

Analisis Klasifikasi Tingkat Infeksi Saluran Pernapasan Akut di Puskesmas Setiabudi dengan Metode *Random Forest*

Intania Dharma Hartarti¹, Fuad Nur Hasan², Lestari Yusuf³

^{1,2,3} Universitas Bina Sarana Informatika, Jl. Kramat Raya No.98, Kwitang, Kota Jakarta Pusat, DKI Jakarta, 10450

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 17-07-2026

Revisi Akhir: 06-08-2026

Diterbitkan Online: 25-09-2026

KATA KUNCI

ISPA

Random Forest

Classification

K-Fold Cross Validation

Machine Learning

KORESPONDENSI

No HP: 089512522985

E-mail: intaniadharmaa@gmail.com

ABSTRACT

Acute Respiratory Infection (ARI) is an infectious disease affecting the respiratory tract that requires appropriate treatment according to its severity. This study aims to implement the Random Forest algorithm to classify ARI severity based on patients' clinical symptoms at Setiabudi Community Health Center and evaluate its performance using K-Fold Cross Validation. The dataset consisted of 3,736 secondary medical records from 2025 with 14 variables, including 13 clinical features (age, respiratory rate, cough, runny nose, fever, sore throat, nausea and vomiting, diarrhea, contact with ARI patients, body aches, lump, history of asthma, and test results) and one target variable representing ARI severity (Mild ARI and Severe ARI). The research process included data preprocessing, data encoding, feature selection, Random Forest model development, and performance evaluation using Accuracy, Precision, Recall, F1-Score, Classification Report, and Confusion Matrix. Precision, Recall, and F1-Score were calculated using the weighted average method due to class imbalance. The model achieved an Accuracy of 95.45%, Precision of 93.57%, Recall of 95.45%, and F1-Score of 94.08%, exceeding the research target of 80% accuracy. These results indicate that Random Forest combined with K-Fold Cross Validation is effective for classifying ARI severity.

1. PENDAHULUAN

Salah satu penyakit infeksi yang masih memerlukan perhatian serius di bidang kesehatan masyarakat Indonesia adalah Infeksi Saluran Pernapasan Akut (ISPA) [1]. Bagian saluran pernapasan atas, seperti hidung, faring, tenggorokan, dan sinus, merupakan area yang terdampak oleh penyakit ini. Kejadiannya umumnya berkaitan dengan infeksi virus, bakteri, serta paparan lingkungan tertentu [2]. Tingginya angka kejadian ISPA menyebabkan penyakit ini berkontribusi terhadap tingginya jumlah pasien yang datang ke fasilitas pelayanan kesehatan tingkat pertama, termasuk puskesmas [3]. Merujuk pada data resmi Dinas Kesehatan DKI Jakarta, selama bulan Januari – Oktober 2025 tercatat sebanyak 1.966.308 kasus ISPA [4]. Sehingga menunjukkan bahwa penyakit ini masih menjadi beban pelayanan kesehatan, terutama pada kawasan perkotaan yang memiliki populasi padat dan paparan polusi udara yang cukup besar [5].

Di Puskesmas Setiabudi, data rekam medis pasien ISPA tersimpan dalam jumlah yang besar dan memuat berbagai

informasi, seperti umur, gejala klinis, riwayat penyakit, dan hasil pemeriksaan [6]. Namun, data tersebut umumnya masih dimanfaatkan sebagai dokumentasi pelayanan kesehatan sehingga belum dimanfaatkan secara optimal sebagai sumber informasi untuk mendukung proses klasifikasi tingkat keparahan penyakit [7]. Padahal, pemanfaatan data rekam medis melalui pendekatan *machine learning* dapat membantu menghasilkan klasifikasi yang lebih cepat, konsisten, dan objektif berdasarkan pola data yang dimiliki [8].

Penggunaan *machine learning* pada bidang kesehatan semakin berkembang, terutama untuk mendukung proses identifikasi dan klasifikasi berbagai penyakit [9]. Salah satu algoritma yang banyak diterapkan adalah *Random Forest*, karena mampu menggabungkan banyak *decision tree*, memiliki ketahanan terhadap *overfitting*, dirancang untuk mengolah data dengan jumlah atribut yang besar sehingga mampu memberikan hasil klasifikasi yang konsisten pada berbagai jenis data medis [10]. Di samping itu, penerapan K-Fold Cross Validation mampu memberikan proses evaluasi yang lebih objektif karena seluruh

data memperoleh kesempatan menjadi data latih maupun data uji secara bergantian [11].

Beberapa penelitian terdahulu membuktikan bahwa algoritma *Random Forest* memiliki tingkat efektivitas yang tinggi dalam melakukan klasifikasi. Penelitian Miftakhudin menerapkan *Random Forest* pada klasifikasi penyakit jantung coroner, walaupun akurasi mencapai 84%, namun jumlah variabel yang dianalisis masih terbatas [12]. Penelitian oleh Ramadhani menggunakan *Random Forest* untuk klasifikasi risiko gizi buruk pada ibu hamil, tetapi hanya memperoleh akurasi 67% [13]. Sementara itu, penelitian mengenai klasifikasi ISPA masih relatif sedikit dan umumnya menggunakan algoritma lain, misalnya pada penelitian Ani, menerapkan algoritma *Naive Bayes* pada klasifikasi ISPA dengan akurasi sebesar 75% [14].

Inovasi utama penelitian ini terletak pada penerapan algoritma *Random Forest* untuk mengklasifikasikan tingkat keparahan Infeksi Saluran Pernapasan Akut (ISPA) menggunakan data rekam medis pasien secara langsung di Puskesmas Setiabudi, yang masih relatif sedikit dikaji dalam penelitian sebelumnya. Penelitian ini menggunakan dataset tahun 2025 yang terdiri atas 14 variabel, yaitu 13 variabel gejala klinis sebagai fitur dan 1 variabel label berupa tingkat keparahan ISPA. Penggunaan data rekam medis dengan variabel gejala klinis yang lebih lengkap diharapkan mampu menghasilkan model klasifikasi yang lebih representatif terhadap kondisi nyata dibandingkan penelitian terdahulu yang menggunakan jumlah variabel lebih terbatas. Selain itu, kinerja model diuji menggunakan metode *K-Fold Cross Validation* berdasarkan nilai *Accuracy*, *Precision*, *Recall*, dan *F1-Score*.

Melalui pendekatan tersebut, penelitian ini diharapkan mampu menghasilkan model klasifikasi dengan tingkat *Accuracy* di atas 80% serta memiliki keandalan yang baik dalam mengklasifikasikan tingkat keparahan ISPA. Temuan yang diperoleh dari penelitian ini diharapkan dapat mendukung upaya dalam pemanfaatan data rekam medis ISPA di fasilitas pelayanan kesehatan tingkat pertama serta membantu tenaga kesehatan dalam menentukan prioritas penanganan pasien secara lebih cepat dan tepat.

