

Klasifikasi Jenis Warna Dokumen Berdasarkan Ruang Warna Cymk Menggunakan Metode K-Means Clustering

Ilham Teguh Prayudha¹, Herlando Prayitno², Farhan Rizqullah Bagaskara³,
Soffiana Agustin⁴

^{1,2,3,4} Program Studi Teknik Informatika, Universitas Muhammadiyah Gresik

E-mail : ilhamtprayudha@gmail.com, andonando158@gmail.com, farhanbagask@gmail.com,
soffiana@umg.ac.id

Abstrak

Di era digital saat ini, mencetak dokumen masih menjadi bagian penting dalam kegiatan bisnis, akademik, dan industri. Meski banyak yang sudah beralih ke dokumen digital, kebutuhan akan dokumen fisik masih ada untuk berbagai keperluan. Kendala muncul ketika pemeriksaan setiap halaman dilakukan secara manual untuk menentukan apakah halaman tersebut berwarna atau tidak, dimana proses tersebut kurang efisien. Karena itu, peneliti mengusulkan penggunaan k-means clustering dalam menentukan kelompok pewarnaan dokumen. Diharapkan penelitian ini dapat berguna dalam menentukan presentase warna dokumen, terutama dalam bidang usaha percetakan. Adapun tahapan penelitian yang dilakukan, mulai dari melakukan pengkonversian citra kedalam ruang warna Cyan, Magenta, Kuning dan Hitam (CYMK). Hasil dari masing-masing ruang warna ini yang akan menjadi masukan K-Means dalam melakukan pengelompokan dokumen. Penentuan kluster dilakukan berdasarkan persentase warna non-black dengan kondisi jika $\leq 5\%$ masuk kluster warna hitam putih, jika $\leq 35\%$ masuk kluster warna rendah, dan jika $> 35\%$ masuk kluster warna tinggi. Berdasarkan hasil dari pembahasan tersebut, penelitian ini berhasil mengklasifikasikan jenis warna pada dokumen menggunakan ekstraksi ruang warna CMYK dan metode klasifikasi K-Means Clustering dengan akurat, serta dapat menentukan total harga berdasarkan presentase warna dokumen.

Kata kunci : Pengolahan Citra, CYMK, K-Means Clustering, Klasifikasi, Ruang Warna

1. PENDAHULUAN

Di zaman digital sekarang, mencetak dokumen tetap merupakan komponen vital dalam aktivitas bisnis, akademik, dan industri. Meskipun banyak orang telah beralih ke dokumen digital, masih ada banyak yang memerlukan dokumen fisik untuk berbagai tujuan. Hal ini didukung oleh survei dari Two Sides, yang menyatakan bahwa 73% responden merasa mereka lebih memahami dan mengolah informasi yang dibaca dalam bentuk fisik dibandingkan dengan bentuk digital. Oleh karena itu, bisnis percetakan dokumen atau juga dikenal dengan *print* dan *fotocopy* masih penting.

Pada umumnya sebuah dokumen memiliki intensitas warna pada tiap halaman dapat bervariasi. Dalam konteks ini, terdapat tiga jenis intensitas warna yang umum ditemui yaitu hitam putih, warna rendah, dan warna tinggi. Masalah muncul ketika proses pengecekan setiap halaman dilakukan secara manual dalam menentukan halaman tersebut berwarna atau tidak, dimana hal tersebut kurang efisien. Sebagai contoh, untuk memeriksa dokumen dengan 100 halaman, petugas harus memeriksa satu per satu halaman. Hal tersebut membuang waktu berharga yang seharusnya bisa digunakan untuk melayani pelanggan lainnya. Petugas harus cermat memeriksa dokumen

untuk menentukan apakah halaman tersebut berwarna atau tidak, karena hal ini mempengaruhi penentuan harga akhir. Menanggapi masalah ini, pendekatan yang dapat dilakukan adalah dengan memanfaatkan teknologi pengolahan citra digital.

