



INDIAN STATISTICAL INSTITUTE BANGALORE

PATTERN RECOGNITION ASSIGNMENT 2025

Contents:-

- 1. Comparative Analysis to Performance Evaluation of**
 - a. Bayes Classifier on Iris and Vowel Dataset**
 - b. KNN Classifier on Iris and Vowel Dataset**
 - c. MDC on Iris and Vowel Dataset**

- 2. Interpretation of performing KMeans Clustering on**
 - a. Garden Image (3 bands- RGB)**
 - b. Remote Sensing Image (Multispectral – 4 bands)**

BY-

Name: SUHADITYA CHAKRABORTY

Roll: mqms2413

Batch: MS-QMS Batch of 2026

1. Comparative Analysis of Performance evaluation of Classifiers

Programming Language	Python
Environment	VSCode

Following results were obtained when Bayes Classifier, KNN and MDC were applied on Iris (balanced) dataset and Vowel (unbalanced dataset):

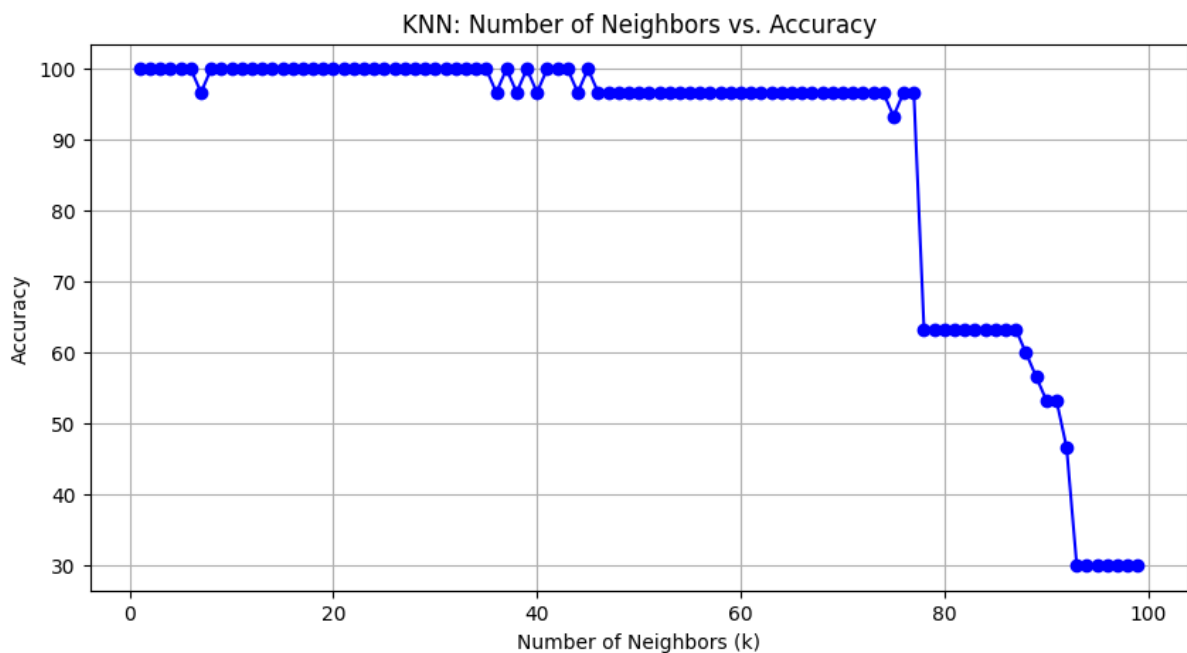
Dataset:	Iris	Algo: MDC	Algo: KNN (k=5)	Algo: Bayes Classifier
Train/ Test Split	Type	Performance	Performance	Performance
100-0	Training accuracy	92.67%	-	98.00%
	Test accuracy	-	-	-
80-20	Training accuracy	93.33%	-	97.50%
	Test accuracy	86.67%	100.00%	96.67%
70-30	Training accuracy	94.29%	-	98.10%
	Test accuracy	88.88%	100.00%	100.00%
60-40	Training accuracy	93.33%	-	98.89%
	Test accuracy	90.00%	98.33%	98.33%
50-50	Training accuracy	94.67%	-	98.67%
	Test accuracy	89.33%	94.67%	97.33%
Average	Training accuracy	93.66%	-	98.23%
	Test accuracy	88.72%	98.25%	98.08%

Dataset:	Vowel	Algo: MDC	Algo: KNN (k=5)	Algo: Bayes Classifier
Train/ Test Split	Type	Performance	Performance	Performance
100-0	Training accuracy	71.41%		80.25%
	Test accuracy	-		-
80-20	Training accuracy	71.41%		80.75%
	Test accuracy	68.57%	84.00%	78.86%
70-30	Training accuracy	70.44%		81.12%
	Test accuracy	70.99%	83.21%	79.39%
60-40	Training accuracy	71.26%		82.38%
	Test accuracy	72.23%	83.95%	79.08%
50-50	Training accuracy	72.87%		84.60%
	Test accuracy	73.39%	84.17%	81.19%
Average	Training accuracy	71.48%	-	81.82%
	Test accuracy	71.29%	83.83%	79.63%

