitian ini menggunakan algoritma klasifikasi *Naïve Bayes Classifier* dengan menambahkan seleksi variabel *Backward Elimination*. Penelitian lain dilakukan oleh Yunitasari (2021) menggunakan klasifikasi *Naïve Bayes Classifier* dengan penambahan *Backward Elimination*, dan diperoleh akurasi sebesar 99,04%.

Pemilihan algoritma *Naïve Bayes Classifier* pada penelitian dikarenakan algoritma ini lebih sederhana dan memiliki keunggulan berupa proses perhitungan yang cepat serta menghasilkan akurasi yang tinggi (Hamzah, 2012). Selain itu, pemilihan algoritma *Naïve Bayes Classifier* ini dikarenakan lima variabel dari dataset pada penelitian ini bertipe kategori, dimana kinerja algoritma ini masih lebih unggul Ketika dilakukan pengujian pada tipe data kategori (Wibawa *et al.*, 2018). *Backward Elimination* merupakan metode seleksi variabel yang menguji semua variabel kemudian menghapus satu per satu variabel

yang dianggap tidak relevan atau tidak berpengaruh dari model (Ghani *et al.*, 2019). Pemilihan *Backward Elimination* didasarkan pada proses perhitungannya yang mempertimbangkan semua subset dari model-model yang memungkinkan sehingga dapat mengurangi resiko melewati atau kehilangan model terbaiknya (Sulaehani, 2016). Penambahan seleksi variabel *Backward Elimination* bertujuan untuk mengetahui variabel apa saja yang berpengaruh dalam model klasifikasi dan dapat meningkatkan akurasi dari algoritma *Naïve Bayes Classifier*. Penelitian ini diharapkan dapat memberikan hasil optimasi dari *Backward Elimination* pada klasifikasi ISPA dan dapat memberikan prediksi dengan nilai akurasi yang maksimal.

## 2. TINJAUAN PUSTAKA

Menurut Kementerian Kesehatan RI, ISPA merupakan infeksi akut pada pernapasan baik dari saluran atas hingga saluran bawah. ISPA ditandai dengan gejala seperti demam, batuk, pilek, dan sakit tenggorokan. ISPA dapat disebabkan oleh virus dan bakteri. Penyakit ini dapat ditularkan melalui udara dan kontak tangan atau permukaan yang terkontaminasi virus. ISPA paling sering menyerang anak terutama anak di bawah lima tahun (balita) apabila ketahanan tubuh menurun. Gejala yang dialami oleh bayi bisa menimbulkan *bronchiolitis* (radang di saluran pernapasan halus di paru-paru) dengan gejala sesak napas (Padila *et al.*, 2019). Menurut Maharani *et al.*, (2017), ISPA dapat dibedakan berdasarkan lokasi infeksi menjadi dua jenis, yaitu:

- a. ISPA atas, yaitu infeksi pernapasan di atas laring seperti rhinitis, rhinosinusitis, faringitis, tonsilitis, dan otitis media.
- b. ISPA bawah, yaitu infeksi pernapasan pada epiglotis, bronchitis, pneumonia, infeksi telinga akut dan kronis.

*World Health Organization* (2007) menyebutkan bahwa ISPA dapat terjadi karena beberapa faktor intrinsik dan faktor ekstrinsik. Faktor intrinsik misalnya usia, berat badan bayi baru lahir, status gizi, pemberian ASI yang rendah, dan status kekebalan. Faktor ekstrinsik misalnya polusi udara, asap rokok, kepadatan anggota keluarga, ventilasi, fasilitas pelayanan kesehatan yang terbatas, dan kurangnya pengetahuan terhadap bahaya ISPA.

