Final Report: Team 6 Predicting Superbowl and College Football Champions

Ankit Arun, Beijie Li, Chetan Naik, and Sindhuri Mamidi

Department of Computer Science, Stony Brook University,
Stony Brook, NY 11794-4400
{aarun,beli,cnaik,smamid}@cs.stonybrook.edu
http://www.cs.stonybrook.edu/~skiena/591/projects

1 Challenge

The goal of our project is to predict the winners of 2015 Super Bowl and the College Football Championship.

"The game isn't over until it's over." It is tough to make prediction, especially about heartfelt fierce games.

Nation Football League: The National Football League (NFL) is a professional American football league that constitutes one of the four major professional sports leagues in North America. It is composed of 32 teams divided equally between the *National Football Conference (NFC)* and *American Football Conference (AFC)*.

NFL runs a 17-week regular season from the week after Labor Day to the week after Christmas, with each team playing sixteen games and having one bye week each season [1].

Out of the league's 32 teams, six (four division winners and two wild-card teams) from each conference compete in the NFL playoffs, a single-elimination tournament culminating in the Super Bowl, played between the champions of the NFC and AFC. The playoff tree Figure 1 below shows this schema. The champions of the Super Bowl are awarded the Vince Lombardi Trophy.

College Level Football: College football is American football played by teams of student athletes fielded by American universities, colleges, and military academies.

Starting in the 2014 season, four Division I Football Bowl Subdivision (FBS) teams will be selected at the end of regular season to compete in a playoff for the FBS national championship. The College Football Playoff will replace the Bowl Championship Series, which had been used as the selection method to determine the national championship game participants starting in the 1998 season.

The 2015 College Football Championship Game is the national championship game of the 2014 college football season. The first-ever College Football Championship Game. It is scheduled to take place on Monday, January 12, 2015, and

Fig. 1: NFL single-elimination tournament Playoff tree.

will be the culmination of the 2014 − 15 bowl games. The national title will be contested via a four-team bracket system, the College Football Playoff.

2 History/Background

Nation Football League: The NFL was formed on August 20, 1920, as the American Professional Football Conference; the league changed its name to the American Professional Football Association (APFA) on September 17, 1920, and changed its name to the National Football League on June 24, 1922, after spending the 1920 and 1921 seasons as the APFA. In 1966, the NFL agreed to merge with the rival American Football League (AFL), effective 1970; the first Super Bowl was held at the end of that same season in January 1967. Today, the NFL has the highest average attendance (67,591) of any professional sports league in the world and is the most popular sports league in the United States. The Super Bowl is among the biggest club sporting events in the world and individual Super Bowl games account for many of the most-watched television programs in American history.

Each team is allowed to have up to 53 players during the regular season, but only 46 can be active (eligible to play) on game days. The champions of the most recent season, the 2013 season, are the Seattle Seahawks, who defeated the Denver Broncos by a score of 43-8 in Super Bowl XLVIII. The team with the most championships is the Green Bay Packers, who have won 13 championships. The team that currently has the most Super Bowl championships is the Pittsburgh Steelers, who have won six.

College Level Football: The first ever intercollegiate football game between two American teams played under rules which would eventually become the rules under which modern American football is governed occurred between Princeton and Rutgers University in 1869. However, this game was far more like that of soccer than what has come to be recognized as American football. The completion of the first ever American football season came as a result of only two total games being played.

A game which modern audiences would more readily recognize as American football occurred six years after the first ever game and occurred between Harvard University and Tufts University on June 4, 1875. The first game ever played that resembles the game as it is known today was played between an American team, Harvard, and a Canadian team, McGill University of Montreal in 1874. This first game was a lot like rugby but much closer to the modern day version of football than soccer.

Formation of the NCAA: College football increased in popularity through the remainder of the 19th century. The Intercollegiate Athletic Association of the United States was formed in 1906. The IAAUS was the original rule making body of college football. The IAAUS got its current name of National Collegiate Athletic Association (NCAA), in 1910.

Point Spread: The point spread remains the favorite way to wager on pro football, regardless of how many new forms of wagering come on stream. It is called the line or spread and is commonly known as betting 'sides'. The common misconception is that Las Vegas sets the spread as its best guess at the margin of victory. But it is just a number they feel that is a perfect balance and will see an equal number of people to bet the underdog as on the favorite [11]. A negative value like -6.5 means that team is favored by 6.5 points. So deduct 6.5 points from their total score. A positive value on the same game would be +6.5 (add 6.5 points to their final score) and would make that team an underdog of 6.5 points. The favorite must win by at least seven points to cover the spread. The underdog can lose by six points and still cover.

