Images Dataset clustering using K-Means Clustering (Unsupervised Learning)

Somesh Sunariwal
AITS
Rajasthan, India
Someshsunariwal@gmail.com

Abstract – **K-Means clustering is an unsupervised learning algorithm which is used in image segmentation or clustering the data which have the similar properties. In this K-Means clustering unsupervised learning project I clustered the images based on their shapes and colors. To start this project I took the dataset from kaggle.com which has 114 classes of images and each class contain approx 400 images. This dataset have 100x100x3 pixel images. K-means clustering algorithm helped to cluster images. I was used the n_clusters = 10 to achieve 10 clusters in this image dataset.**

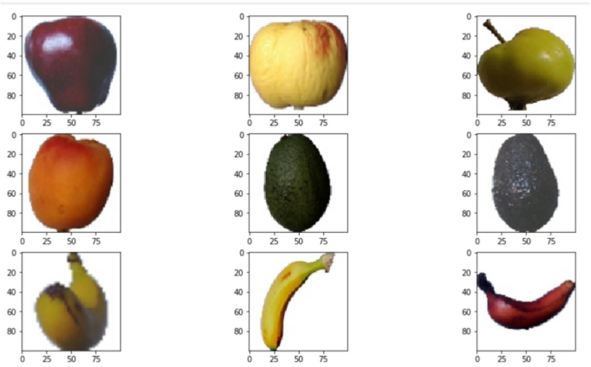
***Keywords*—Machine Learning, K-Means**

I. INTRODUCTION

K-Means clustering is a popular algorithm for unlabelled data. In order to achieve the great result from K-Mean select the value of k precisely. It generates clusters of the similar data points in datasets. In this K-Means project I set the k value to 10 in order to get 10 different clusters which hold the same properties. To perform the clustering on images I got the dataset from the Kaggle.com. This dataset consist of 57,276 images of fruits and these images are divided into 114 classes of fruits and each class contain approx. 400 images. Each image shape is 100x100 pixels with 3 colour channel. For this project I was used this dataset without label because it is unsupervised learning. K-Means was applied to entire dataset with 10 clusters which cluster the entire dataset images into 10 clusters. Result shows that the entire dataset images were clustered based on their shape of fruits and the colour of fruits that will i show you later in report.

II. DATA PRE-PROCESSING

Images are the representation of array of pixels which have the values between 0-1. Colour images have 3 dimensions and each dimension have the array of pixel which holds the value between 0-255. The dataset which I got from kaggle have the colour images which have 3 dimensions.



(**Fig 1**:- Images of dataset)

K-Means works on 2 dimension dataset. In order to apply K-Means I converted this 3 dimension data into 2 dimensions by multiplying the image height, weight and shape that was give me the shape of (19121, 30000). To reduce the time of computing I took the $1/3^{rd}$ images from each class and these extracted images have all type of possible rotation. So model was able to produce precise result.

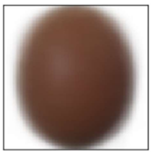
III. MODEL

K-Means model is used from sklearn library. Entire dataset was passed through the K-Means model with value of k=10. Value of k creates 10 clusters finding similarity in dataset. This K-Means model produces the clusters centroid and labels for the image dataset. Clusters centroid shape is (10, 30000) and labels shape is (19121,) respectively. Clusters and labels are help to reproduce images which are clustered on the basis on their shapes and colours.

IV. CLUSTERS PROPERTIES

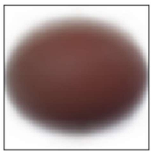
Produce resulted images with clusters on the basis of their shapes and colours shown below:

1. *Brown Oval Shape*



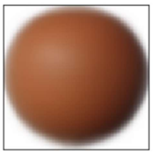
(Fig: - image1)

2. *Brown Rotated Oval Shape*



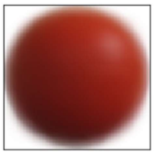
(Fig: - image2)

3. *Brown Round shape*



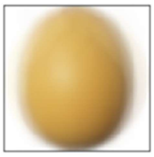
(Fig: - image3)

4. *Red Round shape*



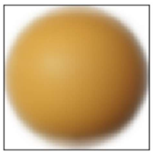
(Fig: - image4)

5. *Yellow Oval shape*



(Fig: - image5)

6. *Yellow Round shape*

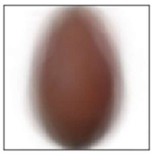


(Fig: - image6)

7. *Darkest Red round*

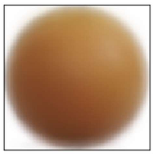
(Fig: - image7)

8. *Brown Long Oval*



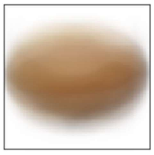
(Fig: - image8)

9. *Dark Yellow Round shape*



(Fig: - image9)

10. *Yellow Rotated Oval shape*

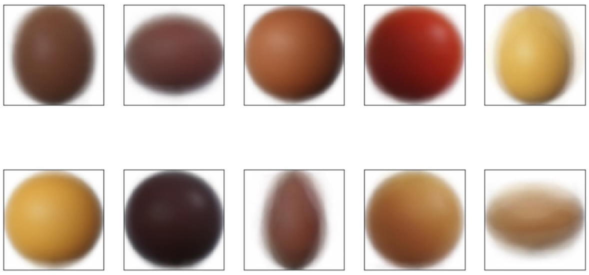


(Fig: - image10)

Above result clearly shows that K-Means algorithm clusters the dataset images on the basis of shapes and colours.

V. RESULT

As I discuss already the K-Means clustering algorithm is a unsupervised learning algorithm. It separates the data according to the similarity in data points.



(Fig:- Result Image after applying K-Means)

And the above result image shows that the dataset images are clustered based on their shapes and colours.