

**PERBANDINGAN *K-NEAREST NEIGHBOR* DAN *NAIVE BAYES* UNTUK KLASIFIKASI COVID-19  
PADA CITRA X-RAY**

**SKRIPSI**



disusun oleh  
**ARIEF RACHMAN HAKIM**  
**20.21.1461**

Kepada  
**PROGRAM SARJANA**  
**PROGRAM STUDI INFORMATIKA**  
**FAKULTAS ILMU KOMPUTER**  
**UNIVERSITAS AMIKOM YOGYAKARTA**  
**YOGYAKARTA**  
**2022**

**PERBANDINGAN *K-NEAREST NEIGHBOR* DAN *NAIVE BAYES* UNTUK KLASIFIKASI COVID-19  
PADA CITRA X-RAY**

**SKRIPSI**

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



disusun oleh

**ARIEF RACHMAN HAKIM**

**20.21.1461**

Kepada

**PROGRAM SARJANA  
PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2022**

**HALAMAN PERSETUJUAN**  
**SKRIPSI**  
**PERBANDINGAN *K-NEAREST NEIGHBOR* DAN *NAIVE BAYES***  
**UNTUK KLASIFIKASI COVID-19 PADA CITRA X-RAY**

yang disusun dan diajukan oleh

**Arlef Rachman Hakim**

**20.21.1461**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 25 Juni 2022

Dosen Pembimbing,

**Rizqi Sukma Kharisma, M.Kom**  
**NIK. 190302215**

**HALAMAN PENGESAHAN**  
**SKRIPSI**  
**PERBANDINGAN *K-NEAREST NEIGHBOR* DAN *NAIVE BAYES***  
**UNTUK KLASIFIKASI COVID-19 PADA CITRA X-RAY**

yang disusun dan diajukan oleh

**Arlef Rachman Hakim**

**20.21.1461**

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 21 Juli 2022

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

**Windha Mega Pradnya D, M.Kom**

**NIK. 190302185**

**Dina Maulna, M.Kom**

**NIK. 190302250**

**Rizqi Sukma Kharisma, M.Kom**

**NIK. 190302215**

Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 21 Juli 2022

**DEKAN FAKULTAS ILMU KOMPUTER**

**Hanif Al Fatta, S.Kom., M.Kom.**

**NIK. 19030209**

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : **Arief Rachman Hakim**

NIM : **20.21.1461**

Menyatakan bahwa Skripsi dengan judul berikut:

### Tuliskan Judul Skripsi

Dosen Pembimbing : **Rizqi Sukma Kharisma, 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, 21 Juli 2022

Yang Menyatakan,



Arief Rachman Hakim

## HALAMAN PERSEMBAHAN

Puji syukur penulis panjatkan kehadiran Allah SWT yang telah melimpahkan rahmat dan hidayah serta karuniaNya kepada penulis dan rekan-rekan sehingga dapat menyelesaikan penelitian ini sesuai dengan target dan mendapatkan hasil yang terbaik.

Tidak lepas dari bantuan beberapa pihak, oleh karena itu penulis ingin menyampaikan terimakasih kepada :

1. Penelitian ini penulis persembahkan kepada kedua orang tua yang selama ini mendukung dan memberi doa sehingga dapat terselesaikannya penelitian ini dengan lancar.
2. Terima kasih kepada dosen pembimbing bapak Rizqi Sukma Kharisma, M.Kom yang telah membimbing, sehingga dapat terselesaikannya penelitian ini dengan lancar.
3. Terima kasih kepada Pakdhe Mat yang telah mendukung secara materil dan immateril terselesaikannya penelitian ini dengan lancar.
4. Terima kasih kepada sanak saudara yang telah mendukung penuh dan memberi doa sehingga dapat terselesaikannya penelitian ini dengan lancar.
5. Terima kasih kepada Nova Arianto yang telah sama-sama berjuang dalam menyelesaikan gelar sarjana
6. Terima kasih kepada teman-teman baik dari kampus UNIVERSITAS AMIKOM Yogyakarta maupun dari luar yang tidak dapat disebutkan satu persatu.