2. TINJAUAN PUSTAKA

2.1 Infeksi Saluran Pernapasan Akut (ISPA)

World Health Organization (WHO) memaparkan bahwa Infeksi Saluran Pernapasan Akut (ISPA) sebagai penyakit infeksi yang menyerang saluran pernapasan atas dan bawah serta bersifat menular. Berdasarkan klasifikasi diagnosis ICD-10 (J00–J06), ISPA memiliki tingkat keparahan yang berkisar dari ringan hingga berat. ISPA saluran pernapasan atas meliputi *rhinitis*, *sinusitis*, *faringitis*, dan *tonsilitis*, sedangkan ISPA saluran bawah meliputi *bronkitis*, *bronkiolitis*, dan *pneumonia* [15]. Gejala yang umum muncul antara lain demam, batuk, hidung tersumbat, sakit tenggorokan, sakit kepala, dan nyeri otot. Diagnosis dilakukan melalui anamnesis, pemeriksaan fisik, serta pemeriksaan penunjang seperti *Polymerase Chain Reaction (PCR)* dan kultur mikrobiologi. Terjadinya penyakit ini dipengaruhi oleh infeksi yang berasal dari virus, bakteri, jamur, maupun parasit, dengan penularan melalui *droplet* atau kontak dengan permukaan yang

terkontaminasi. Selain itu, faktor lingkungan, kualitas udara, kepadatan hunian, usia, status gizi, kebiasaan merokok, dan daya tahan tubuh turut memengaruhi penyebaran serta tingkat keparahan ISPA [16].

2.2 Gejala ISPA

Gejala klinis ISPA diperoleh melalui anamnesis, pemeriksaan fisik, dan pemeriksaan laboratorium bila diperlukan [17]. Gejala yang umum ditemukan meliputi sesak napas yang diukur berdasarkan frekuensi pernapasan sesuai usia dan menjadi salah satu indikator pneumonia apabila melebihi 30 kali per menit pada orang dewasa [18], batuk [19], pilek [20], demam dengan suhu tubuh $\geq 38^{\circ}\text{C}$ [21], sakit tenggorokan, mual muntah, diare [22], nyeri tubuh, serta benjolan berupa pembesaran kelenjar getah bening *servikal*. Selain gejala klinis, terdapat faktor pendukung yang memengaruhi kejadian maupun tingkat keparahan ISPA, yaitu riwayat kontak dengan penderita ISPA, riwayat asma, serta hasil pemeriksaan penunjang seperti foto *toraks*, *Polymerase Chain Reaction (PCR)*, *kultur sputum*, atau pemeriksaan laboratorium lainnya untuk membantu menegakkan diagnosis, terutama pada kasus pneumonia atau ISPA berat [23].

2.3 Klasifikasi ISPA

Tingkat keparahan Infeksi Saluran Pernapasan Akut (ISPA) terbagi menjadi dua klasifikasi, yaitu ISPA Ringan dan ISPA Berat [24].

1. ISPA Ringan, ditandai dengan batuk atau kesulitan bernapas disertai napas cepat tanpa tanda bahaya umum, sehingga pasien masih mampu makan atau minum, tidak mengalami penurunan kesadaran, kejang, maupun tanda kegawatan lainnya, dan umumnya dapat ditangani melalui rawat jalan dengan antibiotik oral serta pemantauan berkala.
2. ISPA Berat, merupakan pneumonia yang disertai satu atau lebih tanda bahaya umum, seperti tidak mampu minum atau menyusu, muntah terus-menerus, kejang, penurunan kesadaran atau letargi, sianosis, serta gangguan pernapasan berat, sehingga memerlukan penanganan segera di rumah sakit, terapi antibiotik parenteral, dan perawatan intensif.

2.4 Machine Learning

Machine Learning termasuk salah satu disiplin ilmu dalam kecerdasan artifisial yang menggunakan algoritma pembelajaran untuk memperoleh pengetahuan data, mengenali pola, serta menghasilkan keputusan atau prediksi secara otomatis tanpa memerlukan banyak intervensi dari manusia. *Machine Learning* didukung oleh bidang matematika, statistika, dan *data mining*, serta bekerja dengan menganalisis data dalam jumlah besar untuk menemukan pola yang dapat meningkatkan akurasi prediksi seiring bertambahnya data yang diproses. Teknologi ini telah banyak diterapkan dalam berbagai bidang, seperti asisten digital, sistem rekomendasi produk, dan layanan hiburan digital [25].

2.5 Data mining

Dalam buku karya Dikananda, *data mining* merupakan proses pengolahan dan analisis kumpulan data berukuran besar untuk mengidentifikasi pola, keterkaitan, serta pengetahuan baru yang

belum teridentifikasi sebelumnya. Pramudiono menjelaskan bahwa *data mining* digunakan untuk menganalisis data secara otomatis guna menemukan pola dan kecenderungan penting, sedangkan Daniel T. Larose mendefinisikannya sebagai proses analisis data untuk mengungkap hubungan tersembunyi serta menyajikan informasi yang relevan sebagai landasan dalam pengambilan keputusan. Proses *data mining* menggabungkan berbagai teknik, seperti *Machine Learning*, statistika, pengenalan pola, basis data, dan visualisasi data, sehingga mampu menghasilkan pengetahuan atau informasi baru dari data yang diolah [26].

2.6 K-Fold Cross validation

K-Fold Cross Validation merupakan metode evaluasi model *Machine Learning* yang membagi dataset menjadi beberapa *fold*, di mana setiap *fold* digunakan secara bergantian sebagai data *training* dan *testing* hingga seluruh data memperoleh kesempatan menjadi data uji. Metode ini bertujuan menghasilkan evaluasi model yang lebih stabil serta mengurangi bias akibat pembagian data [27].

2.7 Klasifikasi

Metode klasifikasi digunakan untuk mengategorikan data ke dalam kelas-kelas tertentu dengan memanfaatkan pola yang dibangun dari data pelatihan (*training data*). Tahapan dalam proses klasifikasi meliputi dua fase, yaitu *learning* untuk membangun model menggunakan algoritma, dan *classification* untuk menguji model menggunakan data *testing* sehingga dapat memprediksi label data baru serta mendukung pengambilan keputusan [28].

2.8 Random Forest

Menurut Supriyadi, Dalam metode Supervised Learning, Random Forest (RF) merupakan algoritma yang digunakan untuk membangun model klasifikasi maupun regresi dengan membangun banyak *decision tree* dari sampel data yang dipilih secara acak, kemudian menggabungkan hasil prediksi setiap pohon menggunakan metode *majority voting* untuk memperoleh prediksi akhir. Semakin banyak pohon yang digunakan, performa model cenderung semakin baik dan stabil. Selain memiliki akurasi yang tinggi, *Random Forest* juga mampu menentukan tingkat kepentingan fitur (*feature importance*) sehingga banyak diterapkan dalam berbagai bidang [29]. Secara matematis, prediksi *Random Forest* dinyatakan dengan rumus:

$$H(x) = \text{mode}\{h_1(x), h_2(x), \dots, h_n(x)\} \quad (1)$$

Di mana $H(x)$ merupakan prediksi akhir, sedangkan $h_i(x)$ hingga $h_n(x)$ adalah prediksi dari setiap *decision tree* yang digabungkan melalui *majority voting* [30].

