

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

www.pro-football-reference.com

product hunt reddit /r/nationals /r/caps xkcd gmail github unsplash 8tracks hypem bleach mu

Sports-Reference: [Baseball](#) • [Basketball \(college\)](#) • [Football \(college\)](#) • [Hockey](#) • [Olympics](#) •

PRO-FOOTBALL-REFERENCE.COM

play index players teams seasons coaches leaders super bowl draft boxscores s

[Mobile Site](#) **You Are Here > [Pro-Football-Reference.com](#) > Pro Football Statistics and H**

News: s-r blog:2015 First Round Picks Up

'Tips' for hints [Tips](#)

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

1. Emmitt Smith	18,355
2. Walter Payton	16,726
3. Barry Sanders	15,269
4. Curtis Martin	14,101
5. LaDainian Tomlinson	
6. Jerome Bettis	13,684
7. Eric Dickerson	13,662
8. Tony Dorsett	13,259
9. Jim Brown	12,739
10. Marshall Faulk	12,312
	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	NFC	W	L	T	SRS
East					East				
NYE (1)	11	3	0	13.2	DAL (3)	10	4	0	1.6
BUF	8	6	0	4.7	PHI	9	5	0	3.5
MIA	7	7	0	4.7	NYG	5	9	0	-3.0
NYJ	3	11	0	-7.4	WAS	3	11	0	-9.1
North					North				
CIN (4)	9	3	1	0.1	DET (2)	10	4	0	1.9
BAL (6)	9	5	0	5.3	GNB (6)	10	4	0	7.7
PIT (5)	9	5	0	0.9	MIN	6	8	0	-2.1
CLE	7	7	0	-4.2	CHI	5	9	0	-7.5
South					South				
IND (3)	10	4	0	6.1	NOB (4)	6	8	0	-2.4
HOU	7	7	0	1.1	CAR	5	7	1	-6.2
JAX	3	12	0	-11.3	ATL	5	9	0	-3.4
TEN	2	13	0	-12.3	TAM	2	12	0	-10.2
West					West				
DEN (2)	11	3	0	10.2	ARI (1)	11	3	0	3.6
SDG	8	6	0	2.3	SEA (5)	10	4	0	7.3
KAN	8	6	0	6.1	SFO	7	7	0	-1.4
OAK	2	12	0	-8.9	STL	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

Used 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~~

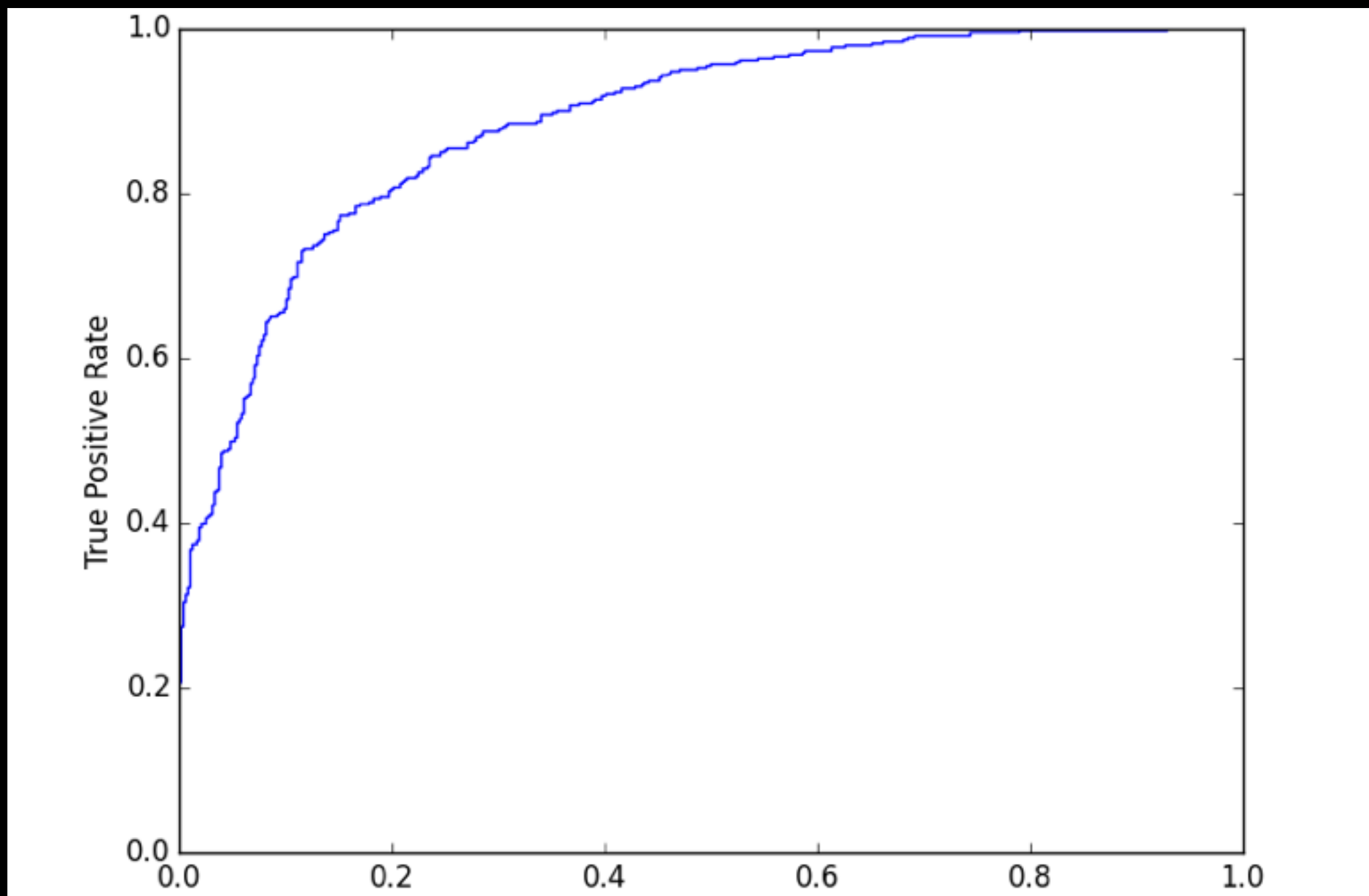
Features

- 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

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.
- Cross Validation Score: 0.81



Validation

Results

Metric:	dn	TotYd	PassY	RushY	TO	dn allowed	TotYd allowed	PassY allowed	RushY allowed	TO_def
Result:	0.0349	0.1397	0.0067	0.0993	0.2288	0.0532	0.1527	0.0	0.09927	0.2062

Feature values in determining outcome

Results

1. Turnovers given up (TO)
2. Turnovers gained (TO_df)
3. Total yards (TY)
4. Total yards allowed (TY_ald)
5. Rush yards (RushY)
6. Rushing yards allowed (RushY_ald)
7. First downs (dn)
8. First downs allowed (dn_ald)
9. Passing yards allowed (PassY_ald)
10. Passing yards (PassY)

Confusion Matrix

	Predicted Yes	Predicted No	
Actual Yes	349	92	441
Actual No	79	397	476
	428	489	

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 ;)
- More data.

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 end zone.

Thank you