

# **UTS DATA MINING**

*Clustering Citra Udara Menggunakan K-means*



## **KELOMPOK 7**

Anggota Kelompok

Alif Al Husaini	140810220036
Darren Christian Liharja	140810220043
Jason Natanael Krisyanto	140810220051

**PROGRAM STUDI S-1 TEKNIK INFORMATIKA**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**UNIVERSITAS PADJADJARAN**

**JATINANGOR**

**2024**

## DAFTAR ISI

<b>DAFTAR ISI</b>	<b>1</b>
<b>PENDAHULUAN</b>	<b>2</b>
1.1 Latar Belakang	2
1.2 Rumusan Masalah	2
1.3 Tujuan	3
1.4 Manfaat	3
<b>METODOLOGI</b>	<b>4</b>
2.1 Business Understanding	4
2.3 Data Preparation	4
2.3.1 Proses Preprocessing	5
2.3.2 Transformasi Data	5
2.3.3 Ekstraksi Fitur	5
2.4 Modelling	5
2.5 Evaluation	6
2.6 Deployment	6
<b>HASIL &amp; PEMBAHASAN</b>	<b>7</b>
3.1 Proses Data Understanding	7
3.22 Transformasi Data	10
3.23 Ekstraksi Fitur	11
3.3 Modeling	12
3.4 Evaluation	15
3.5 Deployment	15
<b>KESIMPULAN &amp; SARAN</b>	<b>19</b>
4.1 Kesimpulan	19
4.2 Saran	19

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Pengolahan citra udara menjadi salah satu bidang yang semakin penting dalam berbagai sektor, mulai dari pemantauan lingkungan, pertanian, perencanaan tata kota, hingga mitigasi bencana. Dengan meningkatnya ketersediaan data citra dari satelit dan drone, analisis terhadap citra ini memungkinkan pengambilan keputusan yang lebih akurat dan efisien. Namun, seiring dengan besarnya volume dan variasi data yang diperoleh, muncul tantangan dalam mengolah dan menganalisis citra tersebut secara efektif.

Salah satu pendekatan yang banyak digunakan dalam pengolahan citra adalah metode *clustering*, yang bertujuan untuk mengelompokkan area berdasarkan kesamaan visual seperti warna, tekstur, dan fitur spasial. Metode ini dapat membantu dalam mengidentifikasi pola-pola serta menganalisis struktur data, seperti pada pengamatan ekosistem atau identifikasi jenis tutupan lahan. Berbagai algoritma *clustering*, seperti K-means, *Hierarchical Clustering*, dan DBSCAN, telah digunakan sebagai metode dalam mengelompokkan data citra dengan berbagai tingkat variasi.

Dengan meningkatnya ketersediaan citra udara dan kebutuhan akan analisis yang lebih tepat, penting untuk memahami penggunaan teknik *clustering* dalam pengolahan citra. Pendekatan ini tidak hanya dapat mendukung pemantauan lingkungan dan perencanaan kota, tetapi juga berkontribusi pada pengembangan aplikasi yang lebih efisien dalam analisis data. Seiring dengan kemajuan teknologi, tantangan baru muncul dalam memahami dan menginterpretasi data citra yang kompleks. Oleh karena itu, penerapan konsep data *preprocessing* dan *clustering* serta pembangunan aplikasi untuk visualisasi hasil clustering menjadi penting.

Untuk mengatasi tantangan ini, penelitian menggunakan metodologi CRISP-DM (*Cross-Industry Standard Process for Data Mining*) sebagai rangka standar. Metodologi ini telah terbukti efektif dalam berbagai penerapan *data mining* yang menyediakan panduan di setiap tahap proses analisis. Dari pemahaman konteks bisnis hingga evaluasi hasil, CRISP-DM membantu memastikan bahwa setiap langkah dilakukan secara proses dan sistematis.

### 1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana cara menerapkan konsep data *preprocessing* pada citra udara agar siap digunakan dalam algoritma *clustering*?
2. Bagaimana membangun aplikasi sederhana yang dapat memvisualisasikan hasil *clustering* dari citra udara dan memungkinkan pengguna untuk mengunggah citra serta mengatur parameter *clustering*?

3. Seberapa besar pengaruh variasi data citra udara terhadap kinerja algoritma clustering yang digunakan, dan metrik apa yang tepat untuk mengevaluasi kinerjanya?
4. Bagaimana hasil clustering dapat dianalisis dan diinterpretasikan dalam konteks aplikasi dunia nyata, serta apa implikasinya untuk pengambilan keputusan di bidang seperti pemantauan lingkungan atau perencanaan tata kota?

### **1.3 Tujuan**

Adapun tujuan dari penelitian ini adalah:

1. Menerapkan konsep data preprocessing dan clustering pada citra udara.
2. Membangun aplikasi sederhana untuk visualisasi hasil clustering.
3. Memahami pengaruh variasi data terhadap kinerja algoritma clustering.
4. Menganalisis dan menginterpretasi hasil clustering dalam konteks aplikasi dunia nyata.

### **1.4 Manfaat**

Beberapa manfaat yang didapatkan dari penelitian ini dapat dijabarkan sebagai berikut:

1. Peningkatan kualitas pengolahan citra dengan pemahaman mendalam tentang proses data preprocessing dan algoritma clustering, sehingga dapat meningkatkan kualitas analisis citra udara yang digunakan dalam berbagai aplikasi.
2. Pengembangan aplikasi praktis untuk visualisasi hasil clustering, memberikan alat yang berguna bagi pengguna untuk memahami dan menganalisis citra udara secara lebih interaktif dan efisien.
3. Peningkatan pengetahuan tentang clustering, membantu peneliti dan praktisi memahami bagaimana variasi dalam data citra udara memengaruhi kinerja algoritma clustering, sehingga dapat memilih pendekatan yang lebih tepat untuk kasus spesifik.
4. Aplikasi dalam berbagai sektor, seperti pemantauan lingkungan, perencanaan tata kota, dan pengelolaan sumber daya alam, memberikan kontribusi signifikan terhadap pengambilan keputusan berbasis data.
5. Kontribusi terhadap ilmu pengetahuan dengan menambah literatur di bidang pengolahan citra dan analisis data, serta menjadi referensi bagi penelitian-penelitian selanjutnya yang berkaitan dengan citra udara dan teknik analisis data.

