Kevin Angotti

DSC 680

Project 3 Proposal

**What domain?**

My project data comes from the Kaggle site; I usually turn to Kaggle for most of my data science projects. The site usually has vast amounts of datasets, and the range of the information is extremely extensive. This particular dataset was uploaded by Savvas T., looking for data analysts to make predictions on NFL combine statistics. The dataset contains the combine results from the year 2000 to 2018.

Please list 10 references to use to make sense of what you're doing with these data.

1. Savvas, T. (2018). Kaggle. Retrieved from <https://www.kaggle.com/savvastj/nfl-combine-data>

2. Kelly, D. (2018). How To View Grit. The Ringer. Retrieved from <https://www.theringer.com/2018/4/12/17227604/nfl-draft-intangibles-scouting-evaluation>

3. Woo, M. (2019). How Much Does The Combine Reveal About Future NFL Player? Inside Science. Retrieved from <https://www.insidescience.org/news/how-much-does-combine-reveal-about-future-nfl-players>

4. Dunlap, A. (2013). Things NFL Scouts Look For At The Combine That Fans Don't. B/R. Retrieved from <https://bleacherreport.com/articles/1535607-things-nfl-scouts-look-for-at-the-combine-that-fans-dont>

5. NFL Draft Combine Testing. Retrieved from <https://www.topendsports.com/sport/gridiron/nfl-draft.htm>

6. Witt, T. (2020). Is it possible to predict the success of NFL Draft picks? FanNatin. Retrieved from <https://www.si.com/nfl/chiefs/gm-report/predicting-nfl-draft-pick-success>

7. King, J. (2020). Using Machine Learning to Predict Fantasy Football Points. Towards Data Science. Retrieved from <https://towardsdatascience.com/using-machine-learning-to-predict-fantasy-football-points-72f77cb0678a>

8. Fridson, M. (2017). NFL Draft Analysis: 30 years of Player Outcomes. NYC Data Science Academy. Retrieved from <https://nycdatascience.com/blog/student-works/nfl-draft-30-years-outcome-analysis/>

9. Moore, K. (2017). Predicting NFL Success with Algorithms. Cognitive. Retrieved from <https://www.cognitivetimes.com/2017/08/predicting-nfl-success-with-algorithms>

10. Robbins, L. (2016). A Data Scientist Dissects the 2016 NFL Draft. WSJ. Retrieved from <https://www.wsj.com/articles/a-data-scientist-dissects-the-2016-nfl-draft-1461793878>

11. Bronshtein, A. (2017). Simple and Multiple Linear Regression in Python. Towards Data Science. Medium. Retrieved from <https://towardsdatascience.com/simple-and-multiple-linear-regression-in-python-c928425168f9>

12. Scouting Combine. (2021). NFL Scouting Combine. NFL Network. Retrieved from <https://www.nfl.com/network/events/nfl-combine>

**What's the data?**

I will examine the National Football League (NFL), scouting combine data that spans from 2000 to 2018. There is 1 column and over 6200 rows. The intention is to use EDA and predictive analytics to see if it is possible to predict the draft position would improve with a data science approach to the variables given.

**Research questions?**

Can a player's AV (actual value) change using predictive analytics?

Which variables contribute most to a prediction change?

Which player's actual value changes the most, a drafted player or undrafted?

**Methods?**

I will start this project with EDA (exploratory data analysis) followed by a simple linear regression for correlations each variable, then dive into multiple regression to predict a player's actual value for comparing their current AV. The goal of my project is to see how prediction can change the outcome of a draft. This type of analysis could be useful for future combines if a player's measurables can e used to predict their value and draft position.

**Anticipated challenges?**

I foresee the main challenges: the dataset only contains 16 variables and less than7000 rows of data. I hope that is enough data for the questions I am asking and the modes I plan to run. The potential for prediction here, I feel, is very positive. The author of the dataset does not provide any thoughts of the type of analysis that this data would be intended for.

For modeling, the biggest challenge will be if multiple regression is the best avenue; I may need to add in other modeling lie decision trees or NN, K-Modes, etc. I will need to get to the modeling stage a bit sooner than previous projects; my time is shorter than this third project and needs time to adjust if that becomes a necessity.

**Concluding remarks.**

I go into this project with an open eye and also with a bit of excitement. I have become more confident with my data analysis skills since this semester has begun and hope time in more storytelling tools and ways tying the data together. I want to use more Power BI for this project if possible. I think an interactive approach to telling the story will be the best way to present the result. If the multiple regression can predict the outcomes as I think it can, an interactive way of presenting it could be very interesting and fun to do.

One of my goals of this project is to determine if this type/way of analyzing NFL scouting data can be useful for future applications. a Power BI interactive tool could be used as a model for television networks or NFL teams to use for college player analysis.