

# Sport analytic regression and the nfl



**ASSIGN  
BUSTER**

## **Sport Analytics**

*Using Open Source Logistic Regression Software to Classify Upcoming Play Type in the NFL.*

Journal Article Review/Module 5

Baker, R. E., & Kwartler, T. (2015). Sport analytics: Using open source logistic regression software to classify upcoming play type in the NFL. *Journal of Applied Sport Management*, 7(2).

### Introduction

The purpose of this research paper was to study the benefits of incorporating data analytics, logistical regression analysis using open source data for play selections of two NFL teams, the Cleveland Browns and the Pittsburgh Steelers between the 2000 and the 2012 seasons. Both teams are in the American Football Conference (AFC) and are only 135 miles from each other. As cities, both Cleveland and Pittsburgh harness a “blue collar” culture and they are both roughly the same market size. However, the on the field success tell a different story. The Cleveland Browns organization has suffered through many losing seasons winning only 71 games from 2000 to 2012 while the Pittsburgh Steelers have won 135 games during this time frame to include Super Bowl victories in 2006 and 2009. By comparing a successful organization regarding victories (Pittsburgh Steelers) versus a less successful organization (Cleveland Browns), logistical regression may be incorporated to assist NFL coordinators assessing the probability of the opposing team installing a run or pass play. The final analysis showed that

the Cleveland Browns offensive play selection was correct 66.9% of the time as opposed to 66.9% for Pittsburgh. The authors concluded that using open source software and logistic regression for the NFL play selection would be beneficial for game time decision making purposes.

### Methodology and Statistical Techniques

Sports, in general, are *big* business garnering approximately \$440 to \$470 billion of the North American economy placing this in the top ten (Fry & Ohlmann, 2012; Plunkett Research, 2014). Therefore any chance of gathering relevant information that can give a team the upper hand would be critical for that organization. Sports analytics is a component that has grown in stature as illustrated in the popular book and movie *Moneyball* (Lewis, 2004) explaining how Oakland A's manager Billy Bean incorporated *Sabermetrics* to determine who to draft, trade and managing the game by using this mathematical and statistical analysis. The Oakland A's (American League) needed to go in this direction since they are typically considered a small market team generating less revenue when compared to their east coast rival's and American League nemesis New York Yankees and Boston Red Sox organizations. The bottom line is that the Oakland A's have less room for error and need to get the most mileage out of their players with a smaller payroll to compete throughout the season and for a possible playoff spot.

According to the authors the general model of statistical analysis in the decision making tree offering feedback is as follows:

- Situation-specific play parameters informs algorithm

- Data collection and statistical analysis
- Application of analysis through decision making
- Resultant performance outcomes of success or failure

The authors provide three examples of data platforms otherwise known as *Big Data, Data Mining, and Spot Analytics*. Big Data captures trillions of bytes of valuable information which is beneficial to the organization (or individual) allowing them to make efficient and effective decisions resulting in a competitive advantage (Bryant, Katz, & Lazowska, 2008). The organization would need to customize their information to capture and analyze as much data as possible; however the downside to Big Data could be the invasion of privacy. Having an understanding of the practical uses of Big Data can have an advantage for understanding trends, communities and individuals (Boyd & Crawford, 2012).

Data Mining's objective is to discover valuable information in large datasets (Hand, 2007). Data mining may allow the organization to have a better understanding of customer relationships which in turn could develop into a strong customer loyalty base as well as assist the coaching staff. By and large, data mining can pay off with positive results for any sports organization whether it is marketing, customer support and the recruitment/drafting of players (Berry & Linoff, 1997). The bottom line is that organizations need to figure out what they are looking for and use the information to their advantage (Berry & Linoff, 1997).

Gathering data may be one thing but interpreting the information into practical use is quite another. There is an open source interactive software called *R*. Developed by the R Foundation, (R, 2014) R is free software which  
<https://assignbuster.com/sport-analytic-regression-and-the-nfl/>

is an open-sourced computing and graphics platform that influences data calculation and graphical display (R Foundation, 2014) which can be used on platforms such as MacIntosh, UNIX, and Windows.

The authors used data from the “ CORE” data file ([www. armchairanalysis. com](http://www.armchairanalysis.com)) which contain over 500, 000 individual records with each record matching up to a single NFL play. Only plays such as *pass* and *run* (offense) were kept for analysis creating a two-fold relationship. Cleveland recorded 12, 187 play records and Pittsburgh recorded 14, 123 respectively. After the coding process was completed, the open source R software was implemented for logistical regression analysis. The authors state that the logistical regression calculates the logistical odds (log odds) of an event occurring. Once the log odds are calculated, the practical outcomes (probability) are known. The probability is the percentage likelihood that an event (perhaps a play) of occurring or not occurring with the range of probability from 0% to 100%. The authors plot the probability and log odds on a scatter plot and by using R, the coefficient table was created to show the Cleveland and Pittsburgh logistic regressions. The data set showed that as the quarter in the game progressed, Cleveland had a higher probability of passing whereas Pittsburgh decreased their passing attempts. The authors state that the model assessed 8, 090 play types correctly for the Cleveland Browns and 4, 097 incorrectly whereas the Pittsburgh Steelers model was 66. 9% accurate classifying 9, 442 correctly and 4, 681 incorrectly.

Conclusion

The authors concluded that there is a place for data analytics for NFL teams and other teams professional or not. In fact, the authors note that given the importance of data analysis for teams, upwards to 190,000 expert analysts and an additional 1.5 million managers to include sport managers will need to be briefed on how to interpret and implement information given the analysis, whether it is the coach (manager) general manager, front office staff and players alike.

My impression is that the amount of data analytics is here to stay and is changing the *culture* of organizations. For example, my employer Pfizer, Inc. has changed the way we look at our business model by implementing new software platforms even down to the district managers and sales representatives. On the one hand there is the propensity of *paralysis by analysis*. Human insight based off of experience, knowledge, and logic has its place and should never be placed aside because of a new software package. On the other hand, if we learn to speak the data language and understand how it can steer us in the right direction, the organization and the customer can benefit greatly. A combination of both is probably the best result for future planning and success.

## References

- Berry, M. J., & Linoff, G. (1997). *Data mining techniques: For marketing, sales, and customer support*. New York, NY: John Wiley & Sons, Inc.
- Boyd, D., & Crawford, K. (2012). Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication, and Society*, 15(5), 662–679.

- Bryant, R. E., Katz, R. H., & Lazowska, E. D. (2008). *Big Data Computing: Creating revolutionary breakthroughs in commerce, science, and society. Computing Research Initiatives for the 21st Century, Computing Research Association* . Retrieved from [http://www.cra.org/ccc/files/docs/init/Big\\_Data.pdf](http://www.cra.org/ccc/files/docs/init/Big_Data.pdf)
- Fry, M. J., & Ohlmann, J. W. (2012). Introduction to the special issue on analytics in sports, part I: General Sports Applications. *Interfaces*, 42(2), 105-108.
- Hand, D. J. (2007). Principles of data mining. *Drug Safety* , 30(7), 621-6.
- Lewis, M. (2004). *Moneyball: The art of winning an unfair game* . New York, NY: W. W. Norton & Company.
- Plunkett Research, Ltd. (2014). *Sports industry market research* . Retrieved from <http://www.plunkettresearch.com/sports-recreation-leisure-market-research/industry-and-business-data>
- R Foundation. (2014). *The R project for statistical computing* . Retrieved from <http://www.r-project.org/>