

**ANALISIS PERBANDINGAN ALGORITMA KNN DAN RANDOM
FOREST CLASSIFICATION UNTUK MEMPREDIKSI COVID-19**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

INGGAR ADI PRASETYO

18.11.1811

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2023**

**ANALISIS PERBANDINGAN ALGORITMA KNN DAN RANDOM
FOREST CLASSIFICATION UNTUK MEMPREDIKSI COVID-19**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

INGGAR ADI PRASETYO

18.11.1811

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA**

YOGYAKARTA

2023

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS PERBANDINGAN ALGORITMA KNN DAN RANDOM
FOREST CLASSIFICATION UNTUK MEMPREDIKSI COVID-19**

yang disusun dan diajukan oleh

Inggar Adi Prasetyo

18.11.1811

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 09 Oktober 2021

Dosen Pembimbing,



Anna Baita, M. Kom
NIK. 190302290

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS PERBANDINGAN ALGORITMA KNN DAN RANDOM
FOREST CLASSIFICATION UNTUK MEMREDIKSI COVID-19**

yang disusun dan diajukan oleh

Inggar Adi Prasetyo

18.11.1811

Telah dipertahankan di depan Dewan Penguji
pada tanggal 27 Juni 2023

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Dr. Ferry Wahyu Wibowo, S.Si, M.Cs
NIK. 190302235

Jeki Kuswanto, M.Kom
NIK. 190302456

Anna Baiha, M.Kom
NIK. 190302290

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 27 Juni 2023

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Inggar Adi Prasetyo
NIM : 18.11.1811

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Perbandingan Algoritma Knn Dan Random Forest Classification Untuk Memprediksi Covid-19

Dosen Pembimbing : Anna Baita, M.Kom

1. Karya tulis ini adalah benar-benar **ASLI** dan **BELUM PERNAH** diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian **SAYA** sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab **SAYA**, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini **SAYA** buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka **SAYA** bersedia menerima **SANKSI AKADEMIK** dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 27 Juni 2023

Yang Menyatakan,



Inggar Adi Prasetyo

HALAMAN PERSEMBAHAN

Puji syukur kepada Tuhan Yang Maha Esa, atas segala rahmat dan juga kesempatan dalam menyelesaikan skripsi saya ini dengan segala kekurangannya yang ada, Segala syukur ku ucapkan kepada-Mu Ya tuhan, karena sudah menghadirkan orang-orang berarti disekeliling saya. Yang selalu memberi semangat dan doa sehingga skripsi saya dapat terselesaikan dengan baik. Skripsi ini saya persembahkan kepada :

1. Persembahkan kecil saya untuk kedua orang tua saya. Ayah dan ibu yang saya sayangi dan cintai sebagai tanda bakti, hormat dan terima kasih yang tak terhingga kupersembahkan karya kecil ini kepada ayah dan ibu yang telah memberikan kasih sayang, segala dukungan, dan cinta kasih yang tak terhingga semoga ini menjadi langkah awal untuk membuat ayah dan ibu bahagia
2. Terima kasih kepada bu Anna Baita, M.Kom selaku dosen pembimbing saya yang paling baik dan bijaksana, terimakasih sudah menjadi orang tua kedua saya di Kampus. Terima kasih atas bantuannya, nasehatnya, dan ilmunya yang selama ini dilimpahkan ke saya dengan rasa tulus dan ikhlas.
3. Tidak lupa saya mengucapkan terimakasih kepada teman-teman disekeliling saya yang juga turut serta membantu skripsi saya

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa, karena atas berkat dan anugerah-Nya penulis dapat menyelesaikan skripsi dengan judul Analisis Perbandingan Algoritma KNN Dan Random Forest Classification Untuk Memprediksi Covid-19 sebagai salah satu syarat mendapatkan gelar sarjana di Universitas Amikom.

