UNIT 5
Wiki Activity: Clustering

Murthy Kanuri
Machine Learning
University of Essex

Table of Contents

1 What is K-Means Algorithm logic

The K-Means Algorithm is an iterative algorithm that splits a data set into non-overlapping subgroups called clusters. The number of clusters created is determined by the value of k, a hyperparameter chosen before the algorithm runs.

Steps for K-Means

1. Choose the number of clusters(k) you wish to put the data into
2. Randomly initialise k points; these are called centroids
3. Identify the points closest to each centroid
4. Calculate the mean of the points in each cluster and move each centroid to that mean point
5. Repeat Step 3and four until the centroid value stays the same

2 Comparing the K-Means Algorithm logic with both Algorithm Logic

2.1 Shabal.in

- Demonstrates the principles of K-Means clustering algorithm logic
- Clicking on any of the points demonstrates the step-by-step visualisation of K-means clustering, focusing on centroid movement and cluster assignment process.
- Manual control allows the users to interact and observe

2.2 Naftali Harris

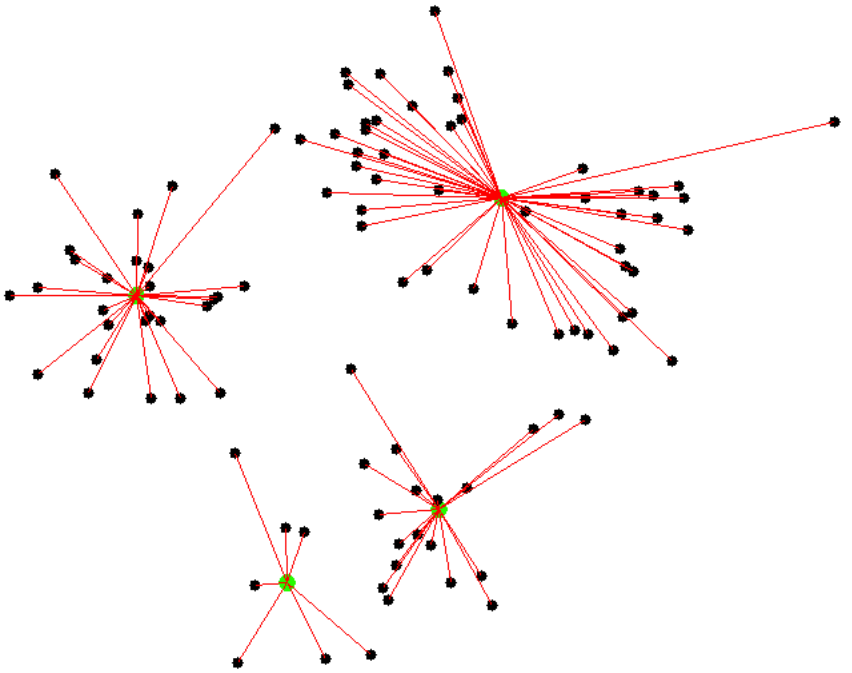
- Demonstrates the principles of K-Means clustering algorithm logic
- Three options are shown – I'll Choose, Randomly and Farthest Point. Select any option that shows further options such as "Uniform Points", Gaussian Mixture etc
- The process is shown in continuous animation and dynamic view (automated and hands-off)

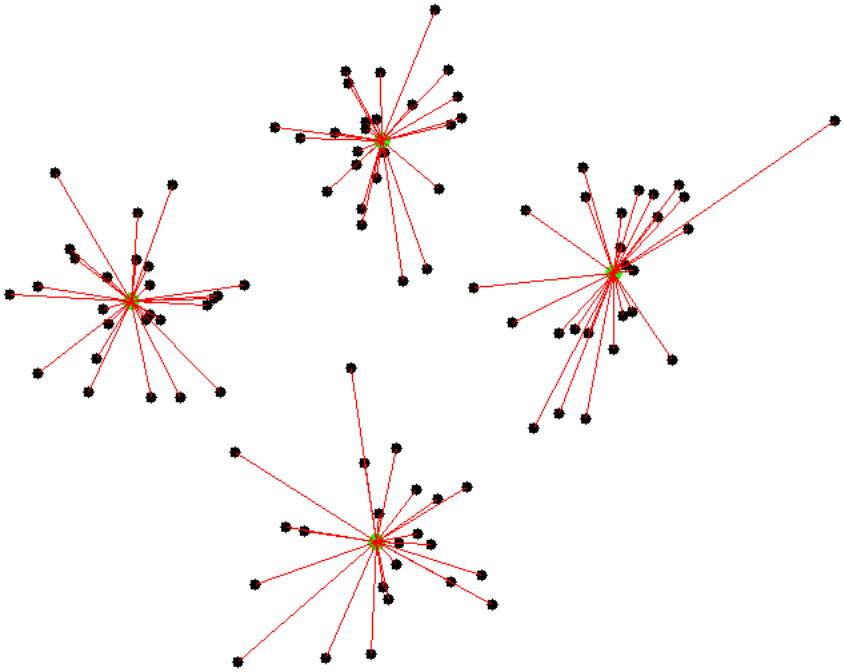
2.3 Legal, Social, Ethical, and Professional Considerations in Machine Learning

