

KLASIFIKASI PENYAKIT PARU-PARU DENGAN ALGORITMA K-NN (K-NEAREST NEIGHBOR) BERDASARKAN FITUR TEKSTUR GLCM

Syahman Syahrir, Samuel Bayulianto, Valerie Febriana Putri Indra Kusumawati, Teguh Muhammad Prasetyo, Abdussalam Amrullah, Andreas Anggono, Riza Ibnu Adam, M.Si

Program Studi Teknik Informatika, Universitas Singaperbangsa Karawang

Jl. HS. Ronggo Waluyo, Puseurjaya

syahman.syahrir18225@student.unsika.ac.id, samuel.bayulianto18068@student.unsika.ac.id,
valerie.febriana18069@student.unsika.ac.id, teguh.prasetyo18139@student.unsika.ac.id,
abdussalam.amrullah18179@student.unsika.ac.id, andreas.anggono18232@student.unsika.ac.id,
riza.adam@staff.unsika.ac.id

Abstract - This study developed a disease classification system in the lungs. The lungs are the organs of respiration (respiration) associated with the respiratory and circulatory systems in the body of air-breathing vertebrates. In order for the identification of lung disease to be optimal, it will be more effective and efficient to create a disease classification application system in the lungs. This application system was built using the K-Nearest Neighbor (K-NN) method. The K-Nearest Neighbor method is used to classify lung diseases. This application system is built using the Matlab App. Matlab is a numerical computing environment and fourth generation computer programming language. The method used is data collection and system design. The result of this application system is in the form of a X-Ray Classification between normal lungs and lungs that have disease, based on the results of the grayscale calculation on the x-ray image. The highest iteration accuracy was obtained at the schema 7 (70% training image & 30% test image) which is 63%.

Keywords – Matlab App, K-Nearest Neighbor Method, X-Ray Classification

Abstrak - Penelitian ini membangun sebuah sistem klasifikasi penyakit pada paru-paru. Paru-paru yaitu organ respirasi (pernapasan) yang berhubungan dengan sistem pernapasan dan sirkulasi (peredaran darah) dalam tubuh vertebrata yang bernapas dengan udara. Agar identifikasi penyakit paru-paru menjadi optimal maka akan lebih efektif dan efisien dibuat sistem aplikasi klasifikasi penyakit pada paru-paru. Sistem aplikasi ini dibangun dengan metode K-Nearest Neighbor (K-NN). Metode K-Nearest Neighbor digunakan untuk melakukan klasifikasi penyakit pada paru-paru. Sistem aplikasi ini dibangun dengan menggunakan Aplikasi Matlab. Matlab merupakan sebuah lingkungan komputasi numerikal dan bahasa pemrograman komputer generasi keempat. Metode yang digunakan ialah pengambilan data dan rancangan sistem. Hasil dari sistem aplikasi ini yaitu Klasifikasi X-Ray antara paru-paru normal dan paru-paru yang memiliki penyakit, berdasarkan hasil dari perhitungan grayscale pada gambar X-Ray. Akurasi tertinggi didapat pada skema 7 (70% citra latih & 30% citra uji) yaitu sebesar 63%.

Kata Kunci – Aplikasi Matlab, Metode K-Nearest Neighbor, Klasifikasi X-Ray

I. PENDAHULUAN

COVID-19 adalah penyakit menular yang diakibatkan infeksi virus corona jenis baru yang muncul pada akhir 2019 pertama kali di Wuhan, Cina yang saat ini menyebabkan pandemi di

dunia. Gejala utama penyakit COVID-19 seperti batuk kering, demam, dan sesak napas[1].

COVID-19 dapat menyebar melalui tetesan kecil (droplet) dari hidung atau mulut pada saat batuk atau bersin. Tetesan kecil tersebut kemudian jatuh pada benda di sekitarnya. Jika ada seseorang

yang menyentuh benda tersebut, lalu langsung menyentuh area wajah seperti mata, hidung, atau mulut, maka orang itu dapat terinfeksi COVID-19. Infeksi COVID-19 juga menyebabkan kematian yang cukup tinggi di berbagai negara[2].

Pandemi COVID-19 menimbulkan terjadinya permasalahan di Indonesia, salah satunya kasus positif COVID-19 yang terus meningkat. Hal ini disebabkan oleh kesalahan dalam penanganan kasus COVID-19 serta buruknya kemampuan pengujian atau kurangnya alat pelindung bagi petugas medis[3].

