

K-Means Clustering on Iris Dataset

Project Overview:

This project applies the K-Means clustering algorithm on the Iris dataset using two features: Sepal Length and Sepal Width. The goal is to divide the dataset into three clusters and analyze grouping behavior.

Dataset:

The Iris dataset contains 150 samples of iris flowers with different measurements.

For this project, we used:

- Sepal Length
- Sepal Width

Objective:

To group the data into 3 clusters using unsupervised learning.

Methodology:

1. Load the Iris dataset.
2. Select Sepal Length and Sepal Width features.
3. Apply K-Means clustering with $K=3$.
4. Fit the model and predict cluster labels.
5. Visualize the clusters using scatter plot.

Algorithm Used:

K-Means Clustering:

- Randomly initialize centroids.
- Assign points to nearest centroid.
- Update centroids.
- Repeat until convergence.

Tools & Technologies:

- Python
- Scikit-learn
- Matplotlib
- Pandas

Conclusion:

The K-Means algorithm successfully grouped the Iris dataset into three distinct clusters based on sepal measurements. Though clustering is unsupervised, results closely resemble the original species separation.