Pengolahan citra merupakan suatu tahapan untuk mengedit serta memodifikasi gambar agar sesuai dengan keinginan pengguna [1]. Pengolahan citra digital mencakup berbagai teknik untuk menganalisis dan memodifikasi gambar digital guna mendapatkan informasi yang berguna atau meningkatkan kualitas gambar. Salah satu komponen penting dalam pengolahan citra digital adalah ekstraksi fitur. Ekstraksi fitur warna adalah salah satu teknik yang kerap diterapkan dalam pengolahan citra digital. Disebutkan bahwa salah satu elemen desain dalam media infografis adalah warna [2]. Warna adalah informasi visual yang sangat penting dan bisa memberikan banyak data tentang objek dalam gambar. Ruang warna merupakan notasi yang digunakan untuk mendefinisikan warna berdasarkan persepsi manusia terhadap spektrum elektromagnetik yang terlihat [3]. Langkah berikutnya dalam pengolahan citra digital adalah klasifikasi. Klasifikasi merupakan tahapan pengelompokan data ke dalam kategori atau kelas berdasarkan fitur-fitur yang telah diekstraksi. Suatu algoritma klasifikasi yang efektif dan sering dimanfaatkan adalah *K-Means Clustering*.

Berdasarkan hasil penelitian yang dilakukan oleh Eni Safitri dengan judul “Implementasi Algoritma *K-Means Clustering* Dalam Menentukan Strategi Marketing Dalam Penjualan Ikan (Studi Kasus: Grosir Ikan Tani Mas Tanjung Morawa)”, Algoritma *K-Means Clustering* dapat mengelompokkan data penjualan ikan berdasarkan cluster tidak laris, laris dan paling laris [4].

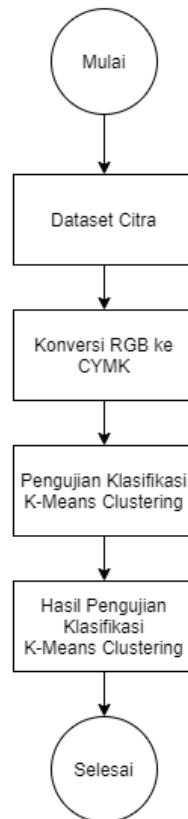
Hal tersebut sejalan dengan penelitian yang dilakukan oleh Hero Diogenes Adoe, dkk., dengan judul “Segmentasi Citra Burung Lovebird Menggunakan *K-Means*, bahwa segmentasi burung lovebird” berdasarkan warna dapat teridentifikasi menggunakan metode *K-Means* dengan akurasi sebesar 96,26% [5].

Selanjutnya dibuktikan juga dari penelitian terdahulu yang dilakukan oleh Khairi Ibnutama, dkk., dengan judul “Seleksi Tingkat Kematangan Citra Buah Belimbing Menggunakan Ruang Warna CMYK”. Penelitian ini berhasil mendeteksi tingkat kematangan buah belimbing dengan menggunakan metode ekstraksi fitur pada ruang warna CMYK dengan mencapai tingkat akurasi yang memuaskan yaitu 86,66% [6].

Berdasarkan ketiga penelitian tersebut, dapat disimpulkan bahwa Algoritma *K-Means Clustering* dan Ruang Warna CYMK dapat digunakan untuk mengklasifikasikan warna. Karena itu, peneliti mengusulkan penggunaan *k-means clustering* dalam menentukan kelompok pewarnaan dokumen. Adapun tahapan penelitian yang dilakukan, mulai dari melakukan pengkonversian citra kedalam ruang warna Cyan, Magenta, Kuning dan Hitam (CYMK). Hasil dari masing-masing ruang warna ini yang akan menjadi masukan *K-Means* dalam melakukan penelompokan dokumen. Diharapkan penelitian ini dapat berguna dalam menentukan presentase warna dokumen, terutama dalam bidang usaha percetakan.

2. METODE PENELITIAN

Metode penelitian yang dipakai untuk membangun aplikasi deteksi warna pada dokumen yaitu menggunakan metode pengolahan citra. Penelitian akan dilakukan sesuai dengan diagram pada Gambar 1.