3 Literature Review

Skiena and Hong [2] have shown relationship between the NFL betting line and public opinion expressed in blogs and microblogs (Twitter). The public sentiment for NFL teams is generated from the Lydia. The original sentiment series are daily positive and negative raw counts for each NFL teams. Based on the raw sentiment counts series, they developed a measure of relative favorableness for the team A over team B as follows:

$$Favorable(A) = \frac{\frac{(PosA+NegB)-(NegA+PosB)}{(PosA+PosB+NegA+NegB)} + 1}{2} \tag{1}$$

The favorable score for team A lies between [0,1] and can be viewed as the winning possibilities of team A for this game suggested by their sentiments. They have done their training and analysis based on the 683 NFL games from 2006 to 2008 and used 2009 data to evaluate their findings. They have found out that sentiment-based models perform much better than any other models in second half of the season. The reason could be that people are not good at correctly interpreting public sentiment. They have demonstrated evidence showing usefulness of sentiment on NFL betting. They demonstrate that a strategy betting roughly 30 games per year identified winner roughly 60% of the time from 2006 to 2009, well beyond what is needed to overcome the bookies typical commission (53%).

Boulier and Stekler [3] compared the performance of the power score method of prediction (measures the relative abilities of teams based on objective criteria such as the teams performance and the strength of its own and its opponents schedules) with the naïve model (home team always wins), the betting market and that of a well-known commentator (from The New York Times) in forecasting the outcomes of American football (NFL) matches. They found that spreads in the betting market offered superior guidance to those of the expert, power score method and the naïve model. Furthermore, when they estimated a model to explain the forecasts of the expert, the fitted values from that model offered superior forecasts to those of the expert himself. This suggests that, when experts depart from the forecasts suggested by their 'normal' or 'average' method of processing quantitative information, they become less rather than more reliable. They concluded with the observation that the Power Score model predicts with 63% probability which is good but not better than the betting model.

David Harville [4] mixed linear models based on the differences in score from past games to develop a procedure for predicting the outcomes of NFL. In detail, he constructed ratings for sports teams based on maximum likelihood estimates in which ratings were random variables. Taking into account the home-field advantage as well as the yearly characteristic performance levels of two teams, the predictions for 1,320 games played between 1971 and 1977 had an average absolute error of 10.68.

David Harville [5] discussed about a procedure for rating high-school or college football teams which is developed by applying linear-model methodology to the point spread for each game. The model includes effects for the home-field advantage and for the mean performance levels of the participating teams. The procedure can be modified to use only win-loss information or to ignore victory margins greater than a given margin.

Andrew D. Blaikie [6] analyzed data analysis is done to identify the most predictive statistics, which are used as data matrix in the model. It is based purely on statistics and used a committee of machines approach for greater consistency. In terms of prediction accuracy, it is found that the college football model performed poorly when compared to the NFL model. They discuss reasons and procedure to overcome the challenges. Afterwards, the models were examined using derivative analysis. The results of the research showed that the NFL model consistently was in the top half compared to other prediction experts, while the college football model tended to be closer to the middle of these rankings.

Soren P. Sorensen [7] reviews the history of ranking systems, and then document a particular open method for ranking sports teams against each other in his report. It illustrates different ranking systems in the context of NCAA Division 1-A football, but the methods described are very general and can be applied to most other sports only with minor modifications.

Apart from the the above discussed literature, we went through various models that can be used in our project. They are listed below:

1. Elo Ratings
In this model each of the players is given rating. Team rating is said to be the cumulative of ratings of all the players in the team. The difference between this cumulative team rating is used to predict the result of the match.

2. Pythagorean Wins

In this method the win rate of team is calculated using following formula.

$$WinRate = \frac{y^\beta}{x^\beta + y^\beta} \tag{2}$$

where,
x = Points against
y = Points for
β = Shrinkage Parameter (2 for NFL)

3. Eigenvector Method
The idea behind this method is to iteratively adjust the strength of a team as the schedule progresses. We assign more credits to victories over teams which are good .

