**Project Title :**

(i). Analysis and Validation of Kmeans and DBSCAN Clustering algorithm

(ii). Crop Recommendation based on soil fertility value

Team ID :

Team Leader Name: Shah Shubham

Team Member Names:

1. Shah Shubham
2. Sahil Dharaviya
3. Marag Modh
4. Harshad Parmar

Team Leader Email : shubhdev481@gmail.com

**Institute name** : Faculty of Technology and Engineering, The Maharaja Sayajirao University of Baroda.

**Abstract:**

We are having a dataset of soil nutrients values (nitrogen, phosphorous and potassium) and the crops suitable according to those nutrients’ values. We want to recommend crop to the farmers according to nutrient values of their soil. The data is unlabeled, so we are trying to generate clusters from this dataset and then predict that nutrients values given by farmer belong to which cluster and recommend crops which belongs to that cluster.

**Introduction**

In the 1st part using the soil fertility dataset we have try to create clusters using nutrients value using Kmean and DBSCAN clustering algorithm. To define the no. of clusters which can be created we have used elbow method in Kmeans algorithm and for it is required to find the optimal value of epsilon() and min points we have used K –distance graph. Using the value of K (no of clusters) which we got in both the clustering algorithms we have validated K value using Silhouette Coefficient we can calculated using below formula.

Silhoutte Score = b-a/max (b, a) Where a = inter cluster distance and b = intra cluster distance. After validating our clusters, we have plotted them using scatter plot.

In the 2nd part using a different soil fertility dataset using elbow method we generated the clusters and applying Kmeans clustering algorithm we have created clusters from dataset having the best silhouette score and plotted them using 3D scatter plot. In GUI section we have created a user interface using Tkinter library of python in which farmers need to enter the nutrients values of their soil and according to those values Kmeans will predict the crops for them which belong a particular cluster.

**Technologies**

* Algorithm : Kmeans Clustering , DBSCAN Clustering
* Dataset : Soil fertility dataset
* Programming language : Python
* Platform : Jupyter Notebook,  VS Code

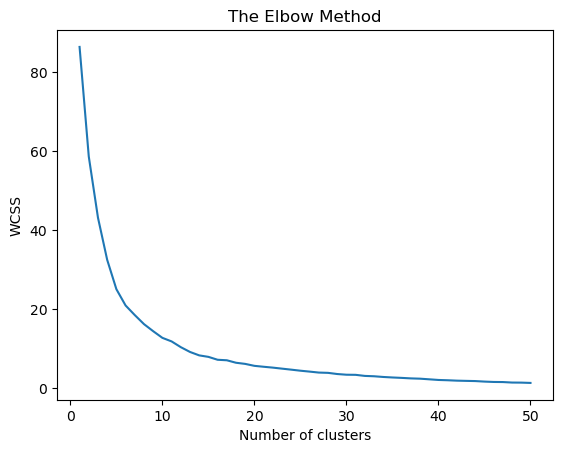
**Results**

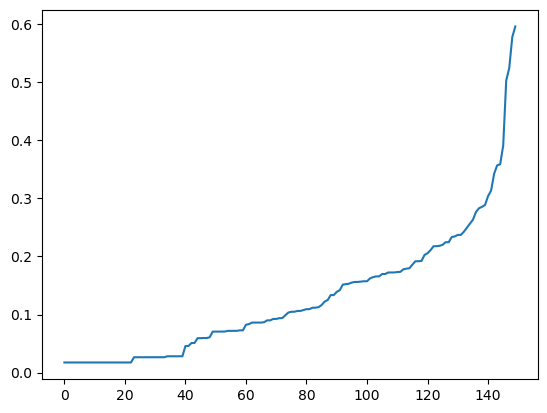
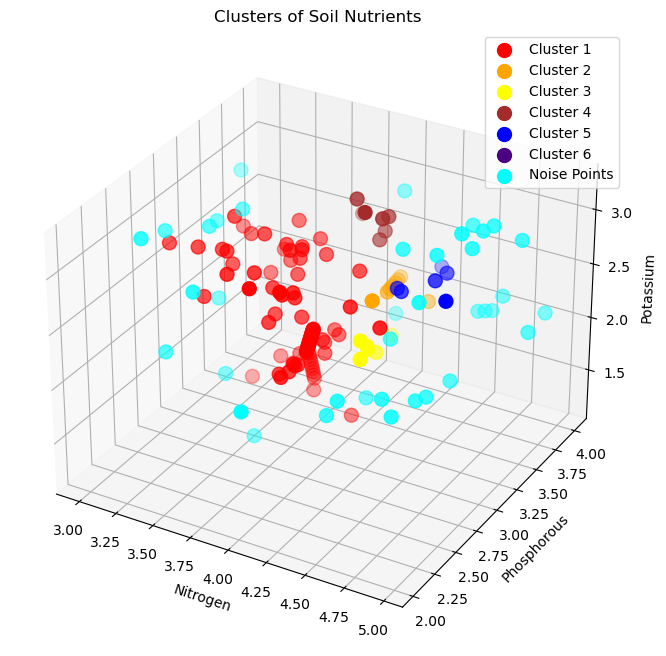
* In the 1st part we have analyze Kmeans and DBSCAN clustering algorithm and validated them using Silhouette Coefficient
* In the 2nd part the clustering model which we have created is giving the accurate value of crops for the nutrient’s values (nitrogen, phosphorous, potassium) we are providing as input in the user interface.