Gambar 1. Diagram Alur Klasifikasi Jenis Warna Dokumen






Berikut adalah penjelasan dari setiap tahap dalam diagram alur tersebut:

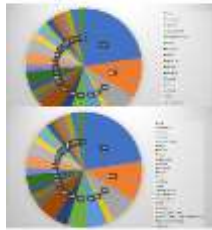
1. **Dataset Citra**
Tahap ini melibatkan persiapan citra atau file yang akan diuji. Setelah file siap, file tersebut diinput ke dalam sistem.
2. **Konversi RGB ke CYMK**
Langkah ini mengonversi citra atau file yang telah diinput dari format warna RGB (Red, Green, Blue) ke format warna CMYK (Cyan, Magenta, Yellow, Key/Black).
3. **Pengujian Klasifikasi K-Means Clustering**
Algoritma K-Means Clustering diterapkan pada citra yang telah dikonversi ke format CMYK. K-Means Clustering adalah teknik pengelompokan yang bertujuan membagi dataset menjadi beberapa cluster berdasarkan karakteristik warna yang mirip.
4. **Hasil Pengujian Klasifikasi K-Means Clustering**
Setelah algoritma K-Means Clustering dijalankan, hasil pengelompokan atau klasifikasi dianalisis. Hasil ini menunjukkan bagaimana citra-citra dalam dataset dikelompokkan berdasarkan fitur warna yang dimiliki.

2.1 Dataset

Pada penelitian ini menggunakan sebuah dokumen dengan format .pdf. Akan tetapi, dalam pengoperasian aplikasi yang akan dibuat, nantinya dapat menggunakan file dengan extension .docx, .doc, .jpg, .jpeg, maupun .png.

Tabel 1. Sampel dataset

No	Dataset	Kelas
1		Hitam Putih
2		Hitam Putih
3		Warna rendah
4		Warna rendah
5		Warna tinggi

6		Warna tinggi
---	---	--------------

2.2 Ruang Warna CYMK

Ruang warna CMYK terdiri dari warna Cyan, Magenta, Kuning, dan Hitam. Nilai untuk Cyan, Magenta, dan Kuning dapat dinyatakan dalam bilangan desimal dari 0 hingga 1, sedangkan nilai untuk warna Hitam dapat dinyatakan dalam bilangan dari 0 hingga 255. [7]. CMYK adalah model warna yang berbasis pada pengurangan sebagian gelombang cahaya (model warna subtraktif) dan umum digunakan dalam pencetakan berwarna. Oleh karena itu, diperlukan minimal 4 warna untuk mereproduksi gambar agar hasilnya relatif sempurna[8]. Berikut rumus untuk menghitung ruang warna CYMK.

$$Black = \min(255 - Red, \quad 255 - Green, \quad 255 - Blue) \quad (1)$$

$$Cyan = \frac{255 - Red - Black}{255 - Black} \quad (2)$$

$$Magenta = \frac{255 - Green - Black}{255 - Black} \quad (3)$$

$$Yellow = \frac{255 - Blue - Black}{255 - Black} \quad (4)$$

2.3 K-Means Clustering

Algoritma *K-Means Clustering* populer dikarenakan kemudahan serta kemampuannya untuk dengan cepat mengklasifikasikan banyak data dan mendeteksi *outlier* [9]. *K-Means Clustering* adalah salah satu metode klasterisasi data non-hirarki [10]. Selain itu, *K-Means Clustering* merupakan teknik unsupervised learning yang mengklasifikasikan data menjadi beberapa klaster (kelompok) berdasarkan kesamaan fitur-fitur mereka [11]. Algoritma ini beroperasi dengan mengiterasikan pusat klaster hingga mencapai konvergensi, yaitu ketika posisi pusat klaster tidak mengalami perubahan yang signifikan [12]. *Clustering* adalah langkah-langkah menyusun sejumlah data atau objek ke dalam kelompok yang disebut klaster, sehingga masing-masing klaster mengandung data yang sangat serupa satu sama lain dan memiliki perbedaan dengan objek dalam klaster lain [13]. Secara konsep clustering digunakan untuk memaksimalkan dan meminimalkan intra antar kelas. Cluster juga dapat diartikan sebagai kelompok sehingga pada dasarnya analisis clustering akan menghasilkan beberapa cluster [14].

