Principles of Satellite Image Processing

GNR 607

NDVI Masking and K-Means Clustering

Esha Yindukuri : 24B2119
Krisha Biradar : 24B3957
Pratusha Pudakalkatti : 24B2112

Project Objectives

Calculate NDVI and isolate vegetation using a user-defined threshold.

Apply K-Means clustering (with user-chosen K) to only the non-vegetated regions.

PROCESSING PIPELINE

Select Blue, Green, Red & NIR TIFFs + Output Folder and specify ‘k’ & NDVI

→ Load the images and normalize RED and NIR

→ Compute NDVI

↓ Threshold NDVI Mask

← Downsampling

← Extract features from non-vegetation pixels and normalize the features

← Apply K-means clustering

↓ Upscale the clusters to full resolution

→ Insert vegetation pixels

→ Show final masked & clustered image and save outputs

NDVI Thresholding

NDVI thresholding uses the Normalized Difference Vegetation Index, computed from the red and near-infrared (NIR) bands, to highlight healthy vegetation.

Plants reflect strongly in NIR and absorb red light, so high NDVI values indicate dense, healthy vegetation, while low values point to sparse or non-vegetated surfaces.

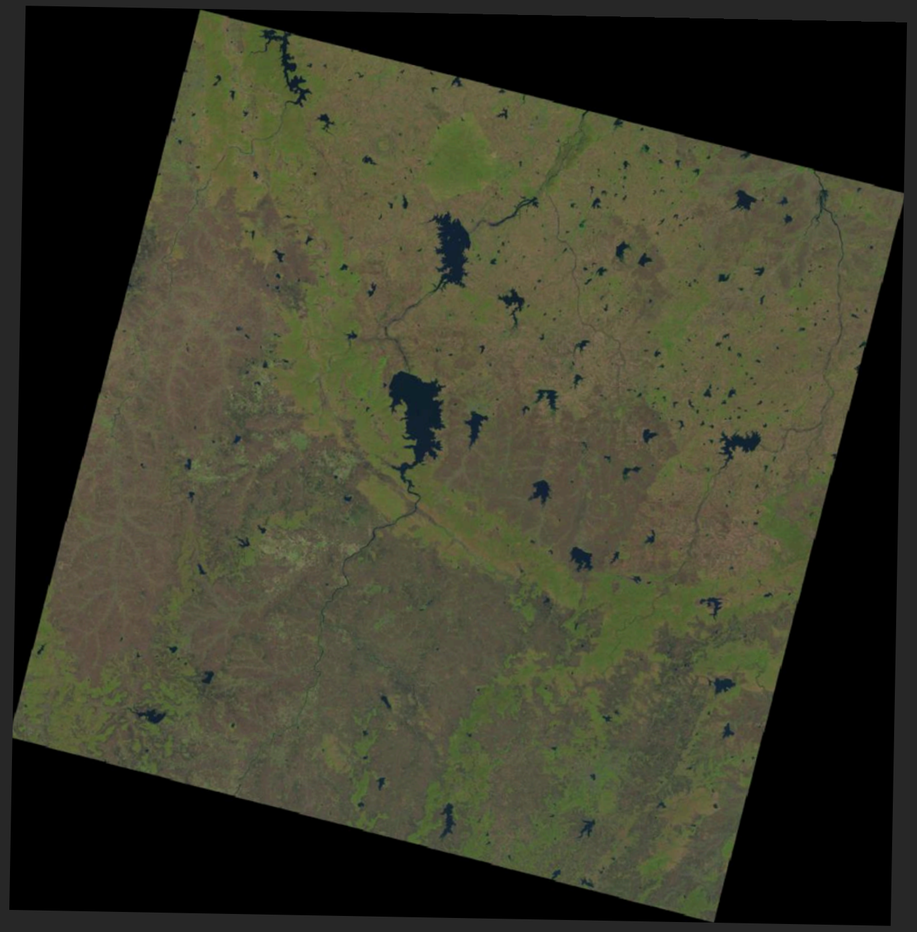
Practically, NDVI is used in agriculture, forestry, drought monitoring, and land-cover mapping to assess crop health, biomass, and vegetation stress. Typical values include ~0.3 for sparse vegetation, ~0.5–0.6 for moderate vegetation, and >0.7 for dense, healthy vegetation.