- Arief Rachman Hakim -



## KATA PENGANTAR

Segala Puji dan syukur penulis panjatkan atas kehadiran Allah SWT yang telah menganugerahkan rahmat serta inayah-Nya, yang karena-Nya, Penulis diberikan kekuatan dan kesabaran untuk menyelesaikan Penelitian. Penyusunan Penelitian ini merupakan syarat untuk memenuhi kelulusan untuk setiap mahasiswa Universitas AMIKOM Yogyakarta. Melalui rangkaian Penelitian ini juga merupakan suatu bukti bahwa mahasiswa telah menyelesaikan kuliah jenjang Strata-I dan untuk memperoleh gelar Sarjana Komputer.

Didalam pengerjaan Penelitian ini telah melibatkan banyak pihak yang sangat membantu dalam banyak hal. Oleh sebab itu, disini penulis sampaikan rasa terima kasih sedalam-dalamnya kepada :

1. Yth. Bapak M. Suyanto, Prof., Dr., M.M. selaku Rektor Universitas AMIKOM Yogyakarta.
2. Yth. Bapak Hanif Al Fatta, S.Kom., M.Kom. selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
3. Yth. Ibu Windha Mega P.D., M.Kom. selaku KaProdi S1 Informatika.
4. Yth. Bapak Rizqi Sukma Kharisma, M.Kom selaku dosen pembimbing
5. Kedua Orang Tua terima kasih yang telah memberikan waktu doa dan dukungannya.

Yogyakarta, 25 Juni 2022

Penulis :

## DAFTAR ISI

HALAMAN PERSETUJUAN .....	ii
HALAMAN PENGESAHAN .....	i
HALAMAN PERNYATAAN KEASLIAN SKRIPSI .....	iii
HALAMAN PERSEMBAHAN .....	v
KATA PENGANTAR .....	vi
DAFTAR ISI .....	vii
DAFTAR TABEL .....	x
DAFTAR GAMBAR .....	xi
DAFTAR LAMPIRAN .....	xiii
DAFTAR LAMBANG DAN SINGKATAN .....	xiv
INTISARI .....	xv
<i>ABSTRACT</i> .....	xvi
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	2
1.3 Batasan Masalah .....	3
1.4 Maksud dan Tujuan Penelitian .....	3
1.5 Manfaat Penelitian .....	3
1.6 Metode Penelitian .....	3
1.6.1 Metode Pengumpulan Data .....	4
1.6.2 Metode Pengembangan Sistem .....	4
1.6.3 Analisis dan Perancangan Sistem .....	4
1.6.4 Implementasi .....	5



1.6.5	Pengujian dan Analisis Hasil .....	5
1.7	Sistematika Penulisan .....	5
<b>BAB II LANDASAN TEORI .....</b>		<b>7</b>
2.1	Penelitian Terkait .....	7
2.2	Corona Virus Disease 2019 (COVID-19) .....	13
2.3	Data Mining .....	13
2.4	Klasifikasi .....	15
2.5	Matlab .....	15
2.6	Citra Digital .....	16
2.7	Pengolahan Citra Digital .....	16
2.8	Citra Warna .....	17
2.9	Segmentasi Citra <i>Thresholding</i> .....	20
2.10	<i>Gray Level Co-occurrence Matrix (GLCM)</i> .....	21
2.11	Algoritma <i>K-Nearest Neighbor</i> .....	23
2.12	Algoritma <i>Naive Bayes</i> .....	25
2.12.1	Teorema Bayes .....	25
2.12.2	Klasifikasi <i>Naive Bayes</i> .....	25
2.13	<i>Confusion Matrix</i> .....	26
<b>BAB III METODOLOGI PENELITIAN .....</b>		<b>28</b>
3.1	Alur Penelitian .....	28
3.2	Pengumpulan Data .....	29
3.3	Pengolahan Data .....	30
3.3.1	<i>Preprocessing</i> .....	30
3.3.2	Proses Ekstraksi Fitur .....	32
3.3.3	Proses Model Klasifikasi .....	34