Penulis menyadari bahwa banyak kekurangan dalam skripsi ini, yang disebabkan adanya keterbatasan penulis. Oleh karena itu, penulis mengharapkan dukungan berupa saran atau kritik membangun sehingga terciptanya perubahan yang lebih baik.

Penyusunan skripsi ini juga tidak lepas dari dukungan berbagai pihak, oleh karena itu dengan penuh rasa syukur penulis juga ingin mengucapkan terima kasih kepada mereka, diantaranya :

1. Bapak Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas Amikom Yogyakarta
2. Bapak Hanif Al Fatta, M.Kom. selaku Dekan Fakultas Ilmu Komputer
3. Ibu Windha Mega Pradnya Duhita, M.Kom. selaku ketua Program Studi Informatika
4. Ibu Anna Baita, M.Kom. selaku Dosen Pembimbing dalam proses pembuatan penyusunan skripsi atas bimbingan, saran dan motivasi yang diberikan
5. Kedua orang tua yang telah memberi semangat, bimbingan serta kasih sayang yang selalu tercurah selama ini
6. Teman teman terdekat saya yang selalu memberi dukungan

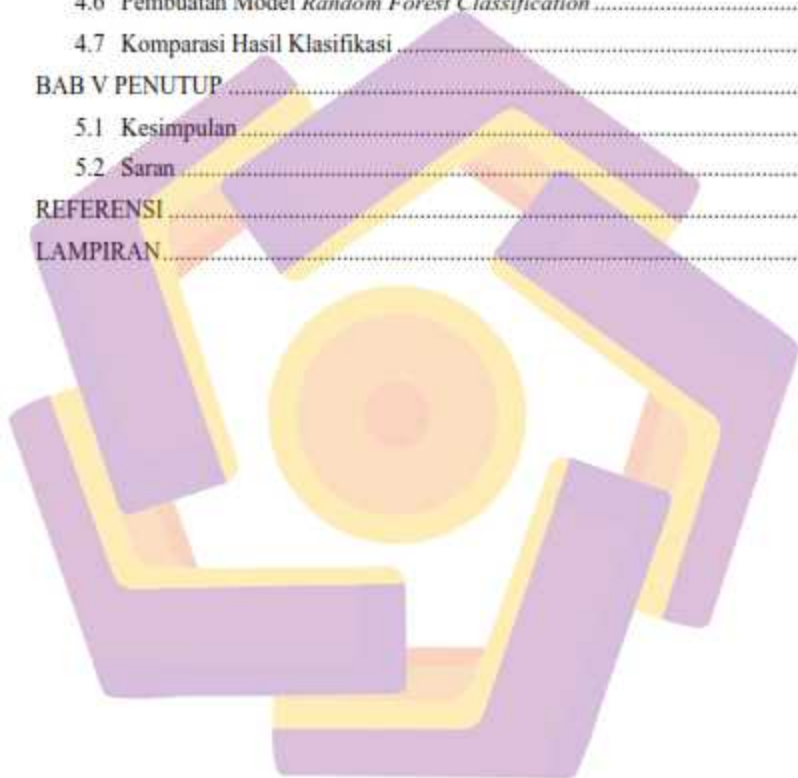
Yogyakarta, 05 Juli 2023

Penulis

DAFTAR ISI

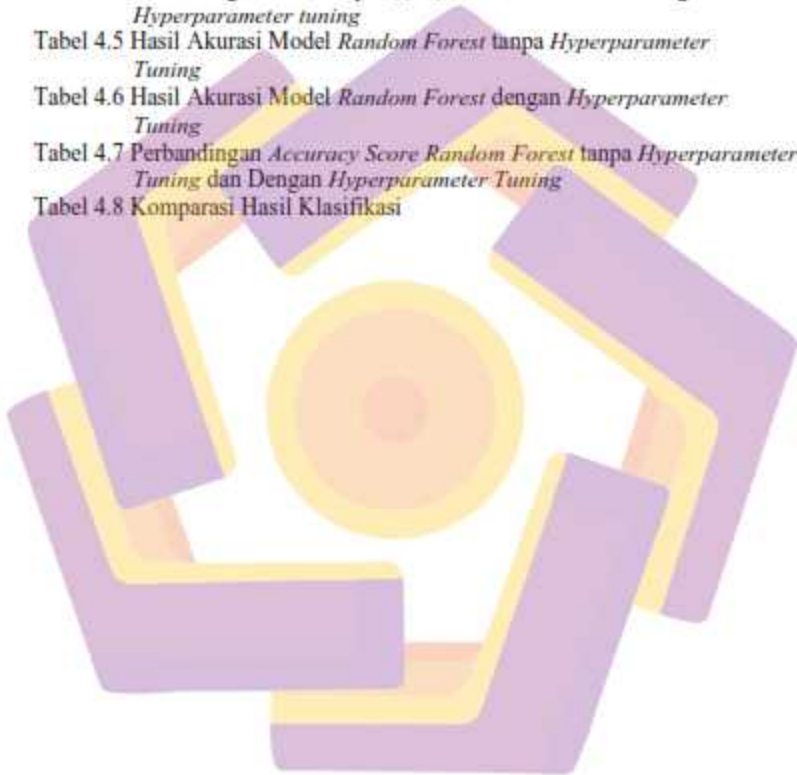
HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR TABEL	ix
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	xi
DAFTAR LAMBANG DAN SINGKATAN	xii
DAFTAR ISTILAH	xiii
INTISARI	xiv