## BAB II

### METODOLOGI

#### 2.1 *Business Understanding*

Pada tahap ini, tujuan utama adalah melakukan segmentasi citra terhadap gambar udara berdasarkan kategori yang telah ditentukan, yaitu *Desert*, *Beach*, *Forest*, dan *Residential*. Segmentasi citra bertujuan untuk mengidentifikasi dan mengelompokkan piksel dalam citra berdasarkan kesamaan warna, sehingga memudahkan analisis lebih lanjut. Dengan adanya kategori yang jelas, hasil segmentasi dapat memberikan wawasan tentang distribusi penggunaan lahan, kondisi lingkungan, serta membantu dalam pengambilan keputusan di berbagai sektor.

#### 2.2 *Data Understanding*

Pada tahap ini, dilakukan pemahaman terhadap data yang digunakan dalam penelitian, yakni Skyview Multi-Landscape Aerial Imagery Dataset. Dataset ini memiliki 15 kategori yang masing-masing terdiri atas 800 gambar. Dataset ini terdiri atas kategori yang mencakup *Agriculture*, *Airport*, *Beach*, *City*, *Desert*, *Forest*, *Grassland*, *Highway*, *Lake*, *Mountain*, *Parking*, *Port*, *Railway*, *Residential*, dan *River*. Setiap gambar memiliki resolusi 256x256 piksel, dengan total 12,000 gambar. Dataset ini merupakan penggabungan dari AID Dataset dan NWPU-Resisc45 Dataset. Setelah analisis awal, dipilih subset gambar yang menarik dan berpotensi mengungkap pola relevan. Dari 15 kategori yang ada, empat kategori utama dipilih, yaitu *Beach*, *Residential*, *Forest*, dan *Desert*. Dari setiap kategori, diambil 10 gambar dengan kemiripan komposisi visual. Subset ini akan digunakan untuk pengembangan model dan analisis lebih lanjut.

Selanjutnya, dilakukan pengecekan terhadap kualitas data. Pertama, setiap gambar dalam dataset diperiksa untuk memastikan dapat dibaca dengan benar. Proses ini melibatkan membaca setiap gambar dan menangani jika terdapat gambar yang bermasalah. Selain itu, jumlah gambar pada setiap kategori diperiksa untuk memastikan distribusi yang seimbang. Dengan menggunakan visualisasi, ditampilkan jumlah gambar per kategori untuk memberikan gambaran jelas mengenai keseimbangan data.

Setelah itu, contoh gambar dari beberapa kategori ditampilkan untuk memberikan gambaran visual terhadap *dataset*. Histogram warna juga dianalisis untuk mengeksplorasi distribusi warna dalam gambar, yang membantu dalam memahami karakteristik visual masing-masing kategori. Analisis ini juga mencari warna dominan dari setiap jenis gambar.

#### 2.3 *Data Preparation*

Tahap Data Preparation bertujuan untuk memastikan data dalam kondisi siap digunakan untuk analisis atau pemodelan lebih lanjut. Pada kasus ini, data yang

digunakan berupa gambar yang perlu diproses agar hasil klasterisasi menjadi optimal. Langkah-langkah *data preparation* yang dilakukan adalah sebagai berikut:

### 2.3.1 Proses Preprocessing

Gambar diproses terlebih dahulu untuk meningkatkan kualitas visualnya. Beberapa teknik *preprocessing* yang dilakukan antara lain *Denoising* untuk mengurangi noise dan memperjelas gambar, *Sharpening* untuk meningkatkan kontras dan detail untuk membantu menciptakan gambar yang lebih tajam, dan *Unsharp Masking* untuk menonjolkan fitur-fitur penting dalam gambar dengan menggabungkan hasil *denoise* dengan gambar aslinya. Hasil dari ketiga proses ini kemudian dibandingkan secara visual untuk menentukan metode mana yang memberikan hasil terbaik. Metode terbaik akan digunakan dalam proses selanjutnya.

### 2.3.2 Transformasi Data

Setelah gambar diproses, langkah berikutnya adalah mengubah gambar menjadi format yang dapat memudahkan dalam proses pengolahan data. Dalam hal ini, gambar yang telah diolah dikonversi menjadi *array* untuk proses analisis lebih lanjut.

### 2.3.3 Ekstraksi Fitur

Dalam tahap ini, fitur dari gambar diekstraksi dengan fokus pada informasi piksel dari gambar. Ekstraksi fitur tekstur dilakukan dengan menggunakan metode GLCM (*Gray Level Co-occurrence Matrix*). Metode ini menganalisis tekstur gambar berdasarkan distribusi spasial dari intensitas piksel yang berbeda. Properti-properti yang diekstrak meliputi *contrast*, *dissimilarity*, *homogeneity*, *energy*, dan *correlation*. GLCM diekstrak dengan beberapa parameter seperti jarak dan sudut untuk memberikan informasi yang lebih detail tentang pola tekstur dalam gambar.

Selain fitur tekstur, fitur warna dari setiap gambar juga diekstrak. Proses ini dilakukan dengan mengubah gambar menjadi array dua dimensi yang terdiri dari nilai RGB setiap piksel. Fitur warna ini kemudian dapat digunakan untuk pengelompokan berdasarkan warna dominan di setiap gambar. Setelah fitur tekstur dan fitur warna berhasil diekstraksi, langkah selanjutnya adalah menggabungkannya menjadi satu set fitur.

## 2.4 Modelling

Pada tahap modeling, algoritma KMeans diterapkan untuk melakukan segmentasi gambar dari dataset yang telah dikumpulkan. Proses dimulai dengan inisialisasi centroid, di mana centroid pertama dipilih secara acak dari data yang ada, dan centroid berikutnya ditentukan berdasarkan jarak dari centroid yang telah ada sebelumnya. Metode ini memastikan bahwa centroid terdistribusi dengan baik di ruang fitur.

Setelah centroid diinisialisasi, proses iterasi dimulai untuk mengelompokkan data. Pada setiap iterasi, setiap titik data dikelompokkan ke cluster terdekat dengan menghitung jarak Euclidean ke setiap centroid. Jarak ini dihitung untuk semua centroid yang ada, dan setiap titik data kemudian ditempatkan pada cluster yang memiliki centroid terdekat. Setelah pengelompokan dilakukan, centroid diperbarui berdasarkan rata-rata posisi semua titik yang terikat dalam cluster masing-masing. Proses ini diulang hingga perubahan posisi centroid berada di bawah ambang batas tertentu, menandakan bahwa model telah mencapai konvergensi.

Setelah model KMeans dilatih, gambar-gambar baru dapat disegmentasi menggunakan centroid yang telah ditentukan. Dalam tahap ini, setiap piksel pada gambar akan diwarnai sesuai dengan warna centroid dari cluster yang relevan, menghasilkan visualisasi segmentasi yang jelas. Hasilnya adalah gambar tersegmentasi yang menunjukkan struktur cluster yang terbentuk selama proses pelatihan.

## 2.5 Evaluation

Pada tahap evaluasi, *Silhouette Score* dihitung untuk menilai kualitas segmentasi yang dihasilkan oleh model KMeans. *Score ini* dapat menunjukkan seberapa baik setiap titik data dikelompokkan ke dalam cluster yang benar dibandingkan dengan cluster lainnya. Nilai Silhouette berkisar antara -1 hingga 1. Semakin tinggi nilainya, semakin baik pemisahan antara titik data dalam cluster tersebut dan titik data di cluster lain.

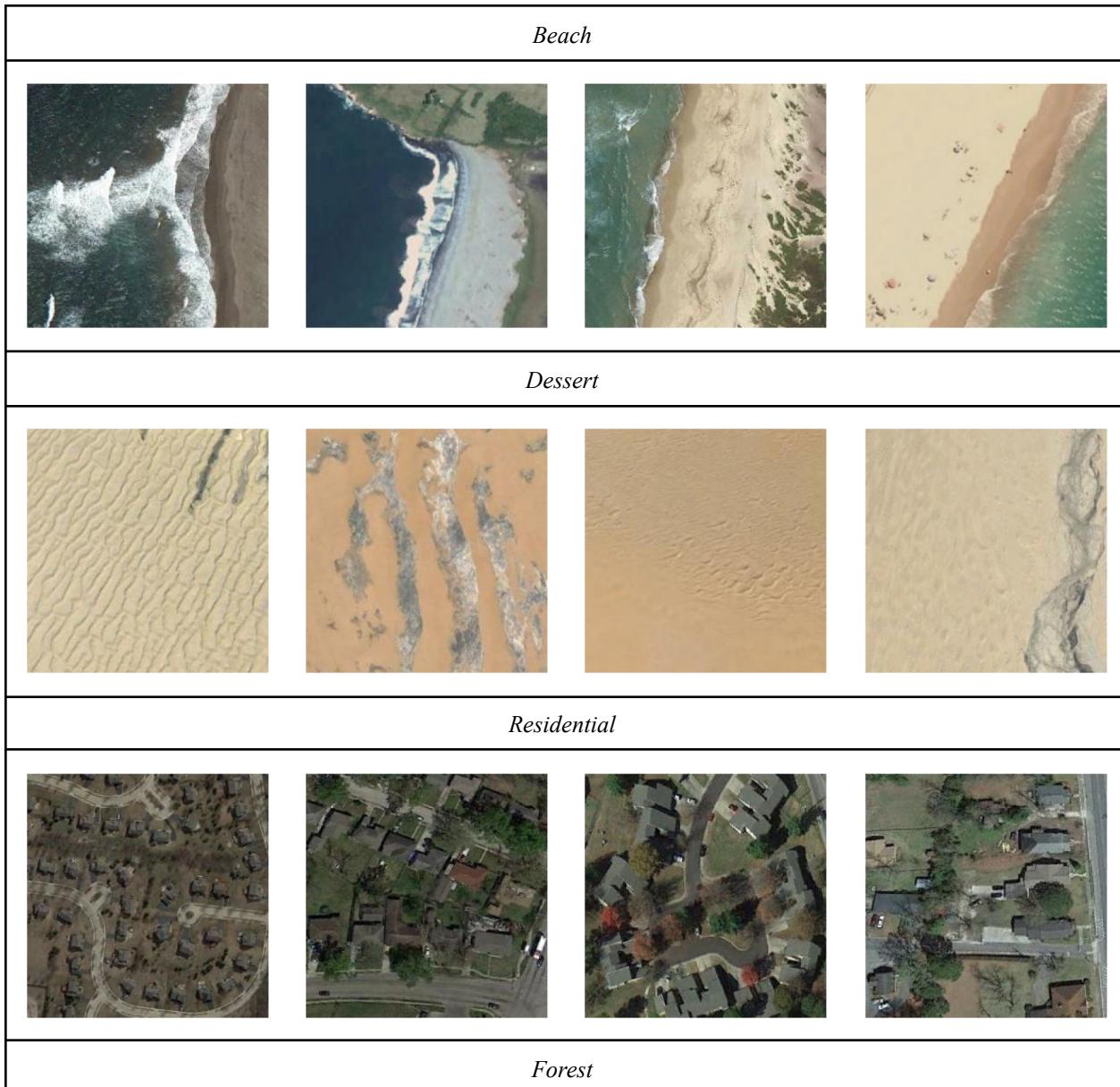
## 2.6 Deployment

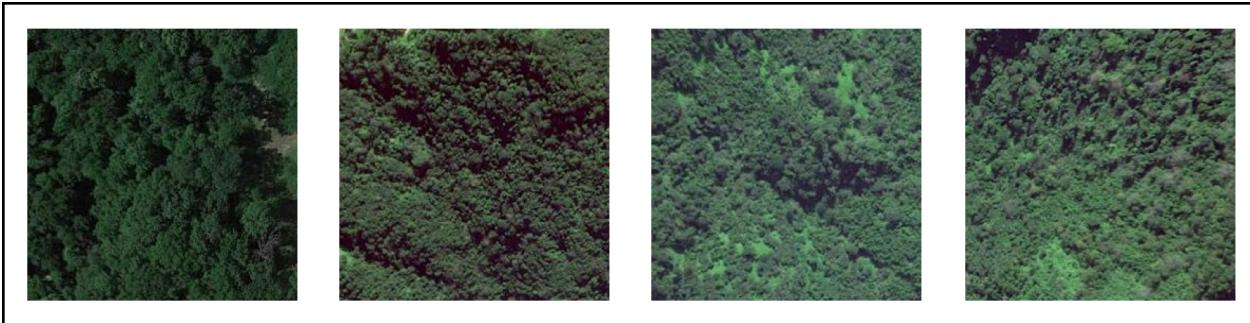
Pada tahap deployment, dashboard interaktif dibangun menggunakan Streamlit untuk memfasilitasi segmentasi gambar. Pengguna dapat mengunggah hingga 10 gambar untuk *training* dan maksimal 5 gambar untuk *testing*. Dashboard ini juga menyediakan slider yang memungkinkan pengguna memilih jumlah cluster yang diinginkan, antara 2 hingga 4, memberikan fleksibilitas dalam menentukan kategori untuk segmentasi gambar. Setelah pengunggahan, gambar-gambar pelatihan dan pengujian ditampilkan dengan jelas, disertai caption untuk memudahkan identifikasi. Setelah proses segmentasi menggunakan algoritma KMeans selesai, hasilnya ditampilkan untuk setiap gambar pengujian dalam bentuk visualisasi yang mencakup gambar yang telah disegmentasi dengan jelas, lengkap dengan detail yang menunjukkan warna dan cluster yang relevan.

## BAB III

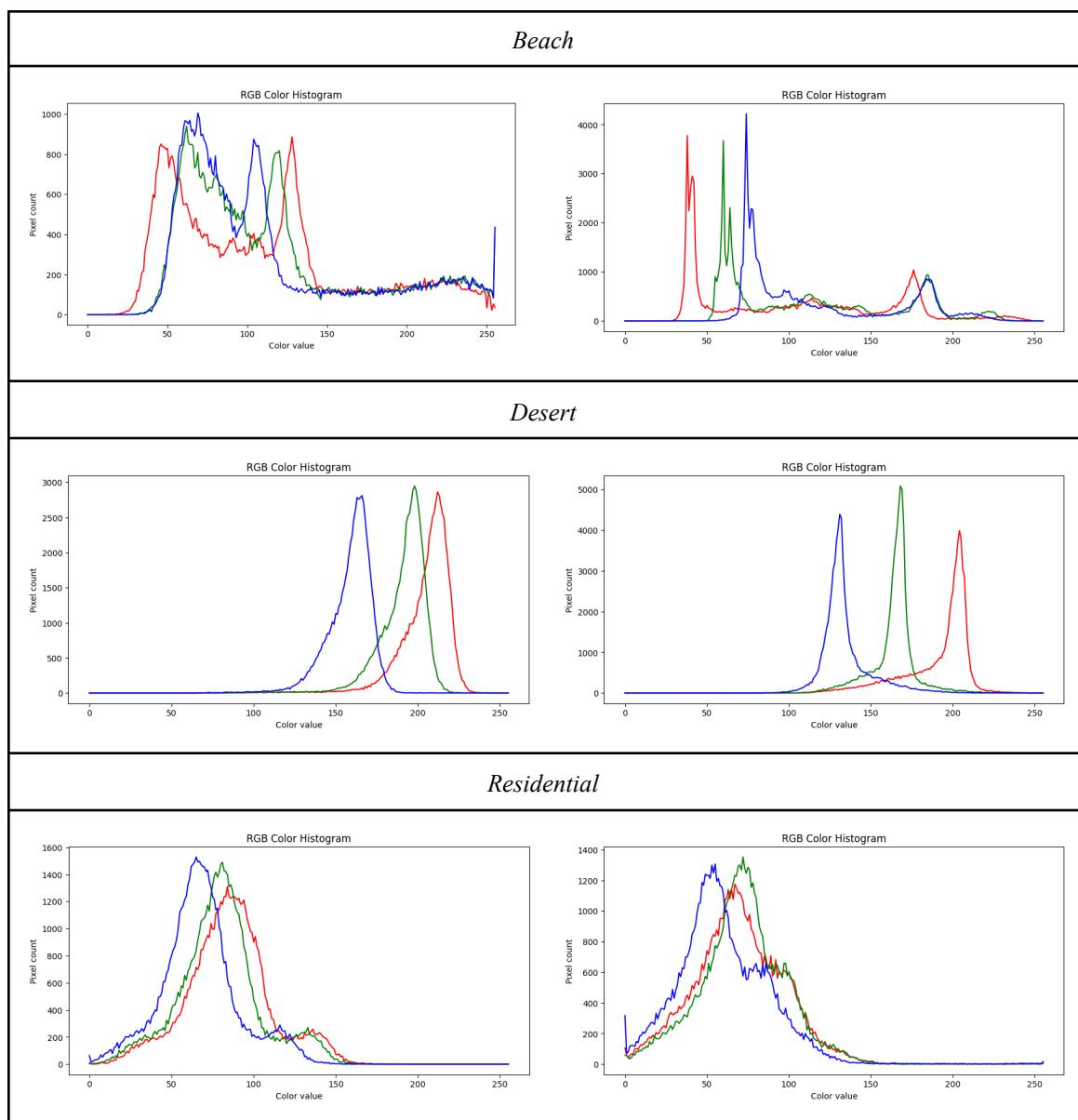
### HASIL & PEMBAHASAN

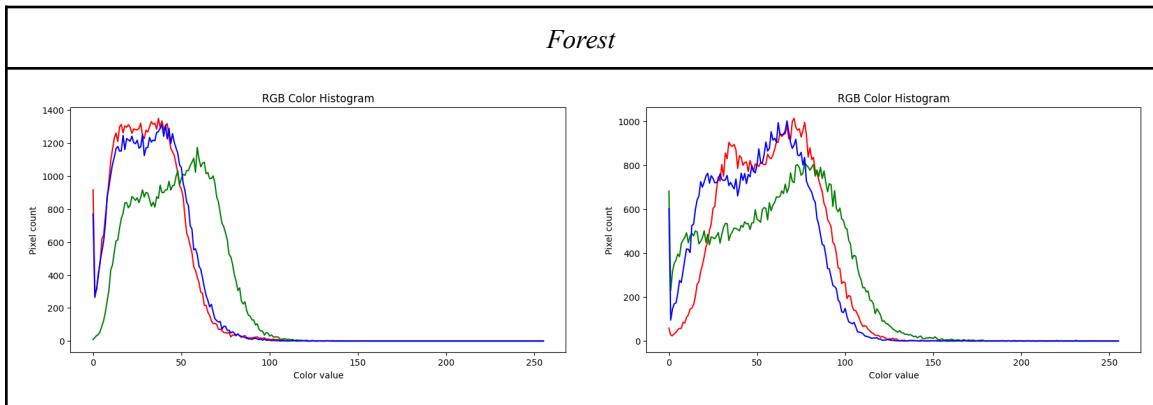
#### 3.1 Proses *Data Understanding*





**Tabel 1.** Sampel Dataset Setiap Kategori





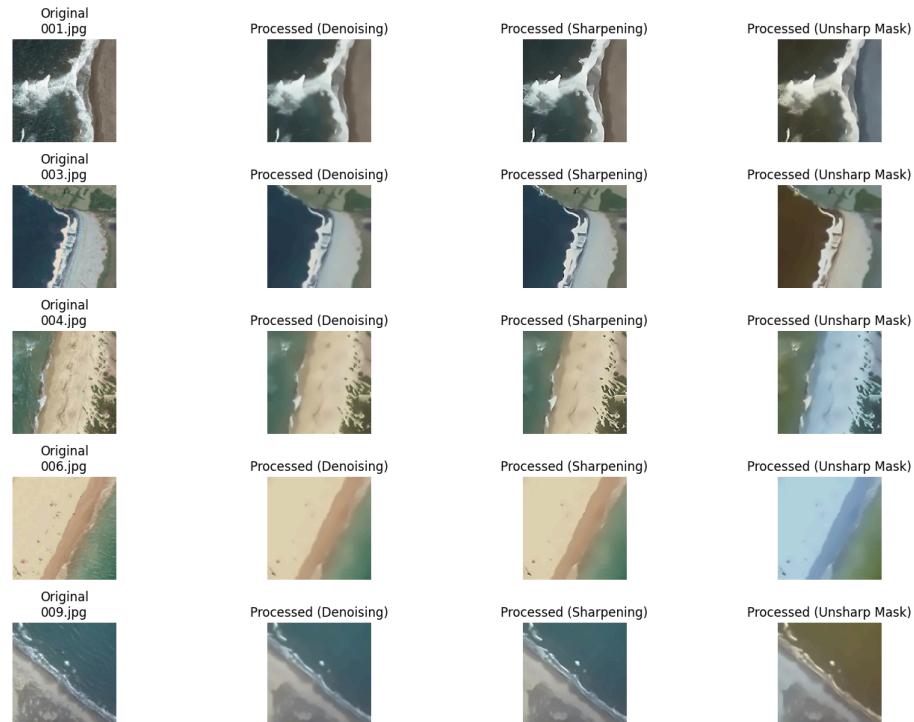
**Tabel 2.** Sampel Histogram Setiap Kategori



**Gambar 1.** Warna Dominan Setiap kategori

### 3.2 Proses Data Preparation

#### 3.2.1 Preprocessing



**Gambar 2.** Perbandingan Proses Preprocessing pada Beach

Dari hasil yang ditunjukkan pada Gambar 2, terlihat bahwa proses *preprocessing* menggunakan teknik sharpening (*sharpen kernel*) menghasilkan gambar dengan ketajaman dan kontras yang lebih baik dibandingkan dengan metode pra-pemrosesan lainnya. Oleh karena itu, keputusan dibuat untuk mengimplementasikan teknik *preprocessing* dengan sharpen kernel guna mencapai hasil klasterisasi yang optimal.

Dalam tahap pelatihan, gambar akan di-*resize* menjadi ukuran 75x75 piksel. Ukuran ini dipilih untuk mempercepat proses pelatihan, sehingga waktu yang dibutuhkan untuk mengolah gambar dalam model klasterisasi dapat diminimalkan. Dengan ukuran yang lebih kecil, model dapat beradaptasi lebih cepat dan melakukan iterasi lebih banyak dalam waktu yang sama.

Namun, saat melakukan segmentasi, gambar akan di-*resize* menjadi ukuran 200x200 piksel. Ukuran ini bertujuan untuk memberikan detail yang lebih baik dalam hasil segmentasi, karena gambar yang lebih besar dapat menangkap fitur dan struktur yang lebih kompleks, yang sangat penting dalam analisis gambar.

### 3.22 Transformasi Data

	Kategori	Gambar	Index	Ukuran	\
0	Beach		1	(75, 75, 3)	
1	Beach		2	(75, 75, 3)	
