

# Implementasi Metode Backward Elimination Pada Algoritma SVM Untuk Klasifikasi Penyakit Ispa

Ahmad Yoga Adi Saputra<sup>1</sup>, Taghfirul Azhima Yoga Siswa<sup>2,\*</sup>, Rofilde Hasudungan<sup>3</sup>

<sup>1,2,3</sup> Fakultas Sains Dan Teknologi, Teknik Informatika, Universitas Muhammadiyah Kalimantan Timur, Samarinda, Indonesia

Email: <sup>1</sup>2011102441145@umkt.ac.id, <sup>2\*</sup>tay758@umkt.ac.id, <sup>3</sup>rh219@umkt.ac.id

\*) Penulis Korespondensi

**Abstrak**— Infeksi Saluran Pernapasan Akut (ISPA) merupakan salah satu penyakit dengan jumlah kasus tinggi di Indonesia. Pada tahun 2023 tercatat 86.364 kasus ISPA, dan di Kota Bontang sendiri terdapat 5.820 kasus pada tahun 2024. Tantangan utama dalam klasifikasi penyakit ISPA adalah rendahnya akurasi prediksi, terutama dalam membedakan antara kategori ringan dan berat. Ketidakseimbangan data antar kelas menyebabkan model klasifikasi cenderung bias terhadap kelas mayoritas. Penelitian ini bertujuan untuk meningkatkan kemampuan klasifikasi penyakit ISPA dengan mengombinasikan algoritma Support Vector Machine (SVM) dan teknik seleksi fitur Backward Elimination. Dataset terdiri dari 1.501 data pasien dari UPT Puskesmas Bontang Barat tahun 2024–2025. Proses penelitian meliputi tahap pre-processing, seleksi fitur, penerapan SVM, validasi menggunakan 10-fold cross validation, dan evaluasi dengan confusion matrix. Hasil penelitian menunjukkan bahwa Backward Elimination berhasil mengidentifikasi 6 fitur yang relevan, yaitu jenis kelamin, sistolik, respiratory rate, berat badan, tinggi badan, dan leukosit polimorfonuklear. Model SVM dengan fitur terpilih menghasilkan akurasi 88,33%, meningkat dari sebelumnya 83,85%. Namun, untuk menghindari bias akibat ketidakseimbangan data, metrik evaluasi tambahan seperti recall, precision, dan F1-score juga dianalisis untuk memastikan model tidak hanya unggul secara akurasi, tetapi juga mampu mengenali kasus ISPA berat secara efektif.

**Kata Kunci:** Support Vector Machine , Backward Elimination , Klasifikasi , Machine Learning , ISPA

**Abstract** - Acute Respiratory Infections (ARI) are among the most prevalent diseases in Indonesia. In 2023, a total of 86,364 ARI cases were reported, with the city of Bontang recording 5,820 cases in 2024 alone. One of the main challenges in ARI classification lies in the low prediction accuracy, particularly in distinguishing between mild and severe cases. The imbalance between class distributions often leads classification models to be biased toward the majority class. This study aims to improve ARI classification performance by combining the Support Vector Machine (SVM) algorithm with the Backward Elimination feature selection technique. The dataset consists of 1,501 patient records collected from UPT Puskesmas Bontang Barat during 2024–2025. The research process includes data pre-processing, feature selection, SVM implementation, 10-fold cross-validation, and performance evaluation using a confusion matrix. The results show that Backward Elimination successfully identified six relevant features: gender, systolic blood pressure, respiratory rate, body weight, height, and polymorphonuclear leukocytes. The SVM model using these selected features achieved an accuracy of 88.33%, an improvement from the previous 83.85%. However, to avoid bias due to data imbalance, additional evaluation metrics such as recall, precision, and F1-score were also analyzed to ensure the model not only performs well in terms of accuracy but also effectively identifies severe ARI cases.

**Keywords:** Support Vector Machine , backward Elimination , Klasifikasi , Machine Learning , ARI

## 1. PENDAHULUAN

Di Indonesia, Penyakit Infeksi Saluran Pernapasan Akut (ISPA) masih menjadi salah satu masalah kesehatan, terutama pada kalangan balita. Penyakit ini masih banyak ditemukan di daerah perkotaan yang di mana paparan polusi udara dapat meningkatkan resiko terkena penyakit ISPA, tercatat pada tahun 2023 sebanyak 86.364 kasus yang terkena penyakit ISPA di Indonesia [1]. Kota Bontang tercatat penyakit ISPA menepati nomor tiga dengan jumlah kasus sebanyak 5.820 kasus (BPS, 2024). Hal ini menunjukkan bahwa kasus ini masih menjadi kasus yang sangat di waspadai. Melihat masih tingginya jumlah kasus yang masih terjadi, diperlukan pendekatan inovatif dalam menangani kasus ISPA, salah satunya dengan menerapkan metode *Machine Learning* guna menganalisis kasus ISPA dalam membantu pencegahan dan penanganan yang lebih efektif.

Dalam *Machine Learning* terdapat beberapa metode salah satunya adalah metode klasifikasi yang dimana metode tersebut dapat mengelompokkan data kedalam kelas tertentu. Banyak penelitian yang telah menerapkan metode klasifikasi di dalam bidang kesehatan seperti klasifikasi penyakit ISPA dengan menggunakan beberapa algoritma seperti *K-Nearest* mencapai akurasi 94%, *Naive Bayes* 98% dan *SVM* 99% [2]. Algoritma klasifikasi juga telah mampu menangani kasus medis lain seperti penyakit diabetes dengan menggunakan berbagai algoritma dengan hasil akurasi sebesar *Logistic Regression* (82%), *Naive bayes* (80%), *Support Vector Machine* (87%) [3],

[4], [5]. Kemudian klasifikasi dalam kasus penyakit jantung menggunakan berbagai Algoritma dengan hasil akurasi sebesar *C4.5* (75%), *K-Nearest* (64%), *Support Vector Machine* (87%) [6], [7], [8]. Dari penelitian yang telah dilakukan sebelumnya yang menggunakan algoritma klasifikasi *Naïve Bayes*, *Logistic Regression*, *K-Nearest*, *C4.5* dan *Support Vector Machine (SVM)* dalam bidang kesehatan terdapat salah satu algoritma yang cukup tinggi nilai akurasinya yaitu algoritma *Support Vector Machine (SVM)*.

