

DATA MINING
Mini Project Clustering



Disusun Oleh:

<i>Tegar Muhamad Rizki</i>	<i>140810220034</i>
<i>Sanjukin Ndube Pinem</i>	<i>140810220050</i>
<i>Fahri Nizar Argubi</i>	<i>140810220054</i>

UNIVERSITAS PADJADJARAN
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
PROGRAM STUDI S1 TEKNIK INFORMATIKA
JATINANGOR
2024

Pendahuluan

Pengolahan citra udara memainkan peran penting dalam pemantauan lingkungan dan perubahan lanskap secara real-time. Salah satu teknik yang umum digunakan untuk memahami pola dan perubahan dalam citra udara adalah **KMeans Clustering**, metode unsupervised learning yang memetakan piksel citra ke dalam beberapa kelompok (cluster) berdasarkan kesamaan warna atau intensitas.

Pada proyek ini, kami menggunakan citra udara yang merepresentasikan dua skenario lingkungan: **pegunungan** dan **deforestasi hutan**. Proses analisis diawali dengan membagi data citra menjadi dua set: **training set** dan **testing set**. Setiap citra diubah ke dalam format warna **RGB** untuk mengekstraksi fitur-fitur warna yang signifikan. Pada training set, algoritma KMeans digunakan untuk membentuk cluster berdasarkan centroid yang merepresentasikan berbagai objek di dalam citra, seperti area hutan, dan wilayah deforestasi.

Setelah training, nilai centroid yang diperoleh digunakan untuk melakukan prediksi pada **testing set**, dimana setiap piksel pada citra testing dikelompokkan ke dalam cluster yang paling sesuai dengan centroid dari training. Pendekatan ini memungkinkan kita untuk mengidentifikasi pola perubahan lanskap, seperti degradasi hutan atau pergeseran area pertanian, dengan menggunakan data berbasis warna dan tekstur pada citra udara.

Proyek ini bertujuan untuk mengembangkan model segmentasi citra yang akurat, yang dapat memetakan perubahan lingkungan, membantu dalam pengambilan keputusan berbasis data, serta memberikan visualisasi yang lebih jelas tentang dampak perubahan lingkungan dalam jangka panjang.

Metodologi

Proyek ini terdiri dari beberapa tahapan utama, yaitu pengumpulan data, praproses data, ekstraksi fitur, proses clustering menggunakan algoritma KMeans, dan evaluasi hasil clustering. Metodologi yang digunakan dirinci sebagai berikut:

1. **Pengumpulan Data**

Data yang digunakan dalam proyek ini berupa citra udara yang diperoleh dari sumber-sumber terbuka di internet. Citra tersebut terdiri dari dua skenario utama, yaitu **pegunungan** dan **deforestasi hutan**. Setiap citra menggambarkan area yang mengalami perubahan lingkungan, yang selanjutnya akan diolah untuk mendeteksi perubahan pola di dalamnya.

2. **Praproses Data**

Sebelum proses clustering dilakukan, setiap citra udara menjalani tahap praproses sebagai berikut:

- **Resize Citra:** Setiap citra diubah ukurannya menjadi resolusi yang seragam untuk memudahkan pemrosesan lebih lanjut. Misalnya, setiap citra diubah menjadi ukuran 256x256 piksel.
- **Konversi ke Array RGB:** Setelah diresize, setiap citra dikonversi ke dalam format **array RGB**. Setiap piksel dalam citra diwakili oleh nilai intensitas warna merah (R), hijau (G), dan biru (B), yang nantinya digunakan sebagai fitur dalam proses clustering.

3. **Ekstraksi Fitur**

Ekstraksi fitur dilakukan dengan mengambil nilai **RGB** dari setiap piksel pada citra. Fitur RGB ini menjadi representasi sederhana dari warna dan tekstur citra, yang akan digunakan untuk memisahkan area yang berbeda di dalam gambar. Dalam tahap ini, tidak dilakukan teknik ekstraksi fitur tambahan, melainkan fokus pada informasi warna yang ada.

4. **Clustering Menggunakan KMeans**

Setelah fitur RGB diperoleh, algoritma **KMeans Clustering** diterapkan untuk membagi citra ke dalam beberapa cluster berdasarkan kesamaan warna piksel. Tahapan clustering meliputi:

- Menentukan jumlah cluster **k** berdasarkan observasi awal dan eksperimen.