Jumlah orang yang terinfeksi COVID-19 di dunia pada tanggal 1 Mei 2021 mencapai 150.989.419 orang dengan angka kematian mencapai 3.173.576 orang di 223 negara. Hingga 28 April 2021, sebanyak 1.011.457.859 dosis vaksin yang telah diberikan[4]. Sementara di Indonesia jumlah orang yang terinfeksi COVID-19 mencapai 1.672.880 orang dengan angka kematian mencapai 45.652 orang[5].

Dalam melakukan diagnosa COVID-19, WHO merekomendasikan pemeriksaan molekuler untuk seluruh pasien yang terduga terinfeksi COVID-19. Metode yang dianjurkan adalah metode deteksi molekuler/NAAT (*Nucleic Acid Amplification Test*) seperti pemeriksaan RT-PCR. Penggunaan *Rapid Diagnostic Test Antigen* (RDT-Ag) tidak digunakan untuk diagnostic. Pada kondisi dengan keterbatasan kapasitas pemeriksaan RT-PCR, *Rapid Test* dapat digunakan untuk skrining pada populasi spesifik dan situasi khusus, seperti pada pelaku perjalanan (termasuk kedatangan Pekerja Migran Indonesia, terutama di wilayah Pos Lintas Batas Darat Negara (PLBDN), serta untuk penguatan pelacakan kontak seperti di lapas, panti jompo, panti rehabilitasi, asrama, pondok pesantren, dan pada kelompok-kelompok rentan). WHO merekomendasikan penggunaan *Rapid Test* untuk tujuan penelitian epidemiologi atau penelitian lain[6].

Selain itu, COVID-19 dapat didiagnosa melalui foto rontgen, MRI, dan CT-Scan. Pengambilan foto rontgen paru-paru seseorang bisa memecahkan dan menganalisis citra supaya bisa mendiagnosis dan juga mengevaluasi penyakit khususnya COVID-19[7].

Klasifikasi penyakit paru-paru menggunakan foto rontgen paru-paru atau lebih dikenal dengan istilah x-ray saat ini sangat bermanfaat di bidang kedokteran. Foto x-ray ini digunakan oleh radiologi untuk melihat kondisi

paru-paru pasien. Namun, masyarakat masih mengalami kesulitan untuk memahami hasil dari foto x-ray, sehingga membutuhkan orang yang ahli dalam bidang kedokteran untuk membacanya[8].

Dari kasus tersebut, maka peneliti mengambil judul penelitian “KLASIFIKASI PENYAKIT PARU-PARU DENGAN ALGORITMA K-NN (K-NEAREST NEIGHBOR) BERDASARKAN FITUR TEKSTUR GLCM”.

A. Data Mining

Data mining adalah suatu proses menemukan pola menarik dan pengetahuan dari sejumlah besar data. Sumber data dapat berupa *database*, *warehouse*, *web*, repositori dan informasi lainnya, atau data yang dialirkan ke sistem secara dinamis[9].

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan[10], antara lain :

- a. Deskripsi
Deskripsi merupakan cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data yang dimiliki.
- b. Klasifikasi
Pada klasifikasi terdapat target variabel kategori, misalnya penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu tinggi, sedang, dan rendah.
- c. Estimasi
Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model yang dibangun menggunakan *record* lengkap yang menyediakan nilai variabel target sebagai nilai prediksi.
- d. Prediksi
Prediksi untuk menerka sebuah nilai yang belum diketahui dan juga memperkirakan nilai untuk masa mendatang.
- e. Pengklasteran
Pengklasteran merupakan golongan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan.
- f. Asosiasi
Asosiasi bertujuan untuk menemukan atribut yang muncul dalam satu waktu.

Dalam dua bisnis lebih umum disebut analisis keranjang belanja.

B. *K-Nearest Neighbor*

Algoritma *K-Nearest Neighbor* (KNN) merupakan algoritma yang sering digunakan untuk klasifikasi, meskipun dapat digunakan juga untuk estimasi data produksi. KNN adalah contoh dari basis pembelajaran dimana data training disimpan sehingga klasifikasi untuk data yang belum terklasifikasi dapat ditemukan dengan cara membandingkannya pada catatan data *training*[11].