*Support Vector Machine (SVM)* merupakan algoritma pembelajaran mesin yang dapat mengelolah data menjadi *hyperplane* untuk mengklasifikasikan data linear berdasarkan kelasnya [9]. Algoritma ini di anggap lebih baik dalam melakukan klasifikasi dibidang kesehatan dibandingkan dengan algoritma yang lain. Dalam kasus kelayakan donor darah algoritma ini mampu memberikan hasil yang cukup baik dengan hasil 96% dibandingkan dengan menggunakan algoritma *Naive Bayes* dengan hasil 93% [10]. Meskipun algoritma *SVM* menunjukkan hasil akurasi yang cukup baik, Namun dari beberapa penelitian sebagian besar masih menggunakan data dengan berdimensi rendah atau jumlah atribut yang terbatas. Seperi penelitian [11]. dengan menggunakan beberapa Algoritma seperti *SVM*, *Decision Tree* dan *Naive Bayes* yang digunakan melakukan klasifikasi data berdimensi tinggi, namun hasil akurasi hanya sebesar 53% hingga 60% . Oleh karena itu penelitian ini menjadi penting untuk menguji efektifitas metode klasifikasi pada data dengan dimensi yang lebih tinggi.

Data berdimensi tinggi merupakan data yang mempunyai banyak atribut atau fitur dalam sebuah dataset untuk digunakan dalam menganalisis. Data berdimensi tinggi memiliki beberapa tantangan, seperti resiko *Overfitting* maupun *Underfitting* serta kesulitan dalam visualisasikan data secara efektif [12]. Dalam beberapa algoritma pun dapat mengalami penurunan peforma dalam nilai akurasi jika dihadapkan dengan data berdimensi tinggi seperti *K- Nearest Neighbor* (52%) dan *Naive Bayes* (67%), [13], [14]. Untuk itu dalam mengatasi masalah yang sering muncul pada data berdimensi tinggi diperlukan metode seleksi fitur dengan tujuan dapat memilih atribut yang benar - benar mempengaruhi terhadap akurasi model sehingga model akan bekerja lebih efektif dan memberikan hasil peforma yang baik dalam proses klasifikasi.

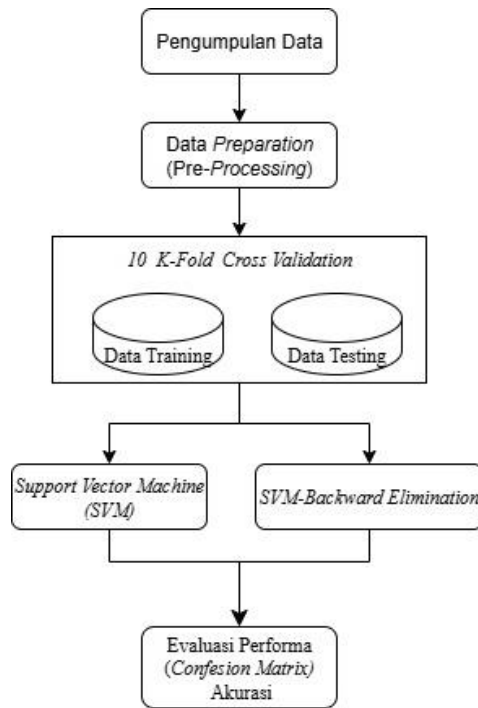
Seleksi Fitur (*Feature Selection*) merupakan teknik yang sering digunakan dalam penerapan *machine learning* yang digunakan pada saat tahap *pre-processing*. Pada tahap ini seleksi fitur dapat bekerja dengan mengurangi jumlah fitur yang tidak relevan untuk di jadikan sebuah hasil yang efisien [15]. Seperti metode *Wrapper* terdapat dua teknik pendekatan yaitu *Forward Selection* dan *Backward Elimination*, metode *filter* juga memiliki dua teknik pendekatan utama yaitu *Information Gain* dan *Gain Ratio*. Pada penelitian [16] yang menggunakan seleksi fitur berbasis *Wrapper* dan *Filter* pada klasifikasi penyakit jantung menunjukkan hasil akurasi yang cukup baik yaitu mencapai 83% dan 90% dari setiap metode seleksi fiturnya. Dalam memilih sebuah metode seleksi fitur yang sesuai dapat memberikan pengaruh yang cukup signifikan terhadap peforma model yang dihasilkan [17]. Oleh karena itu salah satu pendekatan yang sering di gunakan dalam seleksi fitur adalah *Backward Elimination*. Metode tersebut dapat digunakan dalam proses klasifikasi karena dapat mengetahui atribut yang benar – benar berpengaruh. Dikarenakan dalam kemampuan metode tersebut dapat mengevaluasi setiap fitur secara bertahap serta memiliki fungsi dalam mengoptimalkan peforma suatu model dengan sistem kerja pemilihan mundur [18]. Metode *Backward Elimination* juga dapat menghapus fitur yang kurang relevan dan dapat mempertahankan fitur yang memberikan kontribusi dalam hasil klasifikasi.

