k-Means Clustering

What is Clustering?

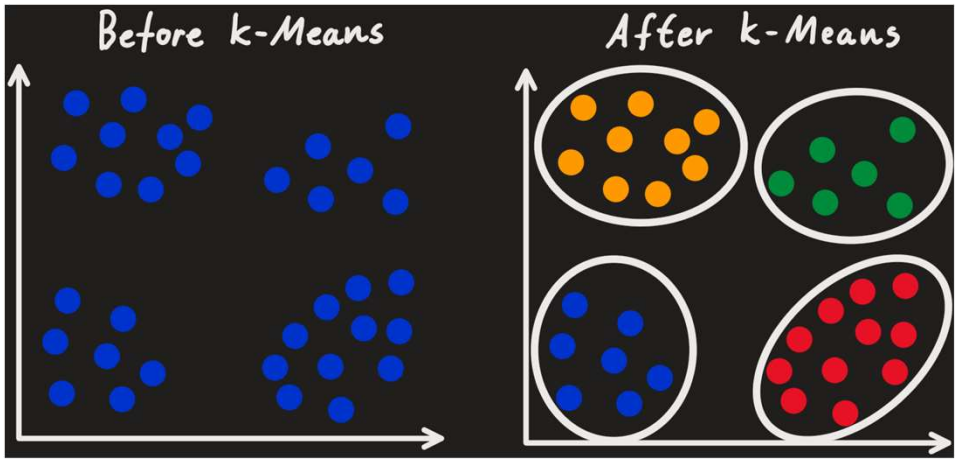
- Clustering is a technique used in unsupervised machine learning to group a set of objects in such a way that objects in the same group, which is called a cluster are more similar to each other than to those in other groups.

The k-Means Algorithm Explained

- k-Means is a popular clustering algorithm that organizes a specified number of data points into a predefined number of clusters. The goal is to minimize the variance within each cluster, where each point belongs to the cluster with the nearest mean.

Example

- Segmenting credit card customers based on spending behavior and usage patterns to tailor marketing strategies and enhance customer service.



The image illustrates the K-Means cluster algorithm. It shows unclassified data points on the left graph and classified clusters on the right graph.

Implementation

1. Data Preparation

- Load the credit card customer dataset.
- Drop unnecessary identifiers and convert data to floats.
- Use features like Total Visits Online and Average Credit Limit.
- Display pair plots and density distributions to identify initial clusters visually.

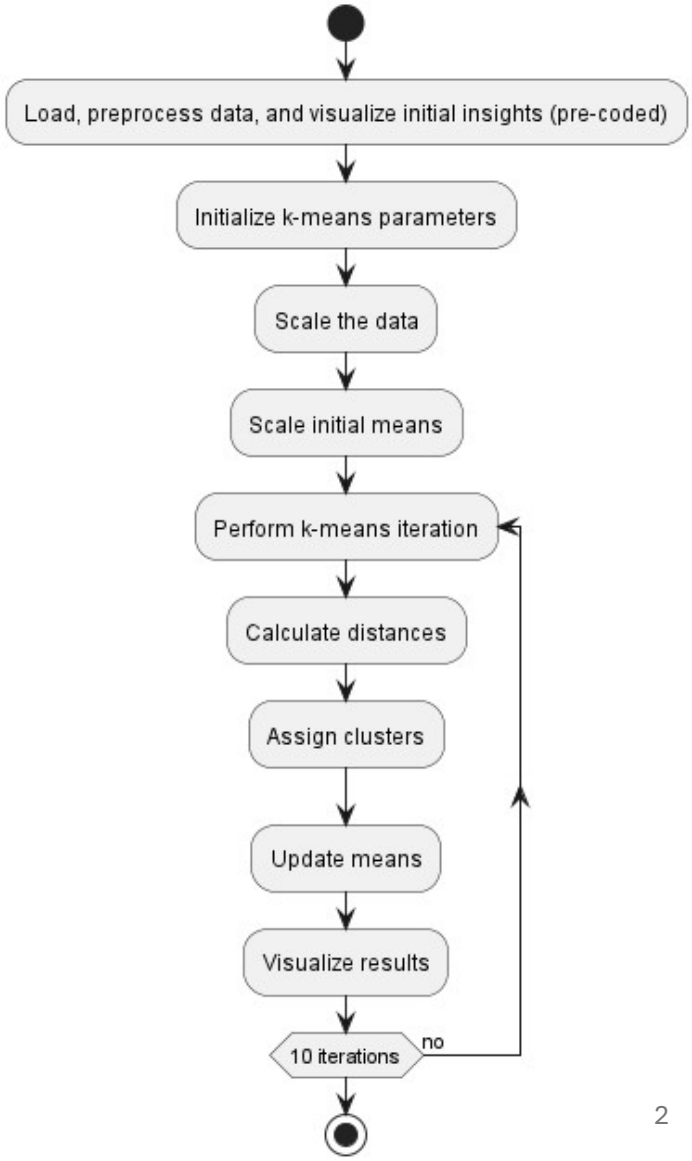
2. Scaling and Initialization

- Set Initial Cluster Centers
- Apply MinMaxScaler scaling to the dataset and cluster centers.

3. k-Means Iterative Process

- Calculate Euclidean distances between each data point and cluster centers.
- Assign each data point to the nearest cluster based on the minimum distance.
- Update the cluster centers based on the mean of the assigned data points.
- Visualize cluster assignments and updated centroids after each iteration.

Clustering - algorithm description



Results and Visualization

Visualization insights

- The Plots depict the progression of k-Means clustering from Iteration 1 to Iteration 2, highlighting changes in centroids' positions and cluster assignments. After the second iteration, there were no further changes.

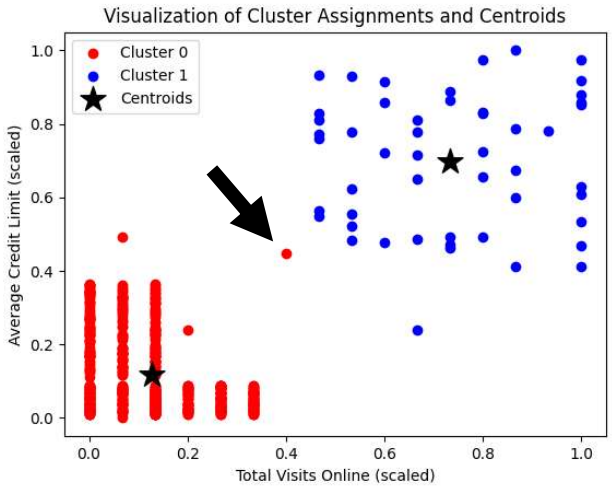
Key Observations

- Centroids move towards denser regions, enhancing cluster accuracy.
- Data points are reassigned between clusters, reflecting refinement in their classification.

Conclusions

- The algorithm adjusts centroids dynamically based on data, optimizing clustering.
- Improved centroid positioning results in more accurate and meaningful clusters, valuable for targeted business strategies.

Plot of the first Iteration:



Plot of the second Iteration: