



# **CPSC-6430 Machine Learning: Implementation & Evaluation**

## **Project 4: K-Means Clustering**

**Dineshchandar Ravichandran  
C19657741  
Email: [dravich@g.clemson.edu](mailto:dravich@g.clemson.edu)**

## Contents

Introduction.....	3
1. Problem Statement.....	3
2. Project Screenshot: .....	3
3. Project Input and Output .....	4
3.1. Input: .....	4
3.2 Output: .....	5
3.2.1 Cluster data plot along with the final centroid points. ....	5
3.2.2 Coordinates for final centroid. ....	6

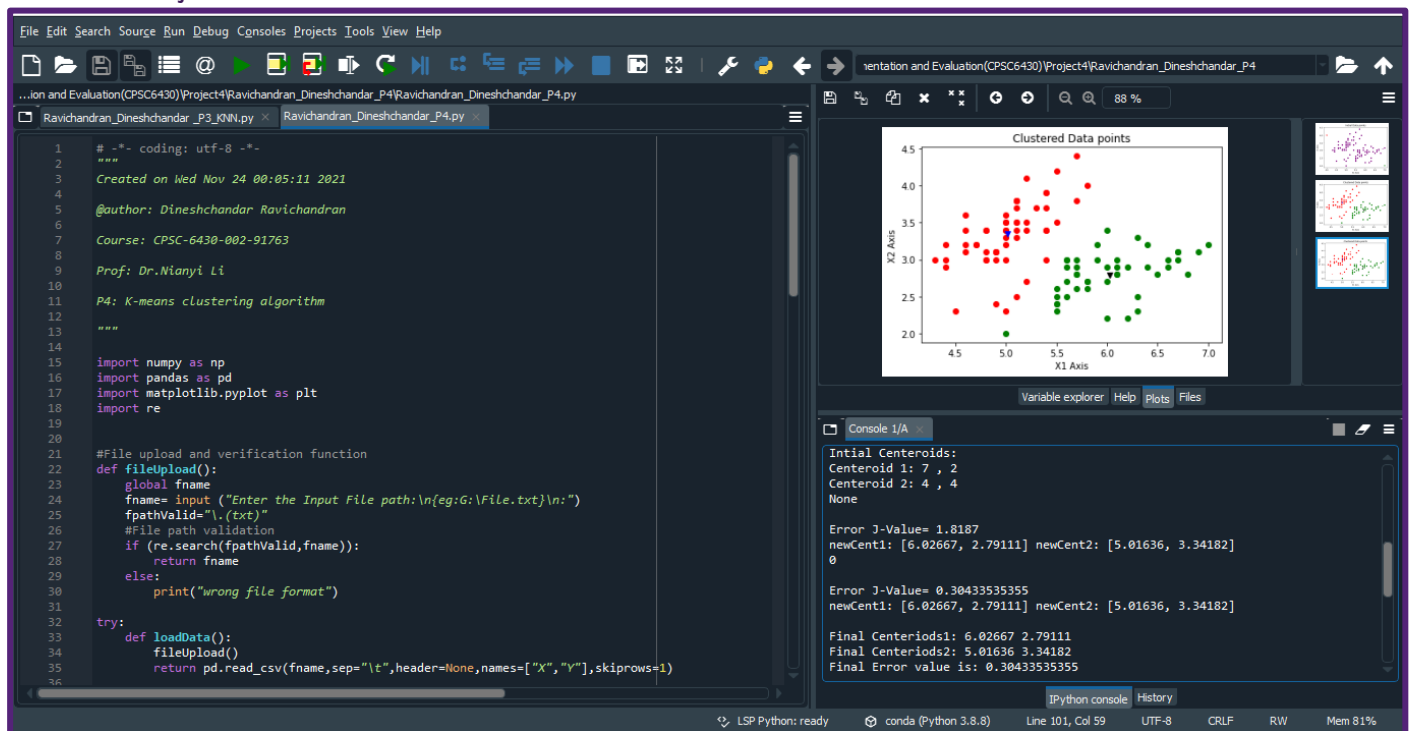
# Introduction

Project 4, to implement a “K-means clustering algorithm” program to perform unsupervised classification on a set of points received from the data file.

## 1. Problem Statement

To implement a Python program to read the data from a “.txt” file containing the data set received from the user. And cluster them in 2 clusters based on the centroids.

