



**PERBANDINGAN ALGORITMA *SUPPORT VECTOR MACHINE (SVM)*
DAN *RANDOM FOREST CLASSIFIER* PADA KLASIFIKASI PENYAKIT
INFEKSI SALURAN PERNAPASAN AKUT (ISPA)**

SKRIPSI

DHEA SYAHIRA JULIANTI

2010511075

**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA
2024**



**PERBANDINGAN ALGORITMA *SUPPORT VECTOR MACHINE (SVM)*
DAN *RANDOM FOREST CLASSIFIER* PADA KLASIFIKASI PENYAKIT
INFEKSI SALURAN PERNAPASAN AKUT (ISPA)**

SKRIPSI

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh
Gelar Sarjana Komputer**

DHEA SYAHIRA JULIANTI

2010511075

**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA
2024**

PERNYATAAN ORISINALITAS

PERNYATAAN ORISINALITAS

Tugas akhir ini adalah hasil karya sendiri, dan semua sumber yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Dhea Syahira Julianti

NIM : 2010511075

Tanggal : 11 Juli 2024

Bilamana dikemudian hari ditemukan ketidaksesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 11 Juli 2024

Yang menyatakan,



(Dhea Syahira Julianti)

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Universitas Pembangunan Nasional "Veteran" Jakarta, saya yang bertanda tangan di bawah ini :

Nama : Dhea Syahira Julianti

NIM : 2010511075

Fakultas : Ilmu Komputer

Program Studi : Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional "Veteran" Jakarta Hak Bebas Royalti Non eksklusif (*Non-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul :

PERBANDINGAN ALGORITMA *SUPPORT VECTOR MACHINE (SVM)* DAN *RANDOM FOREST CLASSIFIER* PADA KLASIFIKASI PENYAKIT INFEKSI SALURAN PERNAPASAN AKUT (ISPA)

Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional "Veteran" Jakarta berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Tugas Akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian Pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada tanggal : 11 Juli 2024

Yang menyatakan,



(Dhea Syahira Julianti)

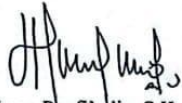
LEMBAR PENGESAHAN

LEMBAR PENGESAHAN

Skripsi diajukan oleh :

Nama : Dhea Syahira Julianti
NIM : 2010511075
Program Studi : S-1 Informatika
Judul Skripsi : PERBANDINGAN ALGORITMA *SUPPORT VECTOR MACHINE*
(*SVM*) DAN *RANDOM FOREST CLASSIFIER* PADA
KLASIFIKASI PENYAKIT INFEKSI SALURAN PERNAPASAN
AKUT (ISPA)

Telah berhasil dipertahankan di hadapan Tim Penguji dan diterima sebagai bagian dari persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi S-1 Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jakarta.



Ridwan Raafi'udin, S.Kom., M.Kom.

Penguji I



Neny Rosmawarni, S.Kom., M.Kom.

Penguji II



Bayu Hananto, S.Kom., M.Kom.

Pembimbing I



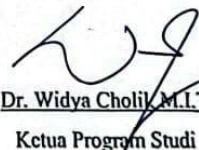
Ika Nurlaili Isnainivah, S.Kom., M.Sc.

