

# Clustering Tingkat Keparahan Tbc Berdasarkan Citra Rontgen Menggunakan Algoritma *K-Medoids* Berbasis Web

Anisa Masnauli Siregar<sup>1</sup>, Insan Taufik<sup>2</sup>

<sup>1,2</sup>Program Studi Ilmu Komputer, Jurusan Matematika, Fakultas Matematika Dan Ilmu Pengetahuan Alam, Universitas Negeri Medan, Jalan William Iskandar Pasar V, Medan 20221, Indonesia  
e-mail: [anisamasnauli05@gmail.com](mailto:anisamasnauli05@gmail.com)<sup>1</sup>, [insantaufik@unimed.ac.id](mailto:insantaufik@unimed.ac.id)<sup>2</sup>

## ABSTRAK

Tuberkulosis (TBC) adalah penyakit menular yang ditumbulkan dari kuman TB (*Mycobacterium tuberculosis*). 90% pasien yang terkonfirmasi TBC paru-paru tidak memberikan gejala (*asymptomatic*). Salah cara untuk mendiagnosis dan menganalisis penyakit yang diderita pasien dengan gambar dua dimensi atau sampel yang disebut dengan citra medis salah satu nya adalah dengan metode foto *rontgen*. Dengan *rontgen* kita dapat mengetahui kerusakan seperti apa yang terjadi pada paru-paru penderita yang diakibatkan bakteri yang menyerang. Dan dari banyaknya data citra *rontgen* tersebut masih belum memiliki label dari tingkat keparahan kerusakan paru-paru yang di derita pasien, dengan adanya pengelompokan tingkat keparahan kerusakan paru-paru *tbc* menggunakan metode data mining yaitu *algoritma K-Medoids clustering* menghasilkan label dari tingkat kerusakan paru-paru dengan mengelompokan data ke beberapa cluster berdasarkan kemiripan datanya. Berdasarkan hasil penelitian, teridentifikasi 3 *cluster*, yaitu cluster 0 dengan tingkat kerusakan paru-paru Lesi Inaktif terdapat 31 data, *cluster* 1 dengan tingkat kerusakan paru-paru Lesi aktif terdapat 120 data, dan cluster 2 dengan tingkat kerusakan paru-paru Luluh Paru terdapat 24 data. Sehingga diharapkan sangat membantu dalam mengetahui label dari tingkat kerusakan paru-paru yang diderita pasien.

**Kata Kunci:** Tuberkulosis, Clustering, *K-Medoids*

## I. PENDAHULUAN

Bakteri *Mycobacterium tuberculosis* merupakan bakteri penyebab penyakit tuberkulosis yang dapat menular [1]. Pada tanggal 2 maret 1882 Robert Koch mengemukakan penemuan dari penyebab panyakit *tuberculosis* (TBC) sejak saat itu telah membuka jalan untuk diagnosa dan pengobatan penyakit ini [2]. Menurut Kementerian Kesehatan (2020), diperkirakan kasus *tuberculosis* di Indonesia mencapai 845.000, dengan sekitar enam puluh sembilan persen atau atau diperkirakan sebanyak lima ratus empat puluh ribu yang terdeteksi. Penularan penyakit tuberkulosis bisa terjadi melalui udara, baik melalui tetesan lendir (*droplet nuclei*) maupun oleh lendir penderita *tuberculosis* yang tertinggal di alat makan pasien, sehingga infeksi melalui udara sangat mudah menular kepada orang lain [3]. Penyakit ini biasanya muncul dengan gejala seperti batuk, demam, nyeri dada, sesak napas dan kelelahan [4]. Hal ini membuktikan bahwa penyakit ini memiliki gejala yang sangat beragam, sehingga diperlukan pemeriksaan radiologis untuk mengetahui apakah seorang pasien pernah terkonfirmasi *tuberculosis*.