2.9 Python

Menurut buku *Machine Learning with Python: Theory and Applications* karya Gui-Rong Liu, dikembangkan oleh Guido van Rossum pada tahun 1991, *Python* merupakan salah satu bahasa pemrograman yang bersifat *open source* dengan filosofi yang menekankan keterbacaan kode (*readability*). *Python* menerapkan pendekatan *object-oriented programming*, sehingga mampu

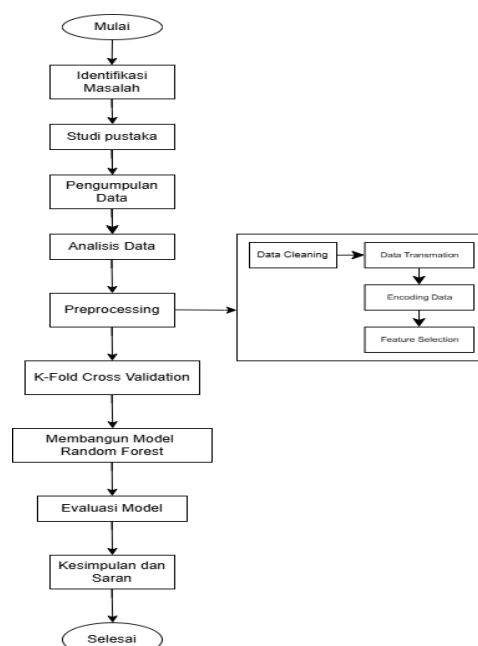
mendukung pengembangan program yang terstruktur, sederhana, dan efisien. Dalam bidang *Machine Learning*, *Python* banyak digunakan karena memiliki sintaks yang mudah dipahami, bersifat fleksibel, didukung berbagai pustaka untuk pengolahan data, analisis, dan pembangunan model, serta dapat dijalankan pada berbagai sistem operasi dengan pengembangan yang terus dilakukan oleh komunitas melalui *CPython* yang dikelola oleh *Python Software Foundation* [31].

2.10 Google Colaboratory

Menurut Wardana, *Google Colaboratory (Google Colab)* merupakan lingkungan kerja berbasis *cloud* yang disediakan oleh *Google* yang memungkinkan pengguna menjalankan bahasa pemrograman *Python* melalui *browser* dengan konsep yang serupa dengan *Jupyter Notebook*. Platform ini memungkinkan pengguna menulis, menjalankan, menyimpan, dan membagikan kode secara online, serta banyak dimanfaatkan dalam bidang *data science* dan *Machine Learning* karena menyediakan sumber daya komputasi *virtual*, terintegrasi dengan *Google Drive*, dan mudah diakses tanpa memerlukan perangkat berspesifikasi tinggi. Meskipun demikian, versi gratis *Google Colab* memiliki keterbatasan, seperti batas waktu penggunaan *notebook* dan penghentian mesin *virtual* saat tidak aktif [32].

3. METODOLOGI

Metode kuantitatif dipilih dalam penelitian ini mengingat data yang dianalisis berupa rekam medis pasien ISPA, seperti usia, gejala klinis, dan hasil laboratorium, dikonversi ke bentuk numerik untuk dianalisis secara statistik dan diolah menggunakan algoritma *Random Forest* dalam proses klasifikasi. Pendekatan ini bertujuan mendeskripsikan data serta membangun model klasifikasi secara objektif dan terukur [33].



Gambar 1. Alur Penelitian

Langkah penelitian dijelaskan sebagai berikut:

3.1 Mulai

Diawali dengan observasi di Puskesmas Setiabudi untuk memahami klasifikasi tingkat ISPA berdasarkan gejala klinis serta memperoleh data melalui dokumentasi rekam medis [34].

3.2 Identifikasi Masalah

Mengidentifikasi permasalahan klasifikasi tingkat ISPA dan menetapkan tujuan penelitian, yaitu membangun model klasifikasi menggunakan algoritma *Random Forest*.

3.3 Studi Pustaka

Mengkaji teori terkait ISPA, *data mining*, *Random Forest*, dan *K-Fold Cross Validation* dari jurnal, buku, serta sumber ilmiah.

3.4 Pengumpulan Data

Menggunakan data sekunder rekam medis pasien ISPA tahun 2025 yang diperoleh dari Puskesmas Setiabudi melalui proses perizinan penelitian [35].

3.5 Analisis Data

Data dipilih menggunakan teknik *purposive sampling*, sehingga hanya data yang memenuhi kriteria penelitian yang digunakan sebagai dataset [36].

3.6 Preprocessing Data

Data dipersiapkan sebelum pemodelan melalui *data cleaning*, *data transformation*, *encoding data*, dan *feature selection* agar siap digunakan dalam proses klasifikasi [37].

3.7 Penerapan K-Fold Cross Validation

Dataset dibagi menjadi 10 fold, kemudian dilakukan proses *training* dan *testing* secara bergantian pada setiap *fold* untuk memperoleh hasil evaluasi yang lebih stabil dan mengurangi bias [38].

3.8 Pembangunan Model Random Forest

Model dibangun menggunakan algoritma *Random Forest* yang menggabungkan banyak *decision tree* melalui *majority voting*. Prosesnya meliputi pemilihan fitur secara acak, pembentukan pohon keputusan, pemisahan *node* menggunakan kriteria terbaik, dan penggabungan hasil prediksi seluruh pohon [39].

3.9 Evaluasi Model

Performa model dievaluasi menggunakan *Confusion Matrix*, *Accuracy*, *Precision*, *Recall*, dan *F1-Score* untuk mengukur ketepatan serta keandalan klasifikasi [40].

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

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (5)$$

3.10 Kesimpulan dan Saran

Merumuskan kesimpulan berdasarkan temuan penelitian serta menyampaikan rekomendasi sebagai bahan pengembangan penelitian pada masa mendatang.

3.11 Selesai

Seluruh tahapan penelitian telah selesai dilaksanakan.

4. HASIL DAN PEMBAHASAN

4.1 Pengumpulan Dataset

Seluruh proses pengolahan dan analisis data dilakukan secara sistematis menggunakan bahasa pemrograman *Python* pada platform *Google Colab*. Dataset pasien ISPA tahun 2025 yang diperoleh dari Puskesmas Setiabudi tersedia dalam format *Microsoft Excel (.xlsx)*.