3. HASIL DAN PEMBAHASAN

Penelitian ini diharapkan dapat mempermudah dan meningkatkan efisiensi penyedia jasa cetak dan fotokopi. Dalam penelitian ini menggunakan aplikasi MATLAB untuk melakukan klasifikasi dan membangun sebuah aplikasi yang menampilkan menu untuk input dan output sesuai dengan klaster yang telah ditentukan [15]. Setelah aplikasi mampu mengklasifikasikan

warna dengan akurat, aplikasi tersebut akan dikonversi menjadi format exe agar dapat dijalankan pada sistem operasi Windows.

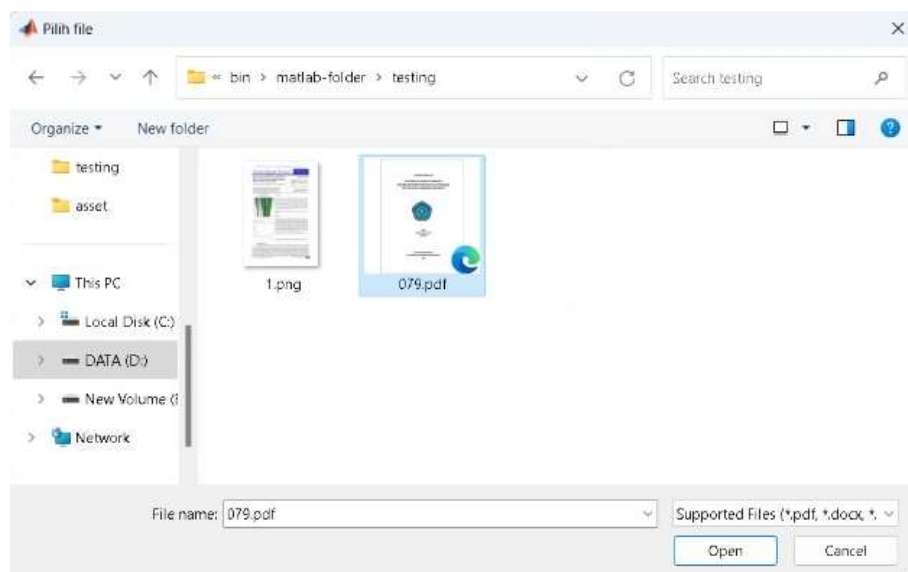
3.1 Implementasi

Menu halaman utama adalah tampilan pertama yang muncul ketika aplikasi ini dijalankan. Tampilan ini hanya berisi tombol untuk mengunggah dokumen yang akan dikelompokkan (clustering) tingkat pewarnaannya. Menu halaman utama ditampilkan pada gambar 2.



Gambar 2. Menu Halaman Utama

Setelah tombol Upload File pada menu utama di klik, aplikasi akan menampilkan file explorer untuk memilih file seperti tampilan pada gambar 3. Adapun file yang akan diuji pengelompokannya dapat berbentuk file dengan extension .pdf, .docx, .doc, .jpg, .jpeg, maupun .png.