2	Beach		3	(75, 75, 3)	
3	Beach		4	(75, 75, 3)	
4	Beach		5	(75, 75, 3)	
5	Beach		6	(75, 75, 3)	
6	Beach		7	(75, 75, 3)	
7	Beach		8	(75, 75, 3)	
8	Beach		9	(75, 75, 3)	
9	Beach		10	(75, 75, 3)	
10	Beach		11	(75, 75, 3)	

	Array
0	[60, 64, 77, 59, 62, 69, 51, 59, 67, 46, 50, 6...
1	[43, 65, 80, 42, 64, 79, 48, 70, 83, 44, 66, 8...
2	[86, 107, 95, 84, 104, 93, 81, 102, 89, 78, 10...
3	[211, 187, 160, 210, 189, 157, 211, 198, 164, ...]
4	[93, 112, 116, 90, 109, 113, 79, 99, 104, 87, ...]
5	[212, 195, 151, 211, 194, 150, 214, 197, 153, ...]
6	[213, 194, 174, 224, 209, 187, 232, 213, 187, ...]
7	[120, 103, 89, 123, 106, 86, 124, 108, 88, 129...]
8	[21, 33, 33, 24, 42, 35, 36, 52, 38, 41, 61, 4...
9	[79, 102, 87, 70, 93, 85, 72, 94, 80, 66, 91, ...]
10	[51, 71, 78, 53, 73, 80, 47, 67, 74, 45, 66, 7...

**Gambar 3.** Contoh Hasil Transformasi Gambar ke Array pada Kategori Beach

Gambar 3 menyajikan informasi mengenai gambar yang telah diproses dan dikategorikan. Setiap gambar termasuk dalam kategori Beach, dengan nomor indeks yang memudahkan identifikasi, mulai dari 1 hingga 11. Ukuran gambar yang terdaftar adalah (75, 75, 3), yang menunjukkan bahwa semua gambar telah di-resize menjadi 75 piksel tinggi dan lebar, serta memiliki 3 saluran warna (RGB). Ukuran ini dipilih untuk mempercepat proses pelatihan model tanpa mengorbankan terlalu banyak informasi visual. Selain itu, kolom representasi array menampilkan nilai piksel setiap gambar, hasil dari pemrosesan yang mencakup teknik pengurangan noise dan penajaman gambar. Data ini akan digunakan dalam analisis lebih lanjut, seperti klasterisasi atau pelatihan model machine learning.