Klasifikasi Bayesian didasarkan pada penerapan teorema Bayes dengan asumsi independensi yang kuat. Teorema Bayes merupakan teorema yang mengacu pada konsep probabilitas bersyarat. Probabilitas bersyarat adalah probabilitas dari terjadinya kejadian yang tidak diketahui akibat terjadinya kejadian yang telah diketahui. Berikut merupakan rumusan umum teorema Bayes (Berrar, 2019):

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

Keterangan:

- $A$  : Sampel data 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 kelas A  
 $P(B)$  : Probabilitas suatu input dari keseluruhan data

Diasumsikan terdapat set data berisi  $n$  kejadian  $x_i, i = 1, 2, \dots, n$ , dimana terdiri dari  $p$  variabel, yaitu  $x_{ik} = (x_{i1}, x_{i2}, \dots, x_{ip})$ . Setiap kejadian diasumsikan milik satu kelas  $y \in \{y_1, y_2, \dots, y_m\}$ . Pembelajaran *Naïve Bayes* mengacu pada konstruksi model probabilistik Bayesian yang memberikan probabilitas kelas posterior ke dalam sebuah kejadian:  $P(Y = y_j | X = x_{ik})$ . Teorema Bayes pada persamaan (1) dapat diterapkan dengan sedikit penyederhanaan notasi, maka didapatkan:

$$P(y_j|x_{ik}) = \frac{P(x_{ik}|y_j)P(y_j)}{P(x_{ik})} \quad (2)$$

Pembilang dalam persamaan (2) merupakan peluang gabungan dari  $x_i$  dan  $y_j$  atau dapat ditulis  $P(x_i, y_j)$ . Pada rumusan berikut hanya akan digunakan  $x$  dan menghilangkan indeks  $i$  untuk lebih sederhana, yaitu:

$$\begin{aligned} P(x|y_j)P(y_j) &= P(x, y_j) \\ &= P(x_1, x_2, \dots, x_p, y_j) \text{ karena } P(a, b) = P(a|b).P(b), \text{ maka} \\ &= P(x_1|x_2, \dots, x_p, y_j)P(x_2, x_3, \dots, x_p, y_j) \\ &= P(x_1|x_2, \dots, x_p, y_j)P(x_2|x_3, x_4, \dots, x_p, y_j)P(x_3, x_4, \dots, x_p, y_j) \\ &= P(x_1|x_2, \dots, x_p, y_j)P(x_2|x_3, x_4, \dots, x_p, y_j) \dots P(x_p|y_j)P(y_j) \end{aligned}$$

Diasumsikan bahwa  $x$  independen satu sama lain. Asumsi tersebut menjelaskan bahwa  $P(x_1|x_2, x_3, \dots, x_p, y_j) = P(x_1|y_j)$ , sehingga peluang gabungan dari  $x$  dan  $y_j$  adalah:

$$\begin{aligned} P(x_1|x_2, \dots, x_p, y_j) &= P(x_1|y_j).P(x_2|y_j) \dots P(x_p|y_j)P(y_j) \\ &= \prod_{k=1}^p P(x_k|y_j)P(y_j) \end{aligned} \quad (3)$$

Jika persamaan (3) disubstitusikan ke dalam persamaan (4), maka diperoleh:

$$P(y_j|x) = \frac{\prod_{k=1}^p P(x_k|y_j)P(y_j)}{P(x)} \quad (4)$$

Keterangan:

- $P(y_j|x)$  : Probabilitas *posterior* data dengan vektor  $x$  pada kelas  $y$
- $\prod_{k=1}^p P(x_k|y_j)$ : Probabilitas *independen* kelas  $y$  dari semua variabel dalam vektor  $x$
- $P(y_j)$  : Probabilitas *prior* kelas  $y$
- $P(x)$  : Probabilitas  $x$ , tidak bergantung pada kelas datanya

Menurut Han *et al.* (2012) jika  $x$  bernilai kontinu, maka perhitungan *likelihood* diasumsikan memiliki distribusi gaussian dengan *mean*  $\mu$  dan standar deviasi  $\sigma$ , sehingga didefinisikan sebagai berikut.

$$P(x_k|y_j) = \frac{1}{\sigma_{y_j}\sqrt{2\pi}} e^{-\frac{(x_k-\mu_j)^2}{2\sigma_{y_j}^2}} \quad (5)$$

Keterangan:

- $P(x_k|y_j)$  : Probabilitas  $x_k$  berdasarkan kondisi  $y_j$  (*likelihood*)
- $x_k$  : Nilai data  $x$  ke- $k$  yang akan diklasifikasi
- $y_j$  : Sub kelas  $y$  yang dicari ( $y_{ISPA}, y_{Pneumonia}, y_{Bukan\ keduanya}$ )
- $\sigma_{y_j}$  : Standar deviasi variabel dengan kelas  $y_j$
- $\mu_{y_j}$  : *Mean*, menyatakan rata-rata variabel dengan kelas  $y_j$