ABSTRACT	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori	13
BAB III METODE PENELITIAN	26
3.1 Objek Penelitian	26
3.2 Alur Penelitian	26
3.3 Alat	35
BAB IV HASIL DAN PEMBAHASAN	38

4.1 Pengumpulan Data.....	38
4.2 <i>Preprocessing</i>	38
4.3 Penentuan Data <i>Training</i> dan Data <i>Testing</i>	40
4.4 Seleksi Atribut	40
4.5 Pembuatan Model <i>K-Nearest Neighbors Classification</i>	42
4.6 Pembuatan Model <i>Random Forest Classification</i>	50
4.7 Komparasi Hasil Klasifikasi	57
BAB V PENUTUP.....	59
5.1 Kesimpulan.....	59
5.2 Saran.....	59
REFERENSI.....	60
LAMPIRAN.....	62



DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	8
Tabel 4.1 Persentase Pembagian <i>Data Training</i> dan <i>Data Testing</i>	40
Tabel 4.2 Hasil Akurasi Model KNN tanpa <i>Hyperparameter Tuning</i>	43
Tabel 4.3 Hasil Akurasi Model KNN dengan <i>Hyperparameter Tuning</i>	46
Tabel 4.4 Perbandingan KNN tanpa <i>hyperparameter</i> dan KNN dengan <i>Hyperparameter tuning</i>	49
Tabel 4.5 Hasil Akurasi Model <i>Random Forest</i> tanpa <i>Hyperparameter Tuning</i>	50
Tabel 4.6 Hasil Akurasi Model <i>Random Forest</i> dengan <i>Hyperparameter Tuning</i>	53
Tabel 4.7 Perbandingan <i>Accuracy Score Random Forest</i> tanpa <i>Hyperparameter Tuning</i> dan Dengan <i>Hyperparameter Tuning</i>	57
Tabel 4.8 Komparasi Hasil Klasifikasi	58