### 3.23 Ekstraksi Fitur

Ekstraksi Fitur untuk Gambar Beach 1:			
	Total number of pixels/features extracted: 5625		
	Red	Green	Blue
0	60	64	77
1	59	62	69
2	51	59	67
3	46	50	62
4	46	51	57
...	...	...	...
5620	125	117	104
5621	122	114	101
5622	124	116	103
5623	121	113	100
5624	127	119	106
[5625 rows x 3 columns]			

Gambar 4. Contoh Hasil Ekstraksi Pixel Gambar pada Gambar 1 Kategori Beach

Angka 5625 pada ekstraksi fitur berasal dari jumlah total piksel yang diekstraksi dari gambar Beach 1 yang telah diproses. Gambar tersebut memiliki ukuran 75x75 piksel, sehingga total piksel dalam gambar dapat dihitung dengan mengalikan lebar dan tinggi gambar:

$$75 \text{ (lebar)} \times 75 \text{ (tinggi)} = 5625 \text{ piksel}$$

Setiap piksel terdiri dari tiga nilai intensitas warna untuk saluran merah, hijau, dan biru (RGB). Oleh karena itu, hasil ekstraksi menghasilkan total 5625 fitur yang mewakili masing-masing piksel dalam gambar. Proses ini penting untuk analisis lebih lanjut, seperti klasterisasi dan pelatihan model machine learning, di mana pola dan karakteristik visual dalam gambar dapat

diidentifikasi dan dipelajari. Dengan mengonversi informasi gambar menjadi format numerik yang terstruktur, langkah-langkah selanjutnya dalam analisis dapat dilakukan secara lebih efisien dan akurat, memungkinkan pemodelan yang lebih baik untuk pengenalan pola dan klasifikasi gambar. Hasil ekstraksi fitur untuk gambar ini dapat dilihat pada Gambar 4.

### 3.3 Modeling

Tahap selanjutnya, kami mencoba menerapkan algoritma KMeans secara manual untuk melatih model dengan gambar pada kategori *Beach* dengan jumlah *cluster* (*k*) sebanyak 3. Pada tahap ini, semua fitur piksel dari setiap gambar dalam kategori tersebut digabungkan, dan centroid dicari berdasarkan nilai *k* yang ditentukan. Fokus utama dari percobaan ini adalah menemukan centroid yang akan digunakan dalam proses segmentasi gambar selanjutnya, sehingga memungkinkan pengelompokan yang lebih efisien berdasarkan karakteristik visual yang ada.

```
Total number of pixels/features extracted: 5625
Initial centroid 1: [223 206 180]
Initial centroid 2: [209 194 174]
Initial centroid 3: [76 96 88]
Initial centroids calculated.
Centroids found:
[[204.60743258 196.56389892 173.04336377]
 [138.19833979 138.59573879 121.56325401]
 [ 68.92670286  85.88677196  81.67566634]]
```

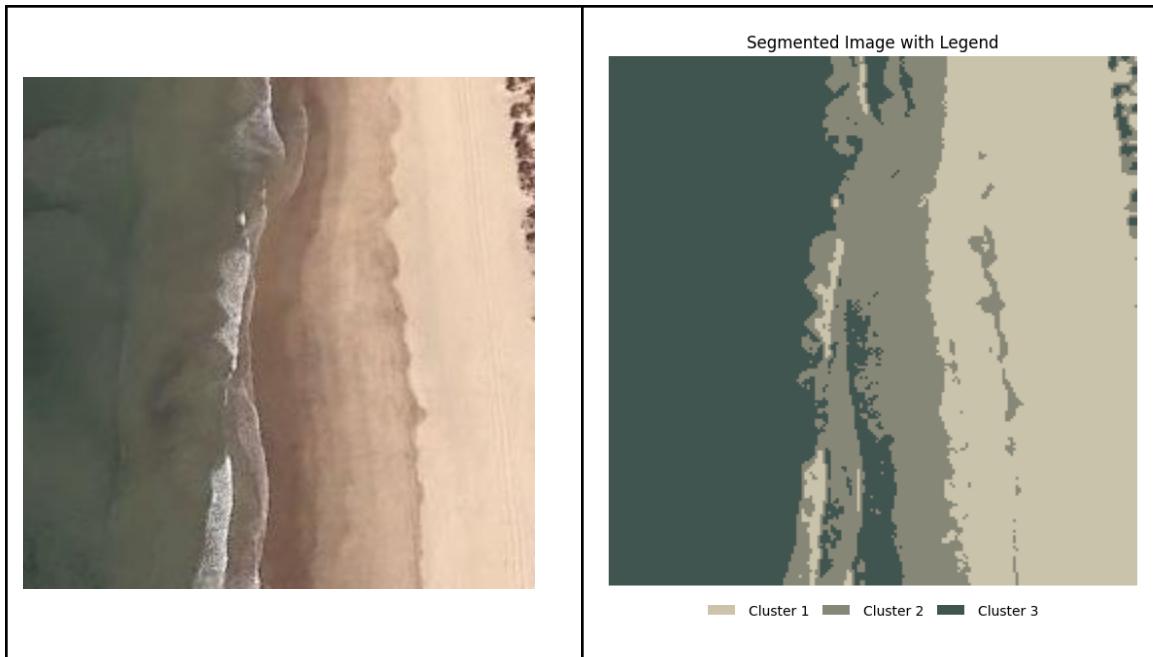
Gambar 5. Hasil Pencarian Centroid untuk 3 Klaster pada Kategori *Beach*

Setelah proses ekstraksi fitur, total 5625 piksel atau fitur berhasil diambil dari tiap gambar dalam kategori *Beach*. Proses ini memberikan informasi penting mengenai nilai intensitas warna untuk setiap saluran merah, hijau, dan biru (RGB). Selanjutnya, dilakukan inisialisasi *centroid* awal yang dihasilkan, yaitu *centroid* 1 dengan nilai [223, 206, 180], *centroid* 2 dengan nilai [209, 194, 174], dan *centroid* 3 dengan nilai [76, 96, 88]. Setelah menjalankan algoritma KMeans manual untuk melatih model dalam mencari *centroid*, *centroid* yang ditemukan adalah [204.61, 196.56, 173.04], [138.20, 138.60, 121.56], dan [68.93, 85.89, 81.68]. Nilai *centroid* ini mencerminkan pusat dari

masing-masing *cluster* yang akan digunakan dalam proses segmentasi gambar lebih lanjut, sehingga memungkinkan pengelompokan gambar berdasarkan kesamaan fitur visualnya.