3.4	Perancangan Sistem.....	35
3.5	Metode Evaluasi.....	36
3.6	Perancangan Antar Muka GUI.....	36
3.7	Analisis Kebutuhan Sistem.....	38
BAB IV HASIL DAN PEMBAHASAN.....		40
4.1	Implementasi Perancangan Sistem.....	40
4.1.1	Pelatihan Data ( <i>Training</i> ).....	41
4.1.2	Pengujian Data ( <i>Testing</i> ).....	60
4.2	Evaluasi Sistem.....	66
4.2.1	Performa Data Latih.....	67
4.2.2	Performa Data Uji.....	68
4.2.3	Performa Model Klasifikasi.....	71
4.3	Implementasi Sistem GUI.....	73
4.4	Implementasi Antarmuka.....	78
4.5	Uji Coba Sistem GUI.....	80
4.6	Faktor Kegagalan.....	83
BAB V KESIMPULAN DAN SARAN.....		84
5.1	Kesimpulan.....	84
5.2	Saran.....	85
DAFTAR PUSTAKA.....		86
LAMPIRAN.....		89

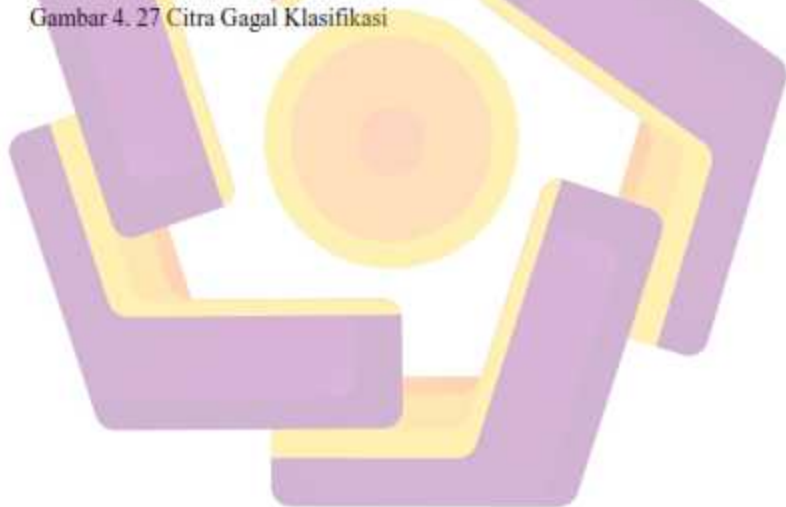
## DAFTAR TABEL

Tabel 1.1 Perbandingan Penelitian Terkait	10
Tabel 2. 1 Warna Dan Nilai Penyusun	17
Tabel 2. 2 Kelebihan Dan Kekurangan Metode <i>K-Nearest Neighbor</i>	24
Tabel 2. 3 Kelebihan Dan Kekurangan Metode <i>Naive Bayes</i>	26
Tabel 2. 4 <i>Confusion Matrix</i> Prediksi Nilai	27
Tabel 3. 1 Minimum <i>Hardware</i>	39
Tabel 3. 2 <i>Hardware</i> Yang Digunakan	39
Tabel 3. 3 <i>Software</i> Yang Digunakan	39
Tabel 4. 1 Hasil Perhitungan Matrix GLCM Data Latih	45
Tabel 4. 2 Ekstraksi Fitur GLCM Pada Data Latih Rotasi $0^{\circ}$	48
Tabel 4. 3 Ekstraksi Fitur GLCM Pada Data Latih Rotasi $45^{\circ}$	49
Tabel 4. 4 Ekstraksi Fitur GLCM Pada Data Latih Rotasi $90^{\circ}$	50
Tabel 4. 5 Ekstraksi Fitur GLCM Pada Data Latih Rotasi $135^{\circ}$	51
Tabel 4. 6 Nilai <i>Mean</i> Ekstraksi Ciri GLCM Pelatihan Data	52
Tabel 4. 7 Hasil Klusifikasi Pelatihan Model <i>Naive Bayes</i>	55
Tabel 4. 8 Hasil Pelatihan Model Klasifikasi KNN	58
Tabel 4. 9 Nilai <i>Mean</i> Ekstraksi Ciri GLCM Pada Data Uji	60
Tabel 4. 10 Hasil Tahap Pengujian Metode <i>Naive Bayes</i>	64
Tabel 4. 11 Hasil Tahap Pengujian Metode KNN	65
Tabel 4. 12 Model <i>Confusion Matrix</i>	67
Tabel 4. 13 Akurasi Pengujian Data Latih	68
Tabel 4. 14 Hasil Percobaan Satu	69
Tabel 4. 15 Hasil Percobaan Dua	69
Tabel 4. 16 Hasil Percobaan Tiga	70
Tabel 4. 17 Performa Model Klasifikasi	71
Tabel 4. 18 Uji Coba Sistem GUI	80
Tabel 4. 19 Pengujian Klasifikasi Sistem GUI	82
Tabel 4. 20 Hasil Pengujian Sistem GUI	82

