**CSE 523 Machine Learning**

**Section 1**

**Group: Decision Makers**

**Project Number 6: Athlete profiling based on similar characteristics**

**Weekly Report**

**Week 4**

**Introduction:**

In todays’ time analysing data has occupied a central role in competing sports like basketball, football, cricket and more, which help in evaluating player at various levels, i.e., individual, team and conference level. By evaluating players, coaches can help players improve their performance of players by incorporating appropriate training methods like skill training or strength training based on their performance. Hence, our goal of the project is to cluster similar players and identify the characteristics of that cluster in order to improve their performance by giving appropriate training where the cluster lags.

**Progress Summary:**

We initiated the discussion on developing which kind of model for clustering athletes. The surveyed papers used the k-means algorithm for grouping of the clusters, which is a centroid based clustering approach. However, we our hypotheses is that the characteristics of a particular cluster is a gaussian. Hence, our approach to solve the above problem will be to use the Distribution based clustering (GMM). The idea is rather than assigning points to clusters, distribution-based approach allocates probability of that point being in different clusters. This gives subtle information about the structure of the data and the uncertainty in cluster assignments. It captures correlation and dependence between the data points. This approach is particularly suitable when the underlying distribution of the data is unknown. The optimal parameters of the distributions are found using machine learning techniques.

**Next Steps:**

We will initiate building model based on our hypotheses. Also, we will continue exploring various existing models for clustering.