

**ANALISIS PERBANDINGAN AKURASI ANTARA ALGORITMA
NAIVE BAYES DAN NEURAL NETWORK UNTUK
KLASIFIKASI STATUS PASIEN COVID-19**

SKRIPSI

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



disusun oleh

IVAN NGUDI RAHAYU

18.61.0150

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

**ANALISIS PERBANDINGAN AKURASI ANTARA
ALGORITMA NAIVE BAYES DAN NEURAL NETWORK
UNTUK MEMPREDIKSI STATUS PASIEN COVID-19**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

IVAN NGUDI RAHAYU

18.61.0150

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

HALAMAN PERSETUJUAN

SKRIPSI

ANALISIS PERBANDINGAN AKURASI ANTARA ALGORITMA NAÏVE BAYES DAN NEURAL NETWORK UNTUK KLASIFIKASI STATUS PASIEN COVID-19

yang disusun dan diajukan oleh

Ivan Ngudi Rahayu

18.61.0150

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 10 Agustus 2023

Dosen Pembimbing,



Theopilus Bayu Sasongko, S.Kom, M.Eng

NIK. 190302375

HALAMAN PENGESAHAN

SKRIPSI

ANALISIS PERBANDINGAN AKURASI ANTARA ALGORITMA NAÏVE BAYES DAN NEURAL NETWORK UNTUK KLASIFIKASI STATUS PASIEN COVID-19

yang disusun dan diajukan oleh

Ivan Ngudi Rahayu

18.61.0150

Telah dipertahankan di depan Dewan Penguji
pada tanggal 10 Agustus 2023

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Anna Baita, M.Kom
NIK. 190302290



Ika Nur Fajri, M.Kom
NIK. 190302268



Theopilus Bayu Sasongko, S.Kom, M.Eng
NIK. 190302375



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 10 Agustus 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 : Ivan Ngudi Rahayu
NIM : 18.61.0150

Menyatakan bahwa Skripsi dengan judul berikut:

ANALISIS PERBANDINGAN AKURASI ANTARA ALGORITMA NAIVE BAYES DAN NEURAL NETWORK UNTUK KLASIFIKASI STATUS PASIEN COVID-19

Dosen Pembimbing : Theopilus Bayu Sasongko, S.Kom, M.Eng

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, 10 Agustus 2023

Yang Menyatakan,



Ivan Ngudi Rahayu

HALAMAN PERSEMBAHAN

Puji dan syukur saya panjatkan kehadirat Allah SWT yang telah memberikan kesempatan saya dalam menyelesaikan tugas skripsi ini. Skripsi ini saya persembahkan kepada:

1. Allah SWT yang telah memberikan nikmat dan karunia-Nya.
2. Kedua orang tua, adik, dan calon istri saya tercinta, Arum Khania yang selalu mendoakan saya dan memberikan saya dukungan penuh di setiap kondisi.
3. Bapak Theopilus Bayu Sasongko, S.Kom, M.Eng selaku dosen pembimbing yang senantiasa membimbing, mengarahkan, dan membantu saya dalam setiap kesulitan dan kemudahan saya dalam menyusun skripsi.
4. Almamater tercintaku kampus ungu, yang telah memberikan saya ilmu dan pengalaman berharga selama saya menjalani perkuliahan.
5. Teman-teman seperjuangan.

KATA PENGANTAR

Puji dan syukur saya panjatkan kehadirat Allah SWT yang telah melimpahkan rahmat dan karunianya sehingga saya dapat menyelesaikan usulan penelitian yang berjudul “Analisis Perbandingan Akurasi antara Algoritma Naive Bayes dan Neural Network untuk Klasifikasi Status Pasien Covid-19”. Skripsi ini membahas tentang hasil implementasi metode naïve bayes dan neural network MLP untuk prediksi status pasien COVID-19.