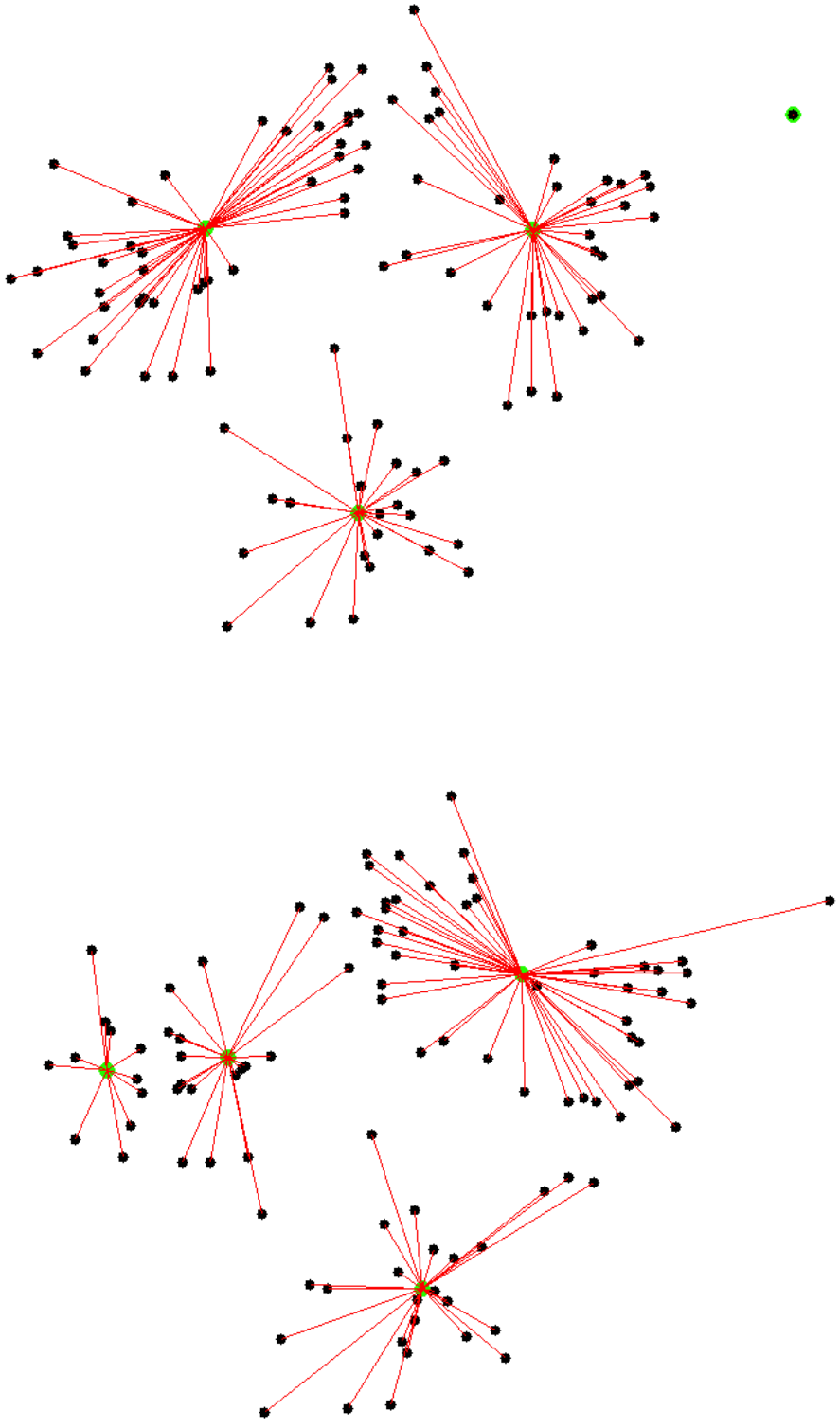
- Ensuring GDPR Compliance or HIPAA in Healthcare clustering
- Data biases can lead to discrimination or unethical profiling
- All the stakeholders should have transparency about the data being used
- Qualified professionals should be made responsible for algorithmic decisions to follow ethical standards

- Ensure that datasets used for clustering are accurate and true representatives.

