

PCA + K-Means Clustering Report

Name: Vaishnavi
SRN: PES2UG23CS367
Section: F

1. Introduction

This report explains the full workflow of the PCA and K-Means clustering notebook, including preprocessing, PCA dimensionality reduction, custom K-means implementation, optimal K selection, and all visual output screenshots.

2. Data Preprocessing

The dataset is loaded, cleaned, and scaled using StandardScaler. Scaling is crucial for PCA and K-means, ensuring all features contribute equally.

3. PCA

PCA reduces multidimensional data into two principal components while preserving maximum variance. This allows effective visualization of patterns.

4. K-Means Algorithm (From Scratch)

A custom implementation initializes centroids, assigns points to clusters using Euclidean distance, recalculates centroids, and repeats until convergence.

5. Finding Optimal K

The Elbow Method and Silhouette Analysis are used to determine the best cluster count by measuring compactness and separation.

6. Results & Visualizations

All output screenshots generated in the notebook are attached below, including PCA scatterplots, elbow curve, silhouette curve, and final clustering visualization.