Pembimbing II



Prof. Dr. Ir. Supriyanto, ST., M.Sc., IPM

Dekan



Dr. Widya Cholik, M.I.T

Ketua Program Studi

Ditetapkan di : Jakarta

Tanggal Ujian : 04 Juli 2024

ABSTRAK

Meningkatnya jumlah kasus infeksi saluran pernapasan akut (ISPA) di Jabodetabek telah dicatat oleh Kementerian Kesehatan RI, dengan rata-rata kasus 200 ribu per bulan. Tingginya tingkat polusi udara di DKI Jakarta memberikan dampak signifikan terhadap kesehatan pernapasan masyarakat, dengan ISPA menjadi salah satu penyakit yang berhubungan dengan polusi udara. Penelitian ini bertujuan untuk membuat dan membandingkan model klasifikasi ISPA menggunakan algoritma *Support Vector Machine (SVM)* dan *Random Forest Classifier*. Data yang digunakan diperoleh dari Puskesmas Kecamatan Matraman selama periode 2021-2023, dengan fokus pada dua jenis ISPA yaitu nasofaringitis dan faringitis. Model dibuat menggunakan berbagai skenario yaitu penggunaan parameter *default* (tanpa *hyperparameter tuning*) dan dengan *hyperparameter tuning*, penggunaan teknik *resampling* seperti *undersampling* dan *oversampling*, pengurangan ukuran jumlah sampel data, serta dengan pembagian data latih dan data uji. Hasil penelitian menunjukkan bahwa kedua model, *SVM* dan *Random Forest*, memiliki akurasi yang tinggi pada berbagai skenario percobaan. Pada pembagian data *training* 90% dan data *testing* 10%, algoritma *SVM* mencapai akurasi tertinggi sebesar 98% pada percobaan dengan ukuran sampel 500 data dan algoritma *Random Forest* mencapai 96,4% pada percobaan 2500 data, meskipun sama dengan akurasi *SVM*. Namun, meskipun *hyperparameter tuning* dilakukan, tidak ada peningkatan signifikan dalam akurasi dibandingkan dengan penggunaan parameter *default* atau tanpa *hyperparameter tuning*. Teknik *resampling* membantu menyeimbangkan kelas, tetapi presisi untuk kelas faringitis tetap rendah, menunjukkan perlunya penyesuaian lebih lanjut untuk meningkatkan performa klasifikasi pada kelas minoritas. Penelitian ini memberikan wawasan tentang performa *SVM* dan *Random Forest* dalam klasifikasi ISPA dan menjadi referensi bagi penelitian selanjutnya untuk topik yang relevan.

Kata Kunci: ISPA, *Support Vector Machine*, *Random Forest Classifier*, *Hyperparameter tuning*

ABSTRACT

The Ministry of Health of the Republic of Indonesia has noted an increase in Acute Respiratory Infection (ARI) cases in the Jabodetabek area, averaging 200 thousand cases per month. High air pollution in the Jakarta Capital Region has a significant impact on respiratory health, with ARI being one of the diseases associated with air pollution. This research aims to develop and compare ARI classification models using Support Vector Machine (SVM) and Random Forest Classifier algorithms. The data used was obtained from the Matraman Sub-district Community Health Center during the period 2021-2023, focusing on two types of ARI: nasopharyngitis and pharyngitis. The model was developed using various scenarios, including using default parameters (without hyperparameter tuning) and with hyperparameter tuning, employing resampling techniques such as undersampling and oversampling, reducing the size of the data sample, and utilizing both training and testing data splits. The research results indicate that both SVM and Random Forest models achieve high accuracy across various experimental scenarios. In the data split of 90% for training and 10% for testing, the SVM algorithm achieved the highest accuracy of 98% in experiment with a sample size of 500 data, and the Random Forest algorithm achieved 96.4% in experiments with 2500 data, even though the accuracy was the same as SVM. However, despite parameter tuning, there was no significant increase in accuracy compared to using default parameters or without hyperparameter tuning. Resampling techniques helped balance the classes, but precision for the pharyngitis class remained low, indicating the need for further adjustments to improve classification performance on minority classes. This study provides insights into the performance of SVM and Random Forest in ARI classification and serves as a reference for further research on the same topic.

Keywords: *ARI, Support Vector Machine, Random Forest Classifier, Hyperparameter tuning*

KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadirat Tuhan Yang Maha Esa atas segala limpahan karunia serta Rahmat-Nya sehingga penulis dapat menyelesaikan penyusunan skripsi ini yang berjudul “Perbandingan Algoritma *Support Vector Machine (SVM)* dan *Random Forest Classifier* Pada Klasifikasi Penyakit Infeksi Saluran Pernapasan Akut (ISPA)”. Dalam penyusunan skripsi ini penulis mendapatkan dukungan dan bantuan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. Ir. Supriyanto, ST., M.Sc., IPM, selaku dekan Fakultas Ilmu Komputer UPN Veteran Jakarta.
2. Ibu Dr. Widya Cholil, M.I.T selaku Kepala Program Studi Informatika.
3. Bapak Bayu Hananto, S.Kom., M.Kom. selaku dosen pembimbing I proposal & skripsi yang telah memberi bimbingan, saran, dan motivasi.
4. Ibu Ika Nurlaili Isnainiyah, S.Kom., M.Sc. selaku dosen pembimbing II skripsi yang telah memberi bimbingan, saran, dan motivasi.
5. Almarhumah Ibu Yuni Widiastiwi, S.Kom., M.Si. selaku dosen pembimbing proposal skripsi.
6. Keluarga tercinta yang selalu memberikan kasih sayang, dukungan, semangat, dan doa yang tiada henti.
7. Teman-teman dan sahabat penulis atas dukungan dan motivasinya.