Cara menentukan kelas pada suatu sampel yaitu dengan melakukan perbandingan nilai *probabilitas posterior* tertinggi dari masing-masing kelas. Penentuan kelas terbaik dapat ditentukan dengan nilai *Maximum a Posteriori* (MAP) kelas  $C_{map}$  melalui persamaan (6):

$$C_{map} = \operatorname{argmax}_{y \in Y} P(y) \prod_{k=1}^p P(x_k|y) \quad (6)$$

*Naïve Bays Classifier* memiliki kelemahan bahwa jika probabilitas bersyarat/*likelihood* adalah 0 (nol) maka tidak berlaku karena probabilitas prediksi (probabilitas *posterior*) akan bernilai 0 juga. Terdapat metode untuk mengatasi permasalahan ini yaitu dengan

menambahkan 1 data pada setiap data di *training set* sehingga probabilitas tidak akan bernilai 0 karena terdapat minimal 1 data. Metode ini disebut dengan *Laplace Correction (Laplacian Estimator)* (Han and Kamber, 2006). *Laplace Correction* dinyatakan dalam persamaan berikut:

$$P(k_i) = \frac{k_i + 1}{n + m} \quad (7)$$

Keterangan:

$P(k_i)$  : probabilitas dari variabel  $k$  pada data ke- $i$ ;  $k = 1, 2, \dots, p$

$k_i$  : jumlah awal sampel dalam kelas variabel  $k_i$

$m$  : jumlah kelas dari variabel  $k_i$

$n$  : jumlah sampel

*Backward Elimination* merupakan metode seleksi variabel dengan melakukan pengujian pada semua variabel kemudian menghapus satu per satu variabel yang dianggap tidak relevan atau tidak berpengaruh dari model (Ghani *et al.*, 2019). Kriteria dalam menentukan variabel yang tidak signifikan dapat diukur melalui nilai AIC (*Akaike Information Criterion*). Metode AIC merupakan metode yang dapat digunakan untuk memilih model terbaik. Menurut metode AIC, model terbaik adalah model yang memiliki nilai AIC terkecil.

*Cross Validation* adalah metode statistik yang digunakan untuk mengevaluasi kinerja model atau algoritma dengan membagi data menjadi data latih dan data uji. *Cross Validation* memiliki beberapa jenis pengujian, salah satunya *K-Fold Cross Validation*. *K-Fold Cross Validation* ini merupakan teknik untuk mengestimasi performansi dari model pelatihan yang telah dibangun (Han and Kamber, 2006). *K-Fold Cross Validation* berfungsi untuk mengestimasi kesalahan prediksi dalam mengevaluasi kinerja model. Pada umumnya, nilai  $K$  pada *K-Fold Cross Validation* yaitu 10. *10-Fold Cross Validation* ini direkomendasikan untuk memperkirakan akurasi karena bias dan variansi yang relatif rendah (Han and Kamber, 2006). Setiap data akan dibagi menjadi 10 partisi data berukuran sama. Masing-masing dari 10 partisi data tersebut, *Cross Validation* menggunakan 9 partisi sebagai data latih dan 1 partisi digunakan sebagai data uji. Nilai rata-rata dari hasil pengujian 10 partisi dihitung sebagai perkiraan akurasi model dan digunakan sebagai indikator kinerja model *K-Fold Cross Validation*.

Kinerja klasifikasi pada umumnya dapat diukur menggunakan *confusion matrix*. *Confusion matrix* merupakan tabel yang menyatakan hasil kinerja klasifikasi (Prasetyo, 2014). Tabel *confusion matrix* berisi banyaknya data uji yang diklasifikasikan benar dan banyaknya data uji yang diklasifikasikan salah. Tabel *confusion matrix* ditunjukkan sebagai berikut:

**Tabel 1.** Tabel *Confusion Matrix*

Kelas Aktual	Kelas Prediksi	
	Positif	Negatif
Positif	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
Negatif	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

Keterangan:

1. *True Positive (TP)* adalah banyaknya data dari kelas aktual yang positif dan diklasifikasikan sebagai kelas positif.
2. *False Positive (FP)* adalah banyaknya data dari kelas aktual yang negatif dan diklasifikasikan sebagai kelas positif.
3. *False Negative (FN)* adalah banyaknya data dari kelas aktual yang positif dan diklasifikasikan sebagai kelas negatif.
4. *True Negative (TN)* adalah banyaknya data dari kelas aktual yang negatif dan diklasifikasikan sebagai kelas negatif.