Beberapa penelitian telah menunjukkan bahwa penggunaan metode seleksi fitur *Backward Elimination* dalam proses klasifikasi dapat meningkatkan akurasi model. Misalnya, dalam penelitian [19] yang menerapkan teknik ini untuk klasifikasi penyakit ISPA, terjadi peningkatan akurasi sebesar 7,84%, dari sebelumnya menjadi 87,25%. Dalam kasus lain, seperti klasifikasi penyakit diabetes, penerapan *Backward Elimination* juga memberikan peningkatan signifikan, di mana akurasi awal sebesar 92% meningkat menjadi 97% setelah dilakukan seleksi fitur [20]. Temuan ini menunjukkan bahwa *Backward Elimination* merupakan teknik yang efektif untuk meningkatkan performa klasifikasi, khususnya dalam bidang kesehatan. Berdasarkan studi literatur yang telah dilakukan, penelitian ini mengusulkan algoritma *SVM* dengan metode seleksi fitur *Backward Elimination* untuk meningkatkan peforma *SVM* dalam menangani klasifikasi kasus penyakit ISPA di Bontang pada dataset berdimensi tinggi. Oleh karena itu dalam penelitian ini masih terbilang inovatif dan diharapkan dapat memberikan peningkatan dalam akurasi klasifikasi dengan kendala dataset berdimensi tinggi pada data Penyakit ISPA.

## 2. METODE PENELITIAN

### 2.2 Prosedur Penelitian

Prosedur penelitian ini terdiri dari serangkaian Langkah-langkah yang dilakukan untuk menyelesaikan masalah dalam penelitian.



Gambar 1 Alur Penelitian

### 2.2.1 Pengumpulan Data

Data yang di gunakan dalam penelitian ini adalah data penyakit ISPA yang di peroleh dari UPT. Puskesmas Kota Bontang. Adapun hasil dari pengumpulan data didapatkan berupa 12 fitur untuk mengklasifikasikan data penyakit ISPA. Berikut adalah fitur-fitur mengenai dataset penyakit ISPA.

Tabel 1 Fitur Data ISPA

No	Atribut	Tipe Data	Keterangan
1	Jenis Kelamin	String	Jenis klamin (Laki-Laki dan Perempuan)
2	Umur	Numeric (Int)	Usia dari seorang pasien
3	Pendidikan	String	Pendidikan terakhir yang ditempuh oleh pasien
4	Sistolik	Numeric (Int)	Tekanan Darah saat jantung berkontraksi berkisar 120/80 mmHg
5	Diastolik	Numeric (Int)	Tekanan Darah saat jantung berelaksasi berkisar 90/60 mmHg
6	Nadi	Numeric (int)	Denyut nadi normal berkisar 60-100 denyut per menit (bpm)
7	RR (Respiratory rate)	Numeric (int)	Laju Pernafasan
8	Temp	Numeric (Float)	Suhu Badan
9	BB	Numeric (Float)	Berat badan
10	TB	Numeric (Float)	Tinggi badan
11	LP	Numeric (Int)	Leukosit Polimorfonuklear atau jenis sel darah putih
12	Disease	String (Class)	target class ( Berat dan Ringan)

Pada Tabel 1 menjelaskan tentang isi atribut – atribut data pada pasien beserta tipe data dan keterangannya yang dimana terdapat 12 atribut yang di gunakan yang diantaranya adalah Jenis Kelamin, Umur, Pendidikan, *Sistolik*, *Diastolik*, Nadi, RR (Laju Pernafasan), Temp (Suhu Badan), BB (Berat Badan), TB (Tinggi Badan), LP (Jumlah *Leukosit Polimorfonuklear*) dan *Disease* Sebagai kelas target. Dalam atribut tersebut terdapat beberapa tipe data seperti *String* dan *Numeric (Int/Float)*. Namun pada atribut *Disease* merupakan label klasifikasi dengan dua kelas yaitu Berat dan Ringan.

### 2.2.2 Data Pre-Processing

Pada Langkah ini akan melakukan 3 langkah, yaitu data *selection*, data *cleaning* dan data *transformation*. Data didapatkan dari UPT. Puskesmas Bontang Barat dengan proses persiapan sebagai berikut:

#### a. Data Selection

Pada tahap ini, dilakukan proses pemilihan fitur yang relevan untuk klasifikasi penyakit ISPA. Fitur yang tidak memberikan kontribusi signifikan dalam proses klasifikasi akan di hapus. Seperti tabel 2 yang menampilkan data dari UPT. Puskesmas Bontang Barat yang awalnya memiliki 12 kolom. Setelah melalui tahap seleksi, ditemukan bahwa 2 kolom kurang efektif dalam klasifikasi penyakit ISPA. Oleh karena itu data yang digunakan setelah seleksi terdiri dari 10 atribut, yang mencakup 9 atribut sebagai fitur dan 1 atribut sebagai label class. Hasil akhir setelah seleksi fitur dapat dilihat pada tabel 2.

**Tabel 2** Data Selection

No	Atribut	Tipe Data	Keterangan
1	Jenis Kelamin	String	Jenis kelamin (Laki-Laki dan Perempuan)
2	Sistolik	Numeric (Int)	Tekanan Darah saat jantung berkontraksi berkisar 120/80 mmHg
3	Diastolik	Numeric (Int)	Tekanan Darah saat jantung berelaksasi berkisar 90/60 mmHg
4	Nadi	Numeric (int)	Denyut nadi normal berkisar 60-100 denyut per menit (bpm)
5	RR ( <i>Respiratory rate</i> )	Numeric (int)	Laju Pernafasan
6	Temp	Numeric (Float)	Suhu Badan
7	BB	Numeric (Float)	Berat badan
8	TB	Numeric (Float)	Tinggi badan
9	LP	Numeric (Int)	<i>Leukosit Polimorfonuklear</i> atau jenis sel darah putih
10	Disease	String (Class)	target class ( Berat dan Ringan)

Pada Tabel 2 merupakan atribut yang tetap digunakan setelah dilakukannya proses seleksi fitur dari total 12 atribut awal menjadi 10 atribut yang di anggap relevan untuk proses klasifikasi yang dimana atribut seperti umur dan pendidikan tidak disertakan karena dinilai tidak memberikan pengaruh yang signifikan terhadap peningkatan akurasi model. Oleh karena itu atribut tersebut dieliminasi guna memaksimalkan kinerja algoritma klasifikasi.