Methodology

$$\text{NDVI} = \frac{NIR - Red}{NIR + Red}$$

We first loaded the satellite image's NIR (b5) and Red (b4) bands.
Using the NDVI formula, we computed an NDVI image where each pixel indicates vegetation strength.
Then applied a user-defined threshold (e.g., 0.1, 0.2) to create a mask: pixels above the threshold are visible and rest are hidden.
This mask separates healthy vegetation from non-vegetated areas, and different thresholds allows us to control how strict we want the vegetation detection to be.

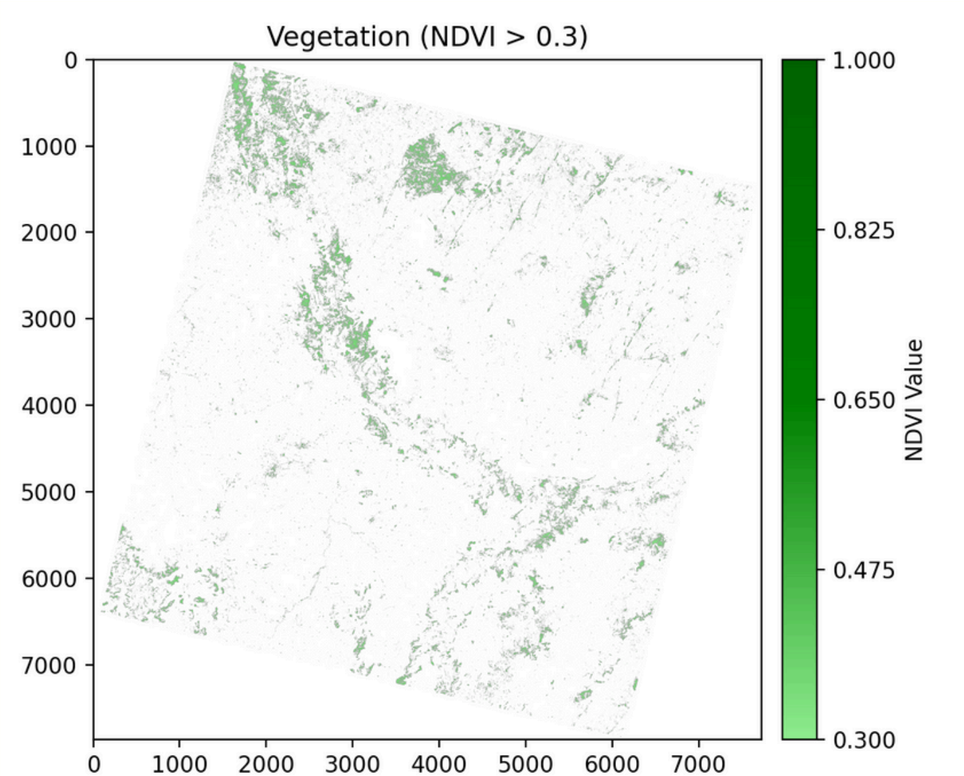
Input data



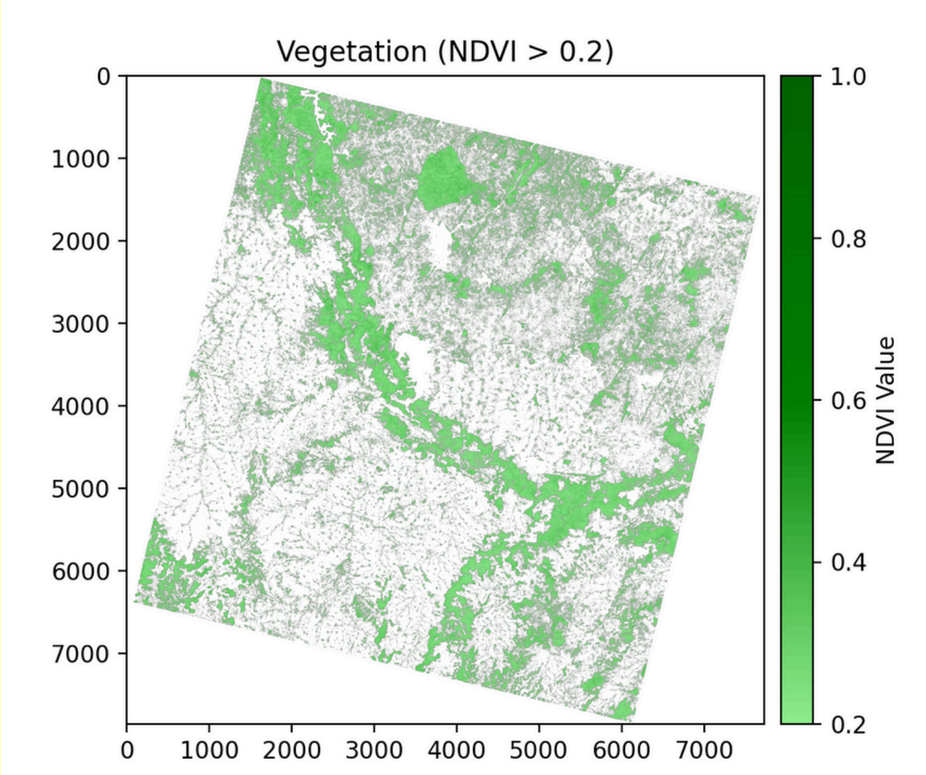
Satellite & Sensor: Landsat 9, OLI/TIRS
Acquisition Date: 7th Nov 2025
Location: ~24.55°N, 78.37°E (In Madhya Pradesh)
Cloud Cover: 0.01%
Bands Used: RGB + NIR