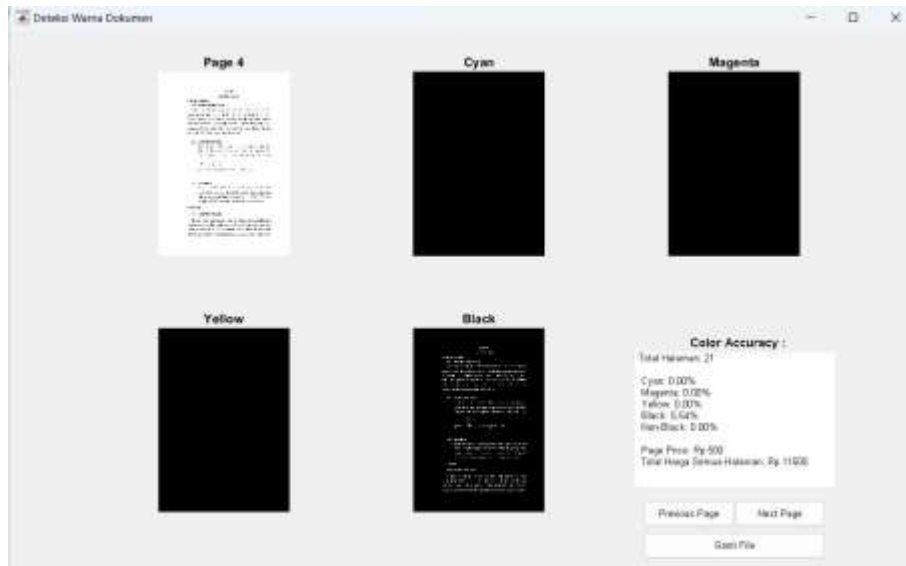


Gambar 2. Menu Upload File

Setelah file dipilih, file tersebut akan dibaca oleh sistem sebagai citra gambar. Citra inilah yang akan dikelompokkan tingkat pewarnaannya. Kemudian citra gambar tersebut dikonversi dari RGB ke CMYK. Setelah itu, dilakukan perhitungan persentase area setiap warna (Cyan, Magenta, Yellow, Black), menentukan harga berdasarkan persentase warna non-black dan mengakumulasi total harga semua halaman. Warna non-black yang dimaksud adalah total dari warna selain black yakni Cyan, Magenta dan Yellow. Setelah itu, muncul jendela UI untuk menampilkan hasil konversi, menampilkan total harga dan jumlah halaman, serta menyediakan navigasi untuk melihat setiap halaman yang dapat dilihat pada gambar 4.

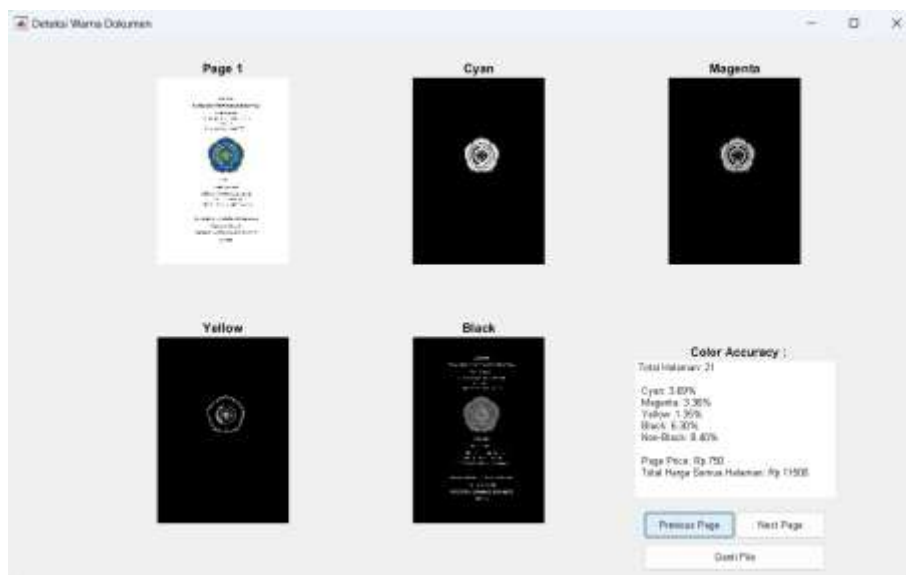
Penentuan kluster dilakukan berdasarkan persentase warna non-black, dengan kondisi jika $\leq 5\%$ masuk kluster warna hitam putih, jika $\leq 35\%$ masuk kluster warna rendah, dan jika $>$

35% masuk kluster warna tinggi. Halaman dokumen pada gambar 4 memiliki warna black sebesar 4,44% dan non-black sebesar 0%. Oleh karena itu, halaman tersebut termasuk dalam kluster hitam putih.



