Assignment 3- Segmentation and Modeling of Iris Species

Fahasena Faisal ID-21068198 **Dataset-https://www.kaggle.com/datasets/arshid/iris-flower-dataset?resource=download**

ABSTRACT:

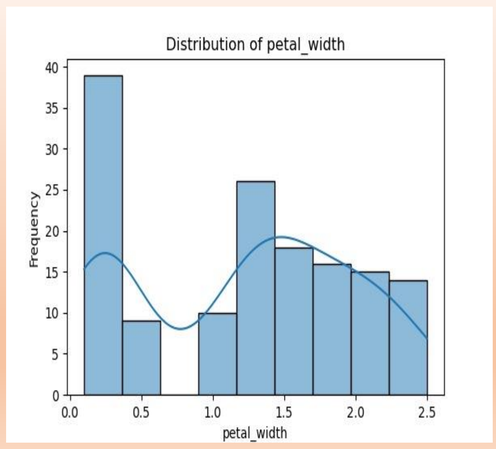
- This analysis of the Iris dataset encompasses data preprocessing, outlier analysis, and feature visualization.
- K-Means clustering and PCA reveal underlying structures, leading to the development and optimization of a K-Nearest Neighbors (KNN) model.
- The model's accuracy is demonstrated through evaluation metrics and visualizations, providing a holistic understanding of the dataset.

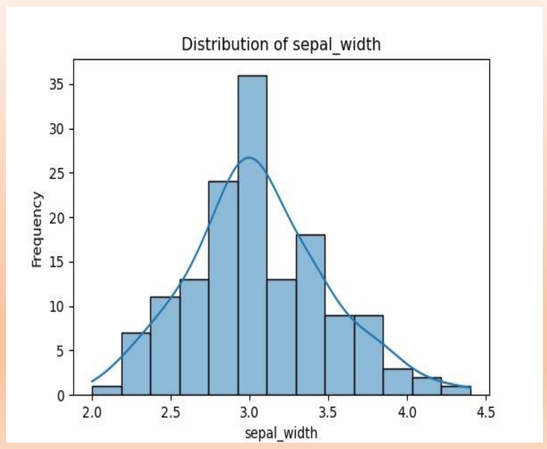
INTRODUCTION:

- In this analysis, we delve into the exploration and classification of the famous Iris dataset, aiming to uncover patterns within the features of iris flowers.
- We embark on a journey of data preprocessing, outlier detection, visualization, and model building to gain insights into the underlying structure of the dataset.
- The ultimate goal is to classify iris flowers accurately using a K-Nearest Neighbors (KNN) model.

RESULT AND IMPLEMENTATION:

HISTOGRAM

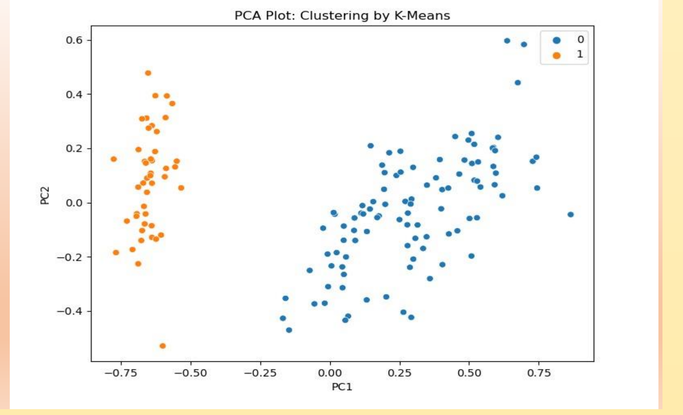


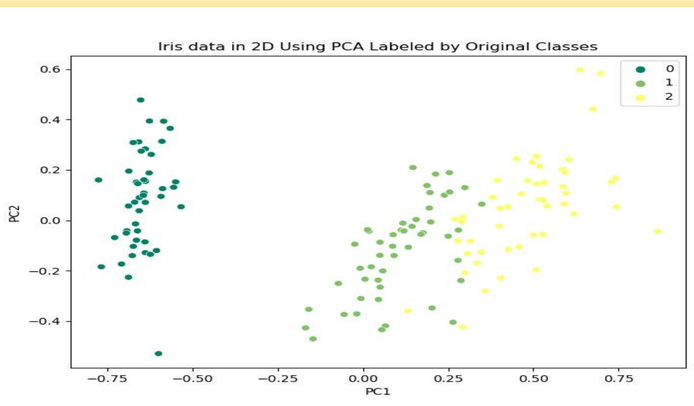


- Histograms illustrate the frequency distribution of each feature, offering insights into the spread and concentration of values.
- These visualizations provide an initial understanding of the data's characteristics.
- picture the dissemination of each feature, including a kernel thickness gauge (kde) for perfection.
- Every histogram shows the recurrence of values for a specific capability

SCATTER PLOT

- Scatter plots depict relationships between features such as sepal length and width, and petal length and width.
- Each point is colored by class, revealing correlations and patterns.
- This code makes a scatterplot utilizing Seaborn ('sns.scatterplot') that plots the connection between seventh length and width in the Iris dataset ('data').
- Each point is hued by its group ("setosa" "versicolor" or "virginica"), which permits the perception of examples or bunches

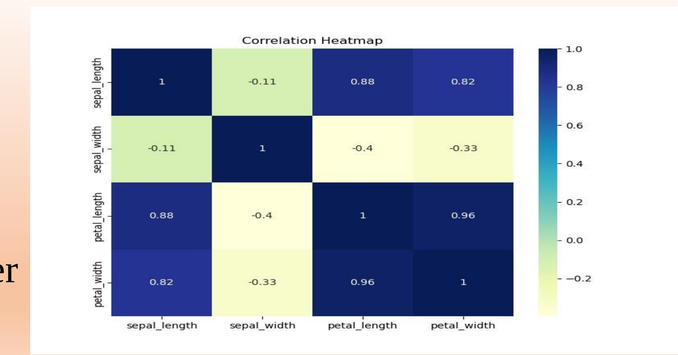


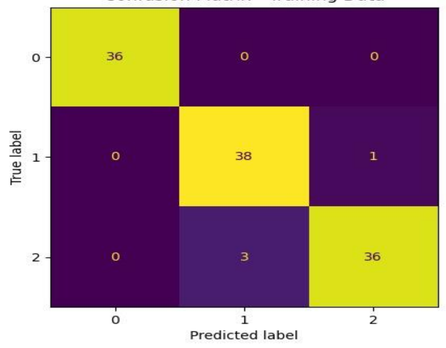


- This code makes a 2D scatterplot utilizing PCA-changed Iris data, with each point hued in view of its unique class utilizing a "summer" variety range.
- This plot visualizes the original class distribution in 2D space using PCA.
- Each point represents an iris flower, colored by its original class, offering a comparison to the clusters identified by K-Means and the KNN model.

CORRELATION HEATMAP

- This computes the relationship grid of the Iris dataset (barring the "class" segment) and afterward delivers it as a heatmap utilizing Seaborn (Yin *et al.* 2023).
- A heatmap colors cells in view of relationship values, which gives insight into how features are connected.
- This provides a comprehensive view of inter-feature relationships. Brighter colors indicate stronger correlations, aiding in feature selection.





- Confusion matrices visually summarize the KNN model's classification performance on training and testing data.
- They reveal true positive, true negative, false positive, and false negative predictions.
- The color-enhanced matrices aid in evaluating the model's accuracy and identifying areas for improvement in classifying iris flowers.

CONCLUSION:

- In conclusion, analysis of the Iris dataset utilizing “K-Nearest Neighbor” (KNN) and “Principal Component Analysis” (PCA) gives significant data to both classification and dimensionality decrease procedures.
- Utilizing KNN, we had the option to characterize iris species in light of their veil and petal measurements, which demonstrated the viability of algorithms in managing multi-class classification tasks.
- The utilization of PCA empowered the perception of the great layered Iris data set in a lower-layered space.

GitHub Repository Link-https://github.com/FahasinaFaisal/Assignment_Clustering