Proses penyusunan hingga selesainya laporan skripsi ini tidak lepas dari bantuan, bimbingan dan dukungan dari berbagai pihak. Maka dari itu, saya tidak lupa mengucapkan terimakasih dan penghargaan kepada :

1. Bapak Prof. Dr. M. Suyanto , M.M, selaku Rektor Universitas AMIKOM Yogyakarta.
2. Bapak Hanif Al Fatta, S. Kom,. M. Kom, selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
3. Bapak Theopilus Bayu Sasongko, S. Kom,. M. Kom, selaku Dosen Pembimbing saya yang telah memberikan bimbingan dan support penuh dalam proses penulisan skripsi ini.
4. Bapak, ibu, adik dan calon istri saya tercinta yang telah memberikan support, motivasi dan doa terhadap saya.
5. Bapak dan Ibu Dosen serta staff Universitas AMIKOM Yogyakarta yang telah banyak memberikan ilmunya selama saya berkuliah.
6. Teman-teman jurusan Informatika Universitas AMIKOM Yogyakarta angkatan 18 khususnya kelas BCI yang telah berjuang bersama semasa perkuliahan.

Yogyakarta, 10 Agustus 2023

Penulis

DAFTAR ISI

ANALISIS PERBANDINGAN AKURASI ANTARA ALGORITMA NAIVE BAYES DAN NEURAL NETWORK UNTUK MEMPREDIKSI STATUS PASIEN COVID-19.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xiii
INTISARI.....	xiv
ABSTRACT.....	xv
BAB I PENDAHULUAN	16
1.1. Latar Belakang	16
1.2. Rumusan Masalah	17
1.3. Batasan Masalah.....	17
1.4. Tujuan Penelitian.....	18
1.5. Manfaat penelitian.....	18
1.6. Sistematika Penelitian	19
BAB II TINJAUAN PUSTAKA.....	20
2.1 Studi Literatur.....	20
2.2 Dasar Teori	22
2.2.1 Data Mining	22

2.2.2	Klasifikasi Data.....	23
2.2.3	One Hot Encoding	24
2.2.4	Min-Max Scaler	25
2.2.5	Synthetic Minority Over-sampling Technique(SMOTE).....	26
2.2.6	SelectKBest.....	26
2.2.7	Naïve Bayes	27
2.2.8	Neural Network.....	28
2.2.9	Multilayer Perceptron	29
2.2.10	Evaluasi Pengujian dengan Confusion Matrix.....	32
2.2.11	K-Fold Cross Validation.....	33
BAB III METODE PENELITIAN.....		35
3.1	Alat dan Bahan Penelitian	35
3.1.1	Alat Penelitian.....	35
3.1.2	Bahan Penelitian	35
3.2	Alur Penelitian.....	37
3.3	Metode Pengolahan Data.....	39
3.3.1	Preprocessing Data	39
3.3.1.1	Drop column	40
3.3.1.2	Handling missing data.....	41
3.3.1.3	Encoding	42
3.3.1.4	Feature Scaling (Min-Max Scaler).....	43
3.3.1.5	Feature Selection (SelectKBest)	44
3.3.2	Balancing Data (SMOTE)	45
3.3.3	Metode Perhitungan Naïve Bayes.....	46
3.3.4	Metode Perhitungan Neural Network MLP.....	46

3.3.5	Confusion Matrix.....	48
3.3.6	Validasi dengan K-Fold Cross Validation.....	48
BAB IV HASIL DAN PEMBAHASAN		49
4.2.1	Drop Column	50
4.2.2	Handling missing data	50
4.2.3	Konversi Data	52
4.2.4	Normalisasi Data.....	53
4.2.5	Feature Selection.....	54
4.4.1	Evaluasi Algoritma Naïve Bayes.....	57
4.4.2	Evaluasi Algoritma Neural Network MLP Classifier.....	61
BAB V PENUTUP.....		68
5.1	Kesimpulan.....	68
5.2	Saran.....	68
DAFTAR PUSTAKA		70