#### b. Data Cleaning

Data *cleaning* (pembersihan data) bertujuan menghilangkan nilai kosong, dataset sering kali mengandung data yang hilang atau kosong, yang dapat menyebabkan terjadinya kesalahan saat membangun model. Data *cleaning* adalah proses penghapusan atau koreksi data yang salah, tidak lengkap dan tidak konsisten. Langkah ini penting untuk memastikan keakuratan pada saat analisis dan prediksi. Dalam penelitian ini, data *cleaning* akan menggunakan fungsi *library pandas* yang bernama *dropna()* untuk menghapus baris yang mengandung nilai Nan atau satu nilai yang hilang di dalam suatu baris.

#### c. Data Transformation

Data *Transformation* merupakan sebuah proses mengubah nilai atribut yang bersifat kategorial (berupa data text) menjadi bentuk numerik. Tahap ini sangat penting dilakukan karena *library scikit-learn* hanya dapat memproses data dengan tipe *numerik*. Beberapa atribut yang perlu di transformasikan meliputi jenis kelamin dan Disease. Proses konversi ini menggunakan *library scikit-learn* dengan memanfaatkan fungsi *labelEncoder*. Fungsi tersebut secara otomatis akan merubah data kategorial menjadi nilai numerik dalam satu kolom data.

### 2.2.3 Pembagian Data

Proses pembagian data dilakukan dengan memisahkan dataset menjadi dua bagian yaitu data latih dan data uji. Data latih digunakan untuk membangun dan mengembangkan model dengan cara mengenali pola serta hubungan



- d) Selanjutnya menentukan nilai *p-value* berdasarkan pada nilai *t* table
- e) Menentukan nilai *p-value* terbesar dari masing-masing atribut
- f) Kemudian dilakukan perbandingan antara nilai *p-value* dengan taraf signifikansi. Jika terdapat atribut yang memiliki *p-value* lebih besar dari taraf signifikansi, maka atribut dengan nilai *p-value* terbesar akan dihapus. Setelah itu, proses diulangi kembali mulai dari langkah pertama hingga seluruh atribut yang tersisa memiliki *p-value* yang lebih kecil dari taraf signifikansi

### 2.2.7 Evaluasi Model

Tahap evaluasi ini merupakan langkah penting setelah pembentukan model yang dimana performa model akan diukur untuk mengevaluasi akurasi. Pengujian ini dilakukan dengan teknik *confusion matrix*. *Confusion matrix* adalah tabel yang menyatakan klasifikasi jumlah data uji yang benar dan jumlah data uji yang salah [24]

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \quad (2)$$

Keterangan :

TP (*True Positive*) : Jumlah prediksi benar untuk kelas positif (kelas minoritas yang benar terklasifikasikan sebagai positif).

TN (*True Negative*) : jumlah prediksi benar untuk kelas negatif (kelas mayoritas yang benar terklasifikasikan sebagai negatif).

FP (*False Positive*) : jumlah prediksi salah dimana kelas negatif diklasifikasikan sebagai positif.

FN (*False Negative*) : jumlah prediksi salah di mana kelas positif diklasifikasikan sebagai negatif

## 3 HASIL DAN PEMBAHASAN

### 3.1 Hasil Penelitian

Pada hasil ini menyajikan hasil dari analisis dan pengujian yang telah dilakukan, disertai dengan penjelasan secara menyeluruh. Analisis ini difokuskan pada evaluasi kinerja sebuah model dalam melakukan proses pemilihan fitur serta hasil yang diperoleh melalui berbagai metode yang digunakan selama penelitian. Sehingga temuan yang dihasilkan akan dibandingkan dengan tujuan awal penelitian untuk menjawab rumusan masalah.

Tahap analisis data mencakup proses pengumpulan, pengolahan, hingga penerapan algoritma *Support Vector Machine (SVM)* yang di kombinasikan dengan teknik Seleksi Fitur *Backward Elimination* dengan menggunakan dataset Penderita Penyakit ISPA di kota Bontang yang terdiri dari 12 fitur dengan 11 fitur sebagai atribut dan 1 fitur sebagai target *class*.

#### 3.1.1 Data Selection

Pada proses ini, dilakukan proses pengambilan data yang dilakukan secara manual untuk memilih fitur-fitur apasaja yang dianggap relevan, sedangkan atribut yang dianggap kurang relevan akan dihilangkan secara manual. Sehingga telah dipilih 9 fitur beserta 1 fitur sebagai target *class* yang ditunjukkan pada tabel 3.1 berikut

Tabel 4 Tampilan data Selection

No	Jns.Kelamin	Sistolik	Diastolik	Nadi	RR	Temp	BB	TB	LP	Disease
1	Perempuan	110	70	89	21	36.5	17	120	79	Berat
2	Perempuan	140	90	80	18	36.0	30	141	87	Berat
3	Laki-Laki	110	70	98	21	36.4	5.7	60	0	Berat
4	Perempuan	110	60	89	18	36.5	28	138	70	Berat
5	Perempuan	110	80	89	23	36.7	25	134	85	Ringan
....	....	....	....	....	....	....	....	....	....	....
149	Laki-Laki	100	70	78	24	36.0	8	73	27	Berat
7										
149	Perempuan	100	70	78	21	36.7	7	69	67	Ringan
8										

149 9	Perempuan	160	70	80	24	36.6	14	110	87	Berat
150 0	Perempuan	100	70	80	24	36.6	12	80	86	Ringan
150 1	Perempuan	120	70	80	24	36.6	6	55	78	Ringan