Menurut (Prasetyo, 2014), terdapat kuantitas lainnya yang dapat digunakan sebagai pengukuran kinerja klasifikasi yaitu sensitivitas dan spesifisitas. Sensitivitas digunakan untuk mengidentifikasi hasil yang terprediksi benar sebagai positif dari sejumlah data yang sebenarnya positif. Persamaan dari sensitivitas adalah sebagai berikut:

$$Sensitivity = \frac{TP}{TP + FN} \quad (8)$$

Spesifisitas digunakan untuk mengidentifikasi hasil yang terprediksi benar sebagai negatif dari sejumlah data yang sebenarnya negatif. Persamaan dari spesifisitas adalah sebagai berikut:

$$Spesificity = \frac{TN}{FP + TN} \quad (9)$$

### 3. METODE PENELITIAN

Data pada penelitian berupa data sekunder yang diambil dari Puskesmas Tanjungrejo, Kecamatan Jekulo, Kabupaten Kudus pada tahun 2022. Data yang digunakan yaitu data penyakit ISPA pada balita dengan variabel yang digunakan yaitu usia, berat badan, gejala batuk, pilek, demam, sesak nafas, muntah, dan klasifikasi penyakit ISPA (ISPA, Pneumonia, dan Bukan ISPA dan pneumonia). Proses pengumpulan data ini diperoleh sejumlah 343 data dalam kurun waktu 1 November 2021 sampai 24 November 2022.

Metodologi penelitian yang dilakukan dalam penelitian ini yaitu sebagai berikut:

1. Mengumpulkan dan memasukkan data sekunder dari Puskesmas Tanjungrejo, yang terdiri dari beberapa variabel yaitu usia, berat badan, gejala batuk, pilek, demam, sesak nafas, muntah, dan klasifikasi penyakit ISPA.
2. Tahap *pre-processing* data meliputi transformasi data.
3. Tahap Klasifikasi *Naïve Bayes Classifier* tanpa seleksi variabel
  - a. Membagi data menjadi *data training* dan *data testing* secara bergantian menggunakan *10-Fold Cross Validation* dengan perbandingan 70:30.
  - b. Tahap klasifikasi data dengan *Naïve Bayes Classifier*.
  - c. Mengevaluasi model klasifikasi menggunakan *Confusion Matrix*.
4. Mengulangi tahap klasifikasi dengan menambahkan seleksi variabel *Backward Elimination* setelah pembagian data pada *data training* dengan langkah-langkah yang sama.
5. Membandingkan hasil akurasi terbaik dari model *Naïve Bayes Classifier* tanpa menggunakan *Backward Elimination* dan dengan menggunakan *Backward Elimination*.
6. Menerapkan model terbaik sebelumnya untuk prediksi data baru.
7. Interpretasi dan kesimpulan.

### 4. HASIL DAN PEMBAHASAN

Selama kurun waktu 1 November 2021 sampai 24 November 2022 Puskesmas Tanjungrejo menerima sebanyak 343 pasien balita dengan gejala penyakit ISPA. Sebelum melakukan klasifikasi, dari data tersebut dapat dilakukan analisis deskriptif dan diketahui bahwa usia dan berat badan penderita ISPA tersebar merata dari usia 1 sampai 59 bulan. Berat badan terendah dan tertinggi berada di luar kategori ideal, dimana menurut World Health Organization (WHO) berat badan terendah pada usia 0 bulan berada pada rentang 3,2 – 3,3 kg dan berat badan tertinggi pada usia 59 bulan berada pada rentang 18,0 – 18,3 kg.

Menurut Kemenkes RI (2018) dalam buku Riskesdas tahun 2018, prevalensi ISPA tertinggi pada balita terjadi usia 12-23 bulan sebesar 9,4% dan prevalensi pneumonia sebesar 2,5%. Hal ini sesuai dengan penemuan kasus ISPA di Puskesmas Tanjungrejo dimana penderita ISPA terbanyak dari kisaran usia 12-23 bulan yakni sebanyak 34 pasien.