Selanjutnya, dari *centroid* yang telah didapatkan, akan dilakukan pengujian dengan gambar kategori *Beach* lain yang tidak digunakan sebagai objek pelatihan dalam pencarian *centroid*. Segmentasi dilakukan dengan memanfaatkan *centroid* yang telah dihitung untuk memberikan warna pada *pixel* gambar sesuai dengan warna dominan dari setiap *cluster*. Proses ini dimulai dengan memuat gambar dan mengekstrak fiturnya, kemudian menghitung jarak setiap piksel ke *centroid* yang ada. Dengan metode ini, setiap piksel akan diberi label berdasarkan *cluster* terdekat, dan gambar hasil segmentasi akan ditampilkan dengan warna yang sesuai dengan *centroid* yang ditentukan. Sebagai contoh, *cluster* 1 pada gambar 1 akan mewakili objek atau area yang sama dengan *cluster* 1 pada gambar 2. Setiap *cluster* akan memberikan representasi visual yang konsisten dan memungkinkan perbandingan antar gambar yang lebih intuitif.

Gambar Asli	Hasil Segmentasi
	<p>Segmented Image with Legend</p>  <p>Cluster 1 Cluster 2 Cluster 3</p>



Tabel 3. Perbandingan Gambar Asli dengan Hasil Segmentasi

Pada Tabel 3, terlihat bahwa kedua gambar berhasil disegmentasi sesuai dengan model pra-latih yang ditetapkan, yaitu menghasilkan tiga *cluster*. Hasil segmentasi menunjukkan bahwa *cluster* 1 pada gambar pertama dan *cluster* 1 pada gambar kedua mewakili area pasir, dengan penampilan yang jelas dan konsisten pada kedua gambar. Selanjutnya, *cluster* 2 pada kedua gambar menunjukkan kesamaan, dengan *cluster* ini mewakili perairan dangkal yang masih memperlihatkan pasir. *Cluster* 3 mencerminkan bagian air yang lebih dalam, tanpa keberadaan pasir yang terlihat. Setiap *cluster* tidak hanya menggambarkan elemen dalam setiap gambar, tetapi juga memberikan representasi visual yang seragam. Hal ini memungkinkan perbandingan antar gambar menjadi lebih intuitif dan efektif, karena *cluster* yang sama pada gambar yang berbeda akan mewakili objek atau area yang serupa.

Namun, penting untuk diperhatikan bahwa data yang digunakan untuk pengujian harus sesuai dengan model yang telah dilatih. Misalnya, jika *centroid* dilatih menggunakan data dari kategori *beach*, maka data uji juga harus berasal dari kategori *beach*. Apabila ingin melakukan segmentasi pada tema lain, seperti *residential*, maka perlu dilakukan pelatihan terlebih dahulu untuk mencari *centroid* yang sesuai dari gambar-gambar *residential* tersebut. Dengan hasil ini, analisis lebih lanjut dapat dilakukan dengan lebih mudah, terutama dalam konteks pengelompokan dan pengenalan pola dalam gambar, berdasarkan *centroid* yang dihasilkan

### 3.4 Evaluation

```
Total number of pixels/features extracted: 40000  
Silhouette Score: 0.581
```

```
Total number of pixels/features extracted: 40000  
Silhouette Score: 0.628
```

Gambar 6. Hasil *Silhouette Score* Terhadap Dua Buah Data Testing

Berdasarkan hasil evaluasi menggunakan *Silhouette Score* terhadap dua data uji yang ditampilkan pada Tabel 3, diperoleh nilai sebesar 0.581 dan 0.628. Nilai ini berada di antara 0 dan 1, yang menunjukkan bahwa model KMeans mampu melakukan segmentasi dengan cukup baik dan gambar-gambar berhasil dikelompokkan ke dalam cluster yang memiliki pemisahan cukup jelas. Dalam proses analisis ini, total 40,000 fitur piksel diekstraksi dari gambar untuk memberikan gambaran yang lebih rinci dan membantu model dalam mendeteksi pola visual yang ada. Jumlah fitur yang besar ini memungkinkan identifikasi yang lebih akurat terhadap area dengan karakteristik visual serupa, yang pada akhirnya meningkatkan kualitas segmentasi.

### 3.5 Deployment

*Deployment* aplikasi ini dilakukan dengan bantuan salah satu library Python yaitu Streamlit yang dapat memudahkan proses pembuatan dan penyebaran aplikasi web secara cepat dan efisien. Untuk *source code* lengkap dari aplikasi ini dapat ditemukan pada *repository* GitHub berikut: [GitHub - Data Mining UTS](#). Pada *repository* tersebut, dapat ditemukan semua *file* dan dokumentasi terkait *development* aplikasi, termasuk dependensi yang digunakan dalam *file requirements.txt*.

Hasil *deploy* dari aplikasi ini dapat diakses melalui link berikut: [UTS Data Mining Kelompok 7](#). Melalui link ini, pengguna dapat mencoba berbagai fitur yang telah disediakan oleh aplikasi. Berikut adalah fitur-fitur yang terdapat pada aplikasi web:

- Upload Gambar Training dan Testing

Pengguna dapat mengunggah hingga 10 gambar untuk *training* dan 5 gambar untuk *testing*. Fitur ini memungkinkan pengguna untuk melatih model *clustering* pada dataset yang lebih besar dan menguji performanya pada data yang terpisah. Gambar yang diunggah dapat berupa format JPG, JPEG, atau PNG.

- Pengaturan KMeans *Clustering*

Aplikasi ini juga menyediakan *slider* interaktif yang memungkinkan pengguna untuk memilih jumlah *cluster* yang diinginkan, mulai dari 2 hingga 4 *cluster*. Dengan pengaturan ini, pengguna dapat menyesuaikan kompleksitas *clustering* sesuai dengan kebutuhan analisis mereka.

