

Big Data Assignment 2 - Player Rank using PySpark

Member Name	USN	Section	Team Name
K Anjali Kamath	PES1201701674	B	BD_749_1674

Date of Evaluation:

Evaluator:

Assignment Submission ID: 138

1. How many iterations did the bowler ranking code run for?

The bowler ranking code ran for 18 iterations.

2. How many iterations did the batsman ranking code run for?

The batsman ranking code ran for 23 iterations.

3. Give the top 3 batsmen and their corresponding ranks according to your output :

RA Jadeja,1.555823474424

JH Kallis,1.503462829488

DJ Bravo,1.476785564149

4. Give the top 3 bowlers and their corresponding ranks according to your output :

DJ Bravo,1.514744880767

SR Watson,1.233336328721

RA Jadeja,1.177397821879

5. What do you think are the shortcomings of ranking players by using this approach?

If a player plays more often, he will have a higher contribution and will in turn have a higher rank.

6. What was your individual contribution to this assignment?

My teammate and I did the tasks together.

7. Any observations/challenges/comments about the assignment.

Getting Spark to work and running the code till convergence was a bit tricky. Also, the immutability of RDDs posed a challenge.