Cara untuk mendiagnosis dan menganalisis penyakit yang diderita pasien dengan gambar dua dimensi atau sampel yang disebut dengan citra medis. Salah satu nya adalah dengan foto *rontgen*. Kendala yang dihadapi adalah belum adanya label untuk mengetahui tingkat keparahan kerusakan paru-paru yang diderita pasien berdasarkan dari banyak nya data citra *rontgen*. Semakin banyak nya masyarakat yang terindikasi

penyakit *tuberculosis* dan sedikitnya informasi mengenai kerusakan paru-paru yang diakibatkan dari bakteri penyebab *tuberculosis* tersebut. Tujuan pelabelan dari data citra rontgen paru-paru tersebut berguna untuk mengetahui letak dari kelas apakah berada paada indeks kanan atau indeks kiri atau indeks lainnya [5]. Apabila diimplementasikan pada penelitian ini akan terbagi pada 3 kelas tingkat lesi aktif, lesi inaktif, dan luluh paru.

Salah satu algoritma yang dipergunakan untuk mengelompokkan data atau membuat label dari data tersebut adalah algoritma *clustering K-Medoids*. Algoritma *K-Medoids* memiliki kemiripan dengan algoritma *K-Means*, namun terdapat perbedaan yang cukup penting, nilai tengah dihitung dengan rata-rata (mean) dan perhitungan jarak dihitung dari data pada masing-masing mean, tetapi pada algoritma ini, data akan digunakan sebagai nilai tengah / disebut dengan medoid, dan jumlah baris dihitung dari baris pertama data [6]. Alasan penggunaan algoritma *k-medoids* pada penelitian ini adalah karena algoritma *k-medoids* menggunakan sebuah objek sebagai pusat cluster yang mewakili setiap cluster. Algoritma *K-Medoids* dirancang untuk mengatasi kelemahan algoritma *K-Means* yang sensitif terhadap outlier, karena objek dengan nilai yang lebih besar menyimpang secara signifikan dari distribusi data.[7].

Dengan tahapan proses menggunakan data rontgen yang diambil dari situs *kaggle* sekitar tahun 2020 – tahun 2022 kemudian menginput data citra untuk

melakukan proses *pre-processing*, *Pre-processing* data ialah langkah penting dalam data mining dengan memperbaiki kualitas citra [8]. Pada penelitian kali ini *pre-processing* dilakukan dengan metode *Histogram Equalization* (HE), alasan digunakannya metode *Histogram Equalization* adalah Pada penelitian “Analisis Perbandingan *He* Dan *Clahe* Pada *Image Enhancement* Dalam Proses Segmenasi Citra Untuk Deteksi Fertilitas Telur” Metode *Histogram Equalization* (HE) mampu memberikan akurasi deteksi dengan persentase 96% [9]. Data yang akan di *pre-processing* berguna untuk memastikan kualitas data yang digunakan baik sebelum dilakukan proses analisis data [8].

Data yang telah di *pre-processing* akan diekstraksi menggunakan ekstraksi fitur tekstur (GLCM) yang memiliki citra berskala keabuan (grayscale) pada penelitian “Evaluasi Ekstraksi Fitur GLCM dan LBP Menggunakan Multikernel SVM untuk Klasifikasi Batik” ekstraksi fitur Gray Level CoOccurrence Matrix (GLCM) memperoleh akurasi tertinggi pada kernel polynomial yaitu sebanyak 100% [10]. Tujuan dari ekstraksi fitur *GLCM* ialah buat mendapatkan karakteristik berasal dari masing-masing citra.

Hasil dari ekstraksi yang dilakukan akan dinormalisasi dengan menggunakan normalisasi data Min-Max seperti pada penelitian “Perbandingan Normalisasi Data Untuk Klasifikasi Wine Menggunakan Algoritma K-Nn” menghasilkan rata-rata akurasi sebesar 59,68% [11]. Setelah normalisasi selesai dilakukan proses selanjutnya adalah *clustering* dan yang akan terjadi selanjutnya adalah membuat aplikasi berbasis web untuk sistem informasi *cluster* dengan fitur validasi. Penelitian ini diharapkan sangat membantu dalam mengetahui label dari tingkat kerusakan paru-paru yang diderita pasien untuk memberikan informasi tentang jenis kerusakan paru-paru yang terindikasi tbc dan dapat memberikan obat yang sesuai agar tidak semakin mempercepat kerusakan yang terdapat di paru-paru.

