Machine Learning (COMP7703)
Table Tennis Project Plan - Semester 1, 2025.

Volter Entoma
44782711

The 3D-PCA visualization of the data for different swing modes reveals that there is quite a lot of overlap between the different modes. This suggests that the swing modes are not easily separable in the feature space, and that a more complex model may be needed to accurately predict the swing mode from the data. For this project I propose the research question: *are there any more styles of swings beyond the three predefined modes?*

For reference, the list of swing modes:

- *test_mode* = 0 (Swing in the air)

- *test_mode* = 1 (Full power stroke)

- *test_mode* = 2 (Stable stroke)

The 3D-PCA visualization of the data with different colors for different swing modes.
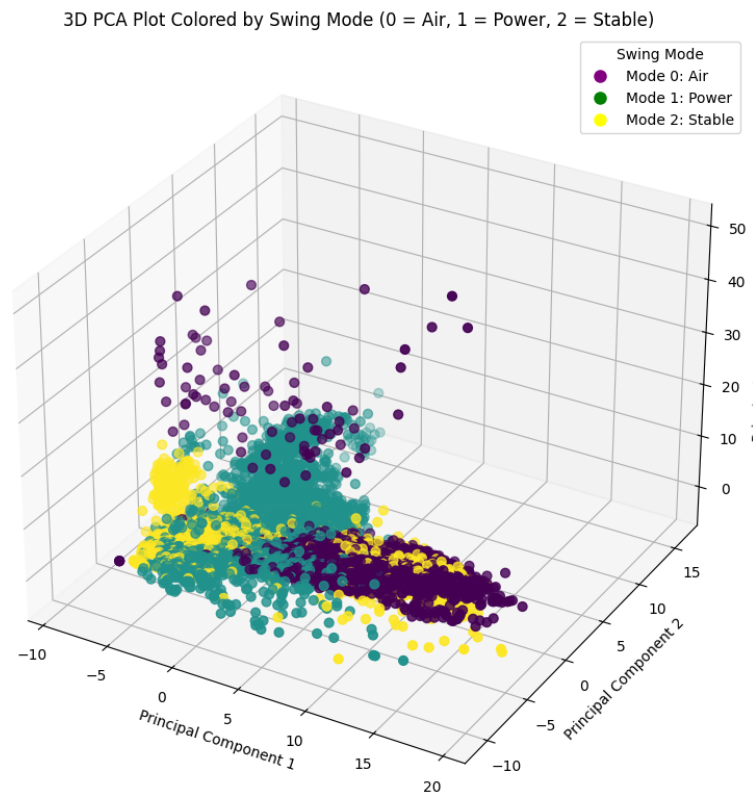

Figure 1: 3D-PCA visualization of the data with different colors for different swing modes.

I theorize that there are four more "intermediate" swing modes that are not explicitly defined in the dataset. These modes are, Air-Power, Air-Stable, Power-Stable, and Air-Power-Stable. These modes are

not explicitly defined in the dataset, but are inferred from strong overlap of some data points from swing mode to another.

