Graphical user interface, text, application

Description automatically generatedGraphical user interface, application

Description automatically generated

Student Name

***Sanatkumar Rajmogali Ippalpalli***

Title of Project Report

***Guided Project 6 – KMeans Clustering Image Segmentation***

Table of Contents

[Introduction 4](#_Toc83555493)

[Eckovation theme & Question 5](#_Toc83555494)

[Prerequisites before starting coding 6](#_Toc83555495)

[program DEVELOPMENT steps 8](#_Toc83555496)

[Dataset/Image requirements 8](#_Toc83555497)

[Technique – KMeans clustering algorithm 8](#_Toc83555498)

[PROGRAM / CODE DEVELOPMENT 9](#_Toc83555499)

[Analysis 18](#_Toc83555500)

[CONCLUSION 20](#_Toc83555501)

[Figure 1 Import libraries and datasets/modules 9](#_Toc83555542)

[Figure 2 Load Datasets/image 10](#_Toc83555543)

[Figure 3 Visualization of Image 12](#_Toc83555544)

[Figure 4 Code for estimate number of optimum clusters 14](#_Toc83555545)

[Figure 5 Model Building to Form clusters plots 15](#_Toc83555546)

[Figure 6 Image Segmentation 18](#_Toc83555547)

***EXECUTIVE SUMMARY***

Science and technology improved many technologies and has guided numerous innovative features which advanced image processing technology.

**Image segmentation** is the process of partitioning a digital image into multiple distinct regions containing each pixel with similar attributes i.e. classification of an image into different groups. There are different methods, and one of the most popular methods is the **k-means clustering algorithm.** K-Means clustering algorithm is an unsupervised algorithm, and it is used to segment the interest area from the background.

As part of guided project, one of the feature is the edge detection technology at identifying **clustering to segment image** as the details mentioned in the question.

KMeans clustering is a fundamental tool in machine learning algorithms.

# Introduction

Science and technology improved many technologies and has guided numerous innovative features which advanced the techniques in deep learning impacting computer vision, image processing.

Clustering is a set of techniques used to partition data into groups, or clusters. Clusters are loosely defined as groups of data objects that are more similar to other objects in their cluster than they are to data objects in other clusters. In practice, clustering helps identify two qualities of data:

* Meaningfulness
* Usefulness

We can perform clustering using many different approaches—so many, in fact, that there are entire categories of clustering algorithms. Each of these categories has its own unique strengths and weaknesses.

The *k***-means clustering** method is an [unsupervised machine learning](https://en.wikipedia.org/wiki/Unsupervised_learning) technique used to identify clusters of data objects in a dataset. There are many different types of clustering methods, but *k*-means is one of the oldest and most approachable. These traits make implementing *k*-means clustering in Python reasonably straightforward, even for novice programmers and data scientists.

Hence Eckovation includes this guided project in the courseware for students to understand, implementation / execute the code themselves.

This report includes the 5W1H about the theme of development of code and running the code with database available over the internet. At the end of the report, the conclusions share the adaptive thresholding & OTSU thresholding features extracted and useful for next course of activities to gain advantages in the edge detection activities development.

# Eckovation theme & Question

**Theme : K-Means Clustering: Image Segmentation**

**Image segmentation** is the process of partitioning a digital image into multiple distinct regions containing each pixel with similar attributes i.e. classification of an image into different groups. There are different methods, and one of the most popular methods is the **k-means clustering algorithm.** K-Means clustering algorithm is an unsupervised algorithm, and it is used to segment the interest area from the **background**..

**Question:**

Take a bright colorful image (Eg: image having fruits in it) and implement image segmentation using K-Means. You can first try to implement K-Means on iris dataset to understand its working and then extend the same logic, using the image pixels as the data points. Hint: All the K centroids will represent a color and therefore, you can initialize all the pixels to belong to a cluster randomly and then start the training of the centroids.

# Prerequisites before starting coding

1. Who - Software needed?
2. What - Version / Release of software?
3. Any Prerequisites
4. How - to install the software
5. Which -libraries are needed to execute the problem statement
6. Where – dataset requirements, path location to include in the code
7. When – to use the above feature extraction
8. Who – Software neeed?

Python

1. What- Version / Release of software?

Python version 3.6 (latest version of python)

1. Any Prerequisites

RAM space availability & hard disk space availability

Admin rights to install the software

1. How - to install the software
2. The following url <https://www.python.org/downloads/>can be referred to download python.
3. Second and easier option is to download anaconda and use its anaconda prompt to run the commands. To install anaconda check this url <https://www.anaconda.com/download/>
4. Which -libraries are needed to execute the problem statement
5. Import cv2 ( pip install OpenCV)
6. Numpy (pip install numpy)
7. Matplotlib (pip install matplotlib)
8. Pandas
9. Seaborn
10. Sklearn datasets
11. Sklearn metrics
12. Sklearn clusters
13. Where – dataset requirements, path location to include in the code
14. Once you have python downloaded and installed, you will need to setup PATH variables (if you want to run python program directly, detail instructions are below in how to run software section). To do that check this: [https://www.pythoncentral.io/add-python-to-path-python-is-not- recognized-as-an-internal-or-external-](https://www.pythoncentral.io/add-python-to-path-python-is-not-recognized-as-an-internal-or-external-command/) [command/](https://www.pythoncentral.io/add-python-to-path-python-is-not-recognized-as-an-internal-or-external-command/).
15. Setting up PATH variable is optional as you can also run program without it and more instruction are given below on this topic.
16. When – to use the above feature extraction
17. When – to use the above technique

There are many other [applications of clustering](https://en.wikipedia.org/wiki/Cluster_analysis#Applications), such as document clustering and social network analysis. These applications are relevant in nearly every industry, making clustering a valuable skill for professionals working with data in any field.

# program DEVELOPMENT steps

* Dataset/Image requirement
* Technique selections
* Program / code development
* Analysis

### Dataset/Image requirements

The image sources used for this project are downloaded / collected from internet.

Iris dataset

All fruits.jpg

### Technique – KMeans clustering algorithm

Conventional *k*-means requires only a few steps.

The first step is to randomly select *k* centroids, where *k* is equal to the number of clusters you choose. **Centroids** are data points representing the center of a cluster.

Text

Description automatically generated

Let us hop to the inscribing carving!

### PROGRAM / CODE DEVELOPMENT

As explained step by step during the lecture by mentor, we would approach steps and understand the basics with brief explanation as needed.

#### Step 1: Import the relevant libraries and applicable datasets/modules

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Figure Import libraries and datasets/modules

Step 2: Load dataset/Image and convert Pandas Dataframe

Table

Description automatically generated

Chart

Description automatically generated with low confidence

Figure Load Datasets/image

Visualize Images

Chart, scatter chart

Description automatically generated

Chart, line chart

Description automatically generated

Graphical user interface, text, application

Description automatically generated

A picture containing chart

Description automatically generated

Chart

Description automatically generated

Figure Visualization of Image

#### Step 3: Estimate optimum number of clusters

Graphical user interface

Description automatically generated with medium confidence

Graphical user interface, text

Description automatically generated with medium confidence

Figure Code for estimate number of optimum clusters

#### Step 4: Model Building

Apply the KMeans algorithm to build model

Graphical user interface, text, application, email

Description automatically generated

Text

Description automatically generated

Graphical user interface, application, scatter chart

Description automatically generated with medium confidence

Graphical user interface, text, application, email

Description automatically generated

Figure Model Building to Form clusters plots

Here are the parameters used in this example:

* **init** controls the initialization technique. The standard version of the *k*-means algorithm is implemented by setting init to "random". Setting this to "k-means++" employs an advanced trick to speed up convergence, which you’ll use later.
* **n\_clusters** sets *k* for the clustering step. This is the most important parameter for *k*-means.
* **n\_init** sets the number of initializations to perform. This is important because two runs can converge on different cluster assignments. The default behavior for the scikit-learn algorithm is to perform ten *k*-means runs and return the results of the one with the lowest SSE.
* **max\_iter** sets the number of maximum iterations for each initialization of the *k*-means algorithm.

#### Step 5: Image Segmentation

Image segmentation is the process of partitioning a digital image into multiple distinct regions containing each pixel with similar attributes i.e. classification of an image into different groups. There are different methods, and one of the most popular methods is the k-means clustering algorithm. K-Means clustering algorithm is an unsupervised algorithm, and it is used to segment the interest area from the background.

Graphical user interface, text, application

Description automatically generatedText

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated

Graphical user interface, text, application

Description automatically generated

A collage of fruit

Description automatically generated with low confidence

Figure Image Segmentation

This is continued to achieve desired results as per requirements.

### Analysis

The KMeans unsupervised learning clustering technique is completed on all fruits image.

We can use the techniques you learned here to cluster your own data, understand how to get the best clustering results, and share insights with others.

This entire program runs within few seconds.

# CONCLUSION

In this guided project, we built KMeans algorithm model to develop segmented image.

This is done in first attempt. Hence, the improvements in the code with time with multiple attempts may be checked and justified for the accuracy score.

This entire program runs within few seconds.

references:

1. [https://towardsdatascience.com/building-a-facial-recognition-model-using-pca-svm-algorithms-c81d870add16](https://towardsdatascience.com/building-a-facial-recognition-model-using-pca-svm-algorithms-c81d870add16 )
2. <https://learnopencv.com/otsu-thresholding-with-opencv/>
3. <https://www.tutorialspoint.com/opencv/opencv_adaptive_threshold.htm>
4. <https://www.tutorialspoint.com/opencv/opencv_colored_images_to_grayscale.htm>
5. <https://www.geeksforgeeks.org/python-thresholding-techniques-using-opencv-set-2-adaptive-thresholding/>