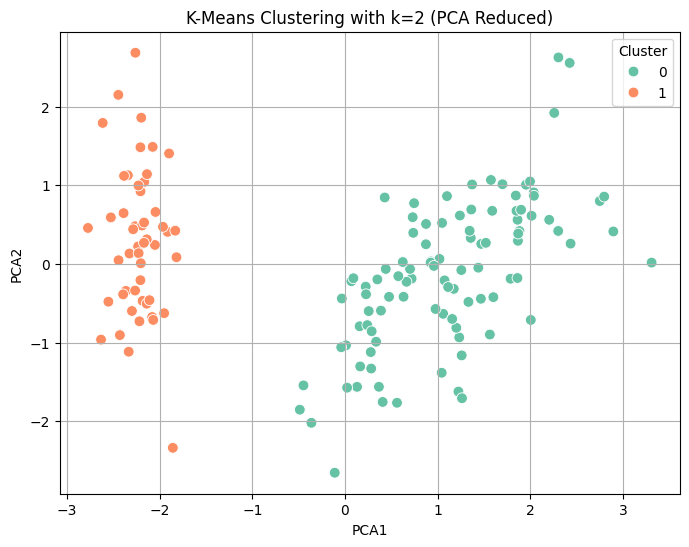
# ML Lab 08

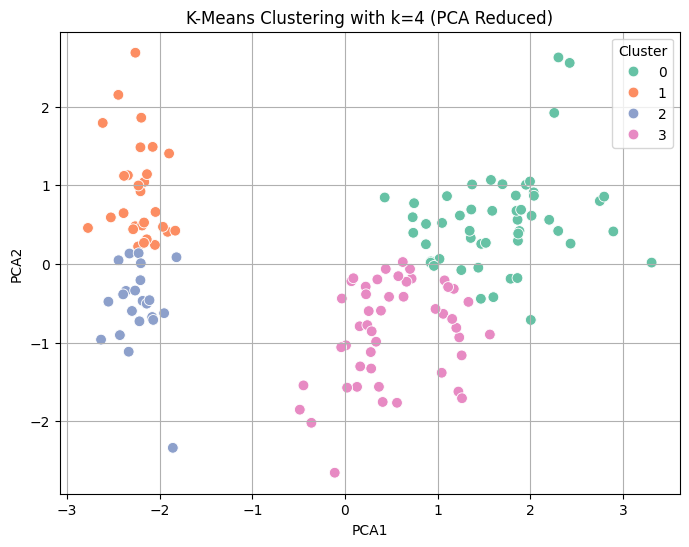
## K-Means Clustering Lab Exercise

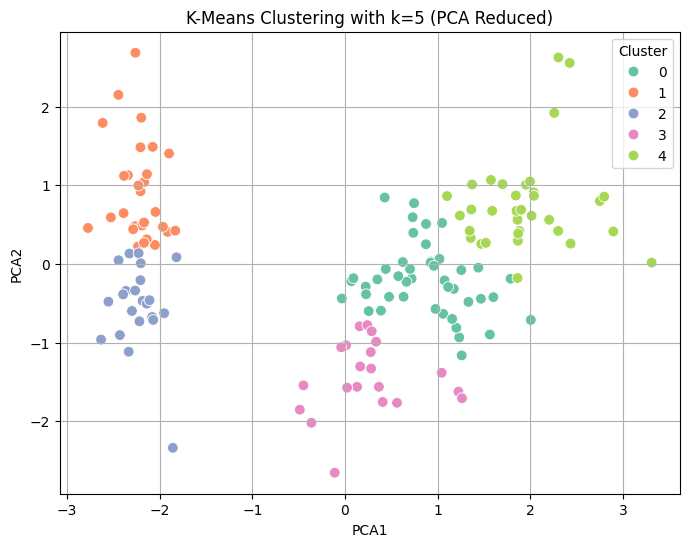
### Task 01

|  |  |  |
| --- | --- | --- |
| **k** | **Silhouette Score** | **Interpretation** |
| 2 | 0.58 | Best separation |
| 3 | 0.48 | Likely reflects some overlap between clusters |
| 4 | 0.39 | Poor clustering, over-segmentation, clusters overlap |
| 5 | 0.35 | Very poor clustering, too many noisy divisions |

* Best silhouette score is for k = 2, but this comes at the cost of biological accuracy as it merges two distinct species into one cluster.
* k = 3 aligns better with domain knowledge (3 species), but the clustering quality is weaker than k = 2.