Results

Threshold = 0.2



Threshold = 0.3



K-means Clustering

K-Means is an unsupervised machine learning algorithm that groups data into K clusters based on similarity.
It works by placing K centroids, assigning each data point to the nearest centroid, and iteratively adjusting the centroids until the clusters stabilize.

It is widely used in image segmentation, pattern recognition, remote sensing classification, customer grouping, etc because it is simple, fast, and effective for large datasets.
Each cluster represents a region where the data points share similar characteristics—for example, in images, clusters might correspond to different land-cover types, colors, textures, or intensity ranges.

Methodology

After extracting the non-vegetated pixels using the NDVI mask, Blue, Green, Red and NIR values are organized into a feature matrix and normalized for stability.

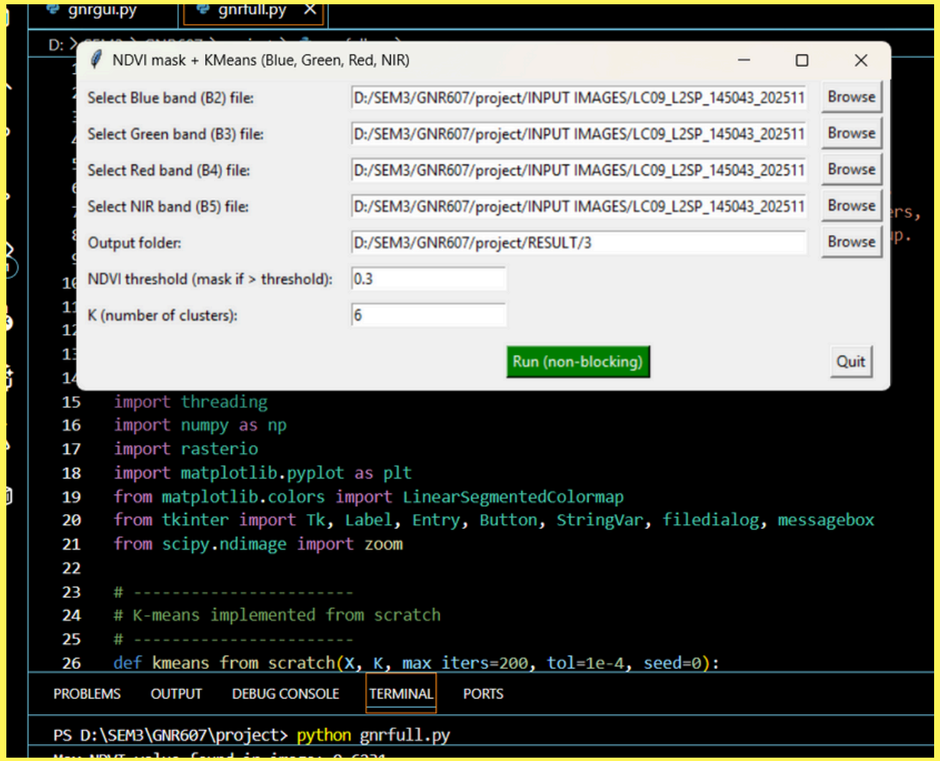
K initial centroids are chosen randomly from these feature points.
Each remaining pixel is then assigned to the nearest centroid using squared Euclidean distance, and new centroids are computed as the mean of the points in each cluster.