4. Bradley - Terry- Luce Model
This is a paired comparison model. The Bradley-Terry model deals with a situation in which n individuals or items are compared to one another in paired contests. The model assumes there are positive quantities $\pi_1, \pi_2, \pi_3 \cdots, \pi_n$, which can be assumed to sum to one, such that:

$$P\{i \ beats \ j\} = \frac{\pi_i}{\pi_i + \pi_j} \tag{3}$$

If the competitions are assumed to be mutually independent, then the probability $p_{ij} = P\{i \text{ beats } j\}$ satisfies the logit model

$$\log \frac{p_{ij}}{1 - p_{ij}} = \phi_i - \phi_j \tag{4}$$

where,

$\phi_i = \log \pi_i$

4 Data Sets

NFL Data We have trained our models for predicting Super Bowl champions using the data purchased from *Armchair Analysis* [8]. They have data for all the games from 2000 to 2013 (607,404 Plays. 3,722 Games. 7,020 Players). This dataset is highly detailed and very accurate. Data is derived through a combination of charting from compressed game footage along with automated processes that check the quality and accuracy of every piece of information at 18 different stages. All plays are individually reviewed by human eyes.

College Football Data *Sunshine Forecast Enterprises* [9] has College Football data for all years from 1978 through 2013. It has extensive set of information such as rushing/passing yards and attempts, fumbles lost, interception throw for both team of each match. Data for 2014 college football games has been taken from *Dr. Wags Blog* [10].

Data Matrices for Baseline and PageRank Model:

The data matrix for predicting champions of NFL and College Football consists of following attributes - year and week of game played, home team, visitor team, home team score, visitor team score and the point difference. A section of data matrix used for NFL prediction is shown in table 1.

Table 1: A section of data matrix used in NFL baseline model.

game_id	year	week	home_team	away_team	home_score	away_score	score_diff
1	2000	1	ATL	SF	36	28	8
2	2000	1	CLE	JAC	7	27	-20
3	2000	1	DAL	PHI	14	41	-27

Data Matrices for Regression Model:

The columns in data matrix for linear regression model consist of game features as shown in table 2 and table 3.

Table 2: Features used in linear regression model.

Features	Description	Features	Description
TID	Team Total ID	SLY	Pass Yardage Short Left
GID	Game ID Number	SMA	Pass Attempts Short Middle
TNAME	Team Name	SMY	Pass Yardage Short Middle
PTS	Points	SRA	Pass Attempts Short Right
1QP	1st Quarter Points	SRY	Pass Yardage Short Right
2QP	2nd Quarter Points	DLA	Pass Attempts Deep Left
3QP	3rd Quarter Points	DLY	Pass Yardage Deep Left
4QP	4th Quarter Points	DMA	Pass Attempts Deep Middle
RFD	1st Downs - Rush	DMY	Pass Yardage Deep Middle
PFD	1st Downs - Pass	DRA	Pass Attempts Deep Right
IFD	1st Downs - Penalty	DRY	Pass Yardage Deep Right
RY	Rush Yards	WR1A	Attempts - WR 1 or 2
RA	Rush Attempts	WR1Y	Yardage - WR 1 or 2
PY	Pass Yards	WR3A	Attempts - WR 3, 4 or 5
PA	Pass Attempts	WR3Y	Yardage - WR 3, 4 or 5
PC	Completions	TEA	Pass Attempts - TE
SK	Sacks (Against)	TEY	Pass Yardage - TE
INT	Intercepting Player's for Defense	RBA	Pass Attempts - RB
FUM	Fumbles Lost	RBY	Pass Yardage - RB
PU	Punts	SGA	Shotgun Attempts
GPY	Gross Punt Yardage	SGY	Shotgun Yardage
PR	Punt Returns	P1A	Pass Attempts - 1st Down
PRY	Punt Return Yardage	P1Y	Pass Yardage - 1st Down
KR	Kick-off Returns	P2A	Pass Attempts - 2nd Down
KRY	Kick-off Returns Yardage	P2Y	Pass Yardage - 2nd Down
IR	Def Intercepting Player Returns	P3A	Pass Attempts - 3/4 Down
IRY	Intercepting Player Return Yardage	P3Y	Pass Yardage - 3/4 Down
PEN	Pen Yardage (Against)	SPC	Short Completion
TOP	Time-of-Possession	MPC	Medium Completion
TD	Touchdowns	LPC	Long Completion
TDR	TD's - Rushing	Q1RA	Rush Attempts - 1st Quarter
TDP	TD's - Passing	Q1RY	Rush Yardage - 1st Quarter
TDT	TD's via Turnovers	Q1PA	Pass Attempts - 1st Quarter
FGM	Field Goals Made	Q1PY	Pass Yardage - 1st Quarter
FGAT	Field Goal Attempts	LCRA	Rush Attempts - Late/Close
FGY	Field Goal Yardage	LCRY	Rush Yardage - Late/Close
RZA	Drives in Red Zone	LCPA	Pass Attempts - Late/Close
RZC	Red Zone Drive TD's	LCPY	Pass Yardage - Late/Close
BRY	Big Rush Yardage	RZRA	Rush Attempts - Red Zone
BPY	Big Pass Yardage	RZRY	Rush Yardage - Red Zone
SRP	Successful Rush Plays	RZPA	Pass Attempts - Red Zone
S1RP	Successful Rush - 1st Down	RZPY	Pass Yardage - Red Zone
S2RP	Successful Rush - 2nd Down	SKY	Total Yards lost to Sacks
S3RP	Successful Rush - 3/4 Down	LBS	Sacks by own LB's
SPP	Successful Pass Plays	DBS	Sacks by own DB's
S1PP	Successful Pass - 1st Down	SFPY	Starting Field Pos
S2PP	Successful Pass - 2nd Down	DRV	Number of Drives on Offense
S3PP	Successful Pass - 3/4 Down	NPY	Net Punt Yardage
LEA	Rush Attempts - Left End	TB	Touchbacks