## II. METODE PENELITIAN

Penelitian ini berdasarkan data paling baru dari tahun 2020-2022 terkait foto *rontgen* tuberculosis (TBC) dengan jumlah data sebanyak 175 data foto *rontgen* yang diambil dari situs resmi *kaggle*. Data tersebut akan digunakan sebagai *input* data untuk di proses pada tahap *pre-processing* dan ekstraksi fitur. Hasil citra *rontgen* paru paru *tuberculosis* (TBC) Dapat dilihat pada gambar dibawah ini.

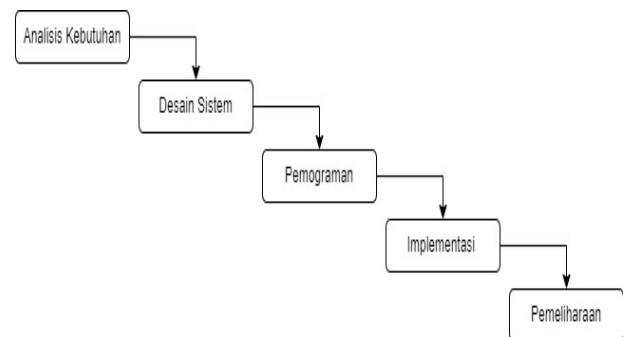
Data citra selanjutnya akan diproses untuk memperbaiki kualitas citra dengan menggunakan *Histogram Equalization* (HE). Setelah citra sudah diperbaiki kualitasnya data citra akan diproses dengan ekstraksi fitur menggunakan metode GLCM dengan menghitung ke 6 fitur yang digunakan yaitu *energy*, *homogeneity*, *contast*, *entropy*, *dissimilarity*, dan *ASM*. Data selanjutnya akan berubah menjadi angka dan dapat

diproses. Data angka yang telah diperoleh kemudian dinormalisasi terlebih dahulu, setelah data dinormalisasi selanjutnya menentukan nilai K yang terbaik untuk proses *clustering* pada penelitian ini menggunakan metode *elbow*. Hasil dari K yang telah diperoleh akan digunakan pada *clustering* dengan menggunakan metode *K-Medoids*.



Gambar 1 Citra *rontgen* paru paru *tuberculosis*

Pada penelitian ini hasil dari *clustering* akan diimplementasikan kedalam sistem informasi berbasis web. metode perancangan aplikasi yang digunakan dengan menggunakan model *waterfall* (air terjun).



Gambar 2 Tahapan metode *Waterfall*

Kemudian, melakukan desain sistem perancangan aplikasi yaitu *use case diagram*, dan DFD. Pada tahap Coding dengan menggunakan bahasa pemrograman *python* dengan memanfaatkan *flask* sebagai *library* dan memanfaatkan database *MySQL*. Kemudian melakukan pengujian sistem dengan menguji coba kualitas sistem agar hasilnya benar-benar sesuai dan melakukan pemeliharaan terhadap sistem.

## III. HASIL DAN PEMBAHASAN

### A. Pengumpulan Data

Penelitian ini data yang digunakan sebanyak 175 data dari citra *Rontgen* paru-paru *Tuberculosis* dari tahun 2020 - 2022, data diperoleh dari situs resmi *Kaggle*.

### B. Processing Data

Pada tahap *Preprocessing* data, data yang telah di ambil dari situs *Kaggle* selanjutnya di proses dengan bantuan perangkat *python* untuk memperbaiki kualitas citra di awal dengan bantuan metode *Histogram Equalization* (HE). dibawah ini

akan menampilkan perbedaan dari beberapa data citra awal sebelum dilakukan perbaikan dengan beberapa data citra yang sudah dilakukan perbaikan menggunakan metode *Histogram Equalization* (HE).



