**🧩 K-Means Clustering Algorithm for Unsupervised Learning Tasks through Storytelling**

## **📖 Overview**

Why not let the data tell its own story? This project explores **K-Means clustering**, an unsupervised machine learning algorithm, in a creative storytelling format. The task involves clustering universities into two categories—**Private** and **Public**—without using their actual labels during clustering. The goal is to demonstrate how K-Means identifies hidden patterns in data and analyze how closely the clusters align with reality when compared to the true labels.

In addition to applying clustering, this project emphasizes **learning through creativity**, blending technical concepts with storytelling to enhance understanding. 🌟

## **🔍 Method**

The project follows these key steps:

1. **Data Collection and Preparation:** 📊 Load the dataset containing information about universities.
2. **Data Preprocessing:** 🔄 Standardize the data using **StandardScaler** to ensure all features are on the same scale.
3. **Clustering with K-Means:**
   * Set the number of clusters to 2 (for Public vs. Private). 🔢
   * Run the K-Means algorithm to group universities based on feature similarity.
4. **Evaluation:** 📈 Although K-Means is unsupervised, the clusters are compared with the actual labels using:
   * **Confusion Matrix**
   * **Classification Report**This is done purely as a learning exercise to observe how well the algorithm performed.

## **📊 Outcome and Results**

* The K-Means algorithm successfully divided the universities into two clusters. ✅
* The confusion matrix revealed how well the clustering aligned with the actual categories.
* Although the clustering wasn’t perfect, it showed that **K-Means can identify meaningful patterns** in the data even without labeled information. This exercise highlights the potential and limitations of clustering algorithms in real-world scenarios. 🔍

## **🔧 Data Processing**

1. **Loading the Dataset:** 📥 The dataset consists of university-related features like student population, faculty ratio, and other key metrics.
2. **Scaling the Data:** ⚖️ Using **StandardScaler**, each feature was transformed to ensure fair comparison and avoid biases from varying magnitudes.
3. **Clustering:** 🧑‍🏫 The K-Means algorithm was applied with 2 clusters to separate universities into **Private** and **Public** categories.
4. **Validation (Optional):** 📋 Compared the generated clusters to true labels, revealing insights into K-Means' performance.

## **📚 References**

* Scikit-learn Documentation: K-Means Clustering
* Medium Article: [K-Means Clustering Algorithm for Unsupervised Learning Tasks](https://medium.com/@sayahfares19/k-means-clustering-algorithm-for-unsupervised-learning-tasks-f761ed7f37c0)
* Dataset Source: (Provide the dataset link if applicable)
* Python Libraries Used:
  + 🐼 pandas
  + 🔢 numpy
  + 📚 scikit-learn
  + 📊 matplotlib
  + 📈 seaborn

## **🎨 Creative Note**

This project was crafted as part of my **learning journey in machine learning and data science**. By telling the story of Professor Geoff and his students, I aimed to make technical concepts more engaging and accessible. ✨

## **🔗 Project Link**

Explore the full project on GitHub: [K-Means Clustering Algorithm through Storytelling](https://github.com/esraalegebaly/K-Means-Clustering-Algorithm-for-Unsupervised-Learning-Tasks-through-Storytelling).