## DAFTAR GAMBAR

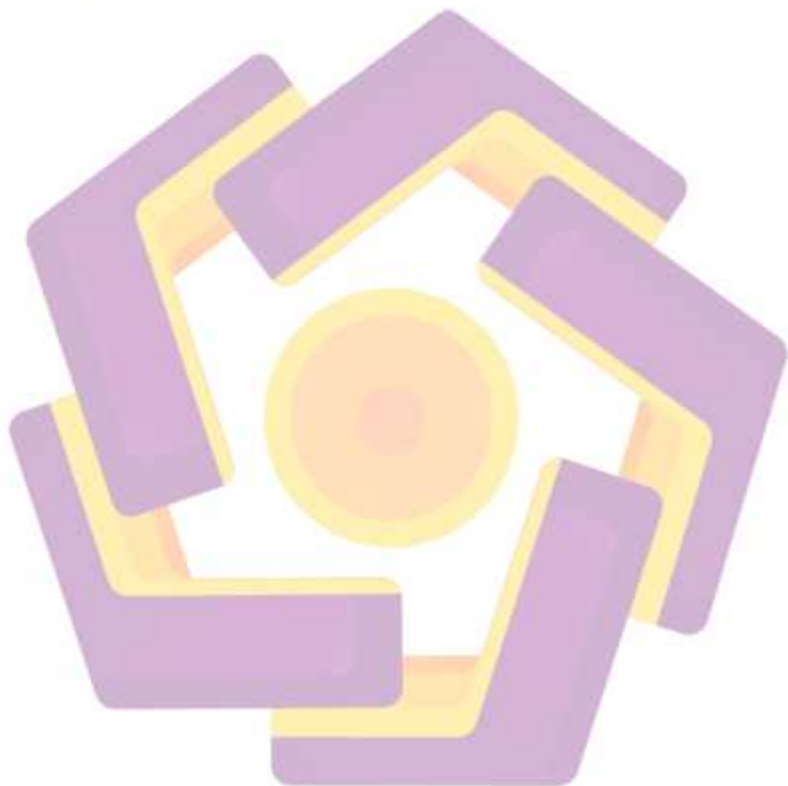
Gambar 2. 1 Citra Dan Pixel Penyusunnya	16
Gambar 2. 2 Representasi Nilai Intensitas Piksel Dengan Kombinasi Warna RGB	18
Gambar 2. 3 Citra Konversi RGB Menjadi <i>Grayscale</i> Dan Pixel Penyusunnya	19
Gambar 2. 4 Citra Hasil Konversi <i>Grayscale</i> Menjadi Biner	19
Gambar 2. 5 Hasil <i>Thresholding</i>	20
Gambar 2. 6 Empat Arah GLCM	21
Gambar 3. 1 Kerangka Alur Penelitian	28
Gambar 3. 2 Sample Citra <i>Chest X-Ray</i>	30
Gambar 3. 3 Alur <i>Preprocessing</i>	30
Gambar 3. 4 Citra Asli(A) Dan Citra Setelah <i>Preprocessing</i> (B)	32
Gambar 3. 5 Alur GLCM	32
Gambar 3. 6 Alur Klasifikasi	34
Gambar 3. 7 Alur Perancangan Sistem	35
Gambar 3. 8 Alur GUI Matlab	37
Gambar 3. 9 Tampilan <i>Graphical User Interfaces</i> Matlab	37
Gambar 4. 1 Struktur File <i>Code</i> Sistem	40
Gambar 4. 2 Kode Program Membaca Data Latih Citra	42
Gambar 4. 3 Kode Program Inisialisasi Variabel Ciri	42
Gambar 4. 4 Perbandingan Struktur <i>Matriks</i> Citra Warna Rgb Dan <i>Grayscale</i>	43
Gambar 4. 5 Kode Program <i>Preprocessing</i>	44
Gambar 4. 6 Kode Program Ekstraksi GLCM	44
Gambar 4. 7 Kode Program Ekstraksi Ciri GLCM	48
Gambar 4. 8 Kode Program Labeling	53
Gambar 4. 9 Kode Program Klasifikasi <i>Naive Bayes</i>	54
Gambar 4. 10 Struktur <i>Database</i> Mdl <i>Naive Bayes</i>	55
Gambar 4. 11 Kode Program Model Klasifikasi KNN	57
Gambar 4. 12 Struktur <i>Database</i> Mdl KNN	58
Gambar 4. 13 Kode Program Load Model Pelatihan	62
Gambar 4. 14 Struktur Variabel Data Tahap Pengujian	63