Sedangkan untuk penemuan kasus pneumonia di Puskesmas Tanjungrejo didominasi oleh balita dengan kisaran usia 48-59 bulan yakni sebanyak 33 pasien.

Data yang diperoleh belum siap diolah sehingga perlu dilakukan *preprocessing data* terlebih dahulu yaitu berupa transformasi data. Langkah awal dalam transformasi data ialah mengubah data kategorik ke dalam bentuk numerik. Langkah berikutnya yaitu mentransformasikan skala data menggunakan metode *min-max normalization*.

Pengujian pertama yaitu klasifikasi *Naïve Bayes Classifier* tanpa seleksi variabel. Pengujian ini bertujuan untuk mengetahui kinerja dari algoritma *Naïve Bayes Classifier* tanpa seleksi variabel dalam mengklasifikasikan data ke dalam kelas yang telah ditentukan. Data dibagi menggunakan metode *10-Fold Cross Validation* dengan 70% dijadikan sebagai data uji dan 30% sebagai data latih sehingga mendapatkan data latih sebanyak 241 data dan data uji sebanyak 102 data. Hasil klasifikasi dari *Naïve Bayes Classifier* tanpa seleksi variabel sebagai berikut.

**Tabel 2.** Hasil Klasifikasi *Naïve Bayes* tanpa Seleksi Variabel

Usia	BB	Batuk	Pilek	Demam	Sesak	Muntah	Klasifikasi Aktual	Prediksi
0,2759	0,3399	Tidak	Tidak	Ya	Tidak	Tidak	Bukan keduanya	Bukan keduanya
1,0000	0,7537	Tidak	Ya	Tidak	Ya	Ya	Pneumonia	Pneumonia
0,8276	0,5567	Ya	Ya	Ya	Tidak	Tidak	ISPA	ISPA
0,3276	0,3103	Ya	Ya	Tidak	Ya	Tidak	ISPA	ISPA
0,4138	0,4089	Tidak	Tidak	Ya	Tidak	Tidak	Bukan keduanya	Bukan keduanya
...	...	...	...	...	...	...	...	...
0,8621	0,6059	Ya	Ya	Ya	Ya	Tidak	Pneumonia	Pneumonia

Berdasarkan output dapat dilihat perbandingan hasil klasifikasi aktual dengan hasil klasifikasi prediksi. Sebanyak 81 data diklasifikasikan dengan benar dan sebanyak 21 data diklasifikasikan salah.

**Tabel 3.** *Confusion Matrix Naïve Bayes Classifier* tanpa Seleksi Variabel

Prediksi	Aktual		
	Bukan Keduanya	ISPA	Pneumonia
Bukan Keduanya	27	1	1
ISPA	1	33	11
Pneumonia	2	5	21

Diperoleh hasil *confusion matrix* yang menunjukkan ketepatan algoritma dalam mengklasifikasikan data pada setiap kelas yang diklasifikasi pada Tabel 2. Kemudian dilakukan perhitungan nilai akurasi untuk algoritma *Naïve Bayes Classifier* tanpa seleksi variabel dan diketahui bahwa sebesar 79,41% tingkat keberhasilan algoritma *Naïve Bayes Classifier* tanpa seleksi variabel dalam memprediksi data klasifikasi ISPA dengan akurat.

Pengujian kedua yaitu melakukan seleksi variabel pada variabel-variabel ISPA. Proses seleksi variabel dapat dilihat pada Tabel 4.

**Tabel 2.** Hasil Akhir Iterasi *Backward Elimination*

Tahap ke-	Variabel	AIC
1	Usia, BB, Batuk, Pilek, Demam, Muntah, Sesak	210,2985
2	BB, Batuk, Pilek, Demam, Muntah, Sesak	208,5663

Diketahui bahwa dengan menghilangkan variabel usia dapat terbentuk model terbaik yang terdiri dari variabel BB, Batuk, Pilek, Demam, Muntah, Sesak dengan nilai AIC yaitu 208,5663. Variabel usia dapat dihilangkan dari model karena tidak cukup kuat untuk menjelaskan munculnya gejala-gejala ISPA, sedangkan variabel berat badan dapat menjadi indikator dari kurang tercukupinya status gizi pada balita. Hal tersebut menunjukkan apabila

balita dengan berat badan yang rendah maka memiliki status gizi yang rendah pula, sehingga dapat menjadi faktor pendukung munculnya gejala-gejala ISPA. Berdasarkan model yang terbentuk dapat diketahui bahwa baik variabel berat badan maupun gejala-gejala ISPA merupakan indikator dari ISPA.

