



CSE523 MACHINE LEARNING

Weekly Report 7: 30 March 2024

WDL

Name	Enrollment No.
Paridhi Jain	AU2120226
Nevil Jobanputra	AU2140209
Saahil Doshi	AU2140106
Jash Parikh	AU2140108

Project 8: Athlete performance in collegiate basketball: Predicting match line-up
(Visualization on Dashboard)

Accomplished:

- Merged the data of season 2 and 3, by keeping only the common features from both the data set.
- Tried imputing through MICE using random-forest classifier.
- Imputed all required values and tried to predict RSI through it, which came out to be a poor result due to below given reasons:
 - Firstly, the data has a lot of missing values, like there are more than 10 features that have more than 80 percent null values and more than 500 rows that have less than 90 percent of missing values.
 - RSI might show certain pattern for different positions of playing, that we are yet to consider.

Next step:

- Now we are trying to figure out a way by which we might cluster the athletes into different buckets based on the position they play from; this might be done by taking the 'game score'.
- What we have planned is to remove missing game score and then taking the mean of the game score for every athlete, they we might be able to cluster the athletes by the game scores.
- Then we would impute values separately for the different clusters of athletes, predict RSI mod for them.
- Then it would be feasible to do time series on those clusters of athletes, and not on every athlete individually.