Jakarta, Juni 2024



Penyusun

Dhea Syahira Julianti

DAFTAR ISI

PERNYATAAN ORISINALITAS.....	iii
PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS.....	iv
LEMBAR PENGESAHAN	v
ABSTRAK	vi
KATA PENGANTAR.....	viii
DAFTAR ISI	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiv
DAFTAR PERSAMAAN	xv
DAFTAR LAMPIRAN	xvi
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Tujuan Penelitian	4
1.4. Manfaat Penelitian.....	4
1.5. Batasan Masalah.....	4
1.6. Luaran yang Diharapkan	4
1.7. Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	7
2.1. Infeksi Saluran Pernapasan Akut (ISPA).....	7
2.2. <i>Machine Learning</i>	8
2.3. <i>Algoritma Support Vector Machine (SVM)</i>	8
2.4. <i>Algoritma Random Forest</i>	12
2.5. <i>Hyperparameter Tuning</i>	13

2.6. <i>Imbalance Data</i>	14
2.7. Akurasi, Presisi, dan <i>Recall</i>	15
2.8. Penelitian Terkait.....	16
BAB III METODOLOGI PENELITIAN.....	19
3.1. Tahapan Penelitian.....	19
3.2. Identifikasi Masalah.....	19
3.3. Studi Literatur.....	20
3.4. Pengumpulan Data.....	20
3.5. Praproses Data.....	21
3.6. Klasifikasi.....	22
3.7. Evaluasi.....	22
3.9. Alat Bantu Penelitian.....	22
3.10. Jadwal Penelitian.....	24
BAB IV HASIL DAN PEMBAHASAN.....	25
4.1. Pengumpulan Data.....	25
4.2. Praproses Data.....	30
4.3. Skenario Percobaan Membangun Model.....	34
4.3.1 Percobaan 1 Semua <i>Dataset</i> dengan Parameter <i>Default</i>	36
4.3.2 Percobaan 2 Semua <i>Dataset</i> dengan <i>Hyperparameter Tuning</i>	39
4.3.3 Percobaan 3 Semua <i>Dataset</i> Parameter <i>Default</i> dengan <i>Undersampling</i>	43
4.3.4 Percobaan 4 Semua <i>Dataset</i> <i>Hyperparameter Tuning</i> dengan <i>Undersampling</i>	46
4.3.5 Percobaan 5 Semua <i>Dataset</i> Parameter <i>Default</i> dengan <i>Oversampling</i>	50
4.3.6 Percobaan 6 Semua <i>Dataset</i> <i>Hyperparameter Tuning</i> dengan <i>Oversampling</i>	53

4.3.7	Percobaan 7 Sampel Data Acak dengan Parameter <i>Default</i>	57
4.3.8	Percobaan 8 Sampel Data Acak dengan <i>Hyperparameter Tuning</i> ...	64
BAB V KESIMPULAN DAN SARAN.....		76
5.1.	Kesimpulan.....	76
5.2.	Saran	77
DAFTAR PUSTAKA.....		78
LAMPIRAN.....		81

DAFTAR TABEL

Tabel 2.1 <i>Confusion Matrix</i>	15
Tabel 2.2 Penelitian Terkait.....	16
Tabel 3.1 Jadwal Penelitian.....	24
Tabel 4.1 Skenario Percobaan Model.....	34
Tabel 4.2 Hasil Percobaan 1	39
Tabel 4.3 Hasil Percobaan 2	42
Tabel 4.4 Hasil Percobaan 3	46
Tabel 4.5 Hasil Percobaan 4	49
Tabel 4.6 Hasil Percobaan 5	52
Tabel 4.7 Hasil Percobaan 6a	55
Tabel 4.8 Hasil Percobaan 6b	57
Tabel 4.9 Rasio Data <i>Split</i> Percobaan 7	59
Tabel 4.10 Percobaan 7 dengan 500 Sampel Data	60
Tabel 4.11 Percobaan 7 dengan 1000 Sampel Data	60
Tabel 4.12 Percobaan 7 dengan 1500 Sampel Data	60
Tabel 4.13 Percobaan 7 dengan 2000 Sampel Data	61
Tabel 4.14 Percobaan 7 dengan 2500 Sampel Data	61
Tabel 4.15 Percobaan 7 dengan 3000 Sampel Data	61
Tabel 4.16 Percobaan 7 dengan 3500 Sampel Data	62
Tabel 4.17 Percobaan 7 dengan 4000 Sampel Data	62
Tabel 4.18 Percobaan 7 dengan 4500 Sampel Data	62
Tabel 4.19 Perbandingan Akurasi Percobaan 7 Tiap Ukuran Data	63
Tabel 4.20 Rasio Data <i>Split</i> Percobaan 8	65
Tabel 4.21 Percobaan 8 dengan 500 Sampel Data	66
Tabel 4.22 Percobaan 8 dengan 1000 Sampel Data	66
Tabel 4.23 Percobaan 8 dengan 1500 Sampel Data	67
Tabel 4.24 Percobaan 8 dengan 2000 Sampel Data	68
Tabel 4.25 Percobaan 8 dengan 2500 Sampel Data	69
Tabel 4.26 Percobaan 8 dengan 3000 Sampel Data	69
Tabel 4.27 Percobaan 8 dengan 3500 Sampel Data	70
Tabel 4.28 Percobaan 8 dengan 4000 Sampel Data	70