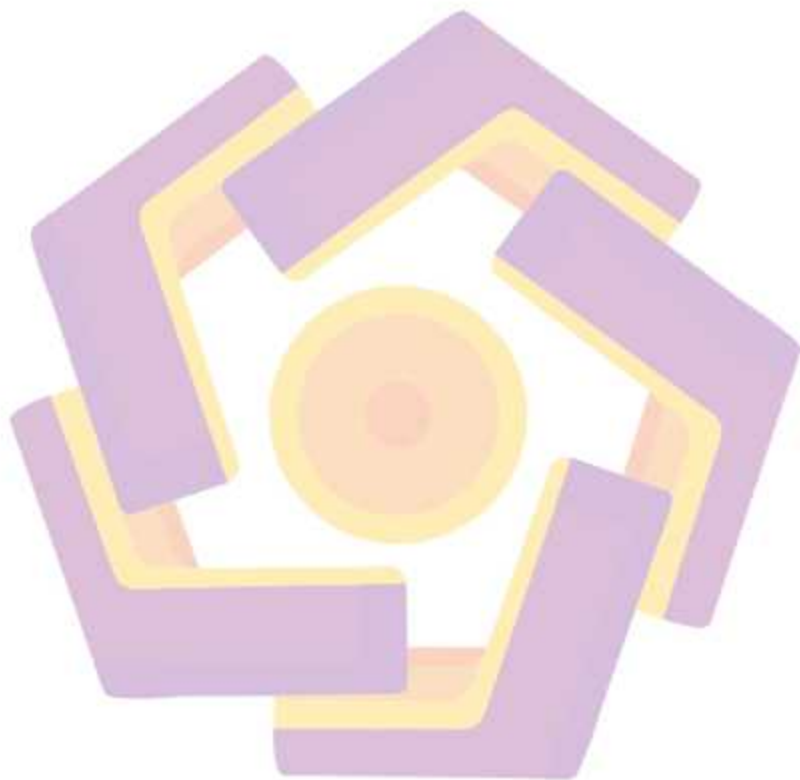
DAFTAR GAMBAR

Gambar 2.1 Model pohon <i>single decision tree</i>	18
Gambar 2.2 <i>Confusion Matrix</i>	22
Gambar 2.3 Grafik <i>Receiver Operating Characteristics (ROC)</i>	24
Gambar 3.1 Alur Penelitian	26
Gambar 3.2 Data Mentah Gejala-Gejala COVID-19	27
Gambar 3.3 Korelasi <i>Heatmap</i> Dari Dataset	28
Gambar 3.4 Flowchart <i>Preprocessing</i> Penelitian	30
Gambar 4.1 Cek Missing Value	39
Gambar 4.2 Konversi Nilai <i>Yes</i> dan <i>No</i> pada Gejala COVID-19	40
Gambar 4.3 <i>Source Code</i> Seleksi Atribut	41
Gambar 4.4 Nilai Skor Setiap Atribut	42
Gambar 4.5 <i>Confusion Matrix</i> KNN Tanpa <i>Tuning</i>	43
Gambar 4.6 <i>Classification Report</i> KNN tanpa <i>Tuning</i>	44
Gambar 4.7 <i>AUC-ROC Curve</i> pada KNN tanpa <i>Hyperparameter Tuning</i>	45
Gambar 4.8 <i>Confusion Matrix</i> KNN dengan <i>Hyperparameter Tuning</i>	46
Gambar 4.9 <i>Classification Report</i> KNN dengan <i>Hyperparameter Tuning</i>	47
Gambar 4.10 <i>AUC-ROC Curve</i> pada KNN dengan <i>Hyperparameter Tuning</i>	49
Gambar 4.11 <i>Classification Report</i> dan <i>Confusion Matrix</i> <i>Random Forest</i> Tanpa <i>Tuning</i>	50
Gambar 4.12 <i>AUC-ROC Curve</i> pada <i>Random Forest</i> Tanpa <i>Hyperparameter Tuning</i>	52
Gambar 4.13 <i>Classification Report</i> dan <i>Confusion Matrix</i> <i>Random Forest</i> dengan <i>Hyperparameter Tuning</i>	53
Gambar 4.14 <i>AUC-ROC Curve</i> pada <i>Random Forest</i> dengan <i>Hyperparameter Tuning</i>	55


