

# Predicting NFL Game Results

Dan Matthews - GADAT3

# Problem Statement

## **Null Hypothesis:**

It is impossible to predict the outcome of an NFL game.

## **Alternative Hypothesis:**

Using game-day statistics, it is possible to predict if a team won a game.

# Why this project?

- Dan really likes sports.
- The 2014 Washington Nationals had a ~80% chance of winning a game when they scored four or more runs.
- Curious if certain aspects of a game are more likely to result in a win (IE: first downs lead to wins).

# DataSet

Data was obtained from [pro-football-reference.com](http://pro-football-reference.com).

Downloaded into CSV by year then aggregated into a single CSV.

The screenshot shows the Pro-Football-Reference.com website. The header includes navigation links for product hunt, reddit, /r/nationals, /r/caps, xkcd, gmail, github, unsplash, 8tracks, hypem, bleach, and more. The main navigation bar lists Sports-Reference: Baseball, Basketball (college), Football (college), Hockey, and Olympics. The Pro-Football-Reference.COM logo is prominently displayed. Below the logo is a navigation bar with links for play index, players, teams, seasons, coaches, leaders, super bowl, draft, boxscores, and more. A green 'Mobile Site' button is visible. The breadcrumb trail reads: You Are Here > Pro-Football-Reference.com > Pro Football Statistics and H. A news banner at the top right says: News: s-r blog:2015 First Round Picks Up.

**Pro-Football-Reference.com News**

- s-r blog: 2015 First Round Picks Up
- s-r blog: Incompletions Added to Player Game/Season Finders
- s-r blog: First Downs Added to Team Game Finder
- s-r blog: Sports Reference is Hiring a Web Developer
- s-r blog: Progressive Leaderboards Added
- s-r blog: More Game Play Finder Additions

**Career Rushing Yards Leaders**

Rank	Player	Yards
1.	<a href="#">Emmitt Smith</a>	18,355
2.	<a href="#">Walter Payton</a>	16,726
3.	<a href="#">Barry Sanders</a>	15,269
4.	<a href="#">Curtis Martin</a>	14,101
5.	<a href="#">LaDainian Tomlinson</a>	13,684
6.	<a href="#">Jerome Bettis</a>	13,662
7.	<a href="#">Eric Dickerson</a>	13,259
8.	<a href="#">Tony Dorsett</a>	12,739
9.	<a href="#">Jim Brown</a>	12,312
10.	<a href="#">Marshall Faulk</a>	12,279

[View full list »](#)

[Previous](#) [Next](#)

**2014 Standings**  
[Full Standings & Stats](#) / [Games](#) / [Leaders](#) / [Milestones](#)  
[Players: Passing](#) / [Rushing](#) / [Receiving](#) / [Fantasy](#)

**Week 16 Standings**  
Current playoff seed in parentheses.

AFC	W	L	T	SRS
East				
<a href="#">NYE</a> (1)	11	3	0	13.2
<a href="#">BUF</a>	8	6	0	4.7
<a href="#">MIA</a>	7	7	0	4.7
<a href="#">NYJ</a>	3	11	0	-7.4
North				
<a href="#">CIN</a> (4)	9	3	1	0.1
<a href="#">BAL</a> (6)	9	5	0	5.3
<a href="#">PIT</a> (5)	9	5	0	0.9
<a href="#">CLE</a>	7	7	0	-4.2
South				
<a href="#">IND</a> (3)	10	4	0	6.1
<a href="#">HOU</a>	7	7	0	1.1
<a href="#">JAX</a>	3	12	0	-11.3
<a href="#">TEN</a>	2	13	0	-12.3
West				
<a href="#">DEN</a> (2)	11	3	0	10.2
<a href="#">SDG</a>	8	6	0	2.3
<a href="#">KAN</a>	8	6	0	6.1
<a href="#">OAK</a>	2	12	0	-8.9