- Menjalankan algoritma KMeans untuk mencari **centroid** dari setiap cluster pada **training set**. Centroid ini merepresentasikan pusat dari masing-masing kelompok piksel yang memiliki kesamaan warna.
- Menggunakan nilai centroid tersebut untuk melakukan prediksi pada **testing set**, di mana setiap piksel akan diberi label berdasarkan kedekatan dengan salah satu centroid.

5. Evaluasi Hasil Clustering

Setelah proses clustering selesai, hasilnya dievaluasi menggunakan **Silhouette Score**, yang merupakan metode evaluasi internal untuk menilai seberapa baik piksel-piksel dalam satu cluster saling berdekatan, dan seberapa jauh mereka dari cluster yang lain. Silhouette Score dihitung untuk memastikan bahwa hasil clustering dapat memisahkan area-area yang berbeda dalam citra dengan baik.

Hasil clustering juga akan divisualisasikan untuk memberikan gambaran visual mengenai segmentasi area penggabungan lahan dan deforestasi hutan pada citra udara.

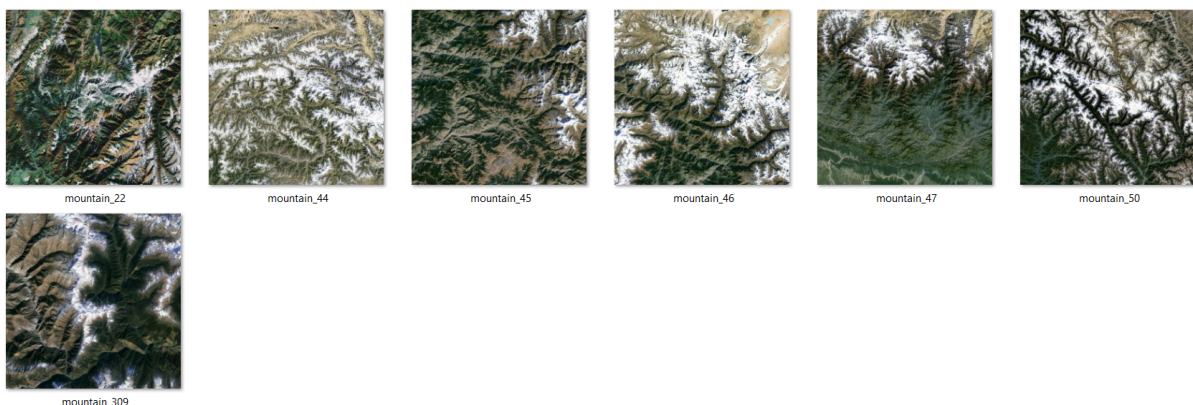
Hasil dan pembahasan

Setelah melakukan proses clustering KMeans yang diterapkan ke citra udara gunung bersalju dan juga hutan yang terdegradasi. Proses KMeans clustering diterapkan dengan jumlah berbeda untuk gunung dan hutan.

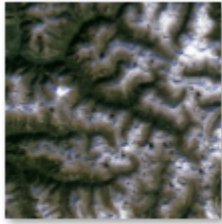
1. Clustering pada citra udara gunung bersalju

Data :

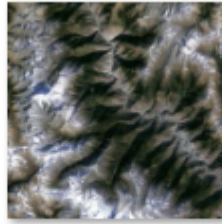
a. Training



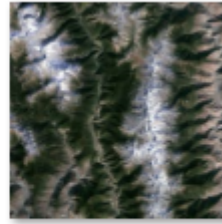
b. Testing



mountain_314



mountain_315



mountain_316

Memisahkan antara area yang bersalju, area hijau atau ada vegetasi, dan area yang tidak ada vegetasi. Berikut ini cluster yang ada:

- Cluster 1: Area bersalju
- Cluster 2: Area yang tidak ada vegetasi
- Cluster 3: Area yang terdapat vegetasi (pepohonan)

Nilai silhouette untuk clustering citra udara gunung berkisar antara: 0.5222, 0.5679, dan 0.6030. Menunjukkan hasil clustering yang cukup baik.