Pada tabel 3.1 merupakan 10 fitur yang sesuai dengan relevansi terhadap kualitas dan signifikan terhadap penelitan fitur-fitur yang dipilih diantaranya adalah Jns.Kelamin, *Sistolik*, *Diastolik*, Nadi, RR, Temp, BB, TB, LP dan *Disease*. Tujuan *Data Selection* adalah untuk memastikan data yang berkualitas dan yang relevan untuk digunakan sehingga dapat meningkatkan efisiensi dalam menganalisis penyakit ISPA

### 3.1.2 Data Cleaning

Pada tahap ini merupakan hasil dari proses pembersihan data pada dataset Penderita penyakit ISPA yang terdiri dari 1.501 baris. Pada proses ini, dilakukan pengecekan dataset yang berisikan nilai kosong atau *NaN* di dalam fitur-fitur menggunakan *missing\_values = data.isna(). Sum()*. jika terdapat nilai Nan di sebuah fitur maka dapat mempengaruhi sebuah hasil analisis.

**Tabel 5** Data Sebelum Tahap Cleaning

No	Fitur	Missing Values
1	Jns.Kelamin	0
2	<i>Sistolik</i>	0
3	<i>Diastolik</i>	0
4	Nadi	0
5	RR	0
6	Temp	13
7	BB	0
8	TB	0
9	LP	4
10	<i>Disease</i>	0

Pada tabel 5 terdapat nilai *NaN* yang dimana nilai *NaN* tersebut Berada pada fitur Temp dengan jumlah *NaN* sebanyak 13 dan pada fitur LP terdapat nilai *NaN* sebanyak 4. Maka diperlukan fungsi *drop\_duplicate()* untuk menghapus baris baris duplikat dan fungsi *dropna()* untuk menghilangkan atau menghapus nilai *NaN* yang ada.

**Tabel 6** Dataset Setelah Tahap Cleaning

No	Fitur	Missing Values
1	Jns.Kelamin	0
2	<i>Sistolik</i>	0
3	<i>Diastolik</i>	0
4	Nadi	0
5	RR	0
6	Temp	0
7	BB	0
8	TB	0
9	LP	0
10	<i>Disease</i>	0

Pada tabel 6 merupakan dataset yang telah di bersihkan nilai *NaN* sebanyak 17 baris sehingga jumlah dataset menjadi berkurang menjadi 1.484 baris. Setelah dilakukannya proses penghapusan tersebut dataset yang sudah bersih dapat di gunakan secara efektif dalam menganalisis penderita penyakit ISPA tersebut.

### 3.1.3 Data Transformation

Tahap *Data Transformation* merupakan tahap yang paling penting dalam proses *Pre-Processing* karena tahap ini dapat mengubah sebuah atribut data menjadi format yang dapat di proses oleh algoritma *Machine Learning*. Penelitian ini mengubah nilai nilai *string* menjadi nilai *numerik* karena sebagian besar algoritma machine learning hanya bisa memproses nilai numerik termasuk algoritma *Support Vector Machine (SVM)* yang hanya bisa mengelola data dalam bentuk *numerik*.

**Tabel 7** *Dataset Sebelum Tahap Transformation*

No	Jns.Kelamin	Disease
1	Perempuan	Berat
2	Perempuan	Berat
3	Laki-Laki	Berat
4	Perempuan	Berat
5	Perempuan	Ringan
....	....	....
1480	Laki-Laki	Berat
1481	Perempuan	Ringan
1482	Perempuan	Berat
1483	Perempuan	Ringan
1484	Perempuan	Ringan

Pada tabel 3.4 merupakan fitur-fitur yang masih memiliki nilai String diantaranya Jns.Kelamin dan Disease oleh karena itu diperlukan proses *Transformation* dengan menggunakan *LabelEncoder* pada *Library Python sklearn*.

**Tabel 8** *Dataset Sesudah Tahap Transformation*

No	Jns.Kelamin	Disease
1	1	0
2	1	0
3	0	0
4	1	0
5	1	1
....	....	....
1480	0	0
1481	1	1
1482	1	0
1483	1	1
1484	1	1

Pada tabel 3.5 merupakan hasil dari proses data *Transformation* yang di mana nilai string berubah menjadi nilai numerik menjadi 1 dan 0 sehingga algoritma *Support Vector Machine (SVM)* dapat mengolah dataset menjadi lebih baik.

### 3.2 Hasil Pembagian Data Dengan *K-Fold Cross Validation*

Pada Penelitian ini, proses pembagian dataset kedalam data *training* dan *testing* merupakan langkah penting yang dapat mempengaruhi kinerja sebuah model dalam *data mining*. Teknik *10-Fold Cross Validation* digunakan untuk

membagi dataset menjadi sepuluh bagian yang sama besarnya pada setiap iterasi yang dilakukan. yang dimana satu bagian menjadi data testing dan sembilan bagian menjadi data training. Pada proses ini dilakukan secara acak dalam setiap *fold*, sehingga setiap bagian dapat memiliki peluang yang sama untuk berperan sebagai data training ataupun data testing. Dengan menggunakan teknik ini dapat mengurangi bias dan variasi dalam kinerja model sehingga hasil evaluasi akan menjadi lebih stabil. Penjelasan yang lebih rinci dalam proses ini akan di bahas pada tahap berikutnya, yaitu hasil pemodelan dan evaluasi.

### 3.3 Hasil Permodelan Dan Evaluasi Algoritma

Pada proses ini merupakan hasil dari tahap pemodelan dan evaluasi dari algoritma yang digunakan dalam penelitian ini. Pada penelitian ini berfokuskan untuk menggunakan algoritma *Support Vector Machine* (SVM) yang di optimalkan menggunakan tambahan metode seperti *Backward Elimination* menggunakan dataset Penderita penyakit ISPA. Proses ini melakukan berbagai tahapan mulai dari pembagian data menggunakan *K fold Cross Validation* hingga menggunakan evaluasi kinerja model dengan metrik – metrik yang cukup baik.