Pengujian ketiga yaitu klasifikasi *Naïve Bayes Classifier* dengan seleksi variabel. Pengujian ini bertujuan untuk mengetahui kinerja dari seleksi variabel *Backward Elimination* dalam mengoptimalkan algoritma *Naïve Bayes Classifier* dalam mengklasifikasikan data ke dalam kelas yang telah ditentukan. Hasil klasifikasi dari *Naïve Bayes Classifier* dengan seleksi variabel sebagai berikut.

**Tabel 3.** Hasil Klasifikasi *Naïve Bayes* dengan Seleksi Variabel

BB	Batuk	Pilek	Demam	Sesak	Muntah	Klasifikasi Aktual	Prediksi
0,3399	Tidak	Tidak	Ya	Tidak	Tidak	Bukan Keduanya	Bukan Keduanya
0,7537	Tidak	Ya	Tidak	Ya	Ya	Pneumonia	Pneumonia
0,5567	Ya	Ya	Ya	Tidak	Tidak	ISPA	ISPA
0,3103	Ya	Ya	Tidak	Ya	Tidak	ISPA	ISPA
0,4089	Tidak	Tidak	Ya	Tidak	Tidak	Bukan Keduanya	Bukan Keduanya
...	...	...	...	...	...	...	...
0,6059	Ya	Ya	Ya	Ya	Tidak	Pneumonia	Pneumonia

Berdasarkan *output* dapat dilihat perbandingan hasil klasifikasi aktual dengan hasil klasifikasi prediksi. Sebanyak 89 data diklasifikasikan dengan benar dan sebanyak 13 data diklasifikasikan salah.

**Tabel 6.** *Confusion Matrix Naïve Bayes Classifier* dengan Seleksi Variabel

Prediksi	Aktual		
	Bukan Keduanya	ISPA	Pneumonia
Bukan Keduanya	27	1	0
ISPA	2	39	9
Pneumonia	1	0	23

Diperoleh hasil *confusion matrix* yang menunjukkan ketepatan algoritma dalam mengklasifikasikan data pada setiap kelas yang diklasifikasi pada Tabel 5. Kemudian dilakukan perhitungan nilai akurasi untuk algoritma *Naïve Bayes Classifier* dengan seleksi variabel dan diketahui bahwa sebesar 87,25% tingkat keberhasilan algoritma *Naïve Bayes Classifier* dengan seleksi variabel *Backward Elimination* dalam memprediksi data klasifikasi ISPA secara akurat.

Berdasarkan ketiga pengujian sebelumnya maka dapat dilakukan evaluasi terhadap algoritma *Naïve Bayes Classifier* tanpa seleksi variabel dan algoritma *Naïve Bayes Classifier* dengan seleksi variabel *Backward Elimination*. Tujuan pada evaluasi ini adalah untuk melihat bagaimana optimasi *Backward Elimination* mampu meningkatkan tingkat akurasi dari 79,41% tanpa seleksi variabel hingga 82,35% dengan seleksi variabel. Optimasi dari seleksi variabel *Backward Elimination* mampu meningkatkan akurasi sebesar 2,94%.

## 5. KESIMPULAN

Berdasarkan hasil analisis dan pembahasan klasifikasi ISPA didapatkan kesimpulan sebagai berikut:

1. Pengujian menggunakan seleksi variabel *Backward Elimination* menghasilkan model terbaik dengan nilai AIC terkecil sebesar 208,5663, sehingga diperoleh variabel yang signifikan yaitu berat badan, batuk, pilek, demam, sesak, dan muntah.