## 2. Project Screenshot:



- The above screenshots represent the code in the SPYDER IDE, along with the response in the console, where the cost value for the final centroid is **0.30433535355**.
- The value of initial centroids as extracted from the user-provided file are:
  - Centroid 1:[7, 2]
  - Centroid 2:[4, 4]
- The value of final centroids are:
  - Final Centroid 1: [ 6.02667, 2.79111]
  - Final Centroid 2: [5.01636 3.34182]
- The above graph illustrates the data points are being classified into clusters:
  - Cluster1: With centroid [6.02667, 2.79111] illustrated in red dots.
  - Cluster2: With centroid [5.01636 3.34182] illustrated in green dots.

- Console Screen Shot for the same:

```

In [12]: runfile('F:/Clemson/COURSE/SEM-1/Machine Learning Implementation and Evaluation(CPSC6430)/Project4/Ravichandran_Dineshchandar_P4/Ravichandran_Dineshchandar_P4.py', wdir='F:/Clemson/COURSE/SEM-1/Machine Learning Implementation and Evaluation(CPSC6430)/Project4/Ravichandran_Dineshchandar_P4')
Enter the name of a Data file

Enter the Input File path:
{eg:G:\File.txt}
:F:/Clemson/COURSE/SEM-1/Machine Learning Implementation and Evaluation(CPSC6430)/Project4/P4Data.txt
Enter the name of a Initial Centroid file

Enter the Input File path:
{eg:G:\File.txt}
:F:/Clemson/COURSE/SEM-1/Machine Learning Implementation and Evaluation(CPSC6430)/Project4/P4Centroids.txt

Initial Centroids:
Centroid 1: 7 , 2
Centroid 2: 4 , 4
None

Error J-Value= 1.8187
newCent1: [6.02667, 2.79111] newCent2: [5.01636, 3.34182]
0

Error J-Value= 0.3043353535
newCent1: [6.02667, 2.79111] newCent2: [5.01636, 3.34182]

Final Centroids1: 6.02667 2.79111
Final Centroids2: 5.01636 3.34182
Final Error value is: 0.3043353535

```

### 3. Project Input and Output

#### 3.1. Input:

- User-provided Data file containing coordinates of 100 samples named “P4Data.txt”.
- User-provided Centroid file containing two centroids named “P4Centroids.txt”.
- Following is the data extracted from the user mentioned above centroid file, presented in the console(highlighted in red):

```

In [12]: runfile('F:/Clemson/COURSE/SEM-1/Machine Learning Implementation and Evaluation(CPSC6430)/Project4/Ravichandran_Dineshchandar_P4/Ravichandran_Dineshchandar_P4.py', wdir='F:/Clemson/COURSE/SEM-1/Machine Learning Implementation and Evaluation(CPSC6430)/Project4/Ravichandran_Dineshchandar_P4')
Enter the name of a Data file

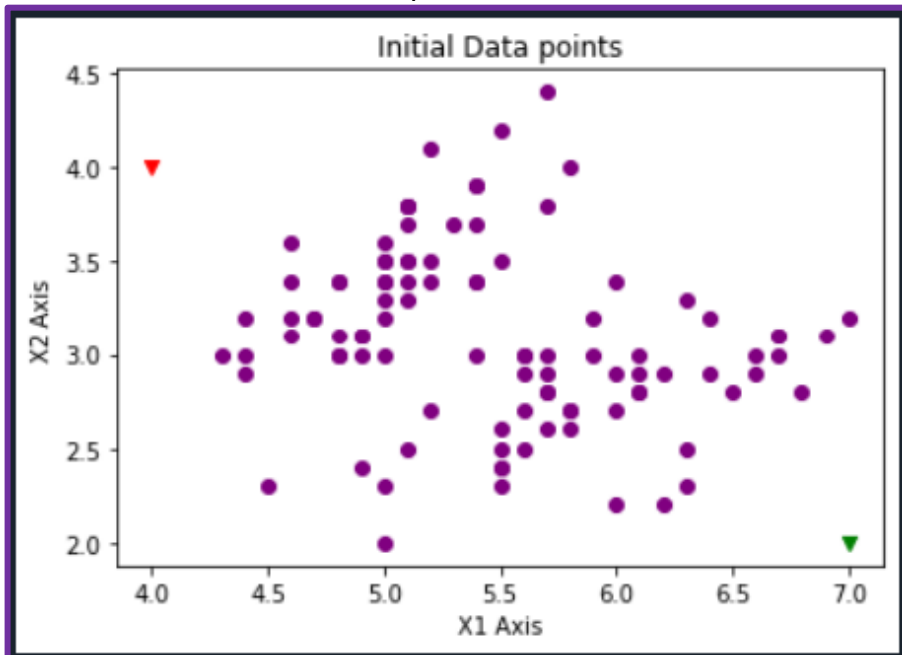
Enter the Input File path:
{eg:G:\File.txt}
:F:/Clemson/COURSE/SEM-1/Machine Learning Implementation and Evaluation(CPSC6430)/Project4/P4Data.txt
Enter the name of a Initial Centroid file

Enter the Input File path:
{eg:G:\File.txt}
:F:/Clemson/COURSE/SEM-1/Machine Learning Implementation and Evaluation(CPSC6430)/Project4/P4Centroids.txt

Initial Centroids:
Centroid 1: 7 , 2
Centroid 2: 4 , 4
None