Tabel 4.29 Percobaan 8 dengan 4500 Sampel Data	71
Tabel 4.30 Perbandingan Akurasi Percobaan 8 Tiap Ukuran Data	71

DAFTAR GAMBAR

Gambar 2.1 <i>SVM</i> (sumber:(Rodríguez-Pérez & Bajorath, 2022))	9
Gambar 3.1 Tahapan Penelitian	19
Gambar 3.2 <i>Flowchart</i> Skenario Percobaan	21
Gambar 4.1 Data Awal Puskesmas Sebelum Pembersihan	25
Gambar 4.2 Data Awal Puskesmas Setelah Pembersihan	26
Gambar 4.3 Data Awal Setelah Ubah Nama Penyakit	26
Gambar 4.4 Perintah untuk Mengambil <i>Dataset</i>	26
Gambar 4.5 Tampilan Dataset dengan <i>Library Pandas</i>	27
Gambar 4.6 Visualisasi Distribusi Kelas	28
Gambar 4.7 Perintah untuk Membuat Plot Distribusi Atribut Data	28
Gambar 4.8 <i>Dataframe</i> setelah <i>encoding</i>	31
Gambar 4.9 Perintah Menghapus Kolom Fitur	32
Gambar 4.10 <i>Dataframe</i> Setelah Penghapusan Fitur	33
Gambar 4.11 <i>Label Encoding</i>	34
Gambar 4.12 <i>Confusion Matrix</i> Percobaan 1	38
Gambar 4.13 <i>Confusion Matrix</i> Percobaan 2	42
Gambar 4.14 <i>Confusion Matrix</i> Percobaan 3	45
Gambar 4.15 <i>Confusion Matrix</i> Percobaan 4	48
Gambar 4.16 <i>Confusion Matrix</i> Percobaan 5	52
Gambar 4.17 <i>Confusion Matrix</i> Percobaan 6a	55
Gambar 4.18 <i>Confusion Matrix</i> Percobaan 6b	56
Gambar 4.19 Grafik Percobaan 1 hingga 8	74

DAFTAR PERSAMAAN

2.1. Persamaan <i>Hyperplane</i>	8
2.2. Persamaan Optimasi <i>Hyperplane</i>	8
2.3. Persamaan <i>Gini</i>	10
2.4. Persamaan <i>Entropy</i>	10
2.5. Persamaan Akurasi	13
2.6. Persamaan Presisi	13
2.7. Persamaan <i>Recall</i>	13

DAFTAR LAMPIRAN

Lampiran 1. Turnitin	81
Lampiran 3. Surat Izin Riset Puskesmas dari Sudinkes	83
Lampiran 4. Surat Persetujuan Etik dari KEP UPN Veteran Jakarta	84
Lampiran 5. Transkrip Wawancara Dokter	85
Lampiran 6. Data Diri Dokter	86
Lampiran 7. Dokumentasi Wawancara Dokter	87
Lampiran 8. Kode Program Percobaan 1	88
Lampiran 9. Kode Program Percobaan 2	89
Lampiran 10. Kode Program Percobaan 3	90
Lampiran 11. Kode Program Percobaan 4	91
Lampiran 12. Kode Program Percobaan 5	93
Lampiran 13. Kode Program Percobaan 6	93
Lampiran 14. Kode Program Percobaan 7	99
Lampiran 15. Kode Program Percobaan 8	98