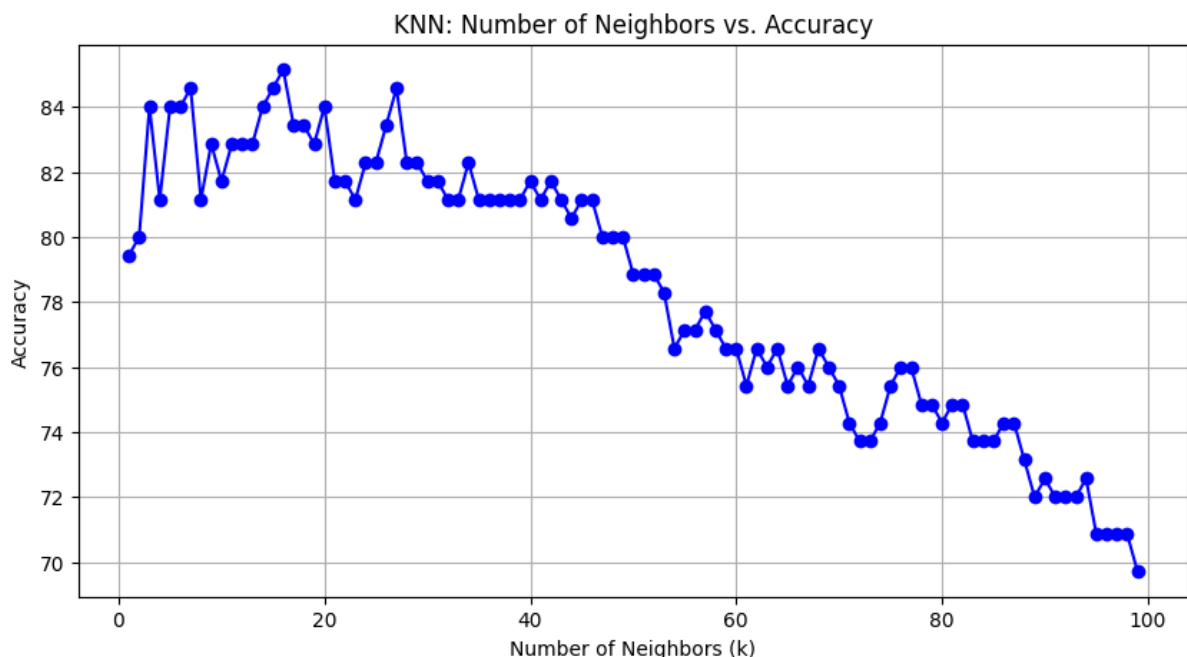
Interpretation: The models work better on Iris dataset than Vowel as the former is a balanced and small dataset which is convenient for them. If we compare the average test accuracy across all kinds of splits, we find that **KNN with k=5 performs the best among the three models for both the datasets.**

k is a hyperparameter which can be tuned further to get the best accuracy. Following k v/s accuracy plot was obtained for both the datasets

Plot for Iris dataset



Plot for Vowel Dataset



Interpretation: For Iris dataset, the accuracy remains higher and more stable for higher values of k until $k > 70$. This can be attributed to the dataset being balanced and well separated. For Vowel dataset, the performance fluctuates as it is an unbalanced dataset and not well separated.

2. Interpretation of performing KMeans Clustering on Images

a. Garden Image

Original Image



Pixels: 200x200

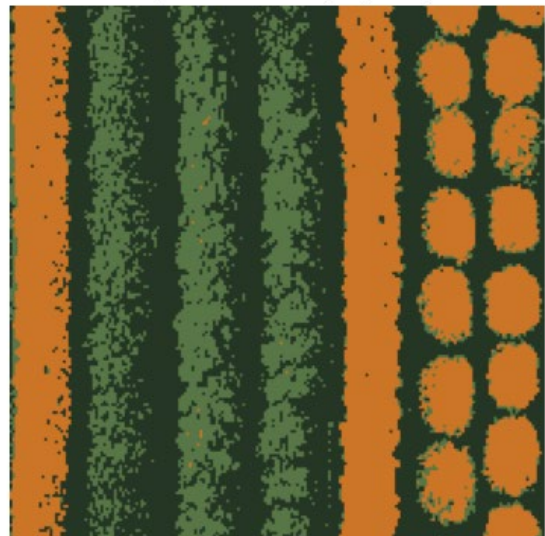
No. of bands: 3 (RGB)

After performing K Means Clustering with $k=3$

Original Image



Segmented Image ($k=3$)

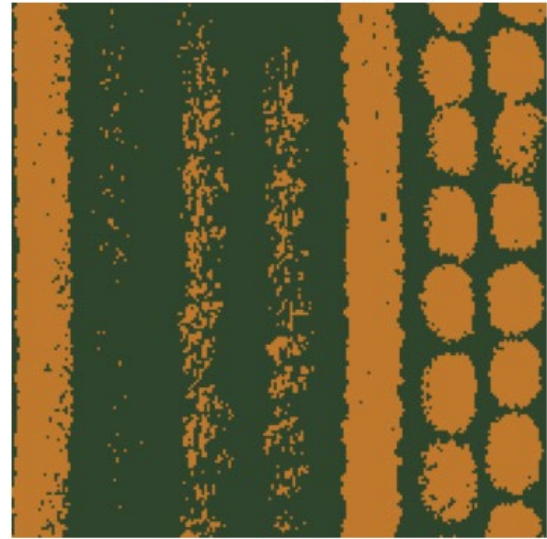


K Means Clustering for different no. of clusters

Original Image



Segmented Image (k=2)