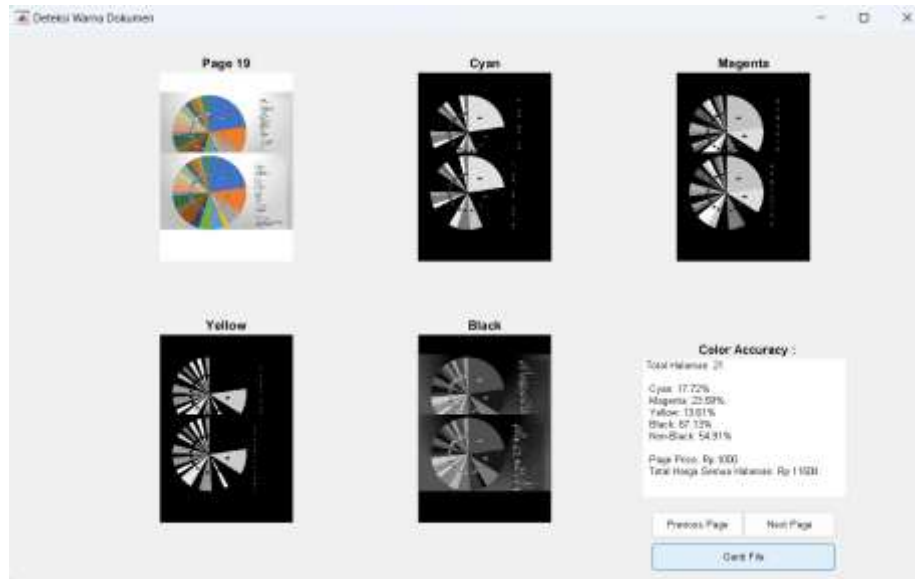
Gambar 4. Tampilan Deteksi Dokumen Berwarna Hitam Putih

Pada gambar 5, halaman dokumen tersebut memiliki warna black sebesar 5,75% dan non-black sebesar 6.16%. Oleh karena itu, halaman tersebut termasuk dalam kluster warna rendah.



Gambar 5. Tampilan Deteksi Dokumen Berwarna Rendah

Pada gambar 6, halaman dokumen tersebut termasuk dalam kluster warna tinggi. Hal ini dikarenakan halaman tersebut memiliki warna black sebesar 24,84% dan non-black sebesar 45.06%.



Gambar 6. Tampilan Deteksi Dokumen Berwarna Tinggi

4. KESIMPULAN

Berdasarkan hasil dari pembahasan tersebut, penelitian ini berhasil mengklasifikasikan jenis warna pada dokumen menggunakan ekstraksi ruang warna CMYK dan metode klasifikasi K-Means Clustering. Selain itu, aplikasi klasifikasi warna yang dibuat menggunakan MATLAB dapat mengklasifikasikan jenis warna pada dokumen dengan akurat, serta dapat menentukan total harga berdasarkan presentase warna dokumen.

5. SARAN

Untuk penelitian berikutnya, disarankan untuk menggunakan ekstraksi fitur warna dan metode klasifikasi yang lain, seperti ekstraksi fitur warna HSV dan metode klasifikasi SVM.

DAFTAR PUSTAKA

- [1] Z. Zidan, "Perbaikan Kualitas Citra Digital Pada Gambar Yang Gelap Untuk Mengidentifikasi Target Dengan Metode Contrast Stretching," UPT PERPUSTAKAAN UIN Ar-Raniry Banda Aceh, 2024. [Online]. Available: <https://repository.ar-raniry.ac.id/id/eprint/35593/>
- [2] Baldomir. Zajc, Marko. Tkalčič, Institute of Electrical and Electronics Engineers., and Institute of Electrical and Electronics Engineers. Region 8., "Colour spaces: perceptual, historical and applicational background," in *The IEEE Region 8 EUROCON 2003. Computer as a Tool*, Ljubljana, Slovenia: IEEE, 2003. doi: 10.1109/EURCON.2003.1248032.
- [3] A. Listya, "Konsep Dan Penggunaan Warna Dalam Infografis," *JURNAL DESAIN*, pp. 10–19, 2019, doi: <http://dx.doi.org/10.30998/jurnaldesain.v6i01.2837>.
- [4] K. Ibutama, M. Gilang Suryanata, R. O. Putri, and A. Al Hafiz, "Seleksi Tingkat Kematangan Citra Buah Belimbing Menggunakan Ruang Warna CMYK," *Agustus*, vol. 22, pp. 302–310, Aug. 2023, [Online]. Available: <https://ojs.trigunadharma.ac.id/index.php/jis/index>
- [5] P. Ayu Lestari, D. Puspita, S. Aminah, and Yadi, "Penerapan Algoritma K-Nearest Neighbor (KNN) Pada Klasifikasi Kualitas Biji Kopi Robusta," vol. 5, no. 2, pp. 605–614, 2024, doi: <https://doi.org/10.30645/kesatria.v5i2.370.g367>.
- [6] S. I. Wahyudi and A. Wibowo, "Implementasi Metode K-Means Clustering Untuk Pengelompokan Data Stok Produk Toko Online Perdagangan Kaos," *Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI) Jakarta-Indonesia*, Sep. 2022, [Online]. Available: <https://senafiti.budiluhur.ac.id/index.php>

