

Inside Out – A Hawk Eye’s View of Cricket

Rajesh Bellamkonda – rb3805 | Srinivas Bontula – ssb597

Description

- A batsman or a bowler in the game of cricket are measured or analyzed based on their averages (for batsman), strike rates (for bowlers).
- But the batsman’s average and bowler’s strike rate are always skewed as they do not take into account the following metrics -
 - Strength of the of the opposition
 - Conditions of the pitch
 - Series being played Home or Away
 - The timeline of the series (based on chronological order)
 - Batsman’s or Bowler’s performance relative to peers in a series.
- A series played on bowler-friendly pitches might deflate the career averages of the batsmen involved.
- Similarly, a series played on a batsmen-friendly pitches or conditions deflates the career averages and strike rate of the bowlers involved.

Approach and Methodology

- A series is a fair unit to use to break down careers.
- We will scrape through the websites to get a detailed report of each test series from 1870 – 2018 [Batsman and Bowlers numbers over a series] to prepare a huge dataset of all series.
- We modify the dataset to include the details of the Host country and the visiting team and the year of the series.
- The numbers of a bowler or batsman over a series are self-contained indicators of performance that are also somewhat normalized for conditions.
- Instead of looking at raw runs, wickets and averages over a series, we will try to look at the relative standing of a player (calculated as “Series Percentile Index”) in the run charts in a series, and average that over his whole career.
- Then we assign weights to the “Series Percentile Index” based on the above-mentioned metrics (Nature of the pitch, opposition, “Home” or “Away”, timeline).
- We plan to do a much more detailed analysis and design an algorithm which takes into account the above-mentioned metrics to assign a value (percentile) to a batsman and a bowler, which gives a complete performance index.

Course Scope

The concepts learnt in the course such as function decorators, function call overheads, itertools, cython and/or numba and other python inbuilt functions and packages will be applied for the following steps –

- a.) WebScraping
- b.) Modifying the datasets
- c.) Iterating the datasets for the players
- d.) Predicting the teams
- e.) Designing the algorithm based on Series Percentile Index.

Expected Results

- Ranking batsman and bowlers by our formula (Average SPV (using multiple metrics)).
- Predicting a team for a series from a rooster (based on the conditions and previous performances).
- All time XIs for each team.
- The batsmen or a bowler who have never been in the lower half of the run-scoring table of a series.

Schedule

- Scraping the web for the data of the series. [03/12/2018 – 04/01/2018]
- Obtain SPVs for the players. [04/02/2018 – 04/06/2018]
- Add weights to the SPVs based on playing conditions. [04/07/2018 – 04/15/2018]
- Conditions -> Nature of the pitch, Home or Away, Year etc.
- Computing the weighted averages for the above. [04/16/2018 – 04/22/2018]
- Generating the outputs based on different metrics. [04/23/2018 – 04/29/2018]
- Final Project Presentation and Report. [04/30/2018 – 05/02/2018]

References

- <http://www.espnricinfo.com/magazine/content/story/461783.html>
- <http://www.cricbuzz.com/cricket-news/99969/wanderers-pitch-india-vs-south-africa-the-rules-for-rating-a-pitch-dangerous-or-unfit>
- <http://www.espnricinfo.com/ci/content/story/981239.html>
- <http://www.espnricinfo.com/magazine/content/story/487445.html>
- http://www.espnricinfo.com/story/_/id/22414318/batsmen-analysed-series-dominance
- <http://www.espnricinfo.com/magazine/content/story/626913.html>
- <http://www.espnricinfo.com/magazine/content/story/626913.html>
- <http://www.howstat.com/cricket/home.asp>