2. Clustering pada citra udara hutan terdegradasi

Data :

a. Training



{FF1CE564-C7CD
-463A-9EBE-BA4
F95800C33}



f1



f2



f3



f4



f5

b. Testing



{60D8C4DA-F603
-479F-A0AE-9E3
7D06391DB}



{B8CF2750-AB3B
-4DBE-8F34-82E
B70EE242A}



f7



f8

Memisahkan antara area yang memiliki vegetasi dengan area yang sudah terdegradasi. Berikut ini cluster yang ada:

- Cluster 1: Area yang terdegradasi

- Cluster 2: Area yang hijau atau padat vegetasi

Nilai silhouette score untuk clustering citra udara hutan terdegradasi berkisar antara: 0.8534, 0.6205, 0.5767, dan 0.7581. Untuk nilai yang 0.8534, itu lebih dari 90% gambarnya berupa area hijau.

Berdasarkan hasil di atas, clustering pada citra hutan terdegradasi memberikan hasil yang lebih baik dibandingkan citra gunung. Hal ini dapat disebabkan oleh perbedaan visual yang lebih tajam antara area yang masih padat vegetasi dengan area yang terdegradasi. Sebaliknya citra gunung memiliki perbedaan yang lebih halus antara area yang bervegetasi, area non-vegetasi, dan area bersalju. Sehingga cluster untuk citra udara gunung lebih sulit dipisahkan secara tegas dibandingkan dengan citra hutan terdegradasi.

Kesimpulan dan saran

Pengujian KMeans clustering pada citra gunung dan hutan menunjukkan hasil yang berbeda. Pada citra gunung menghasilkan silhouette score sekitar 0.5 - 0.6. Sedangkan pada citra hutan terdegradasi, hasil clustering lebih baik dengan silhouette score sekitar 0.55 - 0.85.

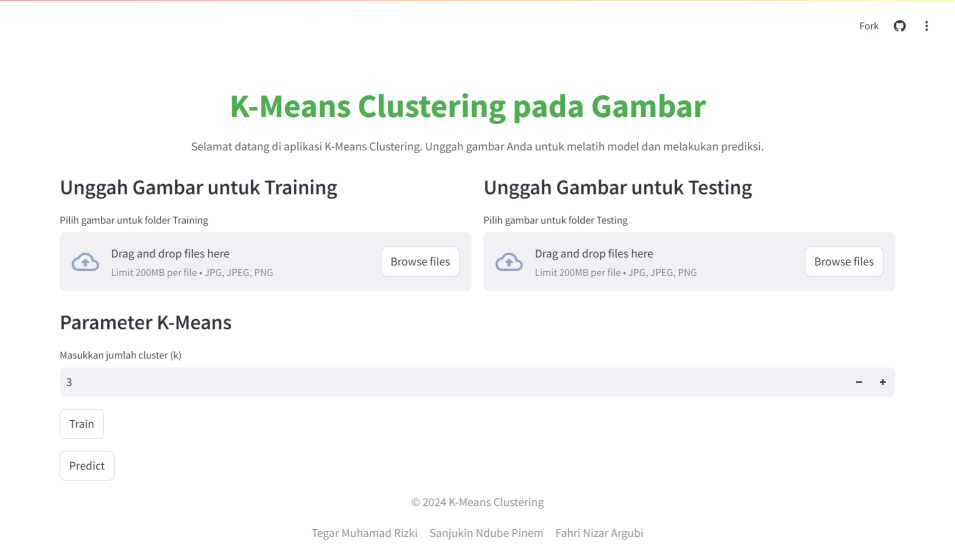
Hal ini menunjukkan perbedaan visual yang lebih jelas antara area dalam gambar. Adanya kontras yang tajam antara area hijau dan area degradasi pada citra hutan memudahkan algoritma KMeans mengelompokkan dengan lebih akurat.

Saran untuk mendapatkan nilai K yang optimal pada metode KMeans clustering, diperlukan pendekatan yang lebih adaptif daripada menentukan K secara manual. Metode yang dapat digunakan, seperti metode elbow.