Empty clusters are reinitialized, and the assignment–update steps repeated until the centroids converge within a small tolerance.

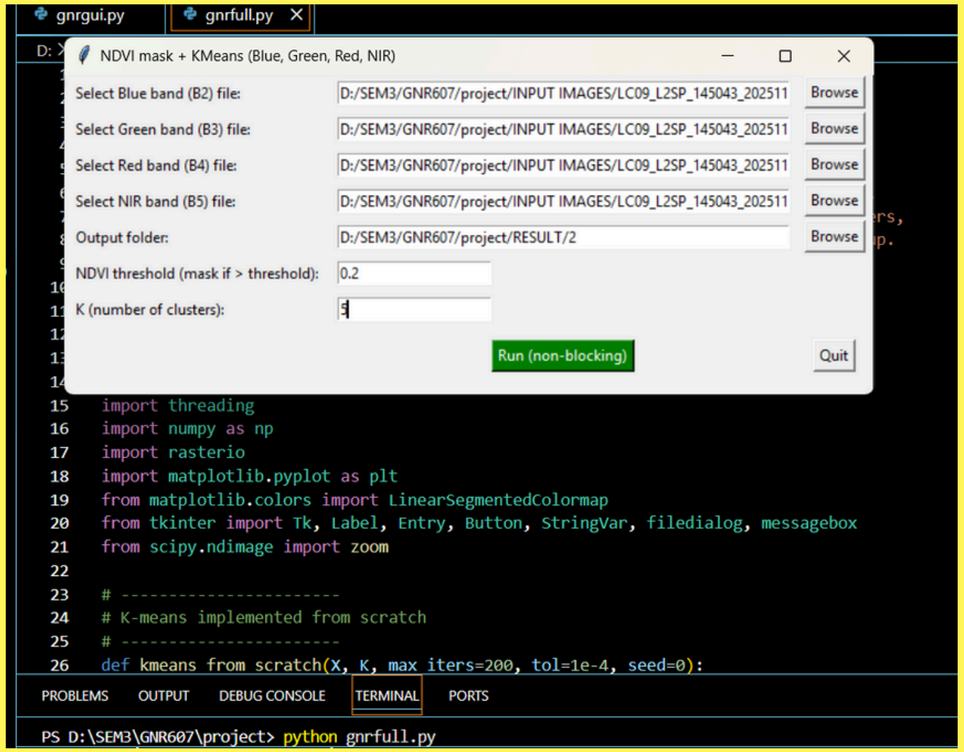
The final output is a cluster label for every non-vegetation pixel, which is then mapped back to the full-resolution image, clearly separating vegetation (from the NDVI mask) and the K spectral clusters of the non-vegetated region.