#### 3.3.1 penerapan menggunakan Algoritma SVM

Pada tahap ini merupakan penerapan metode *SVM* dalam melakukan klasifikasi yang dimana tahap ini mencoba melihat peforma awal dari sebuah Algoritma. Tujuan nya untuk melihat kinerja sebuah algoritma *SVM* dalam mengklasifikasikan Data Penderita Penyakit ISPA di Kota Bontang, Hal ini merupakan sebuah gambaran mengenai seberapa baik Algoritma dalam melakukan tugasnya.

**Tabel 9** Hasil Uji *SVM* Menggunakan *10K – fold Cross Validation*

Iterasi Ke -	Akurasi
1	82,19%
2	83,03%
3	81,03%
4	84,89 %
5	80,89%
6	85,89%
7	83,89%
8	82,89%
9	87,89%
10	82,89%
Rata – rata Akurasi	<b>82,90%</b>

Pada tabel 9 merupakan hasil dari algoritma *SVM* yang menggunakan teknik *10 k fold cross validation* yang dimana mendapatkan nilai akurasi rata rata sebesar 82,90%. Hal ini menunjukkan bahwa Algoritma *Support Vector Machine* (*SVM*) memberikan hasil yang cukup baik dalam hal klasifikasi sebuah penyakit ISPA. Untuk membandingkan hasil evaluasi maka diperlukan perhitungan manual nilai *confussion matrix* untuk mengkonfirmasi hasil dari evaluasi tersebut.

**Tabel 10** *Confussion Matrix SVM*

	<i>Predicted Positive (0)</i>	<i>Predicted Negative (1)</i>
<i>Actual Positive (0)</i>	36	219
<i>Actual negative (1)</i>	21	965

$$Accuracy = \frac{992+36}{992+36+191+21} = 82,90\%$$

Setelah melakukan proses perhitungan manual pada setiap parameternya maka hasil yang diperoleh sama dengan perhitungan menggunakan *Python*. Hal ini membuktikan bahwa model memiliki kemampuan yang baik dalam menerapkan data latih ke data uji dan mampu mengklasifikasikan data dengan akurat.

### 3.3.2 Penerapan Algoritma SVM dengan *Backward Elimination*

Pada proses ini , algoritma SVM diimplementasikan menggunakan kode *python* dengan menggabungkan metode seleksi fitur *Backward Elimination* untuk melakukan evaluasi menggunakan *confussion matrix* di setiap fold dalam 10 – *Fold Cross Validation* agar model algoritma dapat mempelajari data dengan lebih efisien. Berikut hasil dari algoritma SVM dengan menggunakan metode seleksi fitur *Backward Elimination*.

**Tabel 11** Fitur Awal Beserta Nilai *P-Value*

No	Fitur Terpilih	Nilai <i>P-value</i>
1	Jns.Kelamin	0.115911
2	Sistolik	0.345517
3	Diastolik	0.986010
4	Nadi	0.690110
5	RR (Laju Pernafasan)	0.408389
6	Temp	0.871002
7	BB (Berat Badan)	0.210783
8	TB (Tinggi Badan)	0.247140
9	<i>Leukosit Polimorfonuklear (LP)</i>	0.254237

Tabel 11 di atas menunjukkan daftar fitur yang dipertimbangkan dalam proses seleksi menggunakan metode *Backward Elimination*. Metode ini bertujuan untuk mengidentifikasi fitur-fitur yang secara statistik signifikan terhadap variabel target berdasarkan nilai *P-value*. Dalam konteks ini, fitur dengan *P-value* < 0,5 dianggap signifikan dan dipertahankan, sementara fitur dengan *P-value* lebih tinggi dapat dieliminasi secara bertahap.

**Tabel 12** Fitur Terpilih *Backward Elimination*

No	Fitur Terpilih	Nilai <i>P-value</i>
1	Jns.Kelamin	0.115911
2	<i>Sistolik</i>	0.345517
3	RR (Laju Pernafasan)	0.408389
4	BB (Berat Badan)	0.210783
5	TB (Tinggi Badan)	0.247140
6	<i>Leukosit Polimorfonuklear (LP)</i>	0.254237

Pada tabel 12 tersebut menyajikan 6 fitur terpilih yang dipertahankan setelah proses seleksi menggunakan metode *Backward Elimination*, berdasarkan nilai *P-value*. Dalam analisis statistik, *P-value* digunakan untuk mengukur signifikansi suatu fitur terhadap variabel target. Fitur-fitur dengan *P-value* di bawah ambang batas 0,5 dianggap signifikan dan relevan terhadap hasil klasifikasi.

**Tabel 13** Hasil dari algoritma SVM dengan *Backward Elimination*

Iterasi Ke -	Akurasi
1	88,30%
2	87,30%
3	89,30%
4	86,30%
5	88,25%
6	87,17%
7	87,17%
8	88,17%
9	85,17%
10	88,17%
Rata – rata Akurasi	<b>87,37%</b>

Pada tabel 12 merupakan hasil akurasi yang diperoleh dari implementasi algoritma SVM dengan menggunakan metode seleksi fitur *Backward Elimination* menggunakan 10 *fold-cross validation* yang di mana nilai akurasi rata - ratanya adalah 87,37%. Hal ini menunjukkan bahwa algoritma SVM di kombinasikan dengan metode seleksi fitur *Backward Elimination* bekerja dengan cukup baik dalam melakukan klasifikasi penyakit ISPA.