Tabel 1. Deskripsi Dataset

Variabel Data	Deskripsi
Nomor	Nomor Urut sebagai pengganti No. RM atau Nama Pasien
Umur	Usia pasien dalam satuan tahun.
Sesak Napas (RR)	Menunjukkan ada atau tidaknya gejala sesak napas (Ya/No).
Batuk	Menunjukkan ada atau tidaknya gejala batuk (Ya/No).
Pilek	Menunjukkan ada atau tidaknya gejala pilek (Ya/No).
Demam	Menunjukkan ada atau tidaknya gejala demam (Ya/No).
Sakit Tenggorokan	Menunjukkan ada atau tidaknya keluhan sakit tenggorokan (Ya/No).
Mual Muntah	Menunjukkan ada atau tidaknya gejala mual dan muntah (Ya/No).
Diare	Menunjukkan ada atau tidaknya gejala diare (Ya/No).
Kontak dengan Penderita ISPA	Riwayat kontak pasien dengan penderita ISPA (Ya/No).
Nyeri Tubuh	Menunjukkan ada atau tidaknya keluhan nyeri tubuh (Ya/No).
Benjolan	Menunjukkan ada atau tidaknya benjolan pada pasien (Ya/No).
Riwayat Asma	Menunjukkan apakah pasien memiliki riwayat penyakit asma (Ya/No).
Hasil Tes	Hasil pemeriksaan laboratorium/tes penunjang pasien (Pneumonia/No).
Keterangan (Label)	Tingkat infeksi ISPA yang terdiri dari ISPA Ringan dan ISPA Berat, sebagai kelas yang diprediksi oleh algoritma <i>Random Forest</i> .

Selanjutnya menampilkan 5 baris data pertama dari dataset dengan perintah `print (data.head())`.

5 baris data awal

No	Umur	Sesak Napas (RR)	Batuk Pilek	Demam	Sakit Tenggorokan	Mual Muntah	Diare	Kontak dengan penderita ISPA	Nyeri Tubuh	Benjolan	Riwayat Asma	Hasil Tes	Keterangan
0	1	6	Ya	Ya	No	No	No	No	No	No	No	No	ISPA Berat
1	2	50	Ya	No	No	No	No	No	No	No	No	No	ISPA Ringan
2	3	25	Ya	No	No	No	No	No	No	No	No	No	ISPA Ringan
3	4	27	No	Ya	Ya	Ya	Ya	Ya	Ya	Ya	Ya	Ya	ISPA Berat
4	5	43	Ya	No	No	Ya	No	No	No	No	No	No	ISPA Ringan

Diare Kontak dengan penderita ISPA Nyeri Tubuh Benjolan Riwayat Asma \

0	No	No	Ya	No	No	No	No	No	No	No	No	No	No
1	No	No	No	Ya	No	No	No	No	No	No	No	No	No
2	No	No	No	Ya	No	No	No	No	No	No	No	No	No
3	Ya	Ya	No	No	Ya	Ya	Ya	Ya	Ya	Ya	Ya	Ya	Ya
4	No	No	No	Ya	No	No	No	No	No	No	No	No	No

Hasil Tes Keterangan

0	No	ISPA Berat
1	No	ISPA Ringan
2	No	ISPA Ringan
3	Pneumonia	ISPA Berat
4	No	ISPA Ringan

Gambar 2. Hasil Menampilkan 5 Dataset Awal

4.2 Preprocessing

Mengidentifikasi *data cleaning* dengan:

1. Cek *missing value* dengan *data.isnull().sum()*. Hasil pemeriksaan menunjukkan bahwa seluruh variabel tidak memiliki nilai kosong.
2. Cek data duplikat dengan *jumlah_duplikat = data.duplicated().sum()*. Hasil pengecekan menjelaskan bahwa jumlah data duplikat pada dataset adalah 0, yang berarti tidak terdapat baris data yang memiliki nilai identik.
3. Cek tipe data dengan print (*data.dtypes*)

No	int64
Umur	int64
Sesak Napas (RR)	object
Batuk	object
Pilek	object
Demam	object
Sakit Tenggorokan	object
Mual Muntah	object
Diare	object
Kontak dengan penderita ISPA	object
Nyeri Tubuh	object
Benjolan	object
Riwayat Asma	object
Hasil Tes	object
Keterangan	object
dtype:	object

Gambar 3. Hasil cek tipe data

Selanjutnya, mengidentifikasi *Data Transformation* untuk melakukan pengkodean data kategorikal ke bentuk numerik sehingga sesuai dengan kebutuhan algoritma *Random Forest*. Proses diawali dengan mengimpor kelas *LabelEncoder* dari library *scikit-learn*, kemudian membuat objek *LabelEncoder* yang digunakan untuk melakukan transformasi data. Selanjutnya, program melakukan perulangan pada seluruh kolom yang bertipe *object* (teks) dan mengubah setiap nilai kategorikal menjadi nilai numerik. Kolom Keterangan tidak disertakan pada tahap ini karena merupakan variabel label yang akan diproses secara terpisah pada tahap *encoding data*. Setelah proses transformasi selesai, fungsi *print(data.dtypes)* digunakan untuk memastikan bahwa tipe data setiap kolom telah berubah sesuai kebutuhan, sedangkan *data.head()* digunakan untuk melihat lima record pertama dataset sebagai verifikasi hasil transformasi.

No	int64
Umur	int64
Sesak Napas (RR)	int64
Batuk	int64
Pilek	int64
Demam	int64
Sakit Tenggorokan	int64
Mual Muntah	int64
Diare	int64
Kontak dengan penderita ISPA	int64
Nyeri Tubuh	int64
Benjolan	int64
Riwayat Asma	int64
Hasil Tes	int64
Keterangan	object
dtype:	object

Gambar 4. Hasil Mengubah Tipe Data

No	Umur	Sesak Napas (RR)	Batuk	Pilek	Demam	Sakit Tenggorokan	Mual Muntah	Diare	Kontak dengan penderita ISPA	Nyeri Tubuh	Benjolan	Riwayat Asma	Hasil Tes	Keterangan
0	1	6	1	1	0	0	0	0	0	1	0	0	0	ISPA Ringan
1	2	50	1	0	0	0	0	0	0	1	0	0	0	ISPA Ringan
2	3	25	1	0	0	0	0	1	0	0	1	0	0	ISPA Ringan
3	4	27	0	1	1	1	1	0	1	1	0	1	1	ISPA Berat
4	5	43	1	0	0	1	0	0	0	1	0	0	0	ISPA Ringan

Gambar 5. Hasil Transformasi Data

Selanjutnya, tahap *encoding data* dilakukan untuk mengonversi variabel target Keterangan menjadi bentuk numerik menggunakan *LabelEncoder*. Fungsi *fit_transform()* mengubah label kategorikal menjadi nilai numerik yang disimpan pada variabel *y*, sedangkan *classes_* dan *enumerate()* digunakan untuk menampilkan pemetaan kode kelas. Pada penelitian ini diperoleh pemetaan 0 = ISPA Berat dan 1 = ISPA Ringan, yang ditentukan secara otomatis berdasarkan urutan alfabet. Kode numerik tersebut hanya berfungsi sebagai representasi kelas sehingga tidak memengaruhi proses pelatihan maupun hasil evaluasi model *Random Forest*.