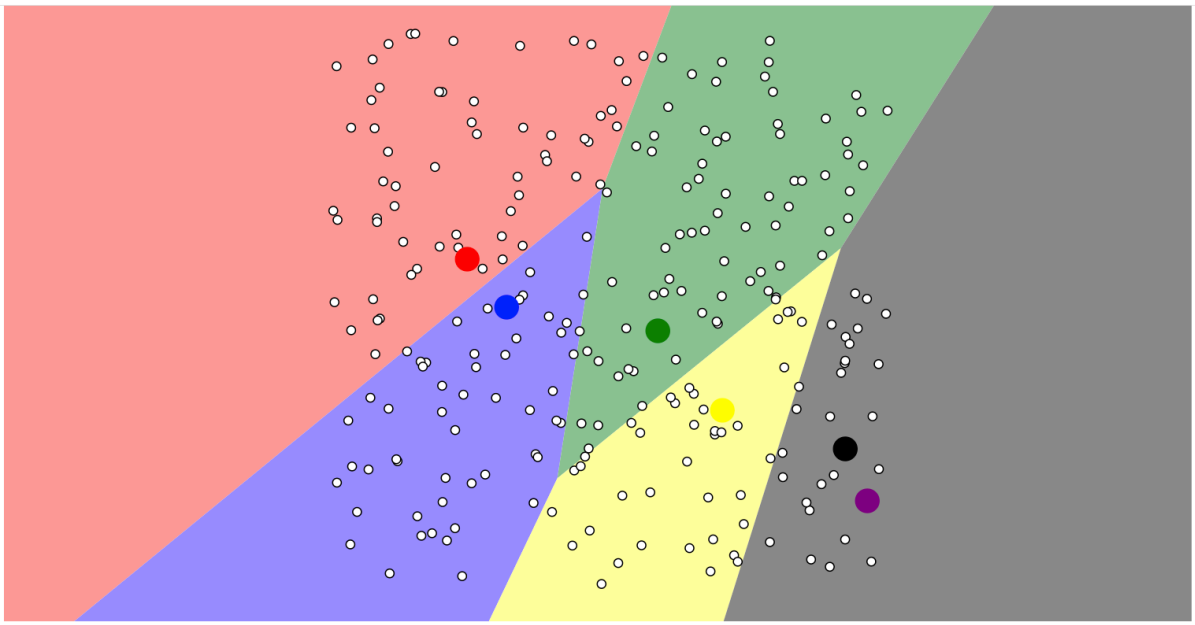
2.4 Images for Reference (Shabal.in)

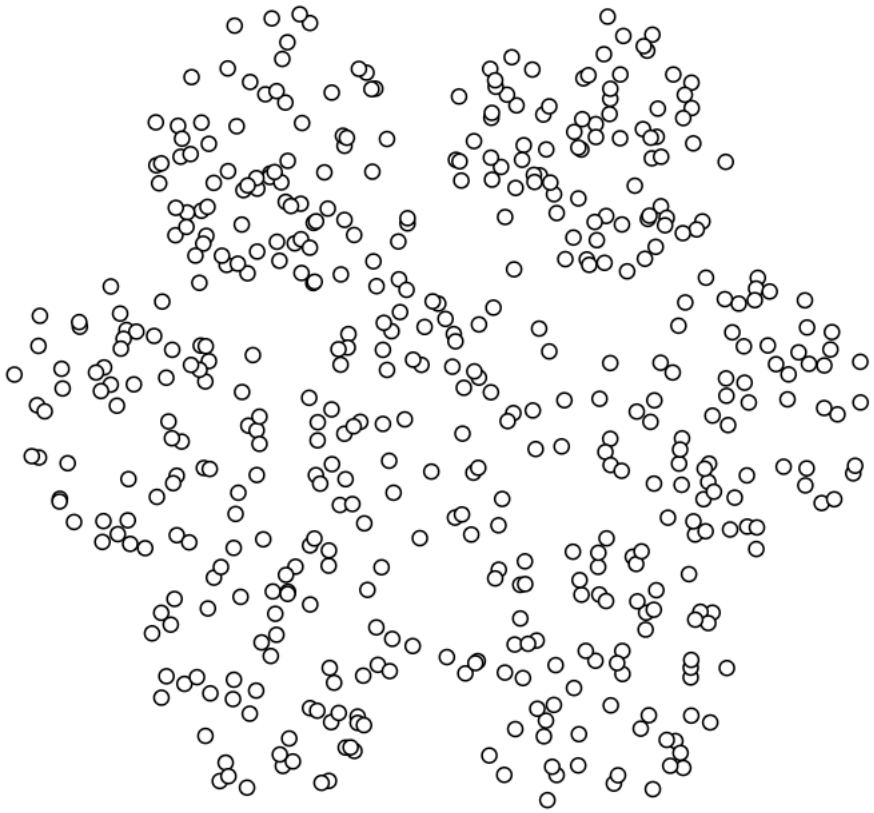






2.5 Images for Reference (Naftali Harris)





References

Shabal. (n.d.) K-Means Clustering Visualization. Available at: https://shabal.in/visuals/kmeans/2.html (Accessed: 10 November 2024).

Harris, N. (n.d.) Visualizing K-Means Clustering. Available at: https://www.naftaliharris.com/blog/visualizing-k-means-clustering/ (Accessed: 10 November 2024).

Patel, S. (2019) K-means Clustering Algorithm: Implementation and Critical Analysis. Saarbrücken: Scholars' Press.

Dwivedi, G. (2023) Optimization of K-Means Clustering Using Genetic Algorithm. Saarbrücken: LAP LAMBERT Academic Publishing.