Table 3: Features used in linear regression model.

Features	Description	Features	
LEY	Rush Yardage - Left End	I20	Number of Punts inside 20
LTA	Rush Attempts Left Tackle	RTD	Punts/Kickoff TD's
LTY	Rush Yardage - Left Tackle	LNR	DL Tackles - Rush
LGA	Rush Attempts Left Guard	LNP	DL Tackles - Pass
LGY	Rush Yardage Left Guard	LBR	LB Tackles - Rush
MDA	Rush Attempts Middle	LBP	LB Tackles - Pass
MDY	Rush Yardage Middle	DBR	DB Tackles - Rush
RGA	Rush Attempts Right Guard	DBP	DB Tackles - Pass
RGY	Rush Yardage Right Guard	NHA	No Huddle Attempts
RTA	Rush Attempts Right Tackle	S3A	3rd & Short Attempts
RTY	Rush Yardage Right Tackle	S3C	3rd & Short Conversion
REA	Rush Attempts Right End	L3A	3rd & Long Attempts
REY	Rush Yardage Right End	L3C	3rd & Long Conversion
R1A	Rush Attempts - 1st Down	STF	Stuffed Runs
R1Y	Rush Yardage - 1st Down	DP	Points by Defense
R2A	Rush Attempts - 2nd Down	FSP	False Starts
R2Y	Rush Yardage - 2nd Down	OHP	Off Holding Penalty
R3A	Rush Attempts - 3/4 Down	PBEP	Play Book Exec. Pen
R3Y	Rush Yardage - 3/4 Down	DLP	Defensive Line Penalty
QBA	QB Rush Attempts	DSP	Def Secondary Penalty
QBY	QB Rush Yardage	DUM	Dumb Penalties
SLA	Pass Attempts Short Left	PFN	Poor Fundamentals Pen

The size of the data-sets that we are using is shown in the Table 4.

Table 4: Size of data-sets

	Source	Years	Rows
NFL	Sunshine regular season data	1983 - 2013	65 X 30 = 1950
NFL	Armchair analysis	2000 - 2013	3722
NCAA	Sunshine forecast data	1978 - 2013	700 X 35 = 24,500

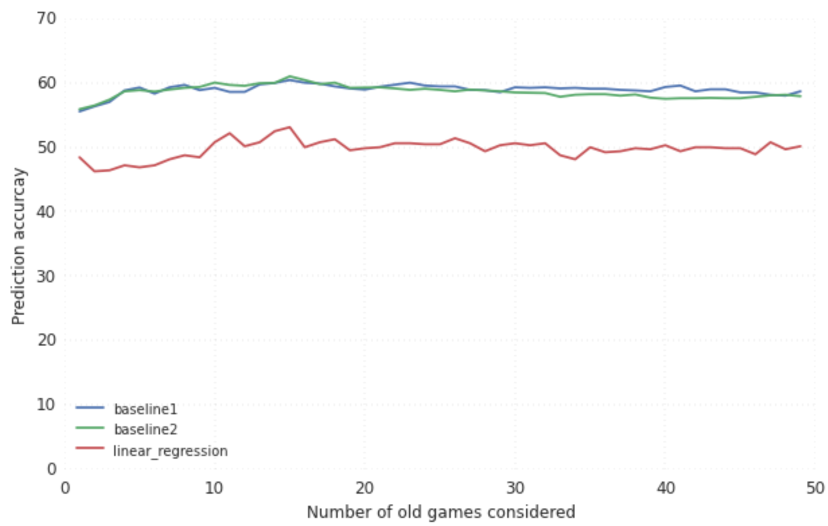
We have divided the data-set into **train and test set** by randomly selecting 60% of the rows as train data while 40% as test. As shown in table 5.

