Ubaid_ur_Rehman

Sir Saood Sarwar

Artificial Intelligence

15 April 2024

ASSIGNMENT # 01

The assignment is about making a 12 players team using the k-mean algorithm of top-order batsman, middle-order batsman as well as all-rounders and bowlers.

Data Preprocessing: The dataset is loaded from a CSV file named "total_data_na.csv" using Pandas. Null values are handled by replacing them with the mean values of their respective columns.

Feature Scaling: StandardScaler from scikit-learn is used to scale the selected features to have a mean of 0 and a standard deviation of 1. This step is crucial for clustering algorithms like KMeans, as they are distance-based and sensitive to the scale of features.

Clustering: KMeans clustering algorithm is applied to the scaled features with varying numbers of clusters from 1 to 9. The elbow method could be employed to determine the optimal number of clusters, which minimizes the within-cluster sum of squares.

Visualization: Scatter plot is created to visualize the clustering results based on the features "Batting Runs" and "Bowling Runs". Each cluster is represented by a different color, allowing for easy visual identification.

Team Formation: Players are selected for different roles based on their cluster memberships. Two top-order batsmen, four middle-order batsmen, three all-rounders, and three bowlers are chosen to form a balanced team.

Conclusion: The analysis successfully clusters the players based on their batting and bowling performances. The balanced team composition ensures a mix of batting strength, all-round abilities, and bowling proficiency.