Sebelum



Sesudah

Gambar 3 Sebelum dan Sesudah *Preprocessing*

C. Ekstraksi Fitur

Data citra yang telah diperbaiki menjadi lebih tajam akan diolah kembali dengan cara ekstraksi fitur memanfaatkan metode *GLCM*. Metode *GLCM* yang digunakan dalam penelitian kali ini adalah *Contrast*, *Correlation*, *Energy* dan *Homogeneity*, *Dissimilarity*, dan *ASM*.

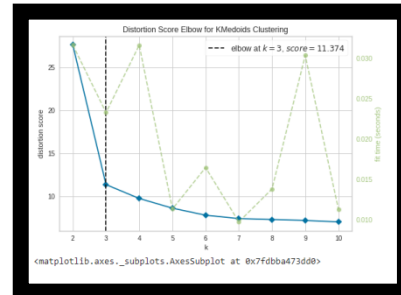
D. Normalisasi Data

Normalisasi yang digunakan pada penelitian ini menggunakan metode normalisasi *Min Max*, metode ini mempermudah dalam melakukan perbandingan. Kegunaan lain dari normalisasi data ini adalah untuk memastikan data berada pada tabel yang tepat.

E. Metode *Elbow*

Sebelum dilakukannya tahapan *clustering*, dilakukannya tes *cluster* dengan metode *Elbow*

terlebih dahulu untuk melihat nilai K terbaik untuk *cluster* tersebut.



Gambar 4 Diagram Kurva Hasil Evaluasi Menggunakan Metode *Elbow*

F. *Clustering K-Medoids*

Pada penelitian kali ini *clustering* di lakukan menggunakan metode *K-Medoids*, metode ini berguna untuk menghasilkan label dari data citra rontgen paru-paru agar dapat mengetahui tingkat dari kerusakan paru-paru. Proses *clustering* akan dijelaskan seperti penelitian sebelumnya [12].

- Langkah yang pertama adalah menentukan data awal yang akan kita gunakan dalam proses *clustering*, data awal yang digunakan pada penelitian ini adalah sebagai berikut :

Tabel 1 Tabel Data Awal

| No  | Contrast                | Correlation             | Energy                  | Homogeneity             | Dissimilarity           | ASM                      |
|-----|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|--------------------------|
| 1   | 0.37427406886<br>533693 | 0.648452494961<br>0835  | 0.0749549323397<br>8434 | 0.140523538353<br>01573 | 0.616404008142<br>7322  | 0.0159562179612<br>06443 |
| 2   | 0.36564399098<br>528494 | 0.661656890313<br>1164  | 0.0651700196470<br>3467 | 0.086730400447<br>41814 | 0.619709608587<br>4756  | 0.0133306247250<br>97509 |
| 3   | 0.04111917153<br>70571  | 0.957496366688<br>4653  | 0.0424437930296<br>5137 | 0.190407938815<br>47952 | 0.110401059536<br>23469 | 0.0078611721423<br>79177 |
| ⋮   | ⋮                       | ⋮                       | ⋮                       | ⋮                       | ⋮                       | ⋮                        |
| 175 | 0.57683832629<br>94452  | 0.405293498225<br>09797 | 0.7801336243380<br>85   | 0.868110199861<br>0425  | 0.403873793101<br>56835 | 0.6341825641536<br>185   |

- Langkah yang kedua adalah inialisasi pusat *cluster* sebanyak 3 karena pada *cluster* ini menggunakan nilai K = 3. Data dipilih secara acak dari data tabel, untuk tabel medoid awal asumsikan sebagai centroid awal yang telah ditentukan terlebih dahulu.
- Langkah ke tiga adalah menghitung jarak dengan memasukan ke enam fitur data awal selanjutnya dikurangi dengan ke enam fitur data acak yang telah dipilih, untuk menghitung centroid 1, 2, dan 3 data awal yang digunakan hanya data yang pertama dari ke enam fitur, Menghitung jarak menggunakan *Euclidian Distance* dengan persamaan (1)

$$d_{ijklmn} = \sqrt{(x_i - y_j)^2 + \dots + (x_i - y_l)^2} \quad (1)$$

- Sedangkan langkah yang keempat adalah dengan menghitung Total Simpangan (S). Setelah didapatkan nilai jarak antara iterasi ke-1 dan iterasi ke-2, hitung total simpangan (S) dengan mencari selisih dari nilai total cost baru - nilai total cost lama. Dengan ketentuan jika  $S < 0$ , maka tukar nilai objek dengan menentukan medoid baru.
- Langkah yang terakhir dilakukan dalam metode ini adalah mengulangi langkah-langkah sebelumnya hingga total simpangan yang dihasilkan tidak lagi  $S < 0$ . Sehingga didapatkan cluster beserta anggota

cluster masing-masing.

Tabel 2 Data Hasil *Clustering* Menggunakan Metode *K-Medoids*

|     | Contrast                | Correlation             | Energy                  | Homogeneity             | Dissimilarity           | ASM                      | Cluster | Nama File                 |
|-----|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|--------------------------|---------|---------------------------|
| 1   | 0.3742740688<br>6533693 | 0.6484524949<br>610835  | 0.0749549323<br>3978434 | 0.1405235383<br>5301573 | 0.6164040081<br>427322  | 0.01595621796<br>1206443 | 0       | Tubercu<br>losis (1)      |
| 2   | 0.3656439909<br>8528494 | 0.6616568903<br>131164  | 0.0651700196<br>4703467 | 0.0867304004<br>4741814 | 0.6197096085<br>874756  | 0.01333062472<br>5097509 | 0       | Tubercu<br>losis<br>(10)  |
| 3   | 0.0411191715<br>370571  | 0.9574963666<br>884653  | 0.0424437930<br>2965137 | 0.1904079388<br>1547952 | 0.1104010595<br>3623469 | 0.00786117214<br>2379177 | 1       | Tubercu<br>losis<br>(100) |
| ⋮   | ⋮                       | ⋮                       | ⋮                       | ⋮                       | ⋮                       | ⋮                        | ⋮       | ⋮                         |
| 175 | 0.5768383262<br>994452  | 0.4052934982<br>2509797 | 0.7801336243<br>38085   | 0.8681101998<br>610425  | 0.4038737931<br>0156835 | 0.63418256415<br>36185   | 2       | Tubercu<br>losis<br>(99)  |

Tabel 3 Jumlah Hasil Dari Cluster Metode *K-Medoids*

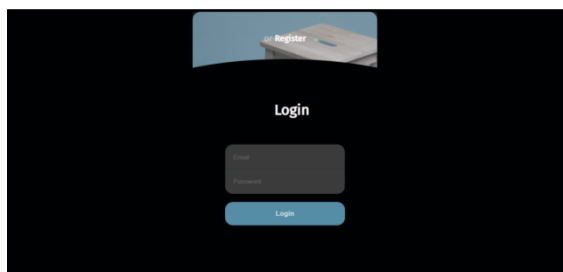
| Cluster | Jumlah Cluster | Nama Label         |
|---------|----------------|--------------------|
| 0       | 31             | Lesi TB Inaktif    |
| 1       | 120            | Lesi TB Luluh Paru |
| 2       | 24             | Lesi TB Aktif      |

Setelah proses clustering dilakukan akan memperoleh banyak data dari tiap cluster yang telah ditentukan.

G. Implementasi Sistem

1. Tampilan Login

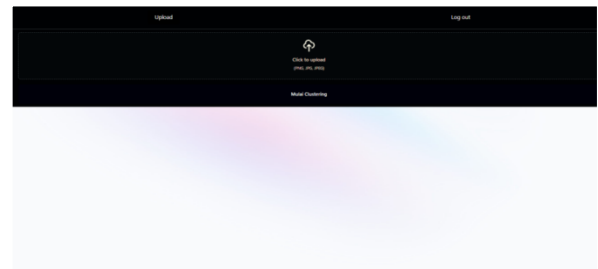
Pada tahapan login akan menampilkan halaman utama yang berisi *email* dan *password* untuk *login*.



Gambar 5 Tampilan Utama (Login)

2. Tampilan Upload Data

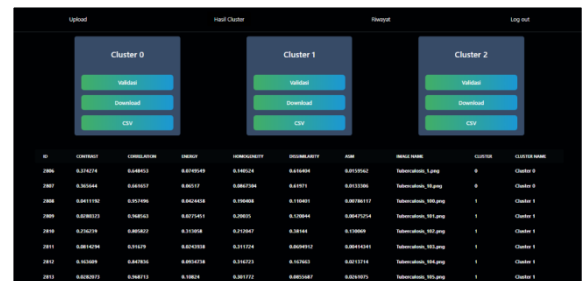
Halaman ini akan menampilkan tampilan dari proses *upload* data yang akan dilakukan. Sebelum memproses data yang akan dicari, data citra rontgen dimasukkan dengan cara menekan menu *upload* yang terdapat pada tampilan dilayar.



Gambar 6 Tampilan Upload Data

3. Tampilan Hasil Cluster

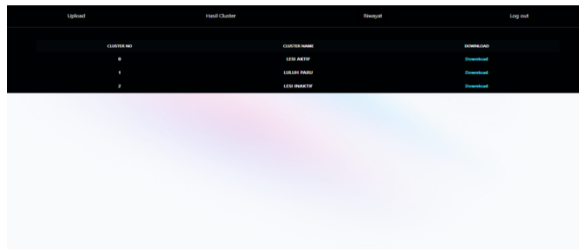
Halaman Hasil *Cluster* adalah halaman yang menampilkan label-label dari setiap *cluster* dan menampilkan hasil dari setiap fitur ekstraksi yang digunakan, dan gambar hasil dari cluster dapat didownload dengan menekan menu *download* yang ada pada layar.



Gambar 7 Tampilan Hasil Cluster

4. Tampilan Hasil Riwayat

Halaman ini menampilkan hasil dari validasi yang telah kita lakukan yang menampilkan hasil nilai cluster beserta nama cluster dan menampilkan gambar hasil validasi yang kemudian dapat didownload untuk disimpan.



Gambar 8 Tampilan Riwayat Admin

#### IV. KESIMPULAN

Hasil dari penelitian yang telah dilakukan adalah *clustering* dengan memanfaatkan metode *K-Medoids* yang bertujuan untuk menentukan label dari data citra yang di gunakan agar dapat mengetahui kelompok dari klaster yang pertama yaitu lesi aktif yang kedua lesi inaktif dan yang terakhir adalah luluh paru. Dari banyaknya data citra *rontgen* yang digunakan telah mendapatkan label hasil dari *cluster* yang berguna untuk mengetahui tingkat dari kerusakan paru-paru pasien, agar mendapatkan penanganan yang lebih tepat sesuai tingkat keparahan kerusakan paru-paru yang dideritanya. Dan selanjutnya akan dibangun sebuah sistem aplikasi berbasis *web*, yang dibuat untuk mempermudah dalam mengetahui label dari data citra rontgen yang digunakan untuk mengetahui tingkat keparahan kerusakan paru-paru *TBC* yang diderita pasien.

#### REFERENSI

- [1] F. Hardin, "Hubungan Pengetahuan Keluarga dengan Pencegahan Penularan TBC Paru Pada Keluarga di Puskesmas Andalas Tahun 2020," vol. 4, no. 2, pp. 446–454, 2021.
- [2] Kemenkes RI, "Tuberkulosis ( TB )," *Tuberkulosis*, vol. 1, no. april, p. 2018, 2018, [Online]. Available: [www.kemendes.go.id](http://www.kemendes.go.id).
- [3] V. Nomor and D. A. N. Demonstrasi, "Jurnal Peduli Masyarakat," vol. 2, pp. 237–242, 2020.
- [4] F. Bangun, J. R. Sagala, P. Studi, T. Informatika, and M. E. Mulyanto, "SISTEM PAKAR MENDIAGNOSA PENYAKIT TBC," vol. 6, 2019.
- [5] P. D. Dyan;, F. Perdana;, and M. T. S. Rizal;, "Klasifikasi Penyakit Kulit Pada Manusia Menggunakan Metode Binary Decision Tree Support Vector Machine (BDTSVM) (Studi Kasus: Puskesmas Dinoyo Kota Malang)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 5, pp. 2919–2926, 2018.
- [6] S. Asmiatun, "Penerapan Metode K-Medoids Untuk Pengelompokkan Kondisi Jalan Di Kota Semarang," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 6, no. 2, pp. 171–180, 2019, doi: 10.35957/jatisi.v6i2.193.
- [7] M. Muslihudin *et al.*, "SISTEM PENDUKUNG KEPUTUSAN PENENTUAN CALON PEMOHON PEMBIAYAAN NASABAH MENGGUNAKAN METODE SIMPLE ADDITIVE WEIGHTING (SAW) (Studi Kasus: BTPN Syariah Kantor Fungsional Operasional Ciawi Kabupaten Tasikmalaya)," *J. TAM (Technology Accept. Model.)*, vol. 4, no. 2, pp. 111–120,

2019, [Online]. Available:

[https://ejournal.diponegara.ac.id/index.php/jusiti/article/view/6%0AIK Hadi - 2019 - eprints.mercubuana-yogya.ac.id%0Ahttp://ojs.stmikpringsewu.ac.id/index.php/proceedingkmsi/article/view/451%0Ahttp://pta.trunojoyo.ac.id/uploads/journals/090451100005/090451](https://ejournal.diponegara.ac.id/index.php/jusiti/article/view/6%0AIK%0A%0Ahttp://ojs.stmikpringsewu.ac.id/index.php/proceedingkmsi/article/view/451%0Ahttp://pta.trunojoyo.ac.id/uploads/journals/090451100005/090451).

- [8] A. Yani, "Analisa Kelayakan Kredit Menggunakan Artificial Neural Network dan Backpropogation (Studi Kasus German Credit Data)," *J. Ilm. Komputasi*, vol. 18, no. 4, pp. 385–390, 2019, doi: 10.32409/jikstik.18.4.2672.
- [9] S. Saifullah, "Analisis Perbandingan He Dan Clahe Pada Image Enhancement Dalam Proses Segmenasi Citra Untuk Deteksi Fertilitas Telur," *J. Nas. Pendidik. Tek. Inform.*, vol. 9, no. 1, p. 134, 2020, doi: 10.23887/janapati.v9i1.23013.
- [10] P. N. Andono and E. H. Rachmawanto, "Evaluasi Ekstraksi Fitur GLCM dan LBP Menggunakan Multikernel SVM untuk Klasifikasi Batik," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 1–9, 2021, doi: 10.29207/resti.v5i1.2615.
- [11] D. A. Nasution, H. H. Khotimah, and N. Chamidah, "Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN," *Comput. Eng. Sci. Syst. J.*, vol. 4, no. 1, p. 78, 2019, doi: 10.24114/cess.v4i1.11458.
- [12] I. Karo Karo, S. Dewi, M. Mardiana, F. Ramadhani, and P. Harliana, "K-Means and K-Medoids Algorithm Comparison for Clustering Forest Fire Location in Indonesia", *JurnalEcotipe*, vol. 10, no. 1, pp. 86-94, Apr. 2023.