Gambar 4. 15 Kode Program Pengujian <i>Confusion Matrix</i>	67
Gambar 4. 16 Perbandingan Performa Data Latih	72
Gambar 4. 17 Perbandingan Performa Data Uji	72
Gambar 4. 18 Perbandingan Performa Waktu Simulasi	73
Gambar 4. 19 Kode Program Pilih Citra <i>Single File</i>	74
Gambar 4. 20 Kode Program <i>Preprocessing</i> Gui	75
Gambar 4. 21 Kode Program Ekstraksi Ciri GLCM	76
Gambar 4. 22 Kode Program Variabel Nilai GLCM	77
Gambar 4. 23 Kode Program <i>Mean</i> Nilai GLCM	77
Gambar 4. 24 Kode Program GUI Klasifikasi KNN	78
Gambar 4. 25 Rancangan Antar Muka Dengan <i>Guide</i>	79
Gambar 4. 26 Tampilan <i>User Interface</i> Sistem Gui	80
Gambar 4. 27 Citra Gagal Klasifikasi	83




## DAFTAR LAMPIRAN

Lampiran 1- Program Pelatihan Matlab	89
Lampiran 2- Program Pengujian Matlab	91
Lampiran 3- <i>Background</i> Aplikasi GUI	93





## DAFTAR LAMBANG DAN SINGKATAN



KNN	<i>K-Nearest Neighbor</i>
NB	<i>Naive Bayes</i>
COVID-19	Coronavirus Disease 2019
GLCM	<i>Gray Level Co-Occurrence Matrix</i>
MATLAB	MATris LABoratory
KDD	<i>Knowledge Discovery in Database</i>
GUI	<i>Graphical User Interfaces</i>
RGB	<i>Red, Green, Blue</i>
TP	<i>True Positive</i>
FP	<i>False Positive</i>
FN	<i>False Negative</i>
TN	<i>True Negative</i>
D	<i>Distance</i>
k	<i>k-neighbor</i>
R	Intensitas warna merah
G	Intensitas warna hijau
B	Intensitas warna biru
$d(x, y)$	Jarak <i>euclidean</i>
$x_i$	Data <i>Testing</i> x pada variabel ke-i
$y_i$	Data <i>Testing</i> y pada variabel ke-i
i	Dimensi data variabel bebas
$P(H X)$	Probabilitas akhir suatu hipotesis H terjadi jika diberikan bukti X
$P(X H)$	Probabilitas sebuah bukti X terjadi akan pengaruh hipotesis
$H P(H)$	Probabilitas awal ( <i>Priori</i> ) hipotesis H
$P(X)$	Probabilitas awal ( <i>Priori</i> ) bukti X
PNG	<i>Portable Network Graphics</i>