NFC	W	L	T	SRS
East				
<a href="#">DAL</a> (3)	10	4	0	1.6
<a href="#">PHI</a>	9	5	0	3.5
<a href="#">NYG</a>	5	9	0	-3.0
<a href="#">WAS</a>	3	11	0	-9.1
North				
<a href="#">DET</a> (2)	10	4	0	1.9
<a href="#">GNB</a> (6)	10	4	0	7.7
<a href="#">MIN</a>	6	8	0	-2.1
<a href="#">CHI</a>	5	9	0	-7.5
South				
<a href="#">NOR</a> (4)	6	8	0	-2.4
<a href="#">CAR</a>	5	7	1	-6.2
<a href="#">ATL</a>	5	9	0	-3.4
<a href="#">TAM</a>	2	12	0	-10.2
West				
<a href="#">ARI</a> (1)	11	3	0	3.6
<a href="#">SEA</a> (5)	10	4	0	7.3
<a href="#">SFO</a>	7	7	0	-1.4
<a href="#">STL</a>	6	8	0	-0.1

**Current NFL playoff standings & matchups**  
**Current 2015 first round draft order**

**Week 16 Games**

**Thursday**  
[Tennessee Titans](#) 13  
[Jacksonville Jaguars](#) 21 [Boxscore](#)

**Saturday**  
[Philadelphia Eagles](#) (9-5) 4:30PM ET  
[Washington Redskins](#) (3-11) [Preview »](#)

**Play Index** answer your stats questions  
**Finders:** [Season](#), [Game](#), [TD](#), [Intro Blog Posts](#)

**Players** stats and more  
[Awards](#), [Manning](#), [Rice](#), [Bradshaw](#), [Game Logs](#), [Big Games](#), [Touchdowns](#), [Quarterbacks](#), ...

**Teams** rosters, stats, and more

# Raw Game Data

- Win
- Day
- Date
- [link to boxscore]
- Overtime (Y/N)
- Opponent
- Points Scored
- Points Allowed
- First downs gained
- Total yards
- Total passing yards
- Total rushing yards
- Turnovers lost
- First downs given up
- Total yards given up
- Total passing yards given up
- Total rushing yards given up
- Turnovers gained by defense
- Offensive rank
- Defensive rank
- Special Teams Rank