**Screenshot**

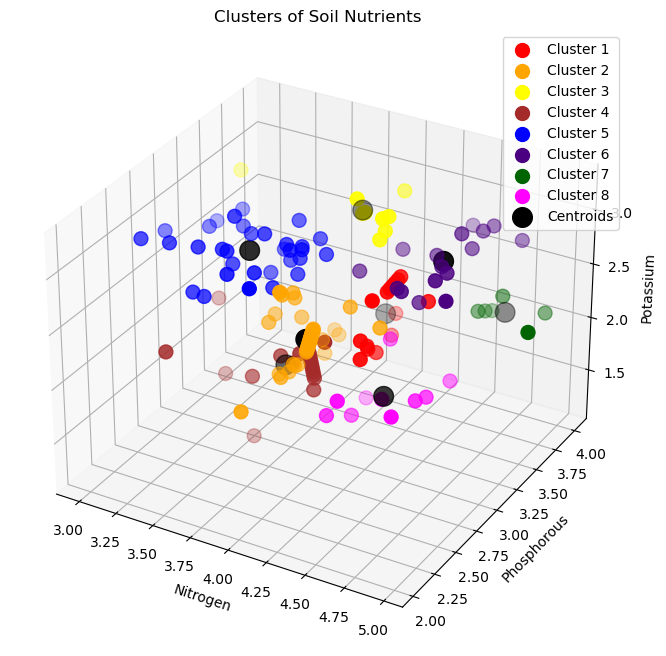
Part 1:

1. Kmeans

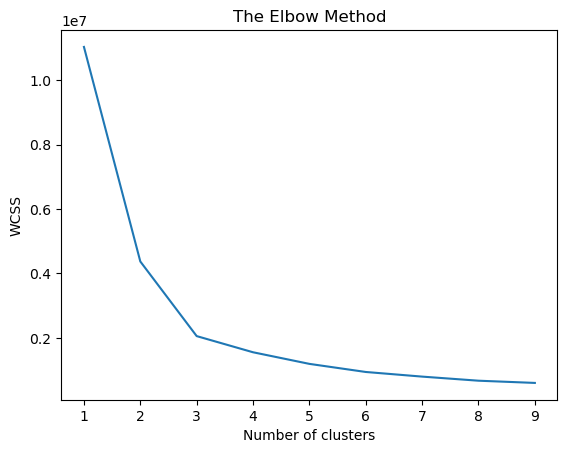


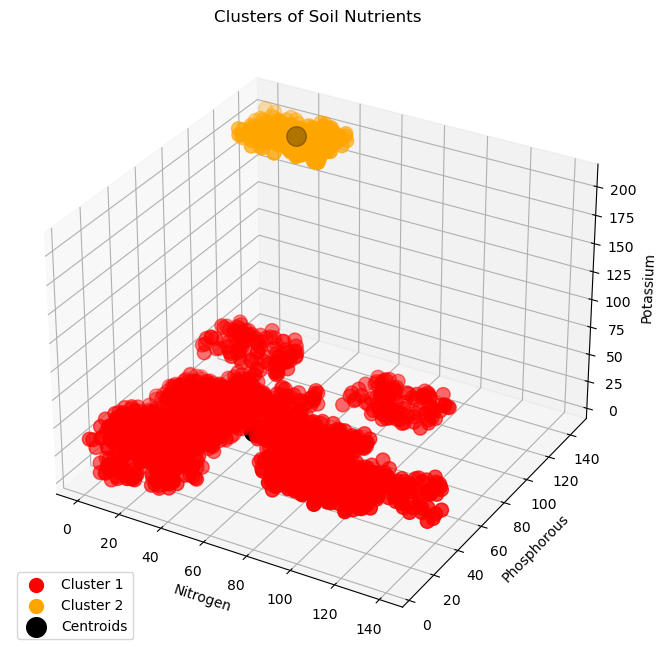


K distance graph for DBSCAN algo to get the optimal value of eps and minpts

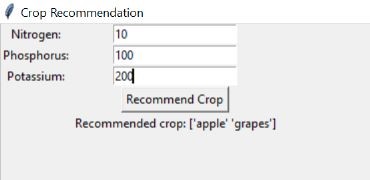


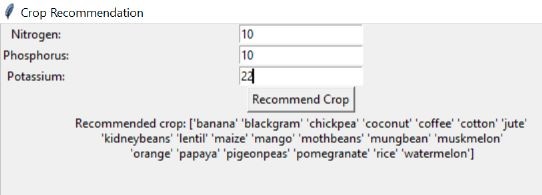
Part 2:





**Output**

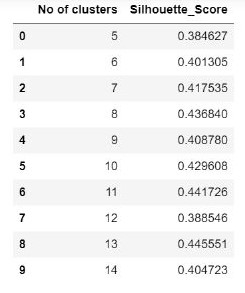
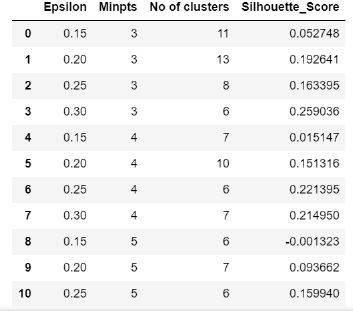
****

****

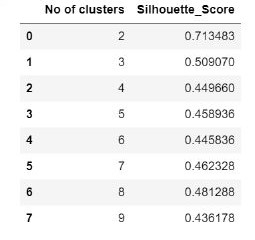
**Accuracy**

Part 1:

1. Kmeans (2) DBSCAN

Part 2:



**Conclusion**

* From the 1st part of clustering analysis, we have concluded that whenever we want to create clusters from a given dataset silhouette score is useful to find the accurate no of clusters that can be created.
* When the shape of clusters we have created is like oval or circle we can clearly say that Kmeans clustering algorithm should be used and when the points are very much scattered everywhere means clusters are not compact at that time we should use DBSCAN algorithm as this algorithm is robust to the noise points which are actually not part of any clusters and separate them from core and border points which Kmeans algorithm cannot separate as it considers all the points to create to generate clusters even if they are scattered to much

**Future Scope :**

1. Here the dataset of soil fertility values which we have used we don’t know about the accuracy of nutrients values and the crops which have been suggested according to their values. But somehow if we get the soil fertility dataset from the authenticate platform i.e., govt of India’s agriculture website then we can accurately predict the crops which are suitable for their soil according to the value of its nutrients and we can also suggest that farmers of a particular taluka of India should grow which crops in their farms if we can have a column having the names of talukas with nutrients value.
2. In this project we have assumed that farmers already have the nutrients value of their soil but in future we are planning to make the user interface in such a way that farmers have to just provide the image of their soil and if we get a dataset having the images of soil according to their nutrients value then we can predict the nutrients value for their soil by processing the provided image using TensorFlow.

Everything depends on dataset so if it is available and authenticate then it is possible otherwise not possible guarantee about the future scope which we have mentioned above.

**Reference:**

 Links:

<https://medium.com/@tarammullin/dbscan-2788cfce9389>

<https://medium.com/@tarammullin/dbscan-parameter-estimation-ff8330e3a3bd>

<https://www.geeksforgeeks.org/dbscan-clustering-in-ml-density-based-clustering/>

<https://www.javatpoint.com/density-based-clustering-in-data-mining>

<https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html>

<https://scikit-learn.org/stable/modules/generated/sklearn.metrics.silhouette_score.html#:~:text=The%20Silhouette%20Coefficient%20is%20calculated,max(a%2C%20b)%20>.

<https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1>

<https://towardsdatascience.com/silhouette-coefficient-validating-clustering-techniques-e976bb81d10c>