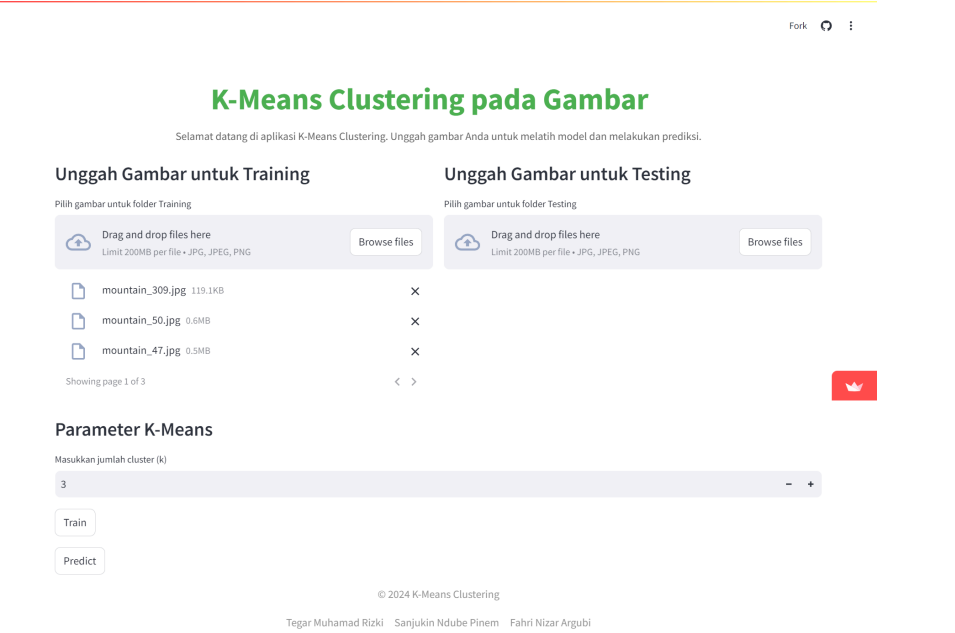
Screenshot implementasi aplikasi

Link : <https://datminproj.streamlit.app/>

1. Tampilan awal website:



2. Unggah gambar untuk training: Klik **browse files** yang ada dibagian training
- a. gunung




- b. Hutan yang terdegradasi

K-Means Clustering pada Gambar


Selamat datang di aplikasi K-Means Clustering. Unggah gambar Anda untuk melatih model dan melakukan prediksi.


Unggah Gambar untuk Training


Pilih gambar untuk folder Training


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


Browse files


 f5.jpg 1.1MB





 f4.jpg 254.7KB



 f3.jpg 134.6KB




Showing page 1 of 2

Unggah Gambar untuk Testing

Pilih gambar untuk folder Testing

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

Browse files

3. Masukkan paramater K-Means dan Klik tombol train untuk train data
- a. gunung

Parameter K-Means

Masukkan jumlah cluster (k)

3

-

+

Train

 Model berhasil dilatih!

Predict

© 2024 K-Means Clustering

Tegar Muhamad Rizki Sanjukin Ndube Pinem Fahri Nizar Argubi

- b. Hutan yang terdegradasi

Parameter K-Means

Masukkan jumlah cluster (k)

2

-

+

Train

 Model berhasil dilatih!

Predict

© 2024 K-Means Clustering

Tegar Muhamad Rizki Sanjukin Ndube Pinem Fahri Nizar Argubi


4. Unggah gambar untuk testing: klik browse file yang ada dibagian testing
- a. gunung

K-Means Clustering pada Gambar


Selamat datang di aplikasi K-Means Clustering. Unggah gambar Anda untuk melatih model dan melakukan prediksi.