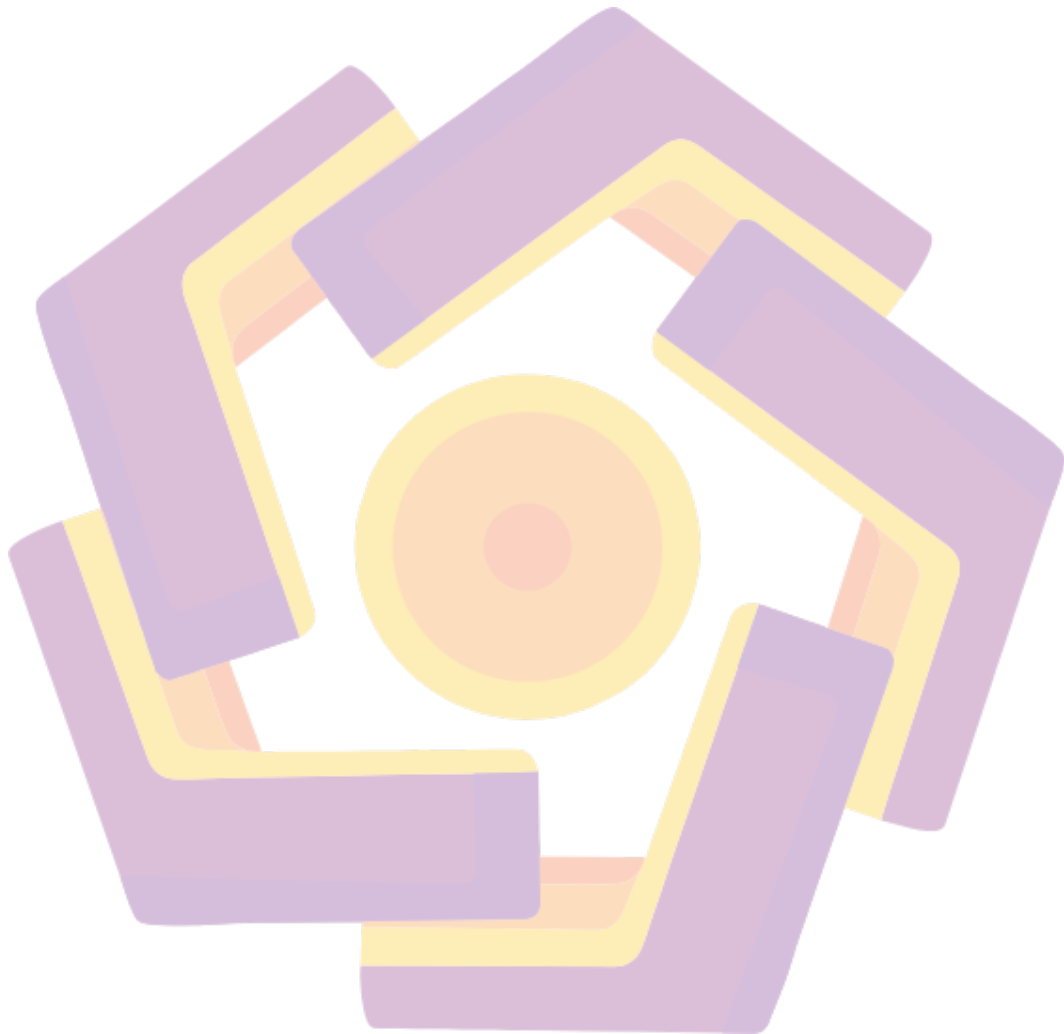
DAFTAR TABEL

Tabel 2.1 Contoh dataset sebelum dilakukan One-Hot Encoding	25
Tabel 2.2 Contoh dataset setelah dilakukan One-Hot Encoding.....	25
Tabel 2.3 Confusion Matrix	32
Tabel 3.1 Spesifikasi Perangkat Keras	35
Tabel 3.2 Daftar Perangkat Lunak yang digunakan	35
Tabel 3.3 Informasi file dataset	36
Tabel 3.4 Daftar Fitur dataset	36
Tabel 4.1 Potongan dataset pasien COVID-19	49
Tabel 4.2 Perbandingan sebelum dan setelah dilakukan handling NaN value	51
Tabel 4.3 Perbandingan sebelum dan sesudah dilakukan Encoding	52
Tabel 4.4 dataset sebelum normalisasi	54
Tabel 4.5 dataset setelah dilakukan normalisasi	54
Tabel 4.6 Confusion matrix tanpa SMOTE dan dengan Feature Selection	57
Tabel 4.7 Evaluasi hasil tanpa SMOTE dan dengan Feature Selection	57
Tabel 4.8 Hasil Validasi k-fold tanpa SMOTE dan dengan Feature Selection	57
Tabel 4.9 Confusion matrix dengan SMOTE dan tanpa Feature Selection	58
Tabel 4.10 Evaluasi hasil Naïve Bayes dengan SMOTE dan tanpa Feature Selection	58
Tabel 4.11 Hasil Validasi k-fold dengan SMOTE dan tanpa Feature Selection	58
Tabel 4.12 Confusion matrix dengan SMOTE dan Feature Selection (test_size = 0.2)	59
Tabel 4.13 Evaluasi hasil dengan SMOTE dan Feature Selection (test_size = 0.2)	59
Tabel 4.14 Hasil Validasi k-fold dengan SMOTE dan Feature Selection (test_size = 0.2)	59
Tabel 4.15 Confusion matrix dengan SMOTE dan Feature Selection (test_size = 0.3)	60

Tabel 4.16 Evaluasi hasil dengan SMOTE dan Feature Selection (test_size = 0.3)	60
Tabel 4.17 Hasil Validasi k-fold dengan SMOTE dan Feature Selection (test_size = 0.3)	60