Table 5: Size of train and test data-sets

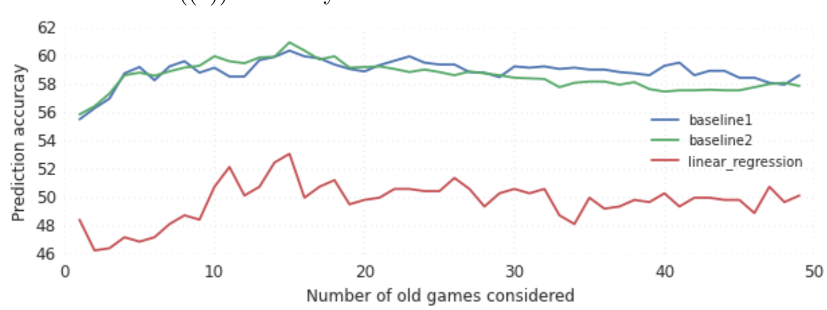
	#Years	#Train rows per year	#Train rows	#Test rows per year	#Test rows
NFL	13	171	2223	115	1495
College Championship	12	495	5940	331	3972

5 Observations

Both NFL and College Football have huge amount of data but all the data is not as useful in predicting the results of current game. The accuracy of a model can vary with number of historic games we are considering to predict the result of and game. We then compared the accuracy of all the models that we have used for different number of historic games. This helped us in finding the optimum number of historic games that gives the highest accuracy. For NFL the optimum number of historic games is 15 for all the models as shown in figure 2 while for College Football the optimum number of games is 17 for baseline models while it is 30 for linear regression model as shown in figure 3. This experiment turned out to be an expensive operation as it took 16 hours to compute the accuracy for all the 50 iterations over 13 years' data.

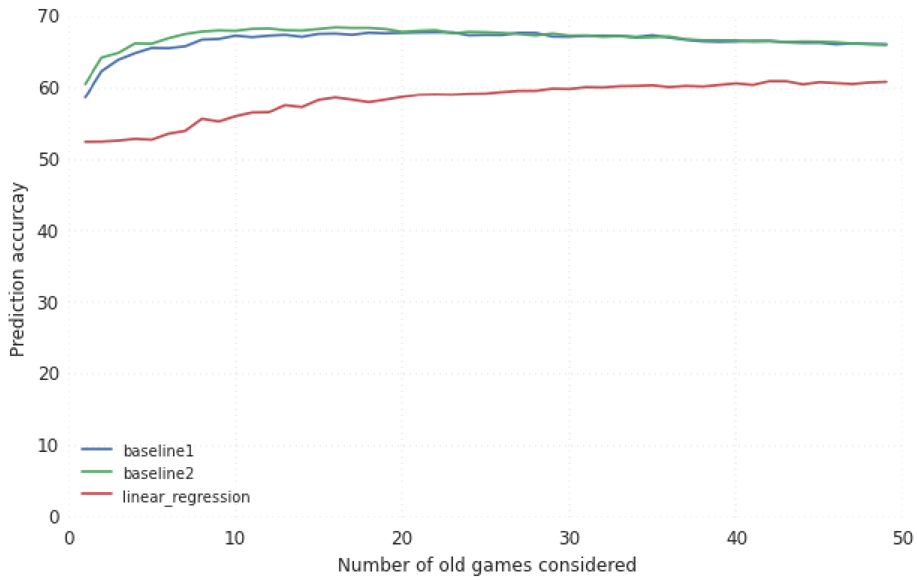


((a)) Accuracy vs Number of historic Games.