Original Image



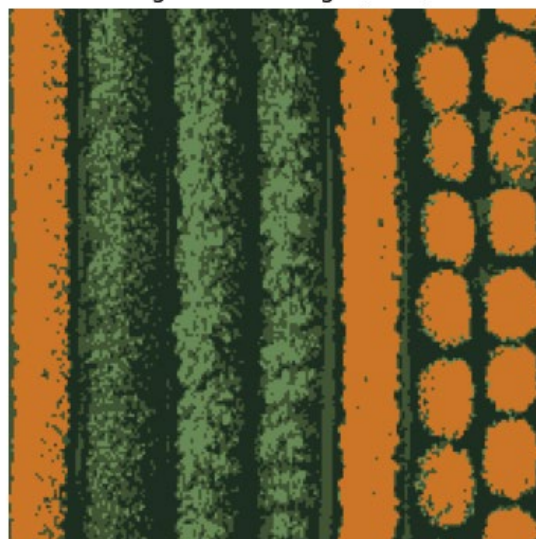
Segmented Image (k=3)



Original Image



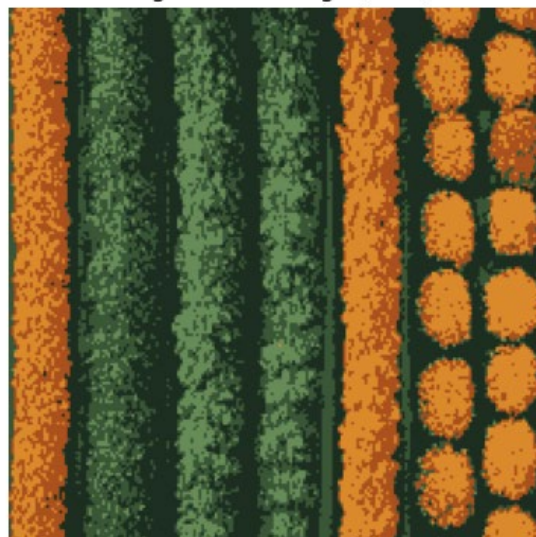
Segmented Image (k=4)



Original Image



Segmented Image (k=5)



Original Image



Segmented Image (k=6)



Original Image



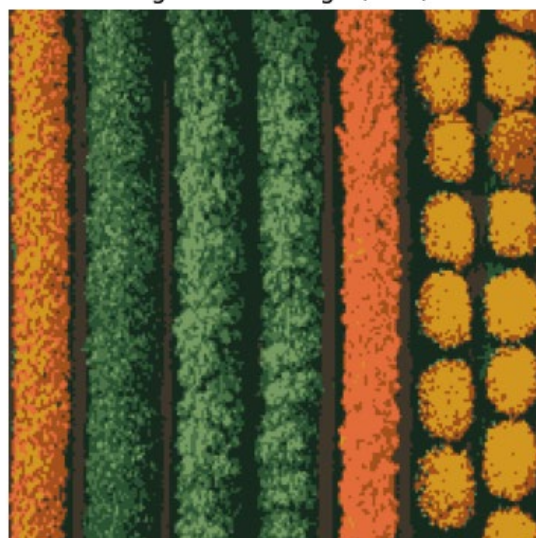
Segmented Image (k=7)



Original Image



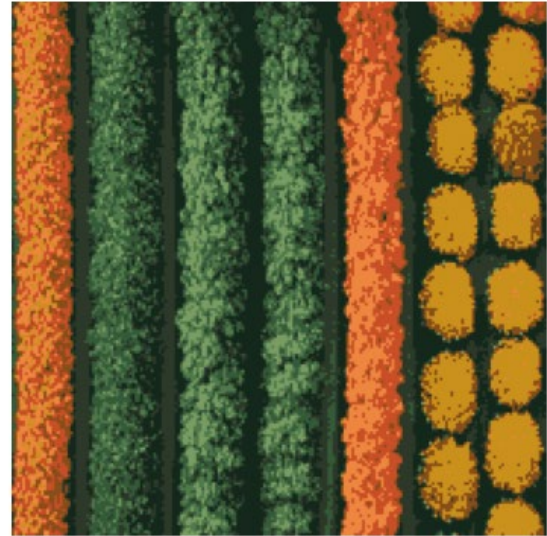
Segmented Image (k=8)



Original Image



Segmented Image (k=9)