Unggah Gambar untuk Training


Pilih gambar untuk folder Training

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


Browse files

 mountain_309.jpg 119.1KB

×

 mountain_50.jpg 0.6MB

×

 mountain_47.jpg 0.5MB


×

Showing page 1 of 3


< >

Unggah Gambar untuk Testing


Pilih gambar untuk folder Testing

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


Browse files

 mountain_316.jpg 119.1KB

×

 mountain_315.jpg 121.5KB

×


 mountain_314.jpg 117.3KB

×


b. Hutan yang terdegradasi

Unggah Gambar untuk Testing


Pilih gambar untuk folder Testing

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


Browse files

 f8.jpg 409.4KB

×

 f7.jpeg 0.5MB

×

 {B8CF2750-AB3B-4DBE-8F34-82EB70EE242A}.png 1.0MB

×

Showing page 1 of 2

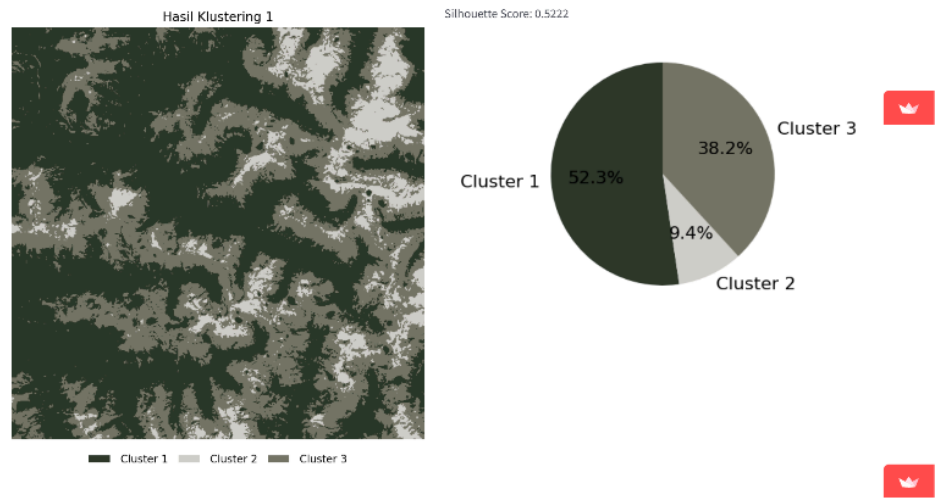
< >

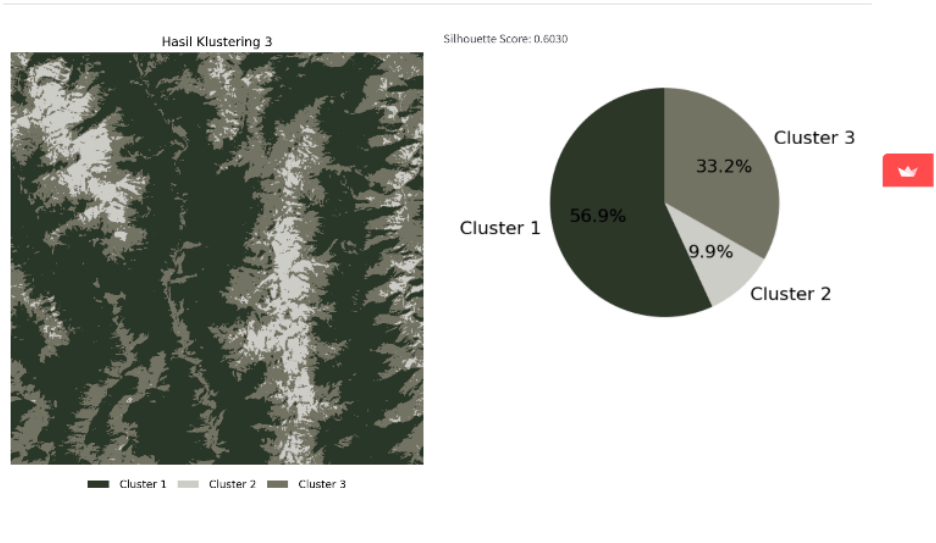
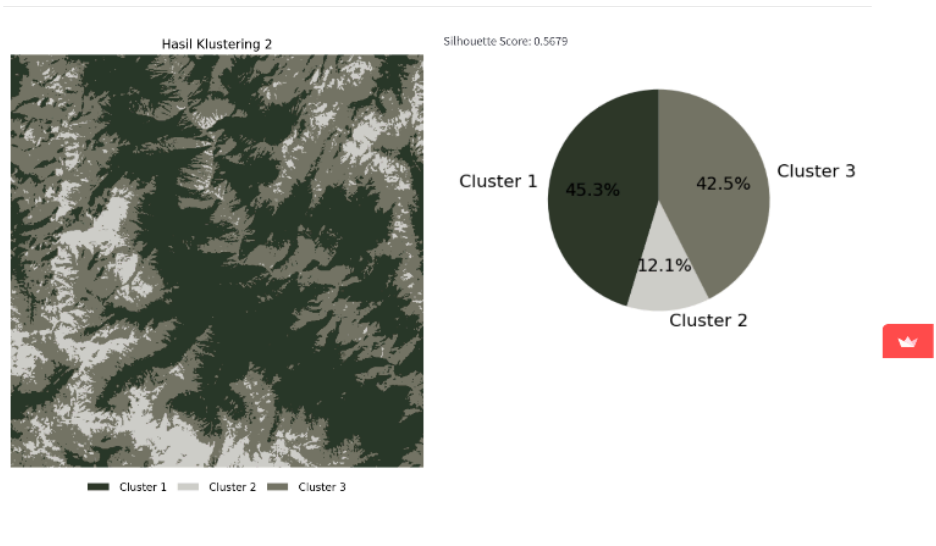
5. Klik Predict