```

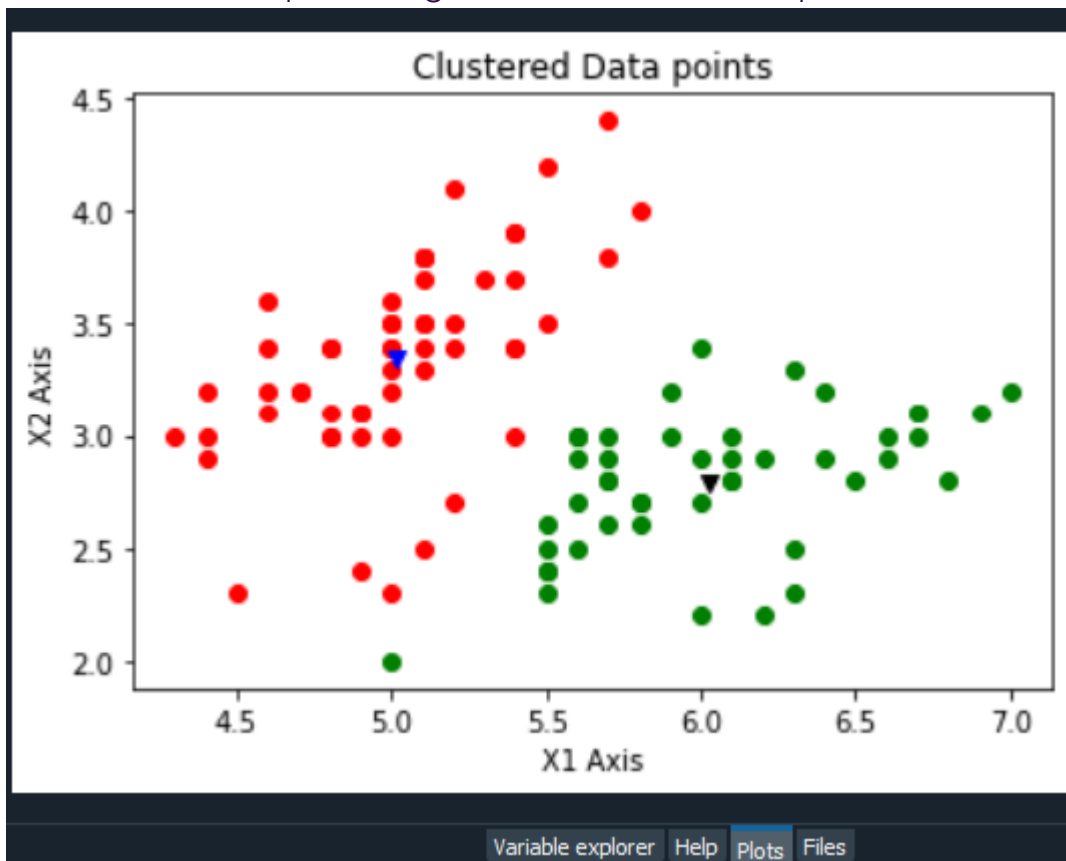
- The printout of the plot based on the data points and centroid points provided by the user before KMean actions are performed:



- Here the data points are presented in purple dots, and the centroid 1 [7,2] is represented as a Green triangle and centroid 2 [4,4] as a red triangle.