Original Image

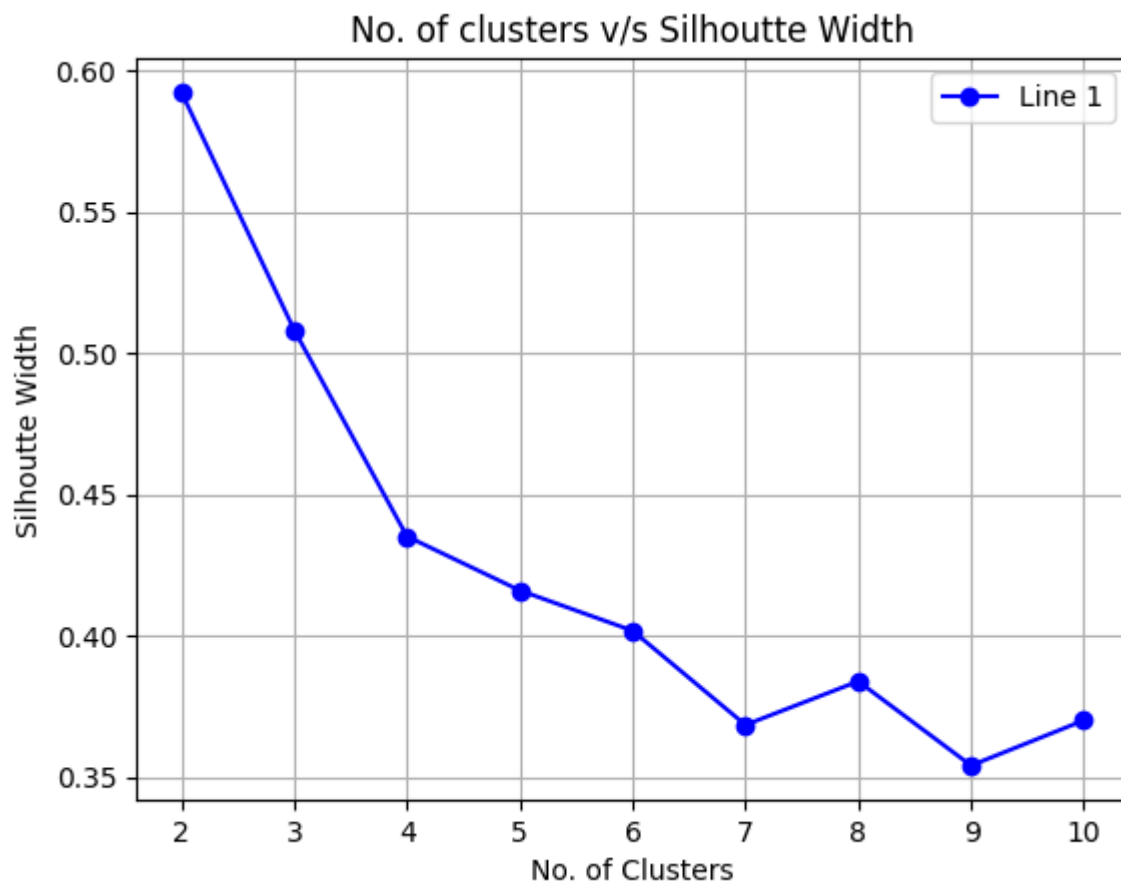


Segmented Image (k=10)



Interpretation: As k increases, segmentation becomes more detailed, capturing finer differences in colour and texture. The segmented image becomes more similar to the original image. However, setting k too high can lead to **over-segmentation**, where noise or small variations are treated as separate regions.

On plotting the Silhouette width v/s no. of clusters, we obtained the following:



The optimal no. of clusters is 2 which basically considers bushes or shrubs (green in colour) and flowers as two different clusters

a. Remote Sensing Image

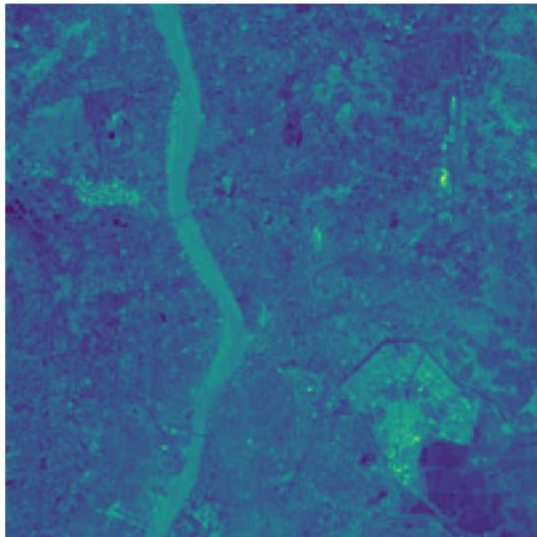


Pixels: 200x200

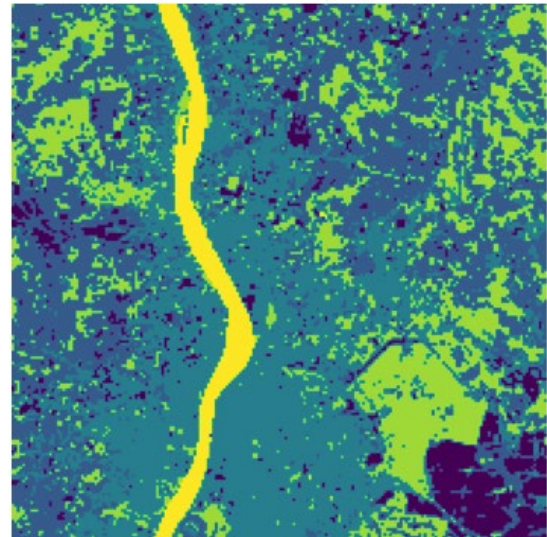
No. of bands: 4

After performing K Means Clustering with $k=5$

Original Image



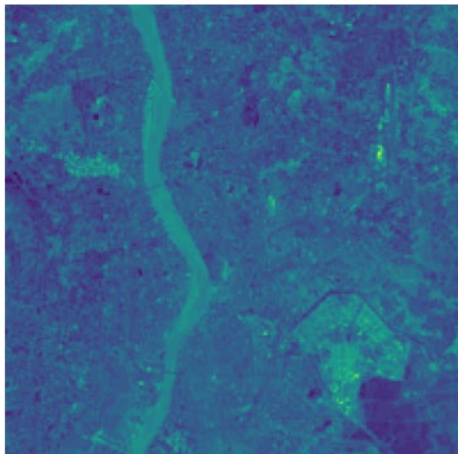
Segmented Image ($k=5$)



