

# Phase 5 – Unsupervised Learning Report

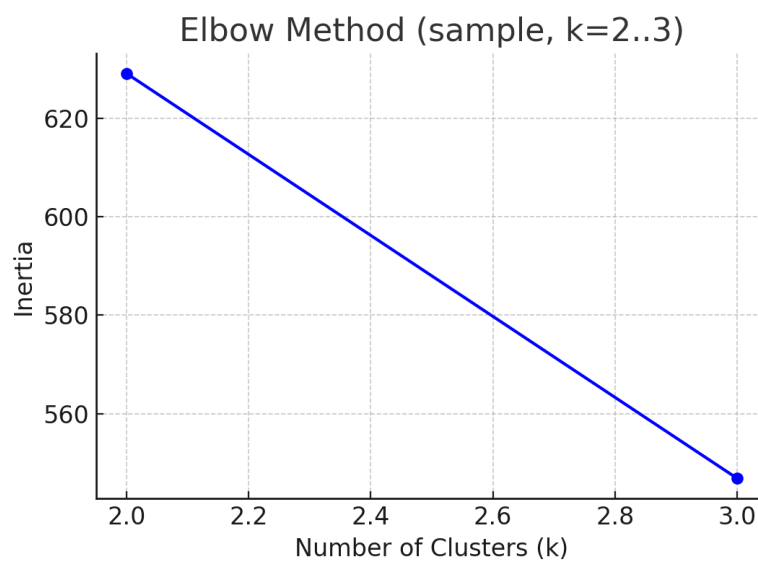
In this phase, I applied unsupervised learning methods to the heart disease dataset. The goal was to use clustering (K-Means) to group the data without labels and compare the clusters to the true target values.

## Data Used

Input file: selected\_features.csv (sample used for speed)

Output file: clusters.csv (with new cluster column).

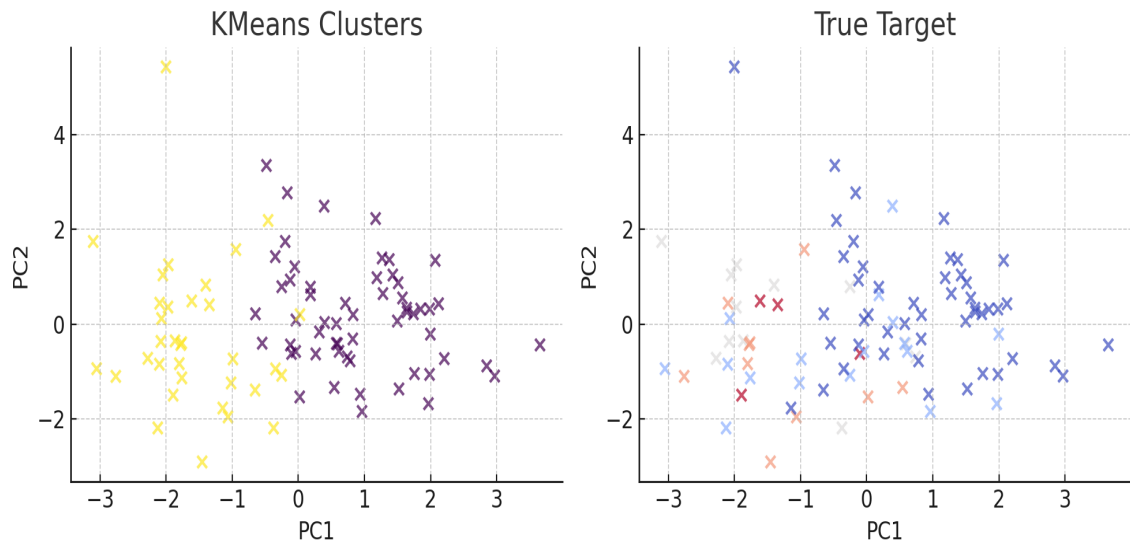
## Elbow Method



## Evaluation Metrics

Metric	Score
Silhouette Score	0.194
Adjusted Rand Index	0.272

## Scatter Plots (PCA)



## Conclusion

K-Means clustering with  $k=2$  was applied to the dataset (sample). The Silhouette Score was 0.194, showing the separation between clusters. The Adjusted Rand Index (ARI) was 0.272, which compares clusters to the true labels. This shows that unsupervised clustering gives some insights, but is less accurate than supervised learning. Overall, Random Forest (from Phase 4) still performs better for prediction.