Terakhir, *Feature Selection* dilakukan untuk menentukan variabel *feature (X)* dan *label (y)* pada proses klasifikasi. Sebanyak 13 variabel gejala klinis, yaitu Umur, Sesak Napas (RR), Batuk, Pilek, Demam, Sakit Tenggorokan, Mual Muntah, Diare, Kontak dengan Penderita ISPA, Nyeri Tubuh, Benjolan, Riwayat Asma, dan Hasil Tes digunakan sebagai *feature*, sedangkan variabel Keterangan digunakan sebagai label yang telah diubah ke bentuk numerik menggunakan *LabelEncoder*. Selanjutnya, program menampilkan jumlah dan distribusi kelas untuk mengetahui karakteristik dataset sebelum pelatihan model. Hasilnya menunjukkan dataset terdiri atas 3.736 data, yaitu 162 data ISPA Berat (4,34%) dan 3.574 data ISPA Ringan (95,66%), sehingga termasuk *imbalanced dataset*.

4.3 Penerapan K-Fold Cross validation

Berdasarkan hasil analisis distribusi kelas, dataset penelitian termasuk *imbalanced dataset*, sehingga digunakan *Stratified K-Fold Cross Validation* untuk mempertahankan ratio setiap kelas pada setiap fold agar hasil evaluasi lebih representatif. Pada penelitian ini digunakan *10 fold (n_splits=10)*, sehingga seluruh data memperoleh kesempatan menjadi data latih dan data uji secara bergantian. Parameter *shuffle=True* digunakan untuk menyusun ulang urutan data sebelum proses pembentukan *fold* sehingga penyebaran data pada setiap *fold* menjadi lebih merata. sedangkan *random_state=42* digunakan untuk menjaga konsistensi pembagian data sehingga hasilnya dapat direplikasi. Tahap ini hanya menginisialisasi konfigurasi validasi silang dan belum melakukan tahapan penyusunan model *Random Forest* dan pengujian kinerjanya.

4.4 Membangun model Random Forest

Pada tahap ini dibangun model *Random Forest* menggunakan *RandomForestClassifier()* dengan *100 decision tree (n_estimators=100)*. Jumlah tersebut merupakan nilai *default* pada *Scikit-learn* dan mampu menghasilkan model yang stabil dengan waktu komputasi yang efisien. Parameter *random_state=42* digunakan agar proses pembentukan model,

termasuk *bootstrap sampling* dan pemilihan fitur secara acak, menghasilkan pembagian yang konsisten setiap kali program dijalankan. Model kemudian dilatih dan diuji menggunakan *Stratified 10-Fold Cross Validation*. Pada setiap putaran validasi, sembilan fold dimanfaatkan sebagai data pelatihan, sedangkan satu fold lainnya digunakan sebagai data pengujian secara bergiliran. Nilai *Accuracy* setiap *fold* disimpan pada variabel *Accuracy_scores*, sedangkan seluruh data aktual (*all y_test*) dan hasil prediksi (*all y_pred*) dikumpulkan sebagai dasar pembentukan *Confusion Matrix*, *Classification Report*, serta perhitungan *Accuracy*, *Precision*, *Recall*, dan *F1-Score* pada tahap evaluasi.

```

=== HASIL 10-FOLD CROSS VALIDATION ===
Fold-1: 0.9599
Fold-2: 0.9519
Fold-3: 0.9545
Fold-4: 0.9599
Fold-5: 0.9519
Fold-6: 0.9572
Fold-7: 0.9491
Fold-8: 0.9571
Fold-9: 0.9517
Fold-10: 0.9517

```

Gambar 6. Hasil K-Fold Cross Validation

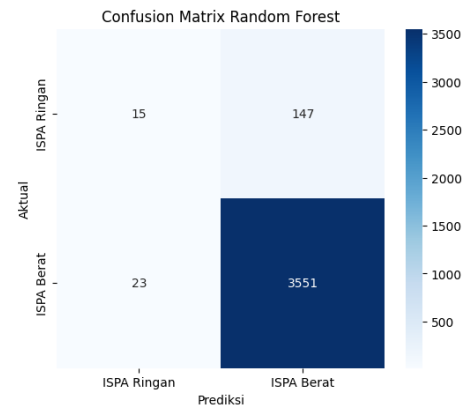
Hasil *10-Fold Cross Validation* menunjukkan bahwa model *Random Forest* menghasilkan nilai *Accuracy* yang relatif konsisten pada setiap fold. Nilai akurasi tertinggi diperoleh pada *Fold-1* dan *Fold-4* sebesar 95,99%, sedangkan nilai terendah terdapat pada *Fold-7* sebesar 94,91%. Perbedaan nilai tersebut wajar karena setiap iterasi menggunakan kombinasi data latih dan data uji yang berbeda. Namun, selisih akurasi yang kecil mengindikasikan bahwa model memiliki kinerja yang konsisten serta mampu menghasilkan hasil klasifikasi yang andal pada setiap proses validasi. Nilai akurasi dari seluruh fold kemudian dirata-ratakan untuk memperoleh performa model secara keseluruhan.

4.5 Evaluasi Model

1. Visualisasi Confusion Matrix

Confusion Matrix dibentuk menggunakan fungsi `confusion_matrix(all_y_test, all_y_pred)` dengan membandingkan data aktual dan hasil prediksi dari seluruh proses *10-Fold Cross Validation*. Hasilnya disimpan pada variabel *cm* dan digunakan untuk mengetahui jumlah prediksi yang benar maupun salah pada setiap kelas sebagai dasar evaluasi performa model.

Pada *Confusion Matrix*, nilai *TP*, *TN*, *FP*, dan *FN* bergantung pada kelas yang dijadikan sebagai kelas positif. Pada penelitian ini, hasil *LabelEncoder* menunjukkan 0 = ISPA Berat dan 1 = ISPA Ringan. Jika ISPA Berat dijadikan kelas positif, diperoleh *TP* = 15, *FN* = 147, *FP* = 23, dan *TN* = 3551. Sebaliknya, jika ISPA Ringan dijadikan kelas positif, diperoleh *TP* = 3551, *FN* = 23, *FP* = 147, dan *TN* = 15. Meskipun interpretasinya berbeda, nilai pada *Confusion Matrix* tetap sama karena hanya sudut pandang kelas positif yang berubah. Pendekatan ini juga digunakan pada *Classification Report* untuk menghitung *Precision*, *Recall*, dan *F1-Score* pada masing-masing kelas.



Gambar 7. Visualisasi Confusion Matrix Random Forest

Visualisasi *Confusion Matrix* dibuat menggunakan *Seaborn* dan *Matplotlib* dalam bentuk *heatmap*. Fungsi `sns.heatmap()` digunakan untuk menampilkan nilai *Confusion Matrix* dengan anotasi angka dan gradasi warna biru, sedangkan `plt.title()`, `plt.xlabel()`, dan `plt.ylabel()` digunakan untuk memberikan judul serta keterangan sumbu. Visualisasi bertujuan agar hasil evaluasi mudah dipahami dibandingkan hanya dalam bentuk tabel angka.

Berdasarkan visualisasi *Confusion Matrix*, model berhasil mengklasifikasikan 3566 dari 3736 data dengan benar, terdiri atas 15 data ISPA Berat dan 3551 data ISPA Ringan, sedangkan 170 data masih salah diklasifikasikan. Warna yang lebih gelap pada diagonal utama menunjukkan bahwa prediksi benar jauh lebih dominan dibandingkan kesalahan klasifikasi. Hasil ini menunjukkan bahwa model sangat baik dalam mengenali ISPA Ringan sebagai kelas mayoritas, namun kemampuan mengenali ISPA Berat masih lebih rendah karena jumlah datanya jauh lebih sedikit (162 data) dibandingkan ISPA Ringan (3574 data).