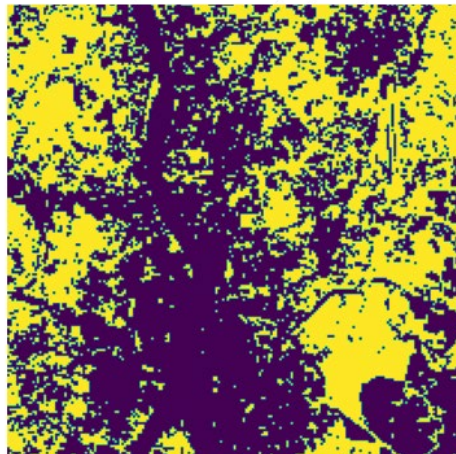
(For proper visualisation, last 3 bands are shown)

K Means Clustering for different no. of clusters

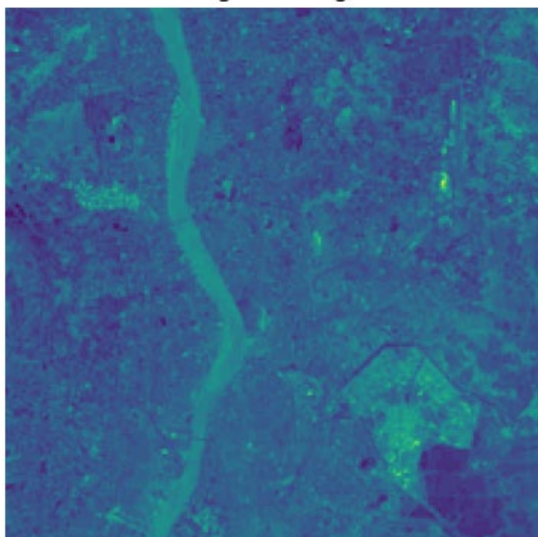
Original Image



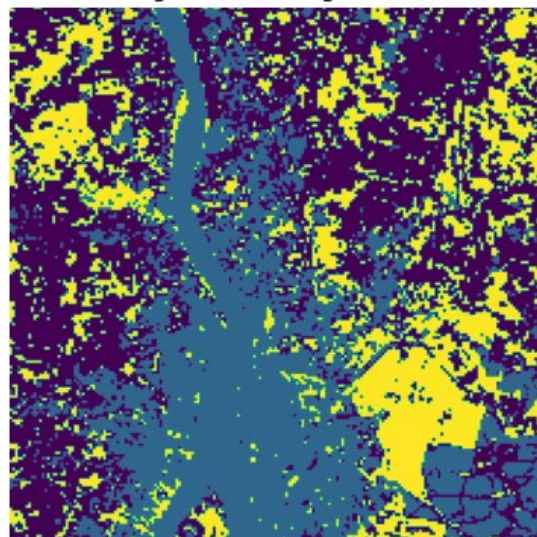
Segmented Image ($k=2$)



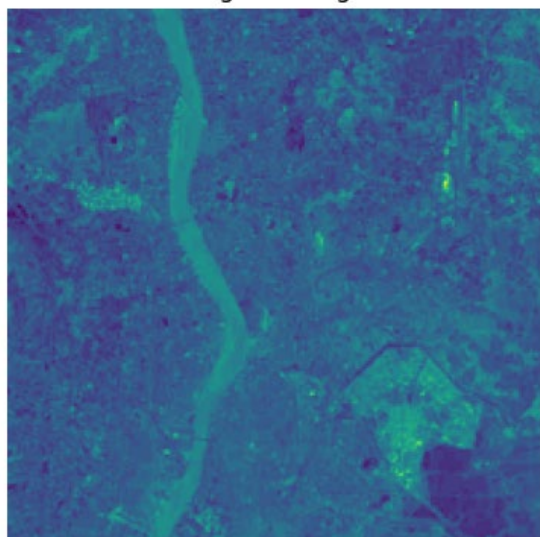
Original Image



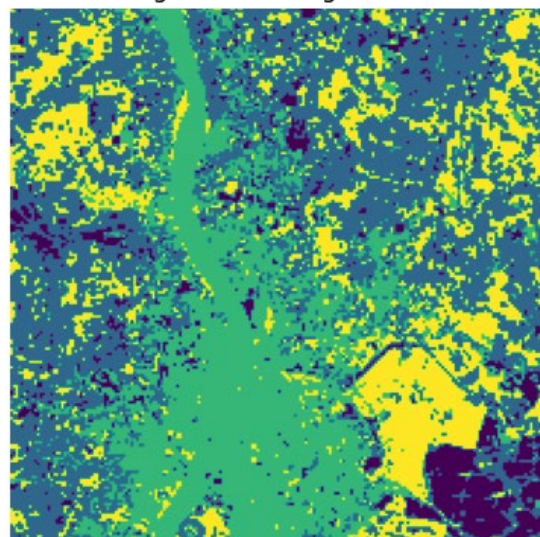
Segmented Image (k=3)