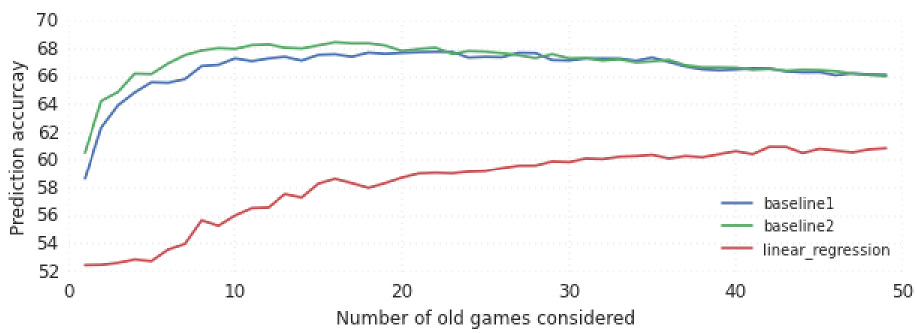


((b)) Detailed view of Accuracy vs Number of historic games.

Fig. 2: Accuracy vs Number of historic games for NFL.



((a)) Accuracy vs Number of historic Games.



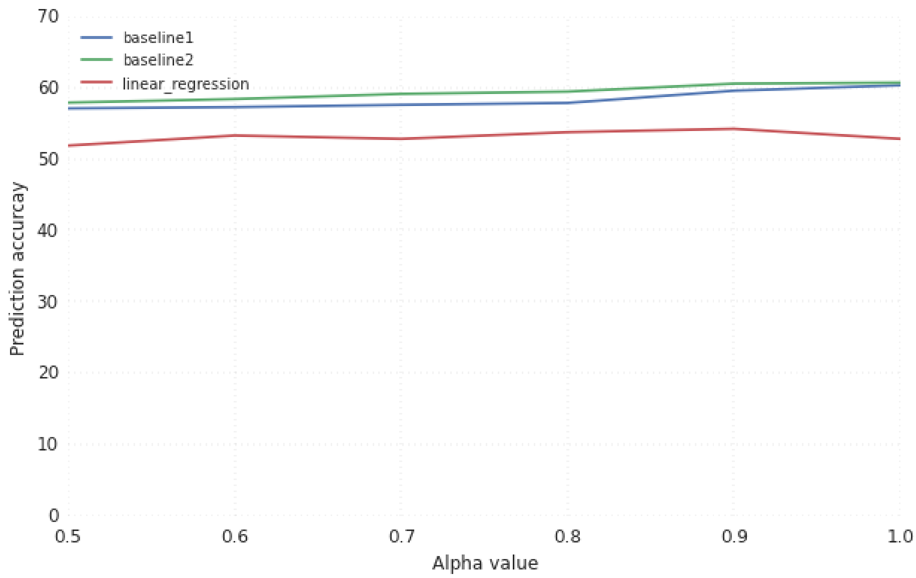
((b)) Detailed view of Accuracy vs Number of historic games.

Fig. 3: Accuracy vs Number of historic games for College Football.

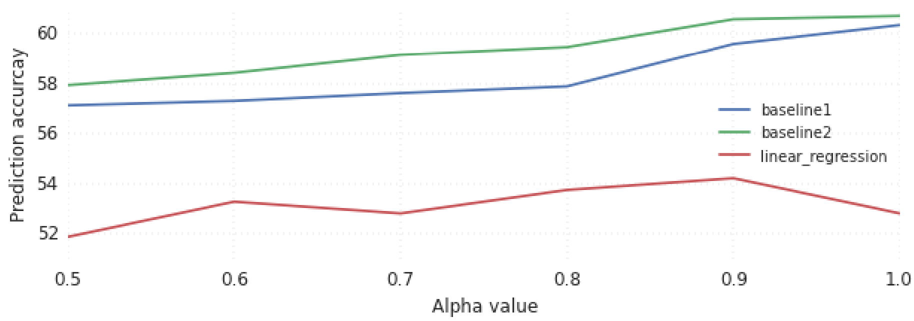
We have also experimented with different weights for historic games using the following formula,

$$weight = alpha^{(n^{th}\ historic\ game)} \tag{5}$$

Figure 4 and figure 5 shows the comparison of accuracies for different values of alpha in the case of NFL and College Football respectively. For NFL, optimum alpha value for baseline models is 1.0 while it is 0.9 for linear regression models. However, for College football alpha value is 1.0 for all the models.

