# **HW 3 (584-Rangwala): Image Clustering**

# Pankaj Kumar Jatav – G10338769 Part 1- Part 2 -

**Goal:**

Implement the K-Means Algorithm

Deal with Image data (processed and stored in vector format)

Think about Best Metrics for Evaluating Clustering Solutions

**Approach:**

I follow the below diagram step to solve this assignment.



**Pre-Processing of Data:**

Below are the steps taken to pre-process that data.

1. Use TSNE to reduce the dimension of the data.

tsne\_data = pd.DataFrame(TSNE(2).fit\_transform(df))

1. Perform feature selection based on different classification models

df = df.loc[:, (df != 0).any(axis=0)]

1. Normalize the data(range from 0 to 1.0)

df = df / 255

**Model Training:**

Below are the steps taken to create model from training data.

1. Wrote the K-Mean algorithm.
2. Find the clustering using K-Means

k\_means = KMeans(k, df, travel\_centroids)

1. Find the Best-Mean based on lowest sum of errors.

for i in range(itr):  
 new\_k\_means = KMeans(k, df, travel\_centroids)  
 travel\_centroids = travel\_centroids + np.round(new\_k\_means.initialize\_centroids, decimals=3).tolist()  
 if (np.sum(k\_means.errors) > np.sum(new\_k\_means.errors)):  
 k\_means = new\_k\_means

**Models:**

I have used the below clustering technique to find the best for this assignment.

1. K-Means
2. Bisect K-Means

**Cross Validation:**

To Cross validate the clustering accuracy. I calculated the sum of errors for each iteration and selected the k-mean cluster with lowest sum of errors. Also performed the bisecting K-means and compare the sum of errors.

Also print the K vs SSEs plot for understand the relationship between K and SSEs.

**Step To Run the program:**

Run main.py file

>>python3 main.py

It will give 7 options as bellow.

Please select a options

1. Part 1 K-Means

2. Part 1 K-Means Bisect

3. Part 1 Plot data graph

4. Part 2 K-Means

5. Part 2 K-Means Bisect

6. Part 2 plot K vs SSEs

7. Exit

Enter your choice: 3

File save as fig.jpeg

Please select a options

1. Part 1 K-Means

2. Part 1 K-Means Bisect

3. Part 1 Plot data graph

4. Part 2 K-Means

5. Part 2 K-Means Bisect

6. Part 2 plot K vs SSEs

7. Exit

Enter your choice: 7

Process finished with exit code 0

**Plots:**

**Part-1 Data:**

**Chart, scatter chart

Description automatically generated**

**Part 2 K vs SSEs:**

**Chart, line chart

Description automatically generated**