2. Hasil Evaluasi Model

Pada penelitian ini, metrik evaluasi yang digunakan meliputi *Accuracy*, *Precision*, *Recall*, dan *F1-Score* agar performa setiap kelas dapat dinilai secara lebih menyeluruh. Karena dataset memiliki distribusi kelas yang tidak seimbang (162 data ISPA Berat dan 3574 data ISPA Ringan), nilai *Precision*, *Recall*, dan *F1-Score* dihitung menggunakan metode *weighted average*, yaitu perhitungan yang mempertimbangkan jumlah data (*support*) pada setiap kelas sehingga lebih merepresentasikan performa model secara keseluruhan [41]. Hasil evaluasi disajikan dalam bentuk *Classification Report* yang menampilkan nilai *Precision*, *Recall*, *F1-Score*, *Support*, *Macro Average*, dan *Weighted Average*.

	precision	recall	f1-score	support
ISPA Berat	0.39	0.09	0.15	162
ISPA Ringan	0.96	0.99	0.98	3574
accuracy			0.95	3736
macro avg	0.68	0.54	0.56	3736
weighted avg	0.94	0.95	0.94	3736
Accuracy	: 95.45%			
Precision	: 93.57%			
Recall	: 95.45%			
F1-Score	: 94.08%			

Gambar 8. Hasil Evaluasi Model

Evaluasi dilakukan menggunakan fungsi `classification_report()`, `accuracy_score()`, `precision_score()`, `recall_score()`, dan

f1_score() dari *sklearn.metrics*. Fungsi *classification_report()* menampilkan nilai evaluasi untuk setiap kelas berdasarkan data aktual (*all_y_test*) dan hasil prediksi (*all_y_pred*), sedangkan *accuracy_score()*, *precision_score()*, *recall_score()*, dan *f1_score()* menghitung performa model secara keseluruhan. Pada perhitungan *Precision*, *Recall*, dan *F1-Score* digunakan parameter *average='weighted'*, kemudian hasilnya ditampilkan dalam bentuk persentase.

Tabel 2. Hasil Evaluasi Model *Random Forest*

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	0.9545	0.9357	0.9545	0.9408

Berdasarkan *Classification Report*, kelas ISPA Berat memperoleh *Precision* 39%, *Recall* 9%, dan *F1-Score* 15% dari 162 data, menunjukkan bahwa model masih sulit mengenali kelas minoritas. Sebaliknya, kelas ISPA Ringan memperoleh *Precision* 96%, *Recall* 99%, dan *F1-Score* 98% dari 3574 data, sehingga model mampu mengklasifikasikan kelas mayoritas dengan sangat baik. Karena dataset tidak seimbang, *Weighted Average* digunakan sebagai acuan utama dibandingkan *Macro Average*. Secara keseluruhan, model menghasilkan *Accuracy* 95,45%, *Precision* 93,57%, *Recall* 95,45%, dan *F1-Score* 94,08%, yang menunjukkan performa klasifikasi sangat baik meskipun kemampuan mengenali ISPA Berat masih lebih rendah akibat ketidakseimbangan distribusi data.

5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian mengenai analisis klasifikasi tingkat Infeksi Saluran Pernapasan Akut (ISPA) menggunakan algoritma *Random Forest* pada data pasien di Puskesmas Setiabudi, dapat disimpulkan bahwa penelitian ini berhasil membangun model klasifikasi melalui tahapan identifikasi masalah, studi pustaka, pengumpulan dan analisis data, *preprocessing* (*Data Cleaning*, *Data Transformation*, *Encoding Data*, dan *Feature Selection*), serta pembagian data menggunakan *K-Fold Cross Validation* sebelum proses pelatihan dan pengujian model. Model yang dibangun mampu mengklasifikasikan tingkat keparahan ISPA menjadi dua kategori, yaitu ISPA Ringan dan ISPA Berat, berdasarkan data gejala klinis pasien. Hasil evaluasi menunjukkan performa model yang sangat baik dengan *Accuracy* sebesar 95,45%, *Precision* sebesar 93,57%, *Recall* sebesar 95,45%, dan *F1-Score* sebesar 94,08%, yang dihitung menggunakan metode *weighted average* karena distribusi data tidak seimbang. Selain itu, hasil *Confusion Matrix* menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan benar, meskipun kemampuan model dalam mengenali kelas ISPA Berat masih lebih rendah dibandingkan ISPA Ringan akibat ketidakseimbangan jumlah data. Dengan demikian, algoritma *Random Forest* yang dikombinasikan dengan *K-Fold Cross Validation* terbukti efektif dalam mengklasifikasikan tingkat keparahan ISPA berdasarkan data gejala klinis pasien di Puskesmas Setiabudi serta berhasil melampaui target *Accuracy* penelitian, yaitu lebih dari 80%.

5.2 Saran

Temuan penelitian ini menjadi dasar dalam penyusunan beberapa saran yang dapat diterapkan pada penelitian selanjutnya. Penelitian berikutnya disarankan membandingkan algoritma *Random Forest* dengan algoritma klasifikasi lainnya, seperti *Support Vector Machine (SVM)*, *Decision Tree*, atau *Naïve Bayes*, untuk memperoleh model dengan performa yang lebih optimal. Selain itu, penambahan variabel klinis yang lebih beragam juga disarankan agar model mampu mengklasifikasikan tingkat keparahan ISPA dengan lebih akurat. Pengembangan model ke dalam bentuk aplikasi berbasis *web* atau *mobile* juga dapat dilakukan sehingga hasil klasifikasi dapat dimanfaatkan sebagai sistem pendukung keputusan yang membantu tenaga kesehatan dalam menentukan prioritas penanganan pasien secara lebih cepat dan tepat.

UCAPAN TERIMA KASIH

Dalam penyusunan penelitian ini, penulis memperoleh dukungan dan bantuan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan apresiasi dan rasa terima kasih kepada:

- Seluruh sivitas akademika Universitas Bina Sarana Informatika selama proses perkuliahan.
- Dosen Pembimbing yang telah meluangkan waktu memberikan arahan dengan baik.