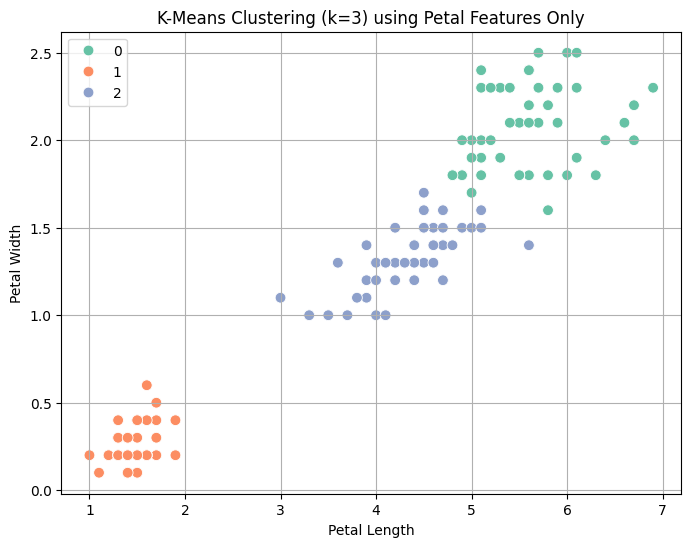




### Task 02

|  |  |  |
| --- | --- | --- |
| **Feature Set** | **Silhouette Score** | **Interpretation** |
| All 4 features (k=3) | 0.48 | Moderate clustering, some overlap |
| petal\_length, petal\_width only (k=3) | 0.67 | Strong separation, compact clusters |

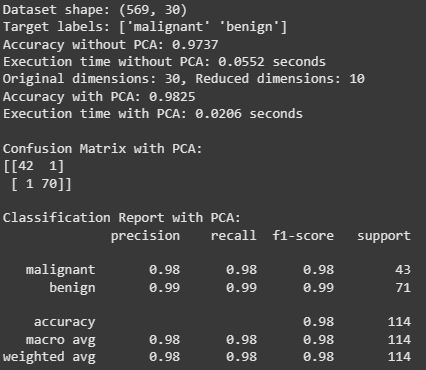
* Using only petal\_length and petal\_width leads to better clustering performance than using all four features.
* This confirms that these two features carry the most discriminative information and dimensionality reduction via feature selection can sometimes outperform using all available dimensions.



## PCA for Classification Lab Exercise

|  |  |  |  |
| --- | --- | --- | --- |
| **Setting** | **Accuracy** | **Execution Time** | **Notes** |
| Without PCA | 97.37% | 0.0552 sec | Full 30 features |
| With PCA (95%) | 98.25% | 0.0206 sec | Reduced to 10 components |

* Applying PCA to retain 95% of the variance reduced the feature set from 30 to 10 principal components. This not only improved execution time significantly (from 0.0552s to 0.0206s) but also increased classification accuracy (from 97.37% to 98.25%).
* The confusion matrix confirms excellent model performance with only two misclassifications.
* This suggests that PCA can enhance performance by eliminating noise and focusing on the most informative dimensions.



### Task 01

|  |  |  |  |
| --- | --- | --- | --- |
| **PCA Setting (Variance Retained)** | **Reduced Dimensions** | **Accuracy** | **Execution Time (seconds)** |
| Without PCA | 30 | 0.9737 | 0.0484 |
| PCA with 85% Variance Retained | 6 | 0.9825 | 0.0040 |
| PCA with 90% Variance Retained | 7 | 0.9825 | 0.0040 |
| PCA with 95% Variance Retained | 10 | 0.9825 | 0.0041 |
| PCA with 98% Variance Retained | 14 | 0.9912 | 0.0100 |

**Observations:**

1. Faster Execution Time with PCA.
2. Increased Time at 98% Variance Retained.

PCA with 98% variance provides the best performance (99.12% accuracy), but 85–95% variance retention achieves nearly the same accuracy (98.25%).

As expected, PCA significantly improves execution time, with the largest reductions seen in 85–95% variance retention settings. However, once you retain 98% variance, the execution time slightly increases due to the higher number of components.