Tabel 4.18 Confusion matrix tanpa SMOTE dan dengan Feature Selection	61
Tabel 4.19 Evaluasi hasil tanpa SMOTE dan dengan Feature Selection	61
Tabel 4.20 Hasil Validasi k-fold tanpa SMOTE dan dengan Feature Selection	61
Tabel 4.21 Confusion matrix dengan SMOTE dan tanpa Feature Selection	62
Tabel 4.22 Evaluasi hasil dengan SMOTE dan tanpa Feature Selection	62
Tabel 4.23 Hasil Validasi k-fold dengan SMOTE dan tanpa Feature Selection	62
Tabel 4.24 Confusion matrix dengan SMOTE dan Feature Selection (test_size = 0.2)	63
Tabel 4.25 Evaluasi hasil dengan SMOTE dan Feature Selection (test_size = 0.2)	63
Tabel 4.26 Hasil Validasi k-fold dengan SMOTE dan Feature Selection (test_size = 0.2)	63
Tabel 4.27 Confusion matrix dengan SMOTE dan Feature Selection (test_size = 0.3)	64
Tabel 4.28 Evaluasi hasil dengan SMOTE dan Feature Selection (test_size = 0.3)	64
Tabel 4.29 Hasil Validasi k-fold dengan SMOTE dan Feature Selection (test_size = 0.3)	64
Tabel 4.30 Confusion matrix dengan SMOTE dan tanpa Feature Selection (test_size = 0.2), (hidden_layer_size=(100,100)	65
Tabel 4.31 Evaluasi hasil dengan SMOTE dan Feature Selection (test_size = 0.2), (hidden_layer_size=(100,100)	65
Tabel 4.32 Hasil Validasi k-fold dengan SMOTE dan Feature Selection (test_size = 0.2), (hidden_layer_size=(100,100)	65
Tabel 4.33 Confusion matrix dengan SMOTE dan Feature Selection (test_size = 0.3), (hidden_layer_size=(100,100)	66

Tabel 4.34 Evaluasi hasil dengan SMOTE dan Feature Selection (test_size = 0.3),
(hidden_layer_size=(100,100) 66

Tabel 4.35 Hasil Validasi k-fold dengan SMOTE dan Feature Selection (test_size
= 0.3), (hidden_layer_size=(100,100) 67