Berikut adalah dokumentasi alur penggunaan aplikasi:

# Dashboard Kluster Citra Udara

## Upload Gambar Training

Pilih hingga 10 gambar

Drag and drop files here  
Limit 200MB per file • JPG, JPEG, PNG

Browse files

	212.jpg 9.0KB	X
	072.jpg 11.4KB	X
	042.jpg 13.6KB	X

Showing page 1 of 4 < >

	001.jpg		003.jpg		004.jpg		006.jpg		009.jpg		010.jpg		028.jpg		042.jpg		072.jpg		212.jpg
--	---------	--	---------	--	---------	--	---------	--	---------	--	---------	--	---------	--	---------	--	---------	--	---------

Gambar 7. Upload Gambar Training

## Upload Gambar Testing

Pilih hingga 5 gambar

Drag and drop files here  
Limit 200MB per file • JPG, JPEG, PNG

Browse files

X X X

	009.jpg 7.1KB	X
	006.jpg 6.1KB	X
	004.jpg 12.2KB	X

Showing page 1 of 2 < >

	001.jpg		003.jpg		004.jpg		006.jpg		009.jpg
--	---------	--	---------	--	---------	--	---------	--	---------

Gambar 8. Upload Gambar Testing

## Pengaturan Klustering KMeans

Pilih jumlah cluster (2-4):

2

•

2

4

Gambar 9. Pengaturan jumlah cluster

Proses training selesai!

### Gambar Testing yang Diupload

Gambar Pengujian 1      Gambar Pengujian 2      Gambar Pengujian 3      Gambar Pengujian 4      Gambar Pengujian 5

### Hasil Segmentasi Gambar Pengujian Berdasarkan Klustering

Segmentasi Gambar Pengujian 1      Segmentasi Gambar Pengujian 2      Segmentasi Gambar Pengujian 3      Segmentasi Gambar Pengujian 4      Segmentasi Gambar Pengujian 5

Cluster 1      Cluster 2

Proses pengujian selesai!

**Gambar 10.** Hasil Segmentasi Menggunakan 2 *Cluster*

### Gambar Testing yang Diupload

Gambar Pengujian 1      Gambar Pengujian 2      Gambar Pengujian 3      Gambar Pengujian 4      Gambar Pengujian 5

### Hasil Segmentasi Gambar Pengujian Berdasarkan Klustering

Segmentasi Gambar Pengujian 1      Segmentasi Gambar Pengujian 2      Segmentasi Gambar Pengujian 3      Segmentasi Gambar Pengujian 4      Segmentasi Gambar Pengujian 5

Cluster 1      Cluster 2      Cluster 3

Proses pengujian selesai!

**Gambar 11.** Hasil Segmentasi Menggunakan 3 *Cluster*

## Gambar Testing yang Diupload



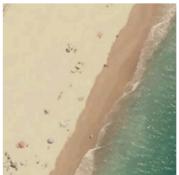
Gambar Pengujian 1



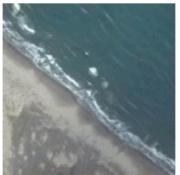
Gambar Pengujian 2



Gambar Pengujian 3

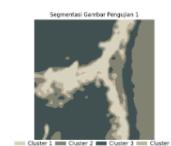


Gambar Pengujian 4

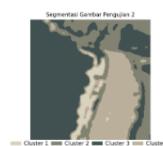


Gambar Pengujian 5

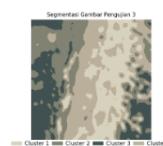
## Hasil Segmentasi Gambar Pengujian Berdasarkan Klustering



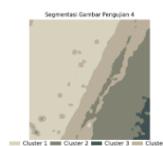
Segmentasi Gambar Pengujian 1



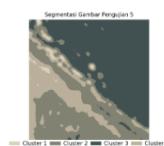
Segmentasi Gambar Pengujian 2



Segmentasi Gambar Pengujian 3



Segmentasi Gambar Pengujian 4



Segmentasi Gambar Pengujian 5

Proses pengujian selesai!

**Gambar 12.** Hasil Segmentasi Menggunakan 4 Cluster

## **BAB IV**

### **KESIMPULAN & SARAN**

#### **4.1 Kesimpulan**

Berdasarkan hasil percobaan, dapat disimpulkan bahwa metode *clustering*, khususnya KMeans, terbukti efektif dalam pengolahan citra untuk mengelompokkan area berdasarkan kesamaan visual. Penggunaan teknik data preprocessing yang tepat, seperti *sharpening* dan *resizing* gambar, memungkinkan model KMeans menghasilkan segmentasi yang baik, sebagaimana terlihat dari nilai *Silhouette Score*. Penerapan *preprocessing* dan *clustering* yang optimal ini membantu meningkatkan akurasi dalam analisis citra, yang sangat penting untuk aplikasi seperti pemantauan lingkungan dan perencanaan tata kota.

Penelitian ini juga menekankan pentingnya memahami dampak variasi data terhadap kinerja algoritma *clustering* dan bagaimana hasil segmentasi dapat diinterpretasikan dalam konteks aplikasi dunia nyata. Penggunaan metodologi CRISP-DM memberikan kerangka kerja yang sistematis dan terstruktur, memastikan setiap langkah dalam proses analisis citra dilakukan dengan efisien dan konsisten. Dengan meningkatnya ketersediaan data citra dari satelit dan *drone*, pendekatan ini menawarkan solusi untuk tantangan besar dalam mengelola dan menganalisis volume data yang kompleks, memungkinkan pengambilan keputusan yang lebih akurat dan berbasis data.

#### **4.2 Saran**

Untuk pengembangan lebih lanjut dalam penelitian dan analisis clustering citra udara, disarankan untuk mempertimbangkan penggunaan algoritma *clustering* alternatif seperti *Agglomerative Hierarchical Clustering (AHC)* dan *DBSCAN (Density-Based Spatial Clustering of Applications with Noise)*. AHC memiliki keunggulan dalam menggambarkan struktur hierarki di antara data citra, sementara DBSCAN efektif dalam mengidentifikasi cluster dengan bentuk yang tidak teratur dan mampu menangani *outliers* dengan baik, yang sering ditemui dalam analisis citra.

Selain itu, penerapan hasil clustering dalam studi kasus nyata seperti pemantauan perubahan vegetasi, pengawasan wilayah perkotaan untuk perencanaan tata kota, serta deteksi dini bencana alam seperti banjir dan kebakaran hutan, dapat memperluas manfaat dari penelitian ini dalam aplikasi dunia nyata. Penelitian lebih lanjut tentang dampak kualitas citra, seperti resolusi dan keberadaan *noise*, terhadap performa algoritma *clustering* juga penting dilakukan, disertai dengan penerapan teknik peningkatan kualitas gambar untuk memastikan data yang digunakan lebih optimal sebelum diproses.