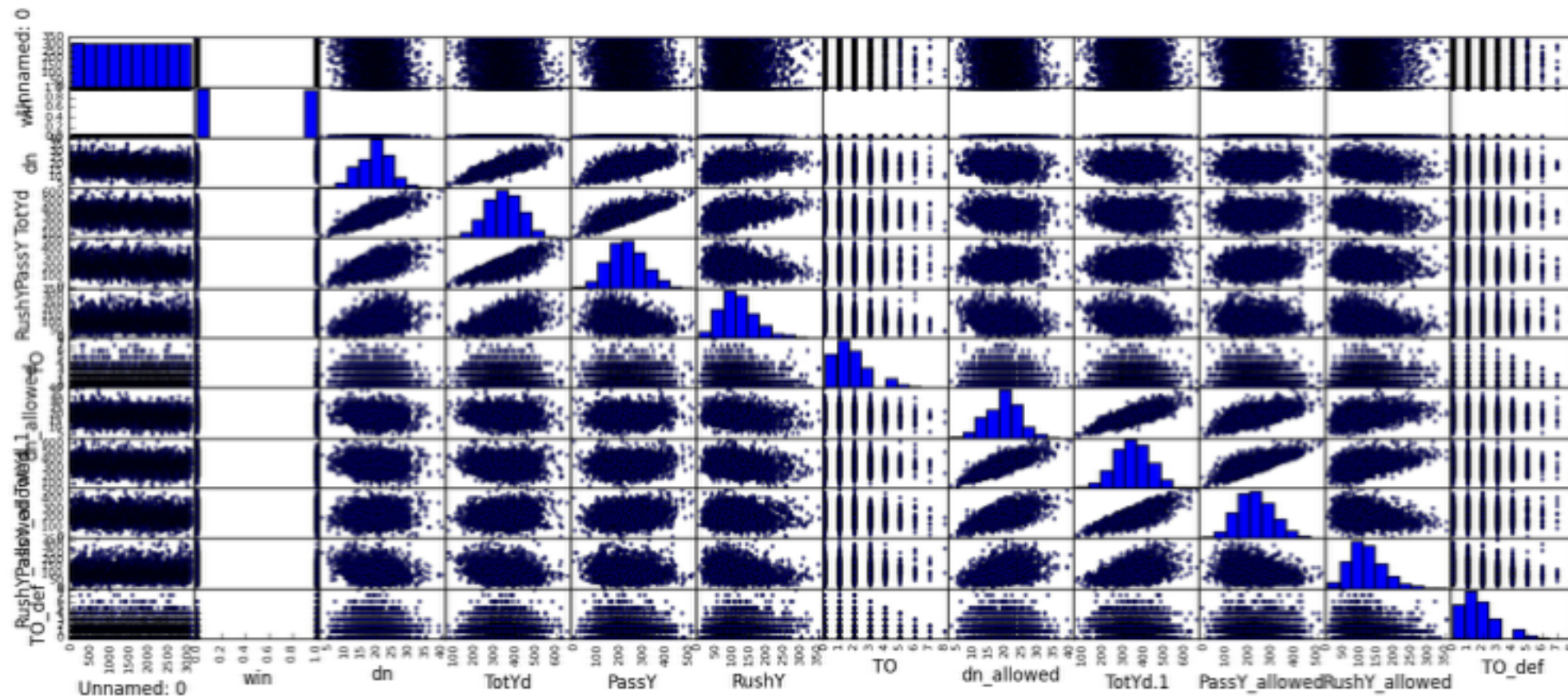
# Usable Game Data

- Win (output variable)
- ~~Day~~
- ~~Date~~
- ~~[link to boxscore]~~
- ~~Overtime (Y/N)~~
- ~~Opponent~~
- ~~Points Scored~~
- ~~Points Allowed~~
- First downs gained
- Total yards
- Total passing yards
- Total rushing yards
- Turnovers lost
- First downs given up
- Total yards given up
- Total passing yards given up
- Total rushing yards given up
- Turnovers gained by defense
- ~~Offensive rank~~
- ~~Defensive rank~~
- ~~Special Teams Rank~~

# Features

- First downs gained
- Total yards
- Total passing yards
- Total rushing yards
- Turnovers lost
- First downs given up
- Total yards given up
- Total passing yards given up
- Total rushing yards given up
- Turnovers gained by defense