2. Pengujian klasifikasi *Naïve Bayes Classifier* tanpa seleksi variabel diperoleh akurasi sebesar 79,41%, sedangkan klasifikasi *Naïve Bayes Classifier* dengan seleksi variabel *Backward Elimination* diperoleh akurasi sebesar 87,25%. Berdasarkan kedua hasil akurasi tersebut, diperoleh bahwa akurasi terbaik didapatkan setelah dilakukan penambahan seleksi variabel *Backward Elimination* dimana mampu meningkatkan akurasi sebesar 7,84%.
3. Hasil klasifikasi yang diperoleh dapat digunakan untuk melakukan prediksi dengan variabel data yang sama, sehingga apabila terdapat data dengan variabel yang sama dapat diolah dengan *GUI-R* untuk mengklasifikasikan data dengan lebih cepat dan efisien.

## DAFTAR PUSTAKA

- Berrar, D. (2019) *Bayes' Theorem and Naive Bayes Classifier. Encyclopedia of Bioinformatics and Computational Biology, Elsevier.*
- Ghani, A. D., Salman, N. and Mustikasari (2019) 'Algoritma k-Nearest Neighbor Berbasis Backward Elimination Pada Client Telemarketing', *Prosiding Seminar Ilmiah Sistem Informasi dan Teknologi Informasi*, 8(2), pp. 141–150.
- Hamzah, A. (2012) 'Klasifikasi Teks Dengan Naïve Bayes Classifier (NBC) Untuk Pengelompokan Teks Berita Dan Abstrak Akademis', *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) Periode III, (2011)*, 269–277.
- Han, J. and Kamber, M. (2006) *Data Mining Concepts and Techniques*. Second Edi. Diane Cerra.
- Kemendes RI (2018) 'RISKESDAS 2018.pdf', *Riset Kesehatan Dasar*.
- Lonang, S. and Normawati, D. (2021) 'Klasifikasi Status Stunting Pada Balita Menggunakan K-Nearest Neighbor Dengan Feature Selection Backward Elimination', *Jurnal Media Informatika Budidarma*, 6(1), p. 49.
- Maharani, D., Yani, F. F. and Lestari, Y. (2017) 'Profil Balita Penderita Infeksi Saluran Nafas Akut Atas di Poliklinik Anak RSUP DR. M. Djamil Padang Tahun 2012-2013', *Jurnal Kesehatan Andalas*, 6(1), p. 152.
- Organization, W. H. (2007) *Infection Prevention and Control of Epidemic-and Pandemic-Prone Acute Respiratory Diseases in Health Care., World Health Organization.*
- Organization, W. H. (no date) *Child growth standards: Standards/Weight-for-age, World Health Organization.* Available at: [www.who.int/toolkits/child-growth-standards/standards/weight-for-age](http://www.who.int/toolkits/child-growth-standards/standards/weight-for-age).
- Padila, Febriawati, H., Andri, J., Dori, R.A. (2019) 'Perawatan Infeksi Saluran Pernafasan Akut (ISPA) pada Balita', *Jurnal Kesmas Asclepius*, 1(1), 25–34.
- Prasetyo, E. (2014) *Data Mining - Mengolah Data Menjadi Informasi Menggunakan Matlab*. Edited by A. Sahala. CV. Andi Offset.
- Sulaehani, R. (2016) 'Prediksi Keputusan Klien Telemarketing Untuk Deposito Pada Bank Menggunakan Algoritma Naïve Bayes Berbasis Backward Elimination', *ILKOM Jurnal Ilmiah*, 8(3), 182–189.
- Tandi, J. Penno, M., Ruterlin, V., Panggeso, A. (2018) 'Kajian Peresepan Obat Antibiotik Penyakit ISPA Pada Anak Di RSUD Anutapura Palu Tahun 2017', *Jurnal Ilmiah Farmasi*, 7(4).
- Wibawa, A. P., Purnama, M. G. A., Akbar, M. F., Dwiyanto, F. A. (2018) 'Metode-Metode Klasifikasi', *Prosiding Seminar Ilmu Komputer dan Teknologi Informasi*, 3(1), 134.
- Yunitasari, Hopipah, H. S. dan Mayasari, R. (2021) 'Optimasi Backward Elimination untuk Klasifikasi Kepuasan Pelanggan Menggunakan Algoritme k-nearest neighbor (k-NN\_ and Naïve Bayes', *Technomedia Journal*, 6(1), 99–110.