I ran a quick pilot test using unsupervised learning using KMeans clustering to see if it could classify the overlap between the different swing modes. The results are shown in the figure below.
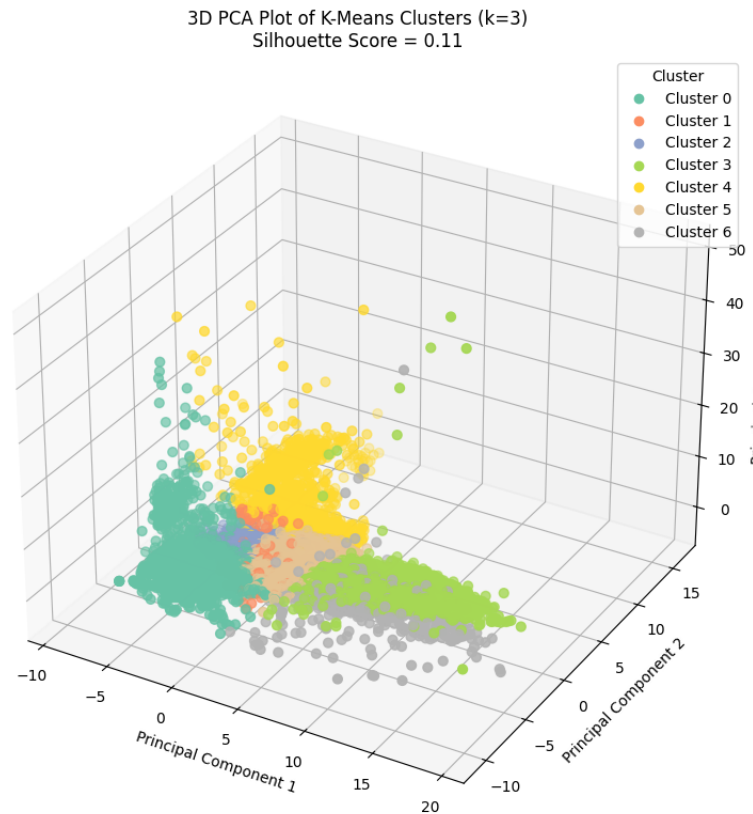

Figure 2: KMeans clustering of the data.

A quick look into the KMeans clustering results shows that the model was able to classify the data into 7 clusters. In particular, "Cluster 0" seems to correlate with the "Stable", "Cluster 6" with the "Air", and "Cluster 4" with the "Power" swing modes. The other clusters are not as clear, however they seem to correlate with the intermediate swing modes I proposed.

While the original labels identify three swing types, the unsupervised analysis using 3D PCA and KMeans ($k = 7$) reveals a more complex structure. Overlaps between modes suggest transitional forms, which align with the idea of intermediate swing styles. Although the silhouette score is quite low (0.11), indicating that the clusters are not well separated, a visual analysis of the clustering suggest that there are indeed more than three swing modes.

The question now is whether these clusters are latent structures or if they are artifacts of the clustering algorithm.

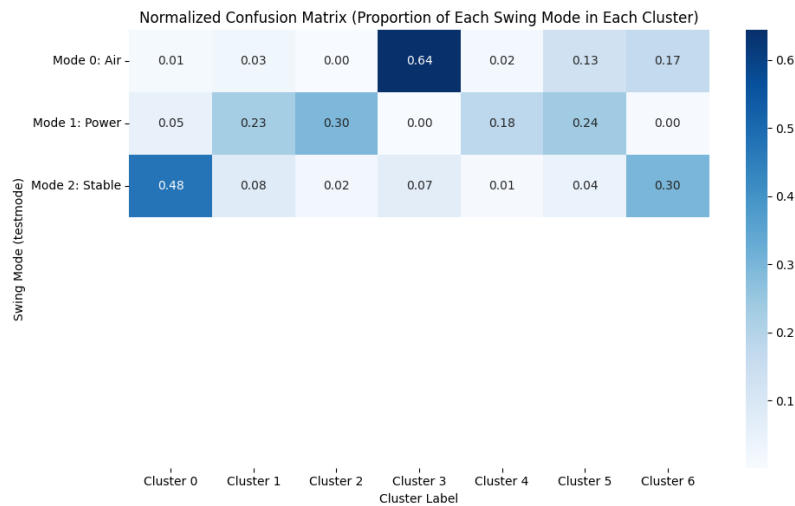The confusion-matrix below shows the results of the KMeans clustering.

Figure 3: Normalized confusion matrix of the KMeans clustering.

Things-to-do:

Cluster Stability Across Different Methods

- Use different k values to see if the clusters are stable across different k values.

- Use different clustering methods (e.g., DBSCAN, Agglomerative Clustering) to see if the clusters are stable across different methods.

- Use different dimensionality reduction methods (though I'm not sure if this is a good idea) to see if the clusters are stable across different methods.