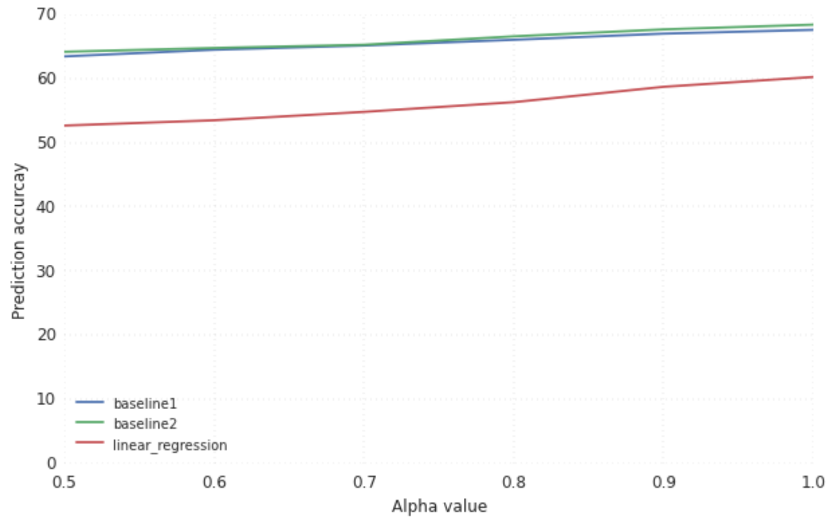


((a)) Accuracy vs Alpha value.

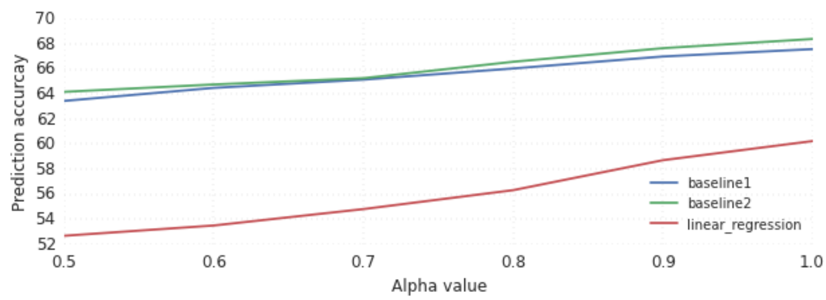


((b)) Detailed view of Accuracy vs Alpha value.

Fig. 4: Accuracy vs Alpha value for NFL. Alpha value decides the weight of historic games considered.



((a)) Accuracy vs Alpha value.



((b)) Detailed view of Accuracy vs Alpha value.

Fig. 5: Accuracy vs Alpha value for College Football. Alpha value decides the weight of historic games considered.

6 Baseline Models

Baseline 1: Our baseline 1 model is a *Point Score Difference Model*, where we take the point difference (difference of total point scored and total points allowed) to predict the winner of a game. *Point Score Difference Model* can be mathematically represented as:

$$M.r = p \tag{6}$$

where,

- M has the number of games played by each team on the diagonal, and the negation of the number of times each team played head-to-head on the off-diagonal spots. The size of matrix M is $Number\ of\ Teams \times Number\ of\ Teams$

– r vector is a list of the average end-of-regular season point differentials for each team.

$$r = \frac{(points\ scored - points\ allowed)}{total\ number\ of\ games\ played\ by\ the\ team}$$

.

– p is the prediction vector of size $Number\ of\ Teams$.

Baseline 2: The second baseline model that we have used is an extension of the baseline 1 model. We have uses home team advantage in addition to the average point difference to predict the winners in this model, following formula describes baseline 2:

$$\begin{aligned} Point\ Difference\,[X - Y] = &\frac{(Avg\ points\ for\ X\ +\ Avg\ points\ against\ Y)}{2} \\ &- \frac{(Avg\ points\ for\ Y\ +\ Avg\ points\ against\ X)}{2} \\ &+ 3\ *\ (X == Home\ Team) \\ &- 3\ *\ (Y == Home\ Team) \end{aligned}$$

The baseline 1 model predicts 57.76% of the NFL games and 67.44% of the College Football games correctly while baseline 2 model predicts 59.70% of the NFL games and 68.40% of the College Football games correctly. This is a good accuracy threshold which we want to beat using advanced models.

7 Advanced Models

Intuition from baseline to advanced models:

Ranking systems generally fall in either predictive or earned ranking category. Earned ranking ranks the teams based on their past performance while predictive ranking provides the probability of a team winning against another.

Our baseline model, *Point Score Difference Model*, categorizes as an earned ranking system. Consider a case where a team A beats B, B beats C, but C beats A. An earned ranking system would not be able to compare such non-linear relations, while a predictive model can compare such relations effectively. In order to make an accurate predictive system we can use extensive set of game features described in tables 2 and 3. This allows a more precise extrapolation of the next weeks' games.

