PROJECT REPORT

```
/***
 * CS6375.001 Machine Learning
 * Machine Learning
 * Vidya Sri Mani
 * Implementaion of KMeans Compression
 */
```

General Program Description

Program : KMeans.java

The program implements k-means clustering for image compression.

The Program is executed taking two given images i.e., Penguins.jpg and Koala.jpg.
For each file, the clustering is done for multiple values of k.
K Values: Specific K Value(given as command line input), 2,5,10,15,20

For each of these k values, a compressed image file is generated. They are stored in the same folder as the input file(Penguins.jpg and Koala.jpg).

Each of these output files are prefixed with the k value number
Example '2-Penguin.jpg' when k=2

• Is there a tradeoff between image quality and degree of compression? What would be a good value of K for each of the two images?

Yes. 'K' represents the degree of compression. A smaller value of k could indicate fewer clusters, and hence fewer colors to represent the image.
Therefore for smaller 'K' values, a lot of details in the image is compromised, which produces a lower image quality.
Higher 'K' values show more colors due to more number of clusters, and hence produces a better image quality. However, higher values of K will take longer to execute.
For the given images, Koala.jpg and Penguin.jpg, reasonably good images are seen when k lies between 15 and 20. The output file shows the reasonable compression ratio and most colors which are identifiable.

For Koala.jpg, k= 10 gives a compression ratio of `4.5, which is close to that for k=15 and 20.Hence k=10 would be a good choice for k, as it has almost the same compression ratio as of greater values of k.`

For Penguin.jpg, k= 15 gives a compression ratio of `6.8, 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.`

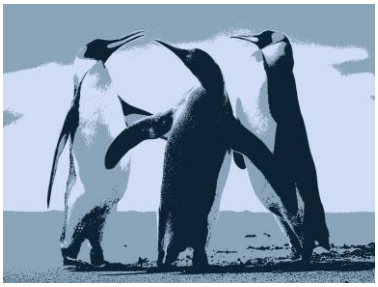
Compression Ratio = Old File size/ New file size

Dataset : Koala.jpg, Penguin.jpg

Koala.jpg

 Original Image : Koala.jpg	Compression Ratios: K=2 : 5.81364753182935 K=5 : 5.813560962534993 K=10 : 4.522178528612879 K=15 : 4.533047319930567 K=20 : 4.563061009817671	Average:5.049099070545092 Variance:0.38982479945099
K=2(specific k value given as input)  Koala1.jpg	K=2  2-Koala.jpg	K=5  5-Koala.jpg
K=10  10-Koala.jpg	K=15  15-Koala.jpg	K=20  20-Koala.jpg

Penguins.jpg

 Original Image : Penguins.jpg	Compression Ratios: K=2 : 9.128661628016149 K=5 : 8.786713207717682 K=10 : 7.7701136794997305 K=15 : 6.83829024062173 K=20 : 6.769669277632724	Average: 7.858689606697602 Variance:0.941816515453169
K=2(specific k value given as input)  Penguins1.jpg	K=2  2-Penguins.jpg	K=5  5-Penguins.jpg
K=10  10-Penguins.jpg	K=15  15-Penguins.jpg	K=20  20-Penguins.jpg