**Tabel 14** *confusion matrix SVM dengan Backward Elimination*

	Predicted Positive (0)	Predicted Negative (1)
Actual Positive (0)	37	109
Actual negative (1)	17	825

$$Accuracy = \frac{835+37}{825+37+109+17} = 87,37\%$$

Pada gambar 3 merupakan proses Setelah melakukan perhitungan manual pada setiap parameternya maka hasil yang diperoleh sama dengan perhitungan menggunakan *Python*. Hal ini membuktikan bahwa model memiliki kemampuan yang baik dalam menerapkan teknik seleksi fitur dengan mendapatkan nilai akurasi sebesar 88,33 yang dimana hasil ini lebih baik dari hasil sebelumnya.

### 3.3.3 Perbandingan Hasil Evaluasi

Hasil ini merupakan hasil evaluasi dari perbandingan sebuah penerapan model algoritma SVM dengan hasil dari penerapan model tabahan seleksi fitur *Backward Elimination* dalam melakukan klasifikasi penyakit ISPA. Perbandingan ini bertujuan untuk menilai sejauh mana kontribusi seleksi fitur terhadap peningkatan akurasi.

**Tabel 15** Hasil Perbandingan Evaluasi Metode

K Fold	Support Vector machine (SVM)	SVM – Backward Elimination	Perubahan Peningkatan nilai akurasi
1	82,19%	88,30%	+6,11%
2	83,03%	87,30%	+4,27%
3	81,03%	89,30%	+8,27%
4	84,89%	86,30%	+1,41%
5	80,89%	88,25%	+7,36%
6	85,89%	87,17%	+1,28%
7	83,89%	87,17%	+3,28%
8	82,89%	88,17%	+5,28%
9	87,89%	85,17%	-2,72%
10	82,89%	88,17%	+5,28%
<b>Rata – Rata Akurasi</b>	<b>82,90%</b>	<b>87,37%</b>	<b>+4,47%</b>

Pada tabel 13 merupakan hasil dari dua pendekatan dalam melakukan klasifikasi penyakit ISPA yaitu algoritma SVM dan SVM menggunakan *Backward Elimination* yang di mana kedua pendekatan tersebut menggunakan 10 k fold. Dari hasil yang didapatkan terlihat akurasi Algoritma SVM hanya mendapatkan sekitar 82,19% hingga 87,89% dengan rata rata akurasi sebesar 82,90%. Sementara hasil dari gabungan algoritma SVM dengan *Backward Elimination* mendapatkan akurasi yang cenderung stabil dengan hasil akurasi sebesar 88,30% hingga 89,30% dengan rata rata 87,37%.

#### 4. KESIMPULAN

Penelitian ini menunjukkan bahwa penerapan teknik seleksi fitur menggunakan *metode Backward Elimination* mampu meningkatkan performa model klasifikasi penyakit ISPA. Seleksi fitur terbukti berperan penting dalam meningkatkan efektivitas model dengan cara menyaring atribut-atribut yang relevan terhadap target klasifikasi. Penerapan algoritma *Support Vector Machine (SVM)* menunjukkan performa yang baik dalam mengklasifikasikan data penyakit ISPA. Kombinasi antara SVM dan teknik seleksi fitur mampu menghasilkan model yang lebih akurat dibandingkan tanpa seleksi fitur. Hal ini menegaskan bahwa pemilihan fitur yang tepat memberikan kontribusi signifikan terhadap peningkatan kinerja model klasifikasi secara keseluruhan

#### REFERENCES

- [1] D. Puspasari, "Survei Kesehatan," *Ski (Survei Kesehat. Indones.*, pp. 1–68, 2023.
- [2] M. Napiah, R. A. Purnama, M. Raharjo, and W. Bismi, "Komparasi Algoritma Untuk Klasifikasi Penyakit ISPA (Infeksi Saluran Pernapasan)," *J. Infortech*, vol. 4, no. 2, pp. 105–110, 2022.
- [3] Erlin, Yulvia Nora Marlim, Junadhi, Laili Suryati, and Nova Agustina, "Deteksi Dini Penyakit Diabetes Menggunakan Machine Learning dengan Algoritma Logistic Regression," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 11, no. 2, pp. 88–96, 2022, doi: 10.22146/jnteti.v11i2.3586.
- [4] N. M. Putry, "Komparasi Algoritma Knn Dan Naïve Bayes Untuk Klasifikasi Diagnosis Penyakit Diabetes Mellitus," *EVOLUSI J. Sains dan Manaj.*, vol. 10, no. 1, 2022, doi: 10.31294/evolusi.v10i1.12514.
- [5] N. Fitriyani, D. R. Amalia, H. H. Handayani, A. Fitri, and N. Masruriyah, "Aplikasi Berbasis Web Berdasarkan Model Klasifikasi Algoritma SVM dan Logistic Regression Terhadap Data Diabetes," *Ris. dan E-Jurnal Manaj. Inform. Komput.*, vol. 7, pp. 1762–1771, 2023, [Online]. Available: <http://doi.org/10.33395/remik.v7i4.13001%0D>
- [6] A. Sepharni, I. E. Hendrawan, and C. Rozikin, "Klasifikasi Penyakit Jantung dengan Menggunakan Algoritma C4.5," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 7, no. 2, p. 117, 2022, doi: 10.30998/string.v7i2.12012.
- [7] A. Yogiarto, A. Homaidi, and Z. Fatah, "Implementasi Metode K-Nearest Neighbors (KNN) untuk Klasifikasi Penyakit Jantung," *G-Tech J. Teknol. Terap.*, vol. 8, no. 3, pp. 1720–1728, 2024, doi: 10.33379/gtech.v8i3.4495.
- [8] M. G. Pradana, P. H. Saputro, and D. P. Wijaya, "Komparasi Metode Support Vector Machine Dan Naïve Bayes Dalam Klasifikasi Peluang Penyakit Serangan Jantung," *Indones. J. Bus. Intell.*, vol. 5, no. 2, p. 87, 2022, doi: 10.21927/ijubi.v5i2.2659.
- [9] N. Pratiwi and Y. Setyawan, "Analisis Akurasi Dari Perbedaan Fungsi Kernel Dan Cost Pada Support Vector Machine Studi Kasus Klasifikasi Curah Hujan Di Jakarta," *J. Fundam. Math. Appl.*, vol. 4, no. 2, pp. 203–212, 2021, doi: 10.14710/jfma.v4i2.11691.
- [10] R. P. Pridiptama, W. Wasono, and F. D. . Amijaya, "Perbandingan Algoritma Support Vector Machine dan Naïve Naïve Bayes pada Klasifikasi Penyakit Tekanan Darah Tinggi (Studi Kasus: Klinik Polresta Samarinda)," *Basis*, vol. 3, no. 1, pp. 1–16, 2024, [Online]. Available: <http://jurnal.fmipa.unmul.ac.id/index.php/Basis>
- [11] M. R. Ahmmed, J. Monir, and S. A. Khushbu, "Analysis of Flood Risk Prediction Using Different Machine Learning Classifiers: A Study of Predicting Flood Risk in Rural Areas, Bangladesh," *2022 13th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2022*, pp. 1–6, 2022, doi: 10.1109/ICCCNT54827.2022.9984449.
- [12] R. A. Saputra *et al.*, "Detecting Alzheimer's Disease by the Decision Tree Methods Based on Particle Swarm Optimization," *J. Phys. Conf. Ser.*, vol. 1641, no. 1, pp. 61–67, 2022, doi: 10.1088/1742-6596/1641/1/012025.
- [13] L. G. Gebeye, E. Y. Dessie, and J. A. Yimam, "Predictors of micronutrient deficiency among children aged 6–23 months in Ethiopia: a machine learning approach," *Front. Nutr.*, vol. 10, no. January, pp. 1–13, 2023, doi: 10.3389/fnut.2023.1277048.