a. gunung

Predict

Hasil Clustering





© 2024 K-Means Clustering

Tegar Muhamad Rizki Sanjukin Ndube Pinem Fahri Nizar Argubi

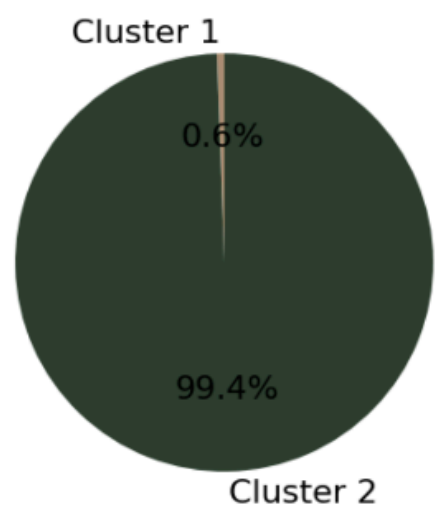
b. Hutan yang terdegradasi

Predict

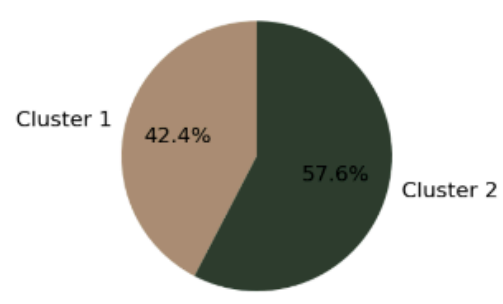
Hasil Clustering

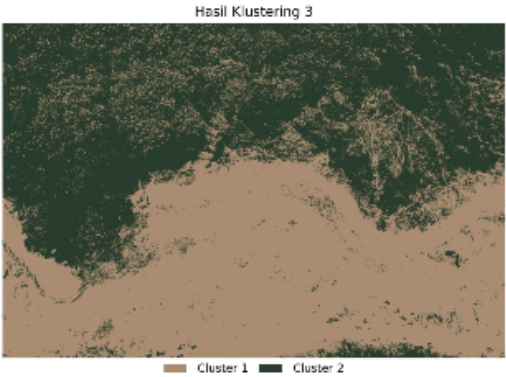


Silhouette Score: 0.8534

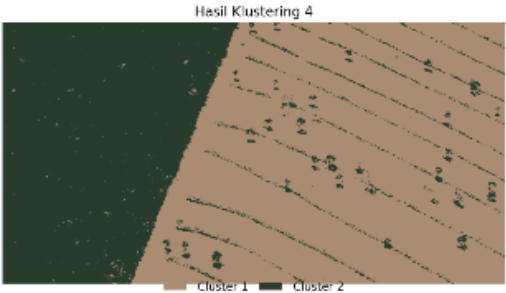
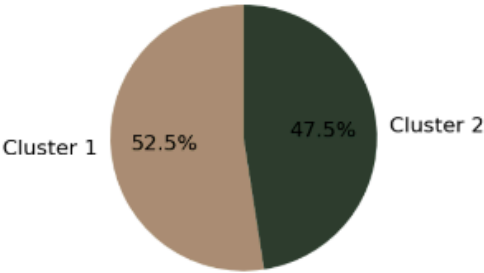


Silhouette Score: 0.6205





Silhouette Score: 0.5767



Silhouette Score: 0.7581

