

Problem Statement Worksheet (Hypothesis Formation)

<Use keyword analysis to predict this years upcoming draft position that is more accurate in predicting a players draft position than using player score alone. Using the key word analysis the R^2 should show a lower variance when applying the predictions to new data >

1 Context

This problem is important to consider because finding novel ways to accurately predict when players will be drafted allows for teams to make more definite plans on who will be available to be drafted when their number is called and how far they may need to move in the draft to select a player that they covet. Also, knowing what traits are valued the most for positions would give prospects and idea of where their focus should be when preparing for the draft.

2 Criteria for success

If a modified score that uses key word analysis can provide less variance to R^2 when compared to using the players score alone.

3 Scope of solution space

<What is the focus of this business initiative? I.e. What are you specific items will you focus on exclusively?>

4 Constraints within solution space

Not all players have text data for their draft profiles. Coveted players tend to have more written about them than players in the later rounds. Also the score given to a player is usually a pretty good indicator of when the player will be drafted so it will be hard to improve upon that.

5 Stakeholders to provide key insight

I think both team scouts, general managers and coaches would have interest in having a more accurate prediction of draft outcomes. I think players would also like to know what descriptors are coveted at the professional level in order to position themselves to get drafted higher and make more money.

6 Key data sources

All of the key data is public and I will be using a kaggle data set. A data scrape could be done of player profile text if it is deemed necessary