DAFTAR PUSTAKA

- [1] W. Wahyudi, A. D. Shabrina, N. B. Kudadiri, N. Z. Komaruddin, and T. Nurlisa, "Faktor Risiko Gejala Penyakit Pernapasan Ispa pada Mahasiswa Fakultas Kesehatan Masyarakat UIN Sumatera Utara," *Jurnal Ilmiah Kedokteran dan Kesehatan*, vol. 4, no. 2, pp. 454–464, May 2025, doi: 10.55606/klinik.v4i2.4099.
- [2] D. Aprilianti, H. D. Sitanggang, M. Syukri, M. B. Butar, and A. Fitri, "DETERMINAN ISPA PADA BALITA DENGAN EKONOMI RENDAH DI INDONESIA (SKI 2023)," *Jurnal Kesehatan Tambusai*, vol. 6, no. 2, pp. 5882–5896, Jun. 2025, doi: 10.31004/jkt.v6i2.44224.
- [3] A. Y. Adi Saputra, T. A. Y. Siswa, and R. Hasudungan, "Implementasi Metode Backward Elimination Pada Algoritma SVM Untuk Klasifikasi Penyakit Ispa," *Jurnal Tekno Kompak*, vol. 20, no. 1, pp. 223–235, Nov. 2025, doi: 10.33365/jtk.v20i1.701.
- [4] L. W. Santosa, "Dinkes DKI catat 1,9 juta kasus ISPA hingga Oktober 2025," ANTARA. Accessed: Apr. 15, 2026. [Online]. Available: <https://www.antaranews.com/berita/5177593/dinkes-dki-catat-19-juta-kasus-isp-hingga-oktober-2025>
- [5] S. S. Mulyati, R. Y. Iriantoi, and N. Hidayah, "Faktor-faktor Risiko yang Berhubungan dengan Kejadian ISPA di Permukiman Sekitar Bandara," *Jurnal Kesehatan Lingkungan Indonesia*, vol. 23, no. 1, pp. 67–72, Feb. 2024, doi: 10.14710/jkli.23.1.67-72.
- [6] I. Romli, "PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA K-MEANS UNTUK KLASIFIKASI PENYAKIT ISPA," *Indonesian Journal of Business Intelligence (IJUBI)*,

- vol. 4, no. 1, p. 10, Jun. 2021, doi: 10.21927/ijubi.v4i1.1727.
- [7] A. Nur Fachmi, F. N. Hasan, A. P. Sari, F. R. Ramadhan, R. D. Patria, and S. Pudjiati, "ANALISIS KLASIFIKASI INDEKS STANDAR PENCEMARAN UDARA JAKARTA TAHUN 2025 MENGGUNAKAN ALGORITMA RANDOM FOREST," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 14, no. 1, Jan. 2026, doi: 10.23960/jitet.v14i1.8477.
- [8] R. K. Hapsari *et al.*, *DATA MINING*. Indie Press, 2024. Accessed: Apr. 10, 2026. [Online]. Available: <https://share.google/gsNjmMlcm9ktU5MKY>
- [9] Hidayat, Andi Sunyoto, and Hanif Al Fatta, "Klasifikasi Penyakit Jantung Menggunakan Random Forest Classifier," *Jurnal SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan)*, vol. 7, no. 1, pp. 31–40, Oct. 2023, doi: 10.47970/siskom-kb.v7i1.464.
- [10] E. W. Fidayanthie, A. Sayfulloh, M. R. Alzena, and N. K. Sari, "Analisis Perbandingan Algoritma Random Forest dan Algoritma Naive Bayes untuk Memprediksi Penyakit Paru-Paru di Indonesia," *Saturnus: Jurnal Teknologi dan Sistem Informasi*, vol. 3, no. 3, pp. 123–135, Jul. 2025, doi: 10.61132/saturnus.v3i3.956.
- [11] N. Blesyova and F. Nur Hasan, "ANALISIS SENTIMEN TERHADAP BEA CUKAI MENGGUNAKAN SUPPORT VECTOR MACHINE DAN K-FOLD CROSS VALIDATION," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 6, pp. 12051–12056, Nov. 2024, doi: 10.36040/jati.v8i6.11774.
- [12] A. Miftakhudin, N. A. Santoso, and B. A. Santoso, "Komparasi Algoritma KNN dan Random Forest untuk Diagnosa Penyakit Jantung Koroner (Studi Kasus: RSUD Dr. Soeselo Slawi)," *RIGGS: Journal of Artificial Intelligence and Digital Business*, vol. 4, no. 3, pp. 2962–2971, Aug. 2025, doi: 10.31004/riggs.v4i3.2424.
- [13] F. Ramadhani, D. Septiana, S. N. Amalia, P. M. Fadilah, and A. Satria, "KLASIFIKASI RISIKO GIZI BURUK PADA IBU HAMIL MENGGUNAKAN METODE RANDOM FOREST," *Djtechno: Jurnal Teknologi Informasi*, vol. 5, no. 2, pp. 370–380, Aug. 2024, doi: 10.46576/djtechno.v5i2.4815.
- [14] B. A. K. P. Ani, T. A. Y. Siswa, and F. Yulianto, "Implementasi Algoritma Naive Bayes Untuk Klasifikasi Penyakit Ispa," *Jurnal Tekno Kompak*, vol. 20, no. 1, pp. 236–249, Nov. 2025, doi: 10.33365/jtk.v20i1.700.
- [15] World Health Organization (WHO), "Infeksi saluran pernapasan atas akut (J00-J06)." Accessed: Apr. 16, 2026. [Online]. Available: <https://icd.who.int/browse10/2019/en#/X>
- [16] S. R. Pakadang, I. Adhayanti, St. Ratnah, and A. M. Salasa, *Buku Ajar Efektivitas Terapi Herbal untuk Infeksi Saluran Pernapasan Akut Mencegah, Mengobati dan Komplementer Pengobatan ISPA*, 1st ed. Nasmedia, 2025.
- [17] Direktorat Promosi Kesehatan dan Kesehatan Komunitas, "Mengenal Gejala ISPA dan Tindakan yang Perlu Dilakukan," *Kemendes*, 2023. Accessed: Jun. 29, 2026. [Online]. Available: <https://ayosehat.kemkes.go.id/mengenal-gejala-ispa-dan-tindakan-yang-perlu-dilakukan>
- [18] Ernstmeyer and Christman, *Nursing Skills*, 2nd edition. National Library of Medicine, 2023. Accessed: Jun. 29, 2026. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK596717/table/ch1survey.T.normal_respiratory_rate_by_a/
- [19] H. R. Amelia, S. Jeniawaty, U. Khasanah, and Sukesu, "Efektivitas Pijat Batuk Pilek dengan Lama Penyembuhan ISPA pada Balita," *Gema Bidan Indonesia*, vol. 13, no. 2, pp. 65–71, Sep. 2024, doi: 10.36568/gebindo.v13i2.209.
- [20] L. Hafidah, A. O. D. Denta, and S. Suraying, "PENANGANAN PERTAMA OLEH IBU PADA BALITA YANG MENGALAMI ISPA DI PONKESDES ANTIKA DI DESA WARU TIMUR WARU PAMEKASAN," *Jurnal Sains dan Teknologi Kesehatan*, vol. 4, no. 2, pp. 60–66, May 2024, doi: 10.52234/jstk.v4i2.231.
- [21] "PNPK 2023 - Tata Laksana Pneumonia Pada Dewasa," *Kemendes*. Accessed: Jun. 29, 2026. [Online]. Available: <https://www.kemkes.go.id/id/pnpk-2023---tata-laksana-pneumonia-pada-dewasa>
- [22] M. K. drg. Emmy Hastuti, "Pneumonia," *Kemendes*. Accessed: Jun. 29, 2026. [Online]. Available: https://keslan.kemkes.go.id/view_artikel/2709/pneumonia
- [23] R. M. Sihombing and I. H. Nasution, "Hubungan Foto Toraks Pasien Pneumonia Dengan Keparahana Klinis Berdasarkan Rekam Medis RS Royal Prima," *JUKEJ: Jurnal Kesehatan Jompa*, vol. 4, no. 3, pp. 1132–1140, Oct. 2025, doi: 10.57218/jkj.Vol4.Iss3.1990.
- [24] Yulia, Rendy, and S. A. Arnomo, "Jaringan Syaraf Tiruan Mendeteksi Penyakit Pneumonia Infeksi Saluran Pernafasan Akut Dengan Algoritma Backpropagation," *The Indonesian Journal of Computer Science*, vol. 12, no. 1, Feb. 2023, doi: 10.33022/ijcs.v12i1.3129.
- [25] E. I. Sela *et al.*, *APLIKASI MACHINE LEARNING PADA ILMU SOSIAL DAN MEDIS*, 1st ed. Deepublish, 2023.
- [26] A. R. Dikananda, Odi Nurdiawan, Ahmad Faqih, Yudhistira Arie Wijaya, Agus Surip, and Dias Bayu Saputra, *ALGORITMA REGRESI LINIER Untuk Prediksi Kemampuan Computational Thinking Pemain Game Dota*, 1st ed. Lakeisha, 2022.
- [27] D. K. Nisa, F. M. Hana, and S. Ulya, "Klasifikasi Kinerja Penjualan Produk Nike Menggunakan Algoritma Random Forest dengan Pendekatan Hold-Out dan K-Fold Cross Validation," *Sainteks*, vol. 23, no. 1, pp. 79–92, May 2026, doi: 10.30595/sainteks.v23i1.29930.
- [28] D. Jollyta, Prihandoko, Alyauma Hajjah, Elin Haerani, and Muhammad Siddik, *ALGORITMA KLASIFIKASI UNTUK PEMULA Solusi Python dan RapidMiner*, 1st ed. Deepublish, 2023.
- [29] E. Supriyadi, *MACHINE LEARNING: DASAR DAN PRAKTIS*, 1st ed. DEEPUBLISH, 2022.
- [30] Avriyan Putra Bintang Pratama and Wahyu Aji Eko Prabowo, "Analisis Perbandingan Model Machine Learning menggunakan Teknik Stratified K-Fold Cross Validation untuk Klasifikasi Penyakit Jantung,"

