

## Final project: NFL by half

### ADVANCED METHODS III 140.753

## 1 Introduction

Have you ever wondered if you should keep watching a football game beyond half-time? You are probably having a busy weekend and need to do other things. Specially if the team that you like has a high probability of winning (or losing if they are doing terrible).

Analyzing NFL data and variables has been done before. For example, Carney and Fenn[10] were interested in identifying variables that affect the Nielsen TV rating of the games. Some of the variables they looked at include the game day winning percentage of the two teams involved. In their analysis, one important variable was whether the local team had gone to the playoffs the previous year which makes sense for explaining the TV rating.

Brian Burke who describes himself as *a former Navy pilot who has taken up the less dangerous hobby of N.F.L. statistical analysis, operates Advanced NFL Stats, a blog about football, math and human behavior*[7] has published many blog posts where he analyzes NFL data. He has devised *narrative statistics* such as a *winning probability model* —it's quite accurate[9]— which he uses to determine the effect specific plays have on the game[8]. Another one that has similar uses is the *expected points*[2] and *expected points added*. By using the *win probability model* he can estimate how exciting the game was[1].

Of special interest, Brian Burke has used logistic regression in a rather exquisite way to predict the probability of winning for each team during the NFL season[3, 4, 7, 5].

The goal of this small project is to use the play-by-play data for the first half of NFL games to predict which team will win. To do so a modification of Brian Burke's game probability model will be implemented. In addition, the resulting prediction model will be deployed on the web so it can be used for the 2013 season.

## 2 Pre-processing

## 3 Exploratory Data Analysis

### 3.1 Training model

## 4 Results

## 5 Deployment on the web

Shiny![11]

```
## This is how you can run the Shiny app
library(shiny)
runUrl("https://github.com/lcolladotor/lcollado753/archive/master.zip",
      subdir = "final/nfl_half/shiny/")
```

## 6 Conclusions

hola [6]

## 7 Reproducibility

The code, data, and report is available at GitHub. Specifically here: [https://github.com/lcolladotor/lcollado753/tree/master/final/nfl\\_half](https://github.com/lcolladotor/lcollado753/tree/master/final/nfl_half). The README file explains the order of the scripts.

This report was generated using the following R packages.

- R version 2.15.2 (2012-10-26), x86\_64-apple-darwin9.8.0
- Locale: en\_US.UTF-8/en\_US.UTF-8/en\_US.UTF-8/C/en\_US.UTF-8/en\_US.UTF-8
- Base packages: base, datasets, graphics, grDevices, methods, stats, utils
- Other packages: knitr 1.1
- Loaded via a namespace (and not attached): digest 0.6.3, evaluate 0.4.3, formatR 0.7, stringr 0.6.2, tools 2.15.2

## References

- [1] B. Burke. *Best Games of 2012 and Best Playoff Games*. URL: <http://www.advancednflstats.com/2013/01/best-games-of-2012-and-best-playoff.html> (visited on 03/17/2013).
- [2] B. Burke. *Expected Points (EP) and Expected Points Added (EPA) Explained*. URL: <http://www.advancednflstats.com/2010/01/expected-points-ep-and-expected-points.html> (visited on 03/17/2013).
- [3] B. Burke. *How the Model Works—A Detailed Example Part 1*. URL: <http://www.advancednflstats.com/2009/01/how-model-works-detailed-example.html> (visited on 03/17/2013).
- [4] B. Burke. *How the Model Works—A Detailed Example Part 2*. URL: <http://www.advancednflstats.com/2009/01/how-model-works-detailed-example-part-2.html> (visited on 03/17/2013).
- [5] B. Burke. *N.F.L. Week 4: Game Probabilities Are Back*. URL: <http://fifthdown.blogs.nytimes.com/2012/09/27/n-f-l-week-4-game-probabilities-are-back/> (visited on 03/17/2013).
- [6] B. Burke. *Play-by-Play Data*. URL: <http://www.advancednflstats.com/2010/04/play-by-play-data.html> (visited on 03/17/2013).
- [7] B. Burke. *Week 4 Game Probabilities, From Advanced N.F.L. Stats*. URL: <http://fifthdown.blogs.nytimes.com/2009/09/30/advanced-nfl-stats-week-4-game-probabilities/> (visited on 03/17/2013).
- [8] B. Burke. *Win Probability Added (WPA) Explained*. URL: <http://www.advancednflstats.com/2010/01/win-probability-added-wpa-explained.html> (visited on 03/17/2013).
- [9] B. Burke. *Win Probability Model Accuracy*. URL: <http://www.advancednflstats.com/2009/07/win-probability-model-accuracy.html> (visited on 03/17/2013).
- [10] S. Carney and A. Fenn. *The Determinants of NFL Viewership: Evidence from Nielsen Ratings*. SSRN Scholarly Paper ID 611721. Rochester, NY: Social Science Research Network, Nov. 2004. URL: <http://papers.ssrn.com/abstract=611721> (visited on 03/18/2013).
- [11] RStudio and Inc. *shiny: Web Application Framework for R*. R package version 0.4.0. 2013. URL: <http://CRAN.R-project.org/package=shiny>.