### 3.2 Output:

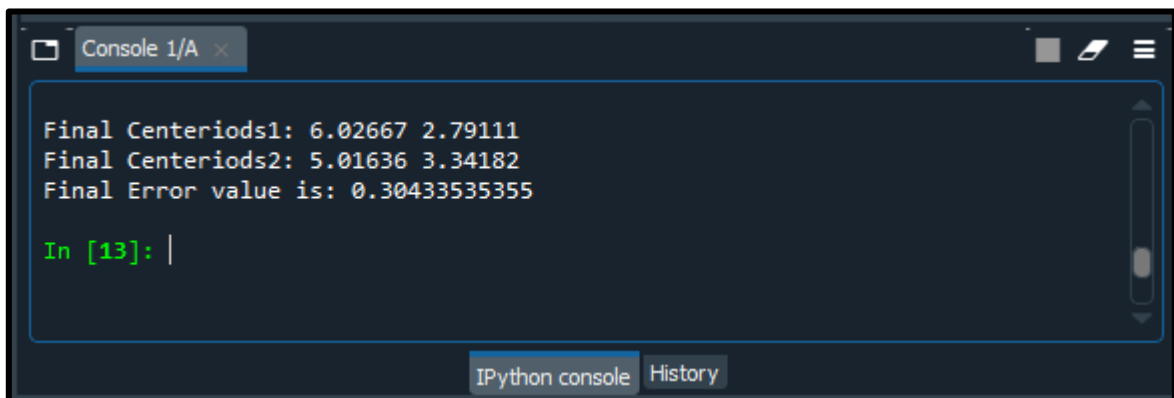
#### 3.2.1 Cluster data plot along with the final centroid points.



- All the data clustered with respect to centroid [5.01636 3.34182] (marked in the blue triangle) is represented as red dots.
- All the data clustered with respect to centroid [6.02667, 2.79111] (marked in the black triangle) is represented as green dots.

### 3.2.2 Coordinates for final centroid.

- The values for the final centroid are [6.02667, 2.79111] and [5.01636 3.34182], as illustrated below:



```
Final Centeriods1: 6.02667 2.79111
Final Centeriods2: 5.01636 3.34182
Final Error value is: 0.30433535355
In [13]: |
```

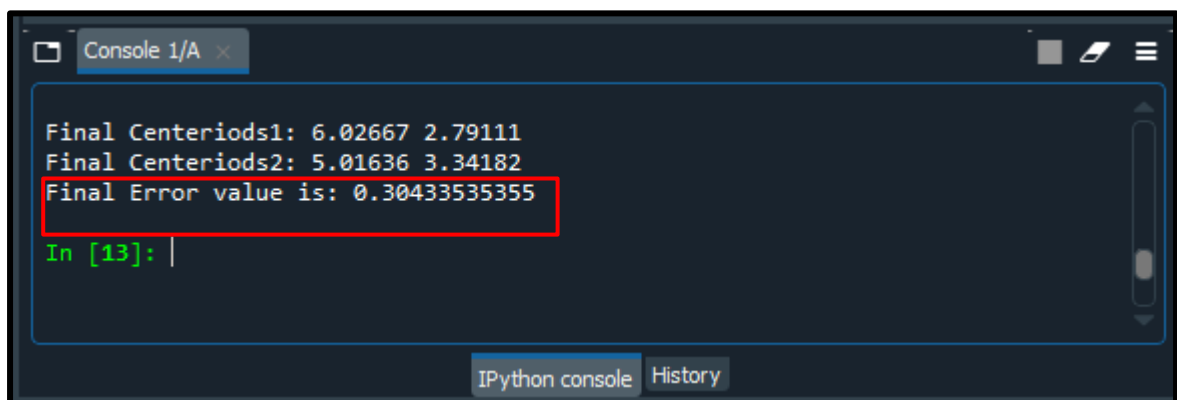
The screenshot shows an IPython console window with a dark background. The output text is as follows:

```
Final Centeriods1: 6.02667 2.79111
Final Centeriods2: 5.01636 3.34182
Final Error value is: 0.30433535355
In [13]: |
```

At the bottom of the console, there are two tabs: "IPython console" and "History".

### 3.2.3 Overall error.

- The overall error = 0.30433535355, as highlighted in red in the below screenshot.



```
Final Centeriods1: 6.02667 2.79111
Final Centeriods2: 5.01636 3.34182
Final Error value is: 0.30433535355
In [13]: |
```

This screenshot is identical to the previous one, but the line "Final Error value is: 0.30433535355" is enclosed in a red rectangular box to highlight it.