KNN termasuk dalam *supervised learning*, dimana hasil *query instance* yang baru diklasifikasikan berdasarkan mayoritas kedekatan jarak dari kategori yang ada dalam KNN. KNN merupakan algoritma yang melakukan klasifikasi berdasarkan kedekatan lokasi jarak suatu data dengan data yang lain. Konsep dasar dari KNN adalah mencari jarak terdekat antara data yang akan dievaluasi dengan K tetangga terdekatnya dalam data pelatihan. Jumlah kelas yang paling banyak dengan jarak terdekat tersebut akan menjadi kelas dimana data evaluasi tersebut berada[12].

Ada beberapa perhitungan pada metode KNN dan yang paling umum digunakan adalah perhitungan berdasarkan jarak terdekat. Fungsi dalam menentukan jarak yang paling umum

digunakan adalah *Euclidean Distance*. Penggunaannya seperti cara manusia berpikir tentang jarak dalam dunia nyata. Contoh rumus *Euclidean Distance*[13].

$$d_{Euclidean}(x,y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Dalam penelitian ini penulis akan menggunakan metode KNN untuk mengklasifikasikan antara paru-paru normal dan paru-paru terinfeksi COVID-19 di Indonesia.

C. *Matlab*

Matlab (Matrix Laboratory) yaitu sebuah lingkungan komputasi numerikal dan bahasa pemrograman komputer generasi keempat. Dikembangkan oleh The MathWorks, MATLAB memungkinkan manipulasi matriks, pemplot-an fungsi dan data, implementasi algoritme, pembuatan antarmuka pengguna, dan peng-antarmuka-an dengan program dalam bahasa lainnya. Meskipun hanya bernuansa numerik, sebuah kotak kakas (toolbox) yang menggunakan mesin simbolik MuPAD, memungkinkan akses

terhadap kemampuan aljabar komputer. Sebuah paket tambahan, Simulink, menambahkan simulasi grafis multiranah dan Desain Berdasar-Model untuk sistem terlekat dan dinamik[14].

D. *Sinar-X*

Sinar-X adalah pancaran gelombang elektromagnetik yang sejenis dengan gelombang radio, panas, cahaya sinar ultraviolet, tetapi mempunyai panjang gelombang yang sangat pendek sehingga dapat menembus benda-benda. Sinar-X ditemukan oleh sarjana fisika berkebangsaan Jerman, yaitu W. C. Rontgen pada tahun 1895[15].

E. *Paru-Paru*

Sistem paru-paru manusia rentan terhadap infeksi karena berkaitan dengan anggota tubuh lain. Aktivitas mata, hidung, atau mulut dapat mempengaruhi tingkat kesehatan paru-paru. Kondisi ini berlaku pada semua orang berapapun usianya. Sebagai contoh udara kotor yang terhirup hidung atau terhisap oleh mulut akan berdampak pada paru-paru. Hal ini yang terjadi pada penyebaran virus corona kepada manusia. Mata, hidung dan mulut yang telah terkontaminasi virus ini mengakibatkan paru-paru menjadi tidak sehat sehingga mudah flu, batuk dan sesak nafas[16].

II. METODE PENELITIAN

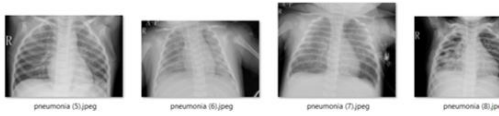
Penelitian ini disusun sebagai penelitian terapan yakni mengidentifikasi penyakit paru-paru menggunakan algoritma KNN berdasarkan citra dari situs kaggle yang kemudian di bagi menjadi dua jenis yaitu citra normal dan citra pneumonia. Penelitian ini memiliki beberapa tahapan diantaranya : Pengumpulan data penyakit paru-paru , preprocessing citra, perancangan sistem, implementasi pembuatan sistem, uji coba sistem

A. *Data Penyakit Paru-Paru*

Pada klasifikasi penyakit paru-paru yang akan dilakukan ini citra yang digunakan merupakan citra yang diperoleh dari situs kaggle. Dalam citra yang didapat hanya terdapat dua klasifikasi yaitu citra paru-paru normal dan citra paru-paru pneumonia, berikut merupakan beberapa contoh citra yang digunakan Seperti pada Gambar 1 dan 2.



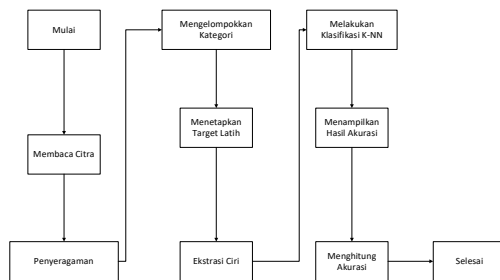
Gambar 1. Contoh Citra Normal



Gambar 2. Contoh Citra Pneumonia

B. Rancangan Sistem

Pada Penelitian ini rancangan sistem klasifikasi penyakit paru-paru menggunakan algoritma K-NN seperti pada gambar 3.



Gambar 3. Rancangan Sistem

a. Penyeragaman Citra

Penyeragaman yang dilakukan dalam proses ini hanya dilakukan konversi citra ke dalam bentuk grayscale dengan memanfaatkan fungsi yang ada di matlab yaitu `rgb2gray`.

b. Ekstraksi Ciri (Pengambilan Fitur)

Inputan yang diambil dari citra dalam program yang akan dibuat ini hanya dengan memanfaatkan pola tekstur yang nanti akan diambil. Pengambilan nilai tekstur disini menggunakan metode GLCM (Gray-Level Co-Occurrence Matrix).

Parameter tekstur yang dapat diekstrak dengan metode GLCM adalah angular second moment, contrast, correlation, variance, inverse difference moment, sum average, sum variance, sum entropy, entropy, difference variance, difference entropy, dan information measures of correlation.

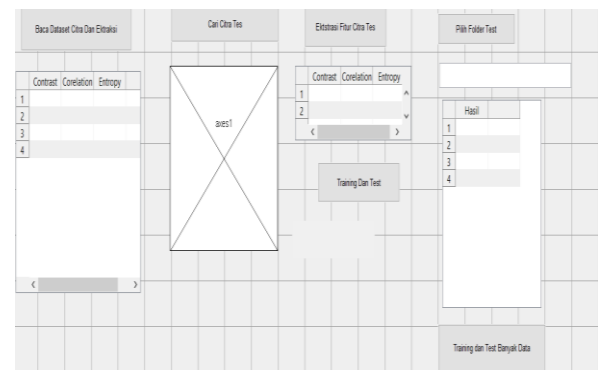
Namun pada program ini, hanya menggunakan dua parameter yaitu contrast dan correlation. Kedua parameter tersebut ditambah satu parameter yaitu entropy dijadikan sebagai nilai masukan dalam algoritma k-nn.

III. HASIL DAN PEMBAHASAN

Berikut merupakan hasil dan pembahasan klasifikasi penyakit paru-paru dengan algoritma K-NN (K-Nearest Neighbor) Berdasarkan fitur tekstur GLCM.

A. Implementasi Sistem Klasifikasi Penyakit paru-paru dengan Matlab

Implementasi dari sistem yang dibangun adalah berupa aplikasi yang dibangun dengan Matlab. Dimana sistem hanya terdapat 1 form, yaitu Tampilan Utama dalam Gambar 4. Di dalam Gambar 4. terdapat menu pilih folder, yang digunakan untuk mengambil citra paru-paru dari folder, Pada baca dataset citra dan ekstraksi akan ditampilkan citra berdasarkan Contrast, Correlation, Entropy, Kemudian pada Training dan Test Banyak Data akan ditampilkan hasil dari klasifikasi menggunakan algoritma KNN. Hasil Pengujian pada penelitian ini dibagi menjadi 9 skema dimana setiap skema berdasarkan citra pelatihan, dan citra pengujian seperti pada Tabel 1.



Gambar 4. Tampilan Utama

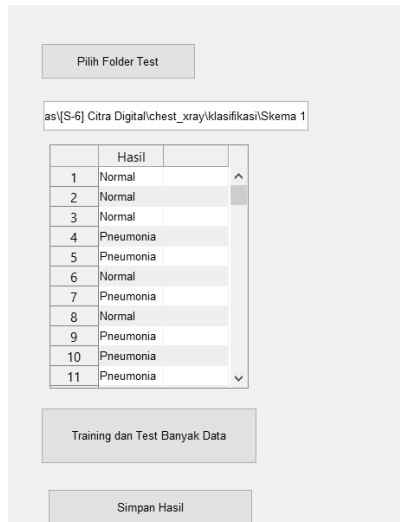
Pada pengujian citra pelatihan dan citra pengujian berdasarkan Tabel 1.

Tabel 1. Skema Pengujian

No.	Keterangan	Citra Pelatihan	Citra Pengujian
1.	Skema 1	100	-
2.	Skema 2	95	5
3.	Skema 3	90	10
4.	Skema 4	85	15
5.	Skema 5	80	20
6.	Skema 6	75	25
7.	Skema 7	70	30
8.	Skema 8	65	35
9.	Skema 9	60	40

B. Pengujian Sistem

Pada pengujian sistem contoh hasil pengujian Skema 1, seperti pada Gambar 5.



Gambar 5. Contoh Hasil Pengujian

Berikut merupakan hasil pengujian berdasarkan 9 skema yang telah dirangkum berdasarkan 100 data citra paru-paru yang dibagi berdasarkan 50 citra normal dan 50 citra pneumonia

a. Hasil pengujian skema 1

Untuk pengujian Skema 1 (100% Citra Latih = 100)
Menggunakan 100 data paru-paru (50 normal & 50 pneumonia)

Didapatkan hasil

1. Kategori normal = 18 benar dan 32 salah
2. Kategori pneumonia = 34 benar dan 16 salah
3. Didapatkan akurasi (jumlah benar/100*100%)
 $52/100*100\% = 52\%$

b. Hasil pengujian skema 2

Untuk pengujian Skema 2 (95% Citra Latih = 95 dan 5% Citra Uji = 5)
Menggunakan 100 data paru-paru (50 normal & 50 pneumonia)

Didapatkan hasil

1. Kategori normal = 19 benar dan 31 salah

2. Kategori pneumonia = 37 benar dan 13 salah
3. Didapatkan akurasi (jumlah benar/100*100%)
 $56/100*100\% = 56\%$

c. Hasil pengujian skema 3

untuk pengujian Skema 3 (90% Citra Latih = 90 dan 10% Citra Uji = 10)
Menggunakan 100 data paru-paru (50 normal & 50 pneumonia)

Didapatkan hasil

1. Kategori normal = 18 benar dan 32 salah
2. Kategori pneumonia = 38 benar dan 12 salah
3. Didapatkan akurasi (jumlah benar/100*100%)
 $56/100*100\% = 56\%$

d. Hasil pengujian skema 4

untuk pengujian Skema 4 (85% Citra Latih = 85 dan 15% Citra Uji = 15)
Menggunakan 100 data paru-paru (50 normal & 50 pneumonia)

Didapatkan hasil

1. Kategori normal = 17 benar dan 33 salah
2. Kategori pneumonia = 39 benar dan 11 salah
3. Didapatkan akurasi (jumlah benar/100*100%)
 $56/100*100\% = 56\%$

e. Hasil pengujian skema 5

untuk pengujian Skema 5 (80% Citra Latih = 80 dan 20% Citra Uji = 20)
Menggunakan 100 data paru-paru (50 normal & 50 pneumonia)

Didapatkan hasil

1. Kategori normal = 19 benar dan 31 salah
2. Kategori pneumonia = 39 benar dan 11 salah
3. Didapatkan akurasi (jumlah benar/100*100%)
 $58/100*100\% = 58\%$

f. Hasil pengujian skema 6

untuk pengujian Skema 6 (75% Citra Latih = 75 dan 25% Citra Uji = 25) Menggunakan 100 data paru-paru (50 normal & 50 pneumonia)

Didapatkan hasil

1. Kategori normal = 20 benar dan 30 salah
2. Kategori pneumonia = 39 benar dan 11 salah
3. Didapatkan akurasi (jumlah benar/100*100%)
 $59/100*100\% = 59\%$

g. Hasil pengujian skema 7

untuk pengujian Skema 7 (70% Citra Latih = 70 dan 30% Citra Uji = 30) Menggunakan 100 data paru-paru (50 normal & 50 pneumonia)

Didapatkan hasil

1. Kategori normal = 22 benar dan 28 salah
2. Kategori pneumonia = 41 benar dan 9 salah
3. Didapatkan akurasi (jumlah benar/100*100%)
 $63/100*100\% = 63\%$

h. Hasil pengujian skema 8

untuk pengujian Skema 8 (65% Citra Latih = 65 dan 35% Citra Uji = 35) Menggunakan 100 data paru-paru (50 normal & 50 pneumonia)

Didapatkan hasil

1. Kategori normal = 22 benar dan 28 salah
2. Kategori pneumonia = 40 benar dan 10 salah
3. Didapatkan akurasi (jumlah benar/100*100%)
 $62/100*100\% = 62\%$

i. Hasil pengujian skema 9

untuk pengujian Skema 9 (60% Citra Latih = 60 dan 40% Citra Uji = 30) Menggunakan 100 data paru-paru (50 normal & 50 pneumonia)