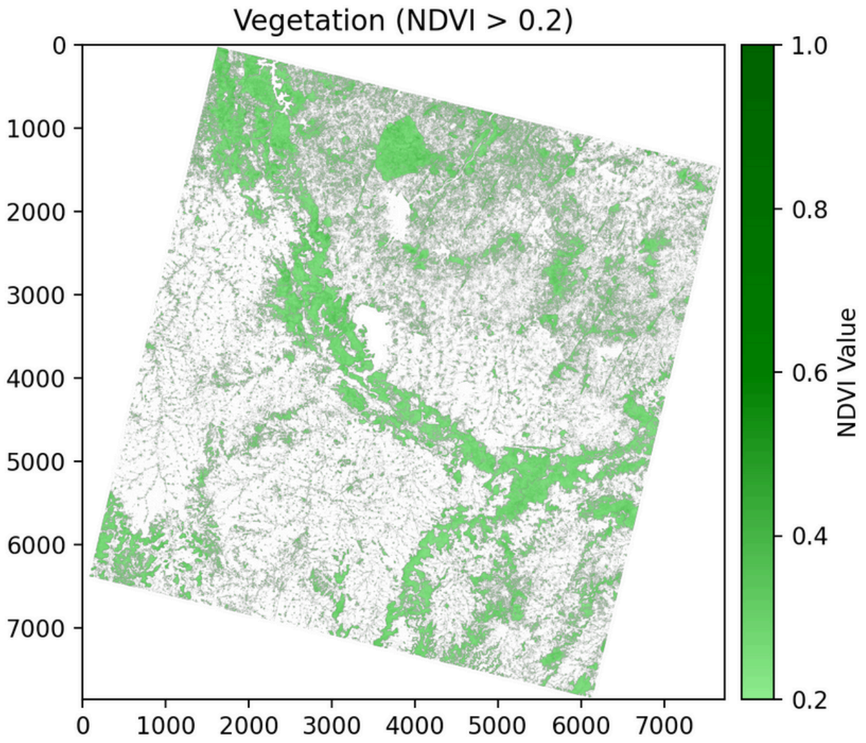
Results

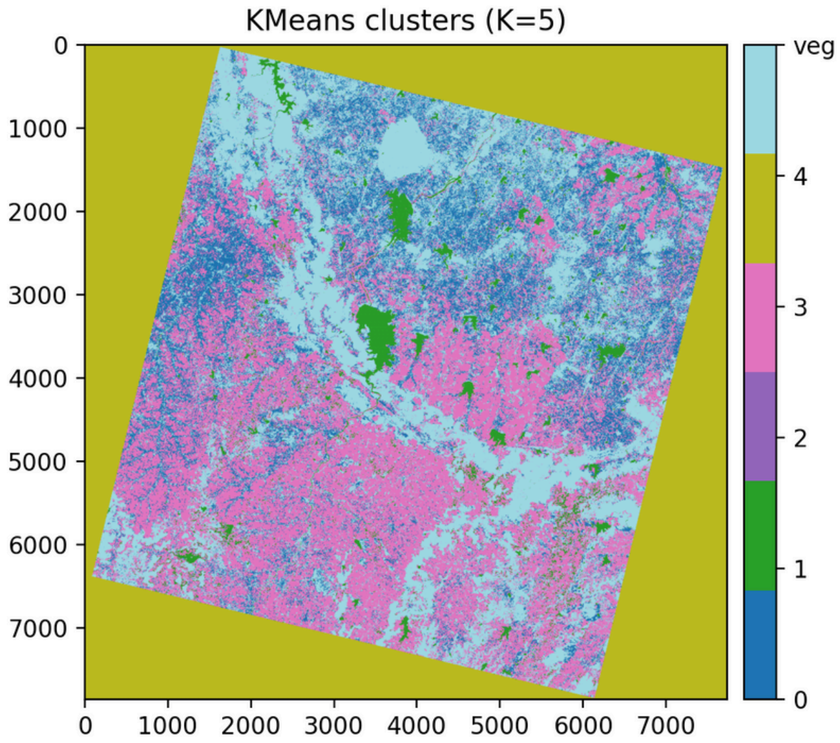
K=6 and NDVI threshold = 0.3

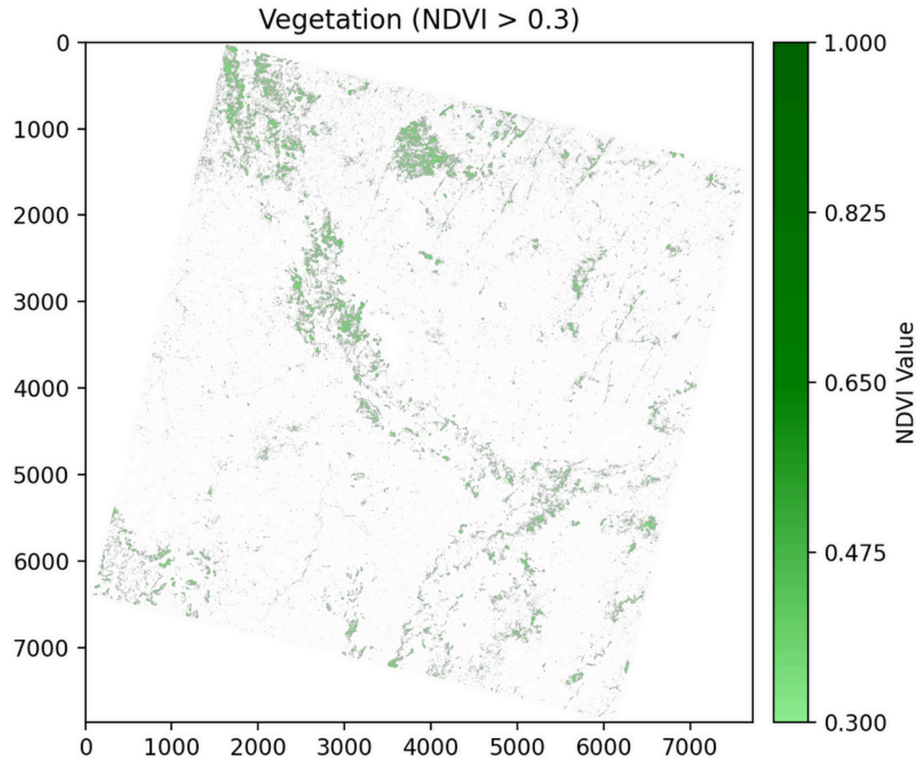


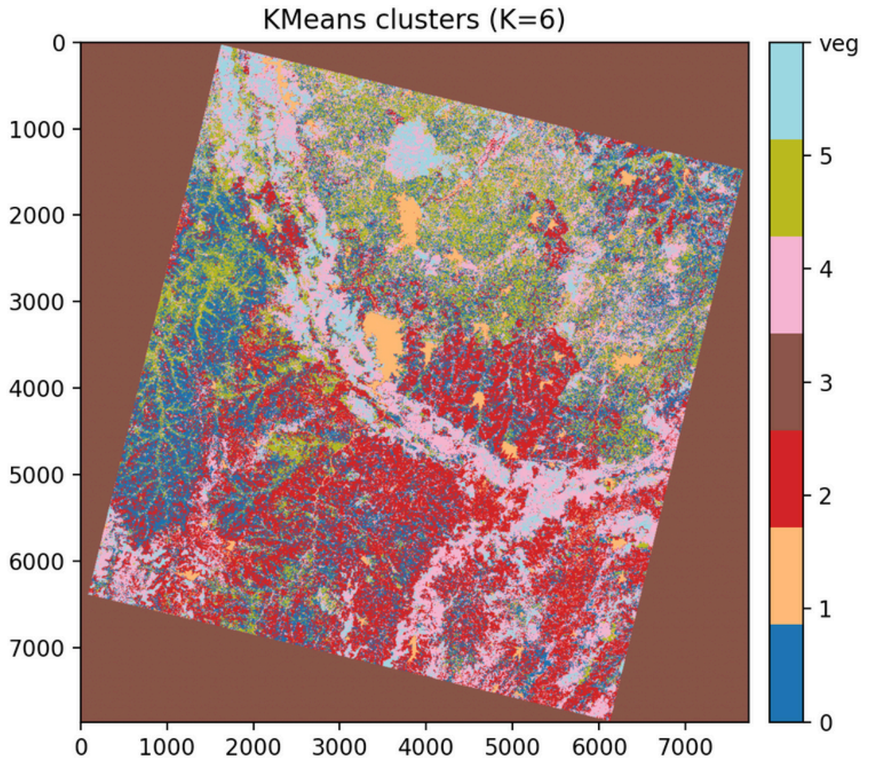
K=5 and NDVI threshold = 0.2




Vegetation (NDVI > 0.2)
0
1000
2000
3000
4000
5000
6000
7000
0
1000
2000
3000
4000
5000
6000
7000
1.0
0.8
0.6
0.4
0.2
NDVI Value


KMeans clusters (K=5)
0
1000
2000
3000
4000
5000
6000
7000
0
1000
2000
3000
4000
5000
6000
7000
veg
4
3
2
1
0


Vegetation (NDVI > 0.3)
0
1000
2000
3000
4000
5000
6000
7000
1.000
0.825
0.650
0.475
0.300
NDVI Value


KMeans clusters (K=6)
0
1000
2000
3000
4000
5000
6000
7000
veg
5
4
3
2
1
0

Bibliography

https://www.sciencedirect.com/science/article/pii/S187705091501944 4 - **Ndvi: Vegetation Change Detection Using Remote Sensing and Gis – A Case Study of Vellore District**

https://towardsdatascience.com/create-your-own-k-means-clustering-algorithm-in-python-d7d4c9077670/ - **Create a K-Means Clustering Algorithm from Scratch in Python**

https://medium.com/analytics-vidhya/image-segmentation-using-k-means-clustering-from-scratch-1545c896e38e - **Image Segmentation using K-means Clustering from Scratch**

THANKYOU