**Name: Manan Dalal**

**NET-ID: MUD200000**

**CS-6375: Machine Learning**

**Assignment-5**

**K-Means Clustering for Image Compression**

**Image-1: Koala.jpg**

A koala bear on a tree

Description automatically generated with medium confidence

**Observations:**

* Results after applying K-Means Algorithm for image Compression on the **Koala.jpg** image.
* Each observation has **20** iterations.

|  |  |  |  |
| --- | --- | --- | --- |
| **K Value** | **Size before compression (in KB)** | **Size after compression (in KB)** | **Ratio** |
| 2 | 762.53 | 128.75 | 5.92 |
| 5 | 762.53 | 169.912 | 4.48 |
| 10 | 762.53 | 172.494 | 4.42 |
| 15 | 762.53 | 167.75 | 4.54 |
| 20 | 762.53 | 157.10 | 4.85 |
| 25 | 762.53 | 155.666 | 4.90 |

**Compressed Images:**

A picture containing text, silhouette

Description automatically generated A picture containing text, dinosaur, close

Description automatically generated A picture containing mammal, stuffed, close, koala

Description automatically generated

K = 2 K = 5 K = 10

A picture containing mammal, koala, close

Description automatically generated A koala bear on a tree

Description automatically generated with low confidence A koala bear on a tree

Description automatically generated with low confidence

K = 15 K = 20 K = 25

**Image-2: Penguins.jpg**

A group of penguins walking on a beach

Description automatically generated with medium confidence

**Observations:**

* Results after applying K-Means Algorithm for image Compression on the **Penguins.jpg** image.
* Each observation has **20** iterations.

|  |  |  |  |
| --- | --- | --- | --- |
| **K Value** | **Size before compression (in KB)** | **Size after compression (in KB)** | **Ratio** |
| 2 | 759.604 | 83.220 | 9.12 |
| 5 | 759.604 | 99.867 | 7.60 |
| 10 | 759.604 | 113.913 | 6.66 |
| 15 | 759.604 | 114.317 | 6.64 |
| 20 | 759.604 | 111.603 | 6.80 |
| 25 | 759.604 | 112.0947 | 6.77 |

**Compressed Images:**

A picture containing outdoor, silhouette

Description automatically generated A group of penguins

Description automatically generated with low confidence A group of penguins

Description automatically generated with medium confidence

K = 2 K = 5 K = 10

A group of penguins walking on a beach

Description automatically generated with low confidence A group of penguins

Description automatically generated with medium confidence A group of penguins

Description automatically generated with medium confidence

K = 15 K = 20 K = 25

**Discussion**

1. **Is there a tradeoff between image quality and degree of compression?**

* Yes. ‘K’ represents the degree of compression.
* Smaller value of K means fewer clusters, and hence fewer colors to represent the image.
* Therefore, for smaller ‘K’ values, a lot of details in the image are compromised, which produces a lower image quality.
* Higher ‘K’ values show more colors due to a larger number of clusters, and hence produces a better-quality image.
* However, higher values of K take longer to execute.

1. **What would be a good value of K for each of the two images?**

* For Koala.jpg, k= 15 gives a compression ratio of 4.54, which is close to that for k=20 and 25. Hence k=15 would be a good choice for k, as it has almost the same compression ratio as of greater values of k.
* For Penguins.jpg, k= 15 gives a compression ratio of 6.64, which is close to that for k=20. Hence k=15 would be a good choice for k, as it has almost the same compression ratio as of greater values of k.