- [14] M. R. Akhmad and T. A. Y. Siswa, "Implementasi K-Nearest Neighbor Dalam Memprediksi Keterlambatan Pembayaran Biaya Kuliah Di Perguruan Tinggi," *Progresif J. Ilm. Komput.*, vol. 18, no. 2, p. 185, 2022, doi: 10.35889/progresif.v18i2.921.
- [15] D. Kurnia, M. Itqan Mazdadi, D. Kartini, R. Adi Nugroho, and F. Abadi, "Seleksi Fitur dengan Particle Swarm Optimization pada Klasifikasi Penyakit Parkinson Menggunakan XGBoost," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 5, pp. 1083–1094, 2023, doi: 10.25126/jtiik.20231057252.
- [16] S. R. Azizah, R. Herteno, A. Farmadi, D. Kartini, and I. Budiman, "Kombinasi Seleksi Fitur Berbasis Filter dan Wrapper Menggunakan Naive Bayes pada Klasifikasi Penyakit Jantung," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 6, pp. 1361–1368, 2023, doi: 10.25126/jtiik.1067467.
- [17] L. Rangga, A. Tarigan, T. Informatika, R. Forest, O. Fitur, and F. Selection, "Optimalisasi Fitur Dengan Forward Selection Pada Estimasi Tingkat Penyakit Paru-Paru Menggunakan Algoritma Klasifikasi Random Forest," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 5, pp. 10341–10348, 2024.
- [18] T. A. Y. Siswa and R. P. Wibowo, "Komparasi Metode Seleksi Fitur Dalam Prediksi Keterlambatan Pembayaran Biaya Kuliah," *Teknika*, vol. 12, no. 1, pp. 73–82, 2023, doi: 10.34148/teknika.v12i1.601.
- [19] A. M. Oktafani, I. T. Utami, and P. Kartikasari, "Optimasi backward elimination pada klasifikasi penyakit ispa menggunakan algoritma naïve bayes classifier 1,2,3," *J. Gaussian*, vol. 14, pp. 23–30, 2025, doi: 10.14710/j.gauss.14.1.23-30.
- [20] M. A. Wiratama and W. M. Pradnya, "Optimasi Algoritma Data Mining Menggunakan Backward Elimination untuk Klasifikasi Penyakit Diabetes," *J. Nas. Pendidik. Tek. Inform.*, vol. 11, no. 1, p. 1, 2022, doi: 10.23887/janapati.v11i1.45282.
- [21] A. Desiani *et al.*, "Perbandingan Klasifikasi Penyakit Kanker Paru-Paru menggunakan Support Vector Machine dan K-Nearest Neighbor," *J. Process.*, vol. 18, no. 1, pp. 54–62, 2023, doi: 10.33998/processor.2023.18.1.700.
- [22] F. Latuconsina and M. S. N. Van Delsen, "Klasifikasi Menggunakan Metode Support Vector Machine ( SVM ) Multiclass pada Data Indeks Desa Membangun ( IDM ) di Provinsi Maluku," *J. Math. adn Stat.*, vol. 7, no. 2, pp. 380–395, 2024.
- [23] R. Resmiati and T. Arifin, "Klasifikasi Pasien Kanker Payudara Menggunakan Metode Support Vector Machine dengan Backward Elimination," *Sistemasi*, vol. 10, no. 2, p. 381, 2021, doi: 10.32520/stmsi.v10i2.1238.
- [24] A. F. Azmi, "Prediksi Churn Nasabah Bank Menggunakan Klasifikasi Random Forest Dan Decision Tree Dengan Evaluasi Confusion Matrix," *Komputa J. Ilm. Komput. dan Inform.*, vol. 13, no. 1, pp. 111–119, 2024, [Online]. Available: <https://ojs.unikom.ac.id/index.php/komputa/article/view/12639>