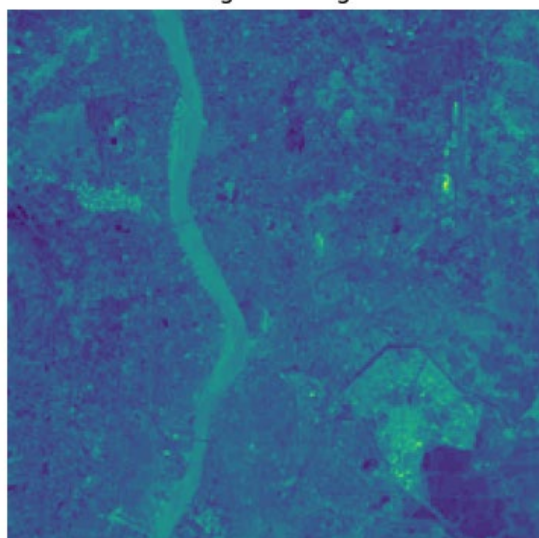
Original Image



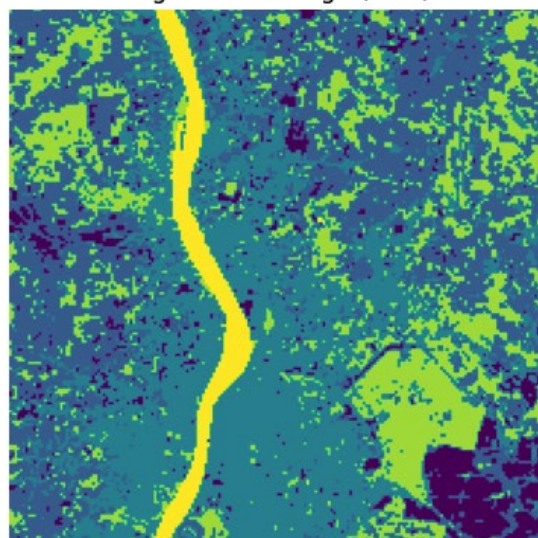
Segmented Image (k=4)



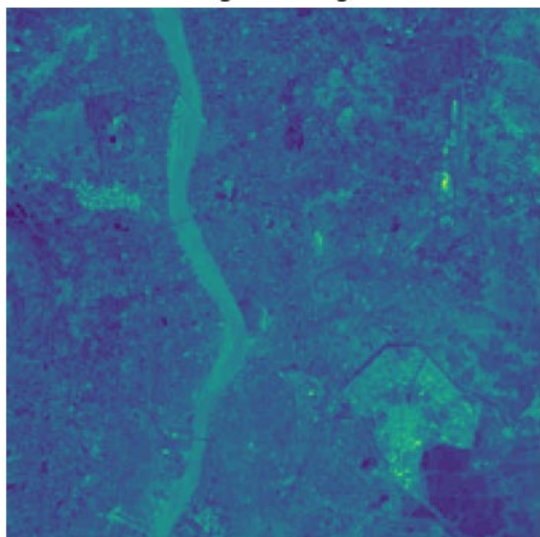
Original Image



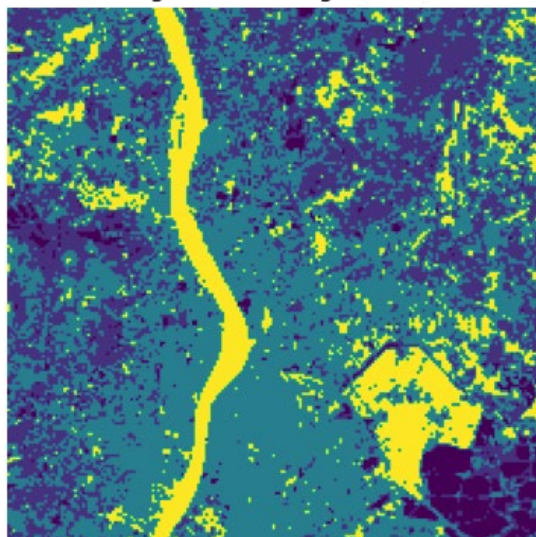
Segmented Image (k=5)



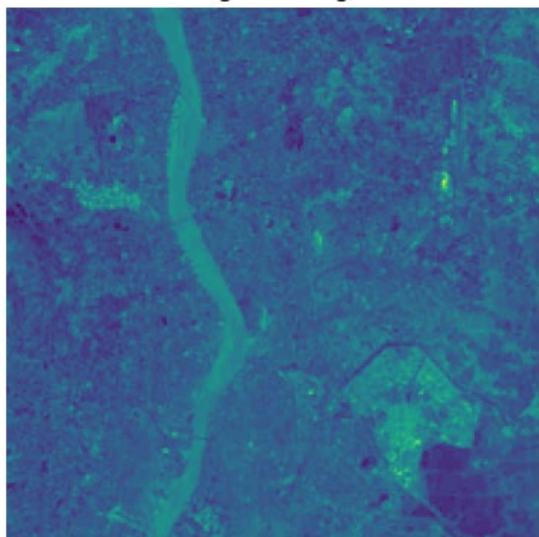
Original Image



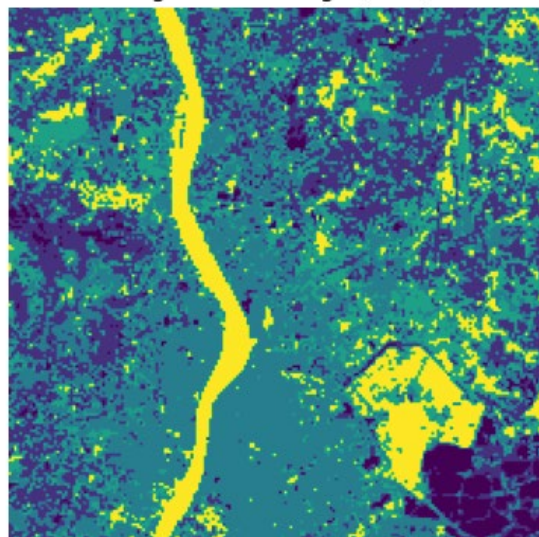
Segmented Image (k=6)



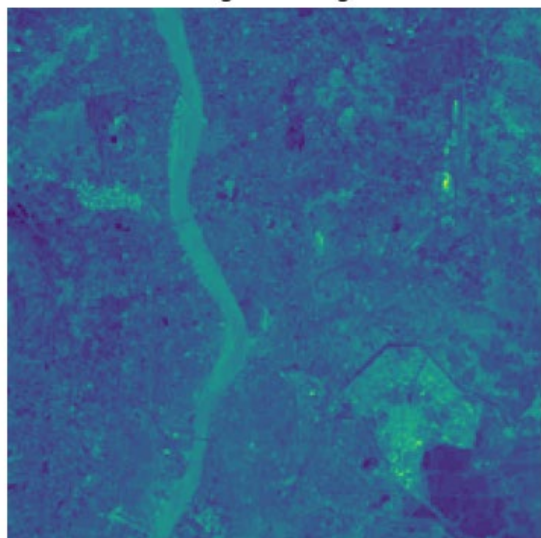
Original Image



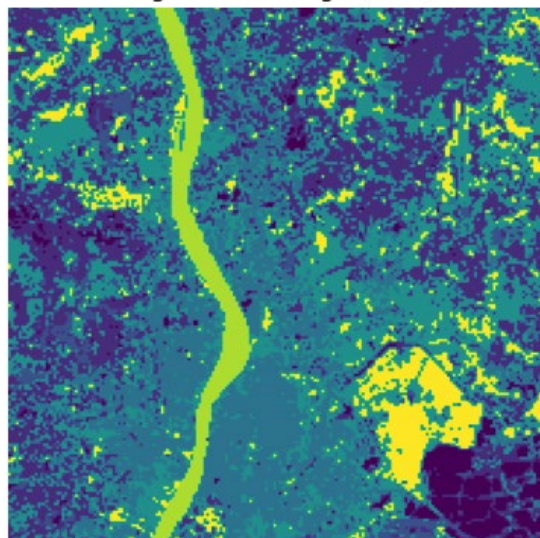
Segmented Image (k=7)



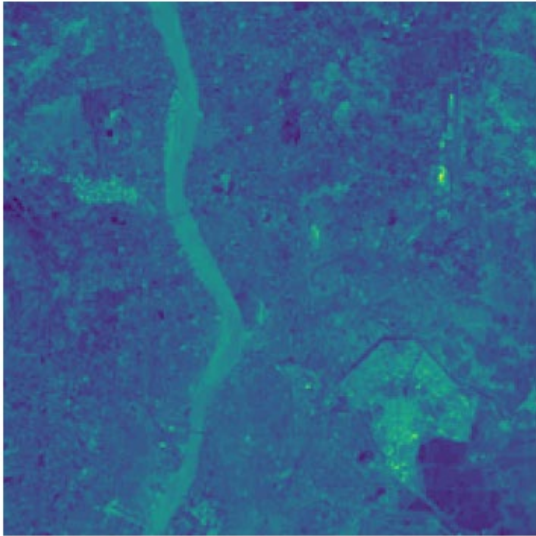
Original Image



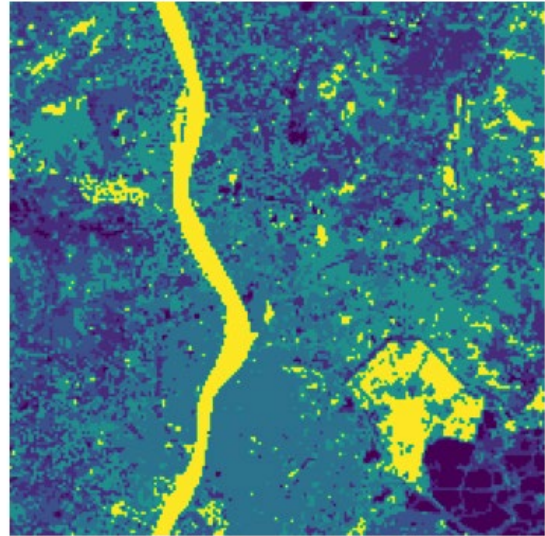
Segmented Image (k=8)



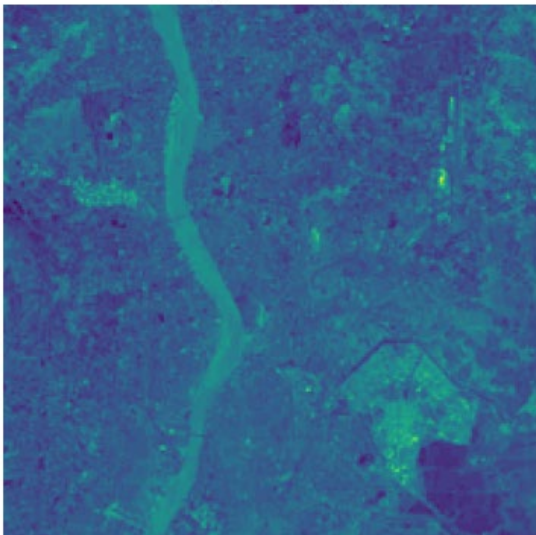
Original Image



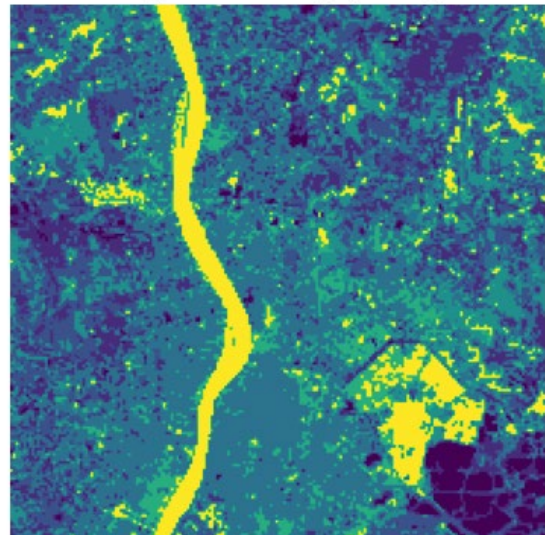
Segmented Image (k=9)



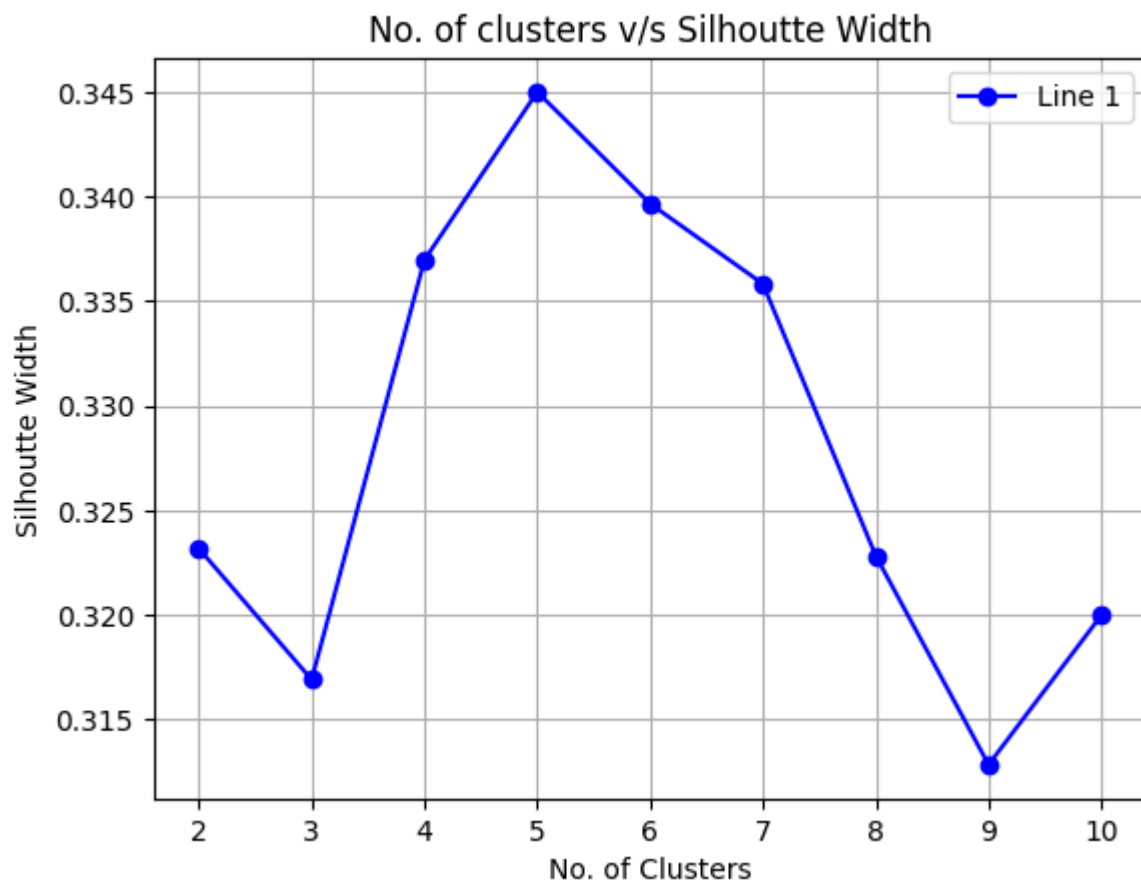
Original Image



Segmented Image (k=10)



On plotting the Silhouette width v/s no. of clusters, we obtained the following:



Interpretation: The optimal number of clusters is 5. The river feature (yellow) becomes clear when k changes from 4 to 5. The 5 different clusters depict different features of the Land cover segmentation

The different colours depict different clusters.