Thus, to better predict the winner of Super Bowl and college football champions with the knowledge of playoff schedule, we decided to use a predictive model which takes more fresh and well-rounded information into consideration.

So efforts should be spent on trying to optimize predictive capabilities. We have used linear regression model trained on historical games as a predictive ranking system.

7.1 Linear Regression Model

We got started with the linear regression method assuming that the dependent variable varies linearly with the independent variable(s).

We are using a linear equation of the form $Y = mX + c$, where,

- Y: Predicted score difference
- X: Features used to predict the score difference
- m: Predictive capabilities of the features
- c: Intercept of the linear model

Features as X

Records in each game which include 142 features such as rush, pass, kick, time of possession, punt and touchdown for each of the teams in the game (home team and away team) is considered as X. We have used two configurations for linear regression models.

- **With separate features -** In this variant, features of both of the playing teams of a game are considered separately in building the model. With this configuration for features we tried the following regression models:
  - *Unconstrained Regression*
  - *Ridge Regression*
  - *Lasso Regression*
- **With paired features -** We have paired the matching features of both of the playing teams of a game to get a single list of features rather than having same feature for both the teams. We have taken the difference of features between home team and away team to generate game data matrix. [3722 games $\times$ 142 features]
  With this configuration also we have tried the following regression models:
  - *Unconstrained Regression*
  - *Ridge Regression*
  - *Lasso Regression*

Point difference as Y

Using the features of both the playing teams, we can predict the point difference between home team and away team.

Training procedure: We train our model by minimizing the cost function to find the right weight for every feature in regression. To train a model which predicts a coming game G, we define the training set to be all previous game in train set played before G. (recording games starting from 2000).

We update the model after every game to get a better fit. To be specific in football context, we are actually trying to build a model which estimates 'what makes a team beat their opponent' to predict the winners of a game. The model is continuously adjusted to follow the current trend in football. For example, the winner of a game played 10 years before might have relied more on passing yard, but as players abilities grow, rushing yard may contribute more to victory.

Predicting procedure: When predicting a coming game G in the test set played by home team T1 and away team T2, we will have an up-to-date model trained as described in the section above. To get the predicted point difference from the model, we would also need all the features of game G. We use the optimum number of historic game as described in the observation section.

Using the the prepared features and updated model, we predict the point difference of game G between teams T1 and T2. After gaining our prediction of point difference, we use the sigmoid function to transfer the difference to the winning probability of T1 for game G.

$$P(d) = \frac{1}{1 + e^{(-s \times d)}} \tag{7}$$

- d: Point difference between home team and away team predicted using Linear Regression model
- $P(d)$: Probability that the home team wins given the point difference d.
- s: Slope of the sigmoid curve

The slope of the sigmoid function determines how quickly the function rises or drops to 1 or 0. A sigmoid function with slope closer to 1 quickly rises/falls to 1/0 while a sigmoid function with slope closer to 0 slowly rises/falls to 1/0.

We did empirical analysis with few slope values to determine the value which gives the best accuracy. We found slope value of 0.2 translated the score difference to probabilities with almost uniform distribution.

Model Validation: We see that the model has high R-squared value (in tab 6). This is one of the indicators of a model with good fit.

We also use residual plot (as shown in figure 6) to validate the regression model. The residuals from a fitted model are the differences between,

1. the responses observed at each combination values of the explanatory variables and
2. the corresponding prediction of the response computed using the regression function

If the residuals appear to behave randomly, it suggests that the model fits the data well.

Table 6: OLS regression results

Dep. Variable:	score_diff	**Covariance Type:**	nonrobust
Model:	OLS	**R-squared:**	0.998
Method:	Least Squares	**Adj. R-squared:**	0.998
Date:	Fri, 12 Dec 2014	**F-statistic:**	1963
Time:	07:40:11	**Prob (F-statistic):**	0
No. Observations:	534	**Log-Likelihood:**	-476.37
Df Residuals:	401	**AIC:**	1219
Df Model:	133	**BIC:**	1788

Evaluating procedure: The error distribution of predicting point difference is shown in figure 7. This distribution is fitting the Gaussian curve (yellow line in the plot) having mean (circular green dot in the plot) and median (circular red dot in the plot) very close to zero. This confirms our observation that the model is behaving normally. The green line in the plot shows the standard deviation and from it we can infer that our predicted point difference can be off from the actