**Project 3- Iris Data KMeans Clustering**

Latheesh Mangeri  
1002116009  
2242-CSE-6363-005

**Problem:**

Classification on the Iris dataset using linear regression and cross-validation. The goal is to build a model that can accurately predict the species of an Iris flower based on its sepal and petal measurements.

**Data:**

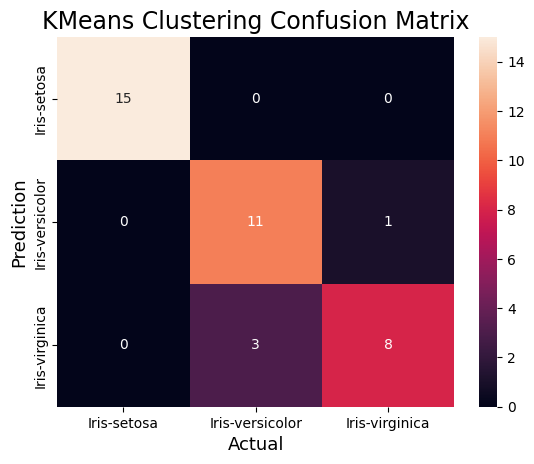
The Iris dataset contains 150 data points, each representing an Iris flower. Each data point has four features: sepal length, sepal width, petal length, and petal width. The label is the species of the flower, which can be Iris setosa, Iris versicolor, or Iris virginica. You can download the data from the provided URL.

**Methodology:**

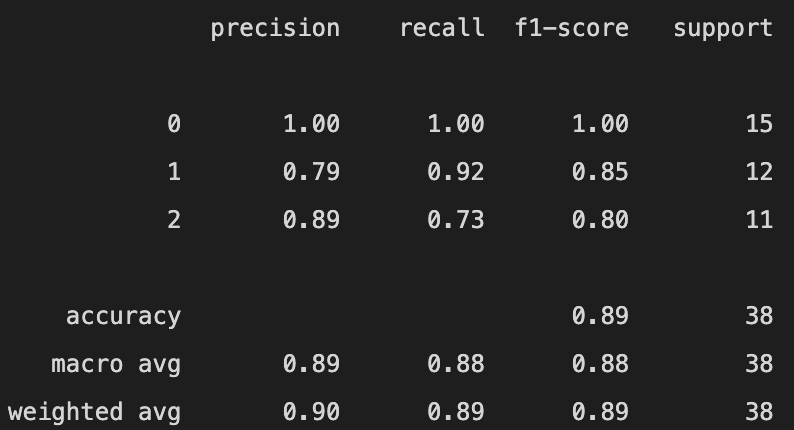
1. **Preprocessing:**
   1. Load the data into your preferred programming language (e.g., Python, R).
   2. Split the data into features and labels.
   3. Consider encoding the categorical label (species) into numerical values.
2. **Model Training:**
   1. Implemented a KMeans Clustering model.
   2. Use (k=3) to train and evaluate the model.
   3. Within each:
      1. Split the data into training and testing sets by 75-25.
      2. Trained the linear regression model on the training set.
      3. Evaluated the model performance on the testing set using metrics like accuracy, precision, recall, and F1-score.

**Results:**

* **Confusion Matrix:**

****

* **Classification Report**

****

**Clusters with n Centroids:**A group of colored dots

Description automatically generated