Didapatkan hasil

1. Kategori normal = 22 benar dan 28 salah
2. Kategori pneumonia = 40 benar dan 10 salah
3. Didapatkan akurasi (jumlah benar/100*100%)
 $62/100*100\% = 62\%$

IV. KESIMPULAN

Aplikasi identifikasi penyakit paru-paru ini mengambil citra dari situs kaggle yang kemudian di bagi menjadi dua jenis yaitu citra normal dan citra pneumonia. Dalam penelitian kali ini menggunakan fitur tekstur yang digunakan untuk dijadikan variabel yaitu Contrast, Corelation & Entropy. Data latih yang digunakan untuk dijadikan target sebanyak 16 citra terdiri dari 8 citra normal dan 8 citra pneumonia. Akurasi tertinggi didapat pada skema 7 (70% citra latih & 30% citra uji) yaitu sebesar 63%. Penelitian sejenis ini diharapkan menambahkan fitur yang digunakan dan juga citra target diperbanyak untuk meningkatkan akurasi ke depannya.

V. DAFTAR PUSTAKA

- [1] World Health Organization (WHO), "Coronavirus," 2020.
- [2] Kementerian Kesehatan Republik Indonesia (Kemkes RI), "Pertanyaan dan Jawaban Terkait COVID-19," 2020.
- [3] A. I. Almuttaqi, "Kekacauan Respons terhadap COVID-19 di Indonesia," *THC Insights.*, vol. 1, no. 13, pp. 1–7, 2020.
- [4] World Health Organization (WHO), "WHO Coronavirus (COVID-19) Dashboard," 2021.
- [5] Kementerian Kesehatan Republik Indonesia (Kemkes RI), "COVID 19," 2021.
- [6] Kementerian Kesehatan Republik Indonesia (Kemkes RI), "Pedoman Pencegahan dan Pengendalian *Coronavirus Disease* (COVID-19)," pp. 1-214, 2020.
- [7] X. Lin, Z. Gong, Z. Xiao, J. Xiong, B. Fan, and J. Liu, "Novel coronavirus pneumonia outbreak in 2019: Computed tomographic findings in two cases," *Korean Journal of Radiology*, vol. 21, no. 3, pp. 365–368, 2020.
- [8] E. Afriandi and Sutikno, "Identifikasi Telapak Tangan Menggunakan Jaringan Syaraf Tiruan *Learning Vector Quantization* (LVQ)," *Jurnal Infotel.*, vol. 8, no. 2, pp. 107-114, 2016.
- [9] J. Han, M. Kamber and J. Pei, *Data Mining Concepts and Techniques Third Edition*,

- United States of America: Morgan Kaufmann*, pp. 1-740, 2012.
- [10] E. F. Wulansari, "Aplikasi Data Mining Market Basket Analysis Penjualan Suku Cadang Sepeda Motor Menggunakan Metode Association Rules Pada PT.Sejahtera Motor Gemilang," 2014.
- [11] Muhammad and B. Larena, "Modified Nearest Neighbor Untuk Prediksi Curah Hujan," *Konferensi Nasional Sistem & Informatika.*, pp. 272-277, 2015.
- [12] X. W. Steinberg, "Top 10 algorithms in data mining," *Knowledge and Information Systems.*, 14:1–37 DOI 10.1007/s10115-007-0114-2, p. 14, 2008.
- [13] D. T. Larose, "Discovering Knowledge In Data An Introduction to Data Mining," *A John Wiley & Sons, INC., Publication.*, 2005.
- [14] itu MATLAB, Apa, Keunggulan-keunggulan MATLAB, Memulai MATLAB, and Command Window. "M-File Logical Operation (Programming)-For-While-Switch."
- [15] E. Souisa, Ratnawati, and B. Sudarsana, "Pengaruh Perubahan Jarak Obyek ke Film Terhadap Pembesaran Obyek Pada Pemanfaatan Pesawat Sinar-X, Tyype CGR," *Buletin Fisika.*, vol. 15, no. 2, pp. 15-21, 2014.
- [16] M. E. El. Zowalaty and J. D. Jarhult, "From SARS to COVID-19: A previously unknown SARS- related coronavirus (SARS-CoV-2) of pandemic potential infecting humans – Call for a One Health approach", *Elsevier B. V.*, vol. 9, pp. 1-6, 2020.