- JURIKOM (*Jurnal Riset Komputer*), vol. 13, no. 2, pp. 679–688, Apr. 2026, doi: 10.30865/jurikom.v13i2.9670.
- [31] Gui-Rong Liu, *Machine learning with Python: theory and applications*. World Scientific Publishing Co. Pte. Ltd, 2023.
- [32] Wardana, *DASAR-DASAR DATA SCIENCE DAN APLIKASINYA DENGAN PYTHON*, 1st ed. Wawasan Ilmu, 2024.
- [33] A. Bandur and H. Prabowo, *PENELITIAN KUANTITATIF Metodologi, Desain, dan Analisis Data Dengan-SPSS, AMOS & Nvivo*, 1st ed. Mitra Wacana Media, 2021.
- [34] Ardiansyah, Risnita, and M. S. Jailani, “Teknik Pengumpulan Data Dan Instrumen Penelitian Ilmiah Pendidikan Pada Pendekatan Kualitatif dan Kuantitatif,” *Jurnal IHSAN : Jurnal Pendidikan Islam*, vol. 1, no. 2, pp. 1–9, Jul. 2023, doi: 10.61104/ihsan.v1i2.57.
- [35] D. N. R. Mastuti, Y. Indriyani, S. N. Anfalla, and S. A. Fazaira, “EVALUASI EFEKTIVITAS PROGRAM PENGENDALIAN ISPA DI PUSKESMAS TIRTO II KABUPATEN PEKALONGAN TAHUN 2024,” *Al-Tamimi Kesmas: Jurnal Ilmu Kesehatan Masyarakat (Journal of Public Health Sciences)*, vol. 14, no. 2, pp. 263–271, Dec. 2025, doi: 10.35328/kesmas.v14i2.3008.
- [36] P. G. Subhaktiyasa, “Menentukan Populasi dan Sampel: Pendekatan Metodologi Penelitian Kuantitatif dan Kualitatif,” *Jurnal Ilmiah Profesi Pendidikan*, vol. 9, no. 4, pp. 2721–2731, Nov. 2024, doi: 10.29303/jipp.v9i4.2657.
- [37] D. Mualfah, W. Fadila, and R. Firdaus, “Teknik SMOTE untuk Mengatasi Imbalance Data pada Deteksi Penyakit Stroke Menggunakan Algoritma Random Forest,” *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 2, pp. 107–113, Aug. 2022, doi: 10.37859/coscitech.v3i2.3912.
- [38] M. Jamil, F. Rozi, and Y. F. Saputra, “Komparasi Kinerja Algoritma Machine Learning Untuk Deteksi Penyakit Infeksi Saluran Pernapasan,” *Jurnal Rekayasa Teknologi Informasi (JURTI)*, vol. 8, no. 1, p. 84, Jun. 2024, doi: 10.30872/jurti.v8i1.16965.
- [39] A. M. A. Rahim, Ingrid Yanuar Risca Pratiwi, and Muhammad Ainul Fikri, “Klasifikasi Penyakit Jantung Menggunakan Metode Synthetic Minority Over-Sampling Technique Dan Random Forest Clasifier,” *Indonesian Journal of Computer Science*, vol. 12, no. 5, Nov. 2023, doi: 10.33022/ijcs.v12i5.3413.
- [40] S. Riyadi, *Machine learning: metode dan evaluasi*. Lembaga Riset dan Inovasi Universitas Muhammadiyah Yogyakarta, 2024. Accessed: Jul. 03, 2026. [Online]. Available: https://ipusnas.perpusnas.go.id/book/a3a10209-f491-47da-8953-81454127c77b?_cf_chl_f_tk=JmOYSrVIDusggmT6EVZbieY1Hyv7J1uUEBh9VLCf5QY-1783050845-1.0.1.1-0zp1Ng3TVtqsU9zpvB2W9btHhTknUrGrOBYCOTASZIE
- [41] scikit-learn, “skor_presisi_rata-rata,” 2026, *scikit-learn*. Accessed: Jul. 03, 2026. [Online]. Available: [https://scikit-](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.average_precision_score.html)

[learn.org/stable/modules/generated/sklearn.metrics.average_precision_score.html](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.average_precision_score.html)

BIODATA PENULIS



Intania Dharma Hartarti

Merupakan Mahasiswa Program Studi Informatika Fakultas Teknik dan Informatika di Universitas Bina Sarana Informatika.



Fuad Nur Hasan

Merupakan Dosen Fakultas Teknik dan Informatika di Universitas Bina Sarana Informatika



Lestari Yusuf

Merupakan Dosen Fakultas Teknik dan Informatika di Universitas Bina Sarana Informatika