

My Skill Outline:

Coding:

- R Studio
 - This will be used for my data cleaning, stat building, visualization, and RBO testing on my final resulting lists of NBA players.
- mobForest Package (Garge et al., 2013)
 - R package for random forests, what I plan to use for variable selection based on player performance.
- gespeR (Webber et al., 2010)
 - R package for performing RBO (ranked based overlap).
- Data cleaning packages
 - TidyR, dplyr, etc.

Data:

- Basketball reference player season data (Basketball-Reference.com, 9/21/2022)
 - Based on the stats I am using, I need seasonal data from 1978-present.
- IRB Certification
 - This allows me to conduct my subject based research
- Questionnaire
 - These are the questions I will be asking my subjects
- Experience Judgment (Desmarais et al., 1995)
 - Partial Order Knowledge Structure

Statistics:

- RBO (Joshi, 2021)

- Rank Based Overlap
- Gamescore (Page et al., 2013)
 - NBA statistic used to calculate a players in game performance
- PER (Page et al., 2013)
 - Player Efficiency rating, used to calculate efficiency performance
- Basic summary statistics
 - Standard deviation, confidence interval, etc.

Annotated Bibliography:

Basketball-Reference.com - Basketball Statistics and History.

<https://www.basketball-reference.com/>. (09/21/2022)

Notes: The website that houses the open source data I will be using.

Desmarais, M. C., Maluf, A., & Liu, J. (1995). User-expertise modeling with empirically derived probabilistic implication networks. *User modeling and user-adapted interaction*, 5(3), 283-315.

Notes: This paper describes what a Partial Order Knowledge Structure is. It also explains how the criteria that are defined for the POKS helps to judge subject experience.

Garge, N. R., Bobashev, G., & Eggleston, B. (2013). Random forest methodology for model-based recursive partitioning: the mobForest package for R. *BMC bioinformatics*, 14(1), 1-8.

Notes: This paper details how to use the mobForest package in R to create random forest machine learning models. I'm using this paper as a refresher for my random forest code.

Joshi, P. (2021, January 12). *RBO V/s Kendall Tau to compare ranked lists of items*. Medium.

Retrieved September 29, 2022, from

<https://towardsdatascience.com/rbo-v-s-kendall-tau-to-compare-ranked-lists-of-items-8776c5182899>

Notes: This Towards Data Science article explains two rank comparison methods. It details why RBO is slightly easier to use due to Kendall's Tau's limitations.

Page, G. L., Barney, B. J., & McGuire, A. T. (2013). Effect of position, usage rate, and per game minutes played on nba player production curves. *Journal of Quantitative Analysis in Sports*, 9(4), 337-345.

Notes: This paper uses gamescore instead of PER to help predict NBA player production curves. It explains the difference between the two stats and why they chose gamescore.

Webber. (2010). *RBO: Rank biased overlap (Webber et al., 2010)*. RDocumentation. Retrieved September 29, 2022, from <https://www.rdocumentation.org/packages/gespeR/versions/1.4.2/topics/rbo>

Notes: This R documentation shows how to use the R package gespeR's RBO function.