AL2002 – Artificial Intelligence Lab
Assignment # 1

Note:
- Please ensure to submit both a PDF document and a Python file containing your code on the classroom.
- Copied assignment will be awarded **zero** marks.
- Submit your file in this format (roll-no-name) i.e (23P-1234-Ali.pdf).

Problem: 1 - 12-player IPL-2018 cricket team using K-mean

K-means Clustering:

K-means clustering is an unsupervised machine learning algorithm used to partition a dataset into 'K' distinct, non-overlapping clusters. The algorithm aims to group similar data points together while keeping them as far apart as possible. It iteratively assigns each data point to the nearest cluster centroid and then recalculates the centroids based on the mean of the data points assigned to each cluster. This process continues until the centroids no longer change significantly, or a specified number of iterations is reached.

Introduction: The **Indian Premier League (IPL) 2018** dataset provides comprehensive batting and bowling statistics of players from that season. In this task, you will delve into the dataset to analyze player performance using clustering techniques and subsequently construct a balanced **12-player cricket team**.

Task Description:

1. **Data Exploration and Preprocessing:**
   - Begin by exploring the IPL 2018 dataset, examining its structure and content.
   - Perform any necessary preprocessing steps, such as handling missing values or encoding categorical variables.

2. **Clustering Analysis:**
   - Apply the K-means clustering algorithm to categorize players based on their batting and bowling statistics.
   - Experiment with different values of 'K' and evaluate the clustering results to determine the optimal number of clusters.

3. **Formation of a Balanced Cricket Team:**
   - Using the insights gained from clustering, construct a balanced **12-player cricket team**.
   - Ensure that the team includes:
     - **Top-order batsmen** (openers),
     - **Middle-order batsmen** (batsmen who typically bat at positions 3 to 6),
     - **All-rounders** (players proficient in both batting and bowling),
     - **Bowlers** (specialists in bowling).

Submission Guidelines:
- Submit your analysis along with the code used for data exploration, preprocessing, clustering, and team formation.
- Justify your selection of players for the cricket team, highlighting their roles and contributions based on the IPL 2018 dataset.