Output: WIN (Y/N) - 1,0

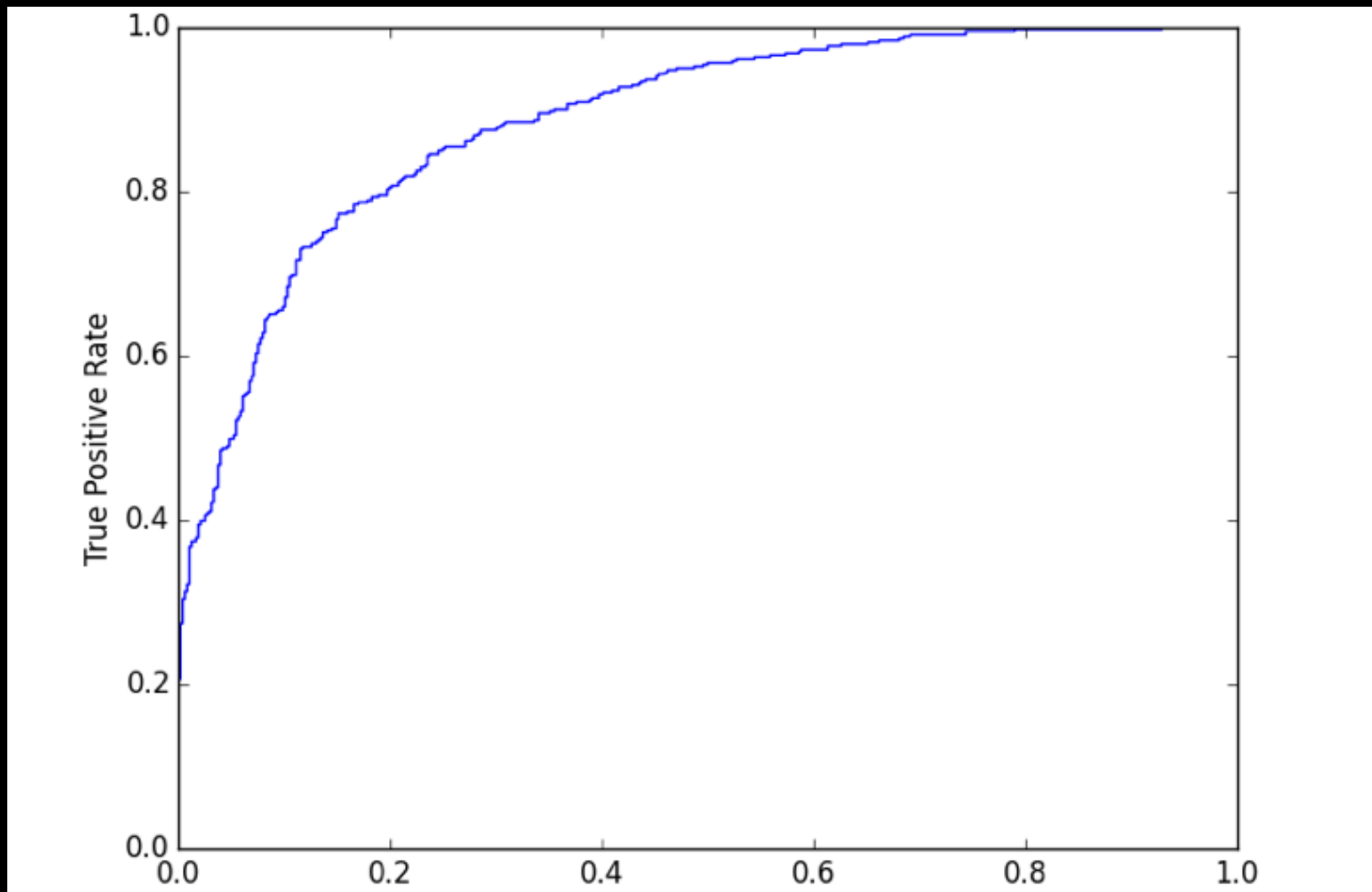


Feature Matrix



# Adaboost

- I chose to use a two-response **Adaboost with  $n\_estimators=200$**  model because individually, none of the factors I chose will be any good at determining a game, but together, they were able to determine the winner of a game with 80% confidence.



Validation

# Challenges

- Determining a project. This is the third project I have attempted.
- Real estate data: the data was too difficult to clean and I wasn't quite sure what my response variable was supposed to be.
- NFL vs. Weather: the data was meaningless. Turns out weather has very little impact on an NFL game.
- Adaboost: we never discussed an actual implementation so actually using the model for validation was slightly difficult.

# Possible Next Steps

- Expanded data:
  - Completion %
  - Third down conversion rate
- Selling this data to an NFL team so they know how to win ;)

# Conclusions and Key Learnings

- Turnovers have the largest impact on a game of any of the variables I tested. If a team wins the turnover battle, they have a very good chance of winning the game.
- This makes sense because for each turnover won, that is another shot for your offense to get into the endzone.

Thank you