

Machine Learning and Its Application in Fantasy Football

Jonathan R. DiQuattro

Thomas Edison State University

Abstract

This paper is a draft chapter 1 of a research paper into the application of using machine learning to determine the top running backs for each season of the last decade with a 60% accuracy.

Keywords: Machine learning, fantasy football, algorithm.

Chapter 1: Introduction

This is a study that explains if it is possible to predict season leading fantasy football running backs through the use of machine learning. Due to the nature of football being a sport with many variables, the primary goal is to use machine learning to predict a top three running back with a 60% accuracy. This chapter will explore the problem with more detail, explain the significance of the study, describe the methodology for determining the results, delve into the delimitations and limitations, and define the terms used throughout the study.

Background

This study attempts to display how machine learning works, by using it to predict the top fantasy football running backs for the last decade of seasons, with over 50 years of data in the National Football League (NFL) it is a primary candidate for displaying how adept machine learning really is.

By the year 2020 the IDC predicts that there will be 5,200 GB of data for every person on earth (Mearian, L.), and by graphing the amount of data produced every year, a logarithmic function emerges, this means that next year we will create more data than we have in the entire history of man and womankind, and there is no end in sight for the next few decades. The only possible way to process this staggering amount of data is through the use machine learning, which is capable of taking many variables and turning them into actual information that can be used in the real world.

Since there are many variables in the NFL, the expectations of this study are that one of the top three running backs in each season can be predicted 60% of the time by using machine learning.

Since there are many variables in the NFL, the expectations of this study are that one of the top three running backs in each season can be predicted 60% of the time by using machine learning.

Problem Statement

To further the dissemination of machine learning concepts and applications, this study asked the following question:

Major Question: Is machine learning is capable of predicting a top three running back in each season for the last decade with a 60% accuracy.

Sub Questions:

1. Can top NFL fantasy players be accurately determined for each season based on previous seasons statistics combined with the application of artificial intelligence?
2. What data will provide an accuracy of over 60%?
3. What Machine learning model/architecture would be most appropriate for determining top running backs?
4. Are there any interesting patterns that emerge from conducting this study?

Professional significance

The amount of machine learning that experts are predicting for the future could quite literally be world altering (Sharma, B.) there is great promise in the fields of automation, self driving cars, augmented reality, and a great number of other things that haven't even been

considered yet, and computer scientists are just barely scratching the surface when it comes to the actual applications of machine learning. It is of crucial importance that researchers find new ways to utilize machine learning to bring awareness to the masses so that future applications have the funding and required workforce to support it, by recreating this study it shows that machine learning can be applied to more than finances and consumer goods, but that it also extends to everyday life to include seemingly mundane practices of fantasy football, and points out the near limitless potential of machine learning, but without a raise in public awareness, it could really prohibit the use of machine learning in the future due to a growing need of experts in the field (Otokiti, E.).

Methodology

To complete this study it took an analysis of machine learning itself to determine the best approach to creating an algorithm that can predict top running backs and a recursive neural network (Gupta, D.) was determined to be the best options for applying machine learning to fantasy football. A strong working knowledge of Yahoo fantasy football scoring and strategy was also required to answer both sub question 1 and sub question 2. The data selected to “train” the model was a combination of teams that consistently lead the NFL in rushing, teams with high winning percentages, and teams that consistently lead the NFL in rushing attempts per game. The following general steps were followed to complete the study:

Step 1 Acquire required data and create a relational database

Step 2 Format data for use in neural network.

Step 2 Code a recursive neural network using python 3.6

Step 3 Compute results

Step 4 Determine conclusion

Delimitations

The following samples were used to “train” the machine:

- Rushing data to include yards per game, winning percentage, and rushing attempts for every team in the NFL for the last 20 years
- For instances where a statistic is not available (not every team was franchised in 1988) A mean average of every other team will not be used, instead a NaN (Also known as a void) will be used in place of an unknown statistic.
- This study will focus on the last decade of top running backs in the NFL.
- This study used Yahoo.com default fantasy football rules located at <https://help.yahoo.com/kb/SLN6489.html>

Limitations

- There are a lot of variables that could affect determining leading rushers, choosing the best ones to predict a top rusher could be very challenging.
- The game of football has changed a lot over the years, and it could potentially be very difficult to account for those changes in the data.

Definition of Terms

The following are relevant definitions and terms for the field of machine learning, and the application on fantasy football:

- Null: Used in place of a zero and represents the non existence of a number (Michalski, K.)
- Fantasy Football: Is an online game where users draft real world players, who get points based off of how they perform during real games.
- Machine Learning: An algorithm that allows for a computer to write their own algorithms to accomplish specific tasks.
- Sample Set: Defines what data will be used to “train” the computer.
- Training: The term used to teach a computer how to write algorithms.
- Running Back: A position in fantasy football that provides the majority of points for a team.
- NFL: National Football League
- Yahoo Fantasy Football: A very popular fantasy football website with a vast number of users.

Summary

This purpose of this study is to predict a top three performing running back for the last decade of NFL seasons using default yahoo fantasy football scoring rules to further the advancement and understanding of machine learning, a data processing technique that is of ever growing importance due to the vast amount of data created each year.

References

Mearian, L. By 2020, there will be 5,200 GB of data for every person on Earth.

<https://www.computerworld.com/article/2493701/data-center/by-2020--there-will-be-5-200-gb-of-data-for-every-person-on-earth.html>. December 11, 2012. Web. Retrieved July 11.

Sharma, B. 'Machine learning is the big one': Deloitte expert looks to the future of tech trends.

<https://www.cnbc.com/2018/04/10/machine-learning-is-the-big-one-deloitte-expert.html>. April 10, 2018. Web. Retrieved July 4, 2018.

Otokiti, E. Experts Claim Machine Learning Talent Shortage Not A Problem.

<https://channels.theinnovationenterprise.com/articles/an-optimistic-look-at-the-machine-learning-talent-shortage>. May 1, 2018. Web. Retrieved July 7, 2018.

Gupta, D. Fundamentals of Deep Learning – Introduction to Recurrent Neural

Networks. <https://www.analyticsvidhya.com/blog/2017/12/introduction-to-recurrent-neural-networks/>. December 7, 2017. Web. Retrieved July 8, 2018.

Michalski, K. Understanding null, undefined and NaN. [https://codeburst.io/understanding](https://codeburst.io/understanding-null-undefined-and-nan-b603cb74b44c)

[-null-undefined-and-nan-b603cb74b44c](https://codeburst.io/understanding-null-undefined-and-nan-b603cb74b44c). March 28, 2018. Web. Retrieved July 8, 2018.