<u>**CS 6375.002 - Machine Learning - Spring '17**</u>

<u>**Assignment 3 - March 26, 2017**</u>

Problem 2. K-means clustering on images

Implementation of K-means clustering algorithm for image compression of the 2 images that have been provided.
- Koala.jpg
- Penguins.jpg

<u>Files:</u>

No.	File Name	Purpose
1	Main.java	Main program to run.
2	KMeans.java	Implementation of the K-Means clustering algorithm.

*Please refer to **Readme.txt** for the steps to compile and run the program, and also for a sample input and output.*

<u>Answers to the questions posed:</u>

❖ <u>Display the images after data compression using K-means clustering for different values of K(2, 5, 10, 15, 20).</u>

The output images have been stored in the following folders:

- Koala_output
- Penguins_putput

❖ <u>What are the compression ratios for different values of K? (Note that we have to repeat the experiment multiple times with different initializations.)</u>

K-value	Compression Ratio	
	for Koala	**for Penguins**
2	5.920	9.080
5	4.427	7.407
10	4.597	6.727
15	4.795	6.520
20	4.865	6.660
50	4.926	6.860
75	4.878	6.451
100	6.802	5.882

❖ <u>Report the average as well as variance in the compression ratio</u>.

	Koala	Penguin
Average of compression ratios	5.15125	6.948375
Variance of compression ratios	0.6402771	0.9231688

❖ <u>Is there a tradeoff between image quality and degree of compression?</u>

Yes, there is trade-off between image quality and compression ratio. As the compression ratio increases, the image quality decreases.

❖ <u>What would be a good value of K for each of the two images?</u>

The best K value for koala: 75

The best K value for penguin: 100

<u>Conclusion</u>:

This assignment was useful in learning and understanding the K-Means clustering algorithm.