DAFTAR LAMPIRAN

Lampiran 1 Data Penelitian

63



DAFTAR LAMBANG DAN SINGKATAN



Σ	Penjumlahan Total
$\sqrt{\quad}$	Akar Kuadrat
CDC	Centers for Disease Control and Prevention
WHO	World Health Organization
SARS	Severe Acute Respiratory Syndrome
MERS	Middle East Respiratory Syndrome
KDD	Knowledge discovery in databases
PSBB	Pembatasan Sosial Berskala Besar
PPKM	Pemberlakuan Pembatasan Kegiatan Masyarakat
ML	Machine Learning
KNN	K-Nearest Neighbors
RF	Random Forest
AUC	Area Under The Curve
ROC	Receiver Operating Characteristics
AUROC	Area Under the Receiver Operating Characteristics

DAFTAR ISTILAH

Algoritma	Sekumpulan instruksi atau langkah-langkah yang dituliskan secara sistematis dan digunakan untuk menyelesaikan suatu masalah untuk menghasilkan suatu output tertentu
Variabel	Suatu nama simbolis yang memiliki nilai dan nilainya bisa diubah-ubah
Klasifikasi	Proses penemuan model (atau fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui



INTISARI

Dalam memprediksi diagnosis COVID-19 berdasarkan gejala akan digunakan teknik klasifikasi yang menghasilkan model prediksi yang akurat dengan menggunakan data *training* gejala COVID-19 untuk terlebih dahulu mempelajari pola-pola yang terbentuk dengan menggunakan algoritma yang akan mempelajari pola yang ada. Algoritma menghasilkan fungsi pendekatan yang menghubungkan variabel *input* dan *output*-nya. Variabel *input* pada penelitian ini akan menerapkan seleksi fitur yang merupakan salah satu tahapan *preprocessing* klasifikasi dimana memilih fitur yang relevan yang mempengaruhi hasil klasifikasi. Variabel *output* adalah akurasi diagnosis COVID-19 melalui gejalanya.

Penelitian ini bertujuan melakukan klasifikasi berdasarkan kriteria gejala COVID-19 menggunakan K-NN dan *Random Forest*. Pada studi ini, dataset yang digunakan adalah gejala COVID-19 yang dikumpulkan dari situs Kaggle, dari tujuan penelitian di atas *research design* yang dipakai oleh penulis ini menggunakan penelitian kuantitatif dengan metode korelasi.

Hasil penelitian didapatkan bahwa algoritma *K-Nearest Neighbor* (K-NN) tanpa *tuning* menghasilkan akurasi sebesar 97,57% dengan *tuning* mengalami kenaikan sebesar 0,07% menjadi 97,64%, sedangkan *Random Forest* (RF) menunjukkan bahwa tanpa *tuning* menghasilkan akurasi sebesar 97,97% namun setelah melalui proses *tuning* parameter, akurasi meningkat menjadi 98,07%.

Kata kunci: Gejala COVID-19, *Classification*, *K-Nearest Neighbor*, *Random Forest*

ABSTRACT

In predicting the diagnosis of COVID-19 based on symptoms, a classification technique will be used which produces an accurate prediction model using COVID-19 symptom training data to first study the patterns that are formed using an algorithm that will study existing patterns. The algorithm produces an approximation function that connects the input and output variables. The input variable in this study will apply feature selection which is one of the classification preprocessing stages where selecting relevant features that affect the classification results. The output variable is the accuracy of the diagnosis of COVID-19 through its symptoms.

This study aims to classify based on the criteria for COVID-19 symptoms using K-NN and Random Forest. In this study, the dataset used was COVID-19 symptoms collected from the Kaggle website. From the research objectives above, the research design used by this author used quantitative research with the correlation method.

The results showed that the K-Nearest Neighbor (K-NN) algorithm without tuning produces an accuracy of 97.57% with tuning increasing by 0.07% to 97.64%, while Random Forest (RF) shows that without tuning it produces an accuracy of 97.97%. but after going through the parameter tuning process, the accuracy increased to 98.07%

Keyword: *Symptoms of COVID-19, Classification, K-Nearest Neighbor, Random Forest*