DAFTAR GAMBAR

Gambar 2.1 Ilustrasi Arsitektur Neural Network.....	29
Gambar 2.2 Ilustrasi Pengujian dengan K-Fold Cross Validation.....	34
Gambar 3.1 Alur Penelitian.....	38
Gambar 3.2 Proses Drop Column	40
Gambar 3.3 Handling Missing Data.....	41
Gambar 3.4 Encoding data.....	42
Gambar 3.5 Feature scaling.....	43
Gambar 3.6 Feature Selection	44
Gambar 3.7 Balancing data.....	45
Gambar 3. 8 Neural Network MLP[28]	46
Gambar 3. 9 Feed Forward.....	47
Gambar 3. 10 Backpropagation.....	48
Gambar 4.1 Drop Column.....	50
Gambar 4.2 handling NaN value.....	51
Gambar 4.3 Normalisasi Data	53
Gambar 4.4 Fitur yang terpilih.....	55
Gambar 4.5 grafik sebelum dilakukan SMOTE.....	56
Gambar 4.6 Komparasi label kelas setelah dilakukan balancing.....	56

INTISARI

COVID-19 adalah penyakit menular yang disebabkan oleh virus SARS-COV-2. Pada manusia, virus ini dapat menimbulkan infeksi pernapasan ringan maupun penyakit seperti SARS atau MERS yang bersifat lebih mematikan. Dengan melihat beberapa atribut yang berpengaruh, kemudian diuraikan dalam penelitian dengan menggunakan naïve bayes dan neural network, dengan tujuan untuk membandingkan akurasi terhadap atribut yang berpengaruh dalam prediksi gejala COVID-19, sehingga hasil dari penelitian ini dapat digunakan sebagai referensi berdasarkan data kesehatan untuk memprediksi status pasien.

Penelitian dilakukan dengan menggunakan dataset pasien COVID-19 yang bersumber dari repository github. Pasien dikategorikan kedalam dua kelas label yaitu positif dan negatif. Metode yang digunakan adalah dengan menggunakan algoritma Naïve Bayes Classifier dan Neural Network MLP untuk dibuat model klasifikasi pasien COVID-19. Penelitian diawali dengan tahap preprocessing data meliputi handling missing data, encoding, normalisasi data, feature selection, dan balancing data. Kemudian dibuat model klasifikasi menggunakan dua algoritma.

Penelitian ini diharapkan dapat memberikan manfaat bagi pihak terkait seperti tenaga medis dan pemerintah dalam screening gejala-gejala COVID-19 secara dini dengan mengimplementasi model klasifikasi yang dibuat. Kemudian diharapkan dapat menjadi acuan untuk penelitian selanjutnya dalam metode klasifikasi data penyakit. Hasil pengujian dari 934 data diperoleh hasil tertinggi menggunakan metode Naïve Bayes dengan perolehan akurasi 68.64%, presisi 75%, dan recall 53%. Kemudian untuk Neural network diperoleh hasil tertinggi dengan akurasi 96.05%, presisi 93%, dan recall 99%.

Kata kunci: *data mining, naïve bayes, neural network, COVID-19, klasifikasi.*

ABSTRACT

COVID-19 is an infectious disease caused by the SARS-COV-2 virus. In humans, this virus can cause mild respiratory infections as well as diseases such as SARS or MERS which are more deadly. By looking at several influential attributes, they are then described in research using naïve Bayes and neural networks, with the aim of comparing the accuracy of the attributes that influence the prediction of COVID-19 symptoms, so that the results of this research can be used as a reference based on health data to predict health status patient.

The research was conducted using a dataset of COVID-19 patients sourced from the github repository. Patients were categorized into two class labels namely positive and negative. The method used is the Naïve Bayes Classifier algorithm and the MLP Neural Network to create a model for classifying COVID-19 patients. The research begins with the data preprocessing stage which includes handling missing data, encoding, data normalization, feature selection, and data balancing. Then a classification model is created using two algorithms.

This research is expected to provide benefits to related parties such as medical personnel and the government in early screening for COVID-19 symptoms by implementing the classification model created. Then it is hoped that it can be a reference for further research in disease data classification methods. The test results from 934 data obtained the highest results using the Naïve Bayes method with 68.64% accuracy, 75% precision, and 53% recall. Then for the Neural network the highest results were obtained with 96.05% accuracy, 93% precision, and 99% recall.

Keyword: *data mining, naïve bayes, neural network, COVID-19, classification.*