## INTISARI

COVID-19 (*Coronavirus Disease 2019*) telah dinyatakan sebagai penyakit menular akut yang menyerang sistem pernapasan manusia. Tenaga medis mendeteksi adanya virus covid-19 di dalam tubuh dapat dilakukan pemeriksaan medis lanjut seperti cek darah, pemeriksaan radiologi rontgen (x-ray) dan swab. Hasil pemeriksaan dilakukan untuk analisa apakah paru-paru pasien terdeteksi dan terinfeksi oleh COVID-19 atau tidak, tindakan ini penting untuk diagnosis klinis COVID-19 bagi penderita menjalankan isolasi dini dan pengobatan serta mencegah penyebaran penyakit. Mendeteksi Covid-19 dari hasil citra pemeriksaan radiologi rontgen (x-ray) pasien menjadi salah satu cara tercepat untuk mengidentifikasi pasien Covid-19.

Salah satu solusi untuk identifikasi penyakit COVID-19 berdasarkan citra x-ray paru-paru adalah dengan metode klasifikasi berbasis *machine learning* yaitu *K-Nearest Neighbor* Dan *Naive Bayes*. Data yang akan digunakan merupakan hasil kelola citra medis citra x-ray paru-paru yang mana citra tersebut diekstraksi menggunakan metode fitur ekstraksi *gray level co-occurrence matrix* (GLCM) yaitu *contrast*, *correlation*, *energy*, dan *homogeneity*. Data akan dibagi berdasarkan data latih dan data uji. Metode pengolahan citra membantu dalam mengidentifikasi dan mengklasifikasikan citra masukan ke dalam kelas yang berbeda.

Hasil dari penelitian untuk membandingkan penerapan dari algoritma *K-Nearest Neighbor* dan *Naive Bayes* dalam klasifikasi COVID-19. Pengolahan citra yang dilakukan meliputi konversi citra RGB menjadi citra *grayscale*, segmentasi citra menggunakan metode *Thresholding*, ekstraksi ciri citra berdasarkan nilai GLCM, dan klasifikasi citra menggunakan algoritma *K-Nearest Neighbor* dan *Naive Bayes*. Prototipe sistem klasifikasi COVID-19 dibangun menggunakan lingkungan platform pengembangan MATLAB.

**Kata kunci:** COVID-19, *K-Nearest Neighbor*, *Naive Bayes*, GLCM, Klasifikasi

## **ABSTRACT**

*COVID-19 (Coronavirus Disease 2019) has been declared an acute infectious disease that attacks the human respiratory system. Medical personnel detect the presence of the Covid-19 virus in the body, further medical examinations such as blood tests, x-rays and swabs can be carried out. The results of the examination are carried out to analyze whether the patient's lungs are detected and infected by COVID-19 or not, this action is important for clinical diagnosis of COVID-19 for patients carrying out early isolation and treatment and preventing the spread of the disease. Detecting Covid-19 from the results of the patient's x-ray radiology (x-ray) images is one of the fastest ways to identify Covid-19 patients.*

*One solution for the identification of COVID-19 disease based on x-ray images of the lungs is a machine learning-based classification method, namely K-Nearest Neighbor and Naive Bayes. The data to be used is the result of medical image management of lung x-ray images where the image is extracted using the gray level co-occurrence matrix (GLCM) feature extraction method, namely contrast, correlation, energy, and homogeneity. Data will be divided based on training data and test data. Image processing methods help in identifying and classifying the input image into different classes.*

*The results of the study were to compare the application of the K-Nearest Neighbor and Naive Bayes algorithms in the COVID-19 classification. Image processing includes converting RGB images into grayscale images, image segmentation using the Thresholding method, image feature extraction based on GLCM values, and image classification using the K-Nearest Neighbor and Naive Bayes algorithms. The prototype of the COVID-19 classification system was built using the MATLAB development platform environment.*

**Keyword:** COVID-19, K-Nearest Neighbor, Naive Bayes, GLCM, Classification