-
- [7] D. Hernando, A. W. Widodo, and C. Dewi, "Pemanfaatan Fitur Warna dan Fitur Tekstur untuk Klasifikasi Jenis Penggunaan Lahan pada Citra Drone," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 4, no. 2, pp. 614–621, 2020, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [8] Juandri and N. Anwar, "Pengenalan Warna Terhadap Objek Dengan Model Analisis Elemen Data Warna Gambar Berbasis Deep Neural Network," *BULLET : Jurnal Multidisiplin Ilmu*, vol. 2, pp. 23–31, 2023, [Online]. Available: <https://journal.mediapublikasi.id/index.php/bullet>
- [9] U. D. Rosiani, C. Rahmad, M. A. Rahmawati, and F. Tupamahu, "SEGMENTASI BERBASIS K-MEANS PADA DETEKSI CITRA PENYAKIT DAUN TANAMAN JAGUNG," *JIP (Jurnal Informatika Polinema)*, vol. 6, no. 3, 2020, [Online]. Available: <https://doi.org/10.33795/jip.v6i3.331>
- [10] A. Andreansyah, "Aplikasi Pengenalan Pola Citra Logo Obat Medis Menggunakan K-Means Clustering," *ELTI Jurnal Elektronika, Listrik dan Teknologi Informasi Terapan*, vol. 2, no. 1, pp. 7–12, 2020, [Online]. Available: <https://ojs.politeknikjambi.ac.id/elti>
- [11] I. R. Nr, E. Prasetyowati, and D. B. Said, "Penerapan Citra Berbasis K-Means Clustering untuk Mendeteksi Penyakit Bulai Pada Komoditas Jagung Madura," 2023. [Online]. Available: <https://jurnal.umj.ac.id/index.php/just-it/index>
- [12] E. Safitri, "Implementasi Algoritma K-Means Clustering Dalam Menentukan Strategi Marketing Dalam Penjualan Ikan (Studi Kasus: Grosir Ikan Tani Mas Tanjung Morawa)," *JIKTEKS : Jurnal Ilmu Komputer dan Teknologi Informasi*, vol. 02, no. 02, pp. 23–33, 2024, [Online]. Available: <jurnal.faatuatua.com/index.php/JIKTEKS>
- [13] H. Diogenes Adoe, A. Yuniar Rahman, and Istiadi, "Segmentasi Citra Burung Lovebird Menggunakan K-Means," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 10, no. 1, pp. 706–718, 2023, [Online]. Available: <http://dx.doi.org/10.30645/j-sakti.v4i2.236>
- [14] N. Silalahi, "Penentuan Strategi Promosi Universitas Budi Darma Menggunakan Algoritma K-Means Clustering," *Terapan Informatika Nusantara*, vol. 1, no. 1, pp. 40–46, 2020.
- [15] S. Handoko, F. Fauziah, and E. T. E. Handayani, "IMPLEMENTASI DATA MINING UNTUK MENENTUKAN TINGKAT PENJUALAN PAKET DATA TELKOMSEL MENGGUNAKAN METODE K-MEANS CLUSTERING," *Jurnal Ilmiah Teknologi dan Rekayasa*, vol. 25, no. 1, pp. 76–88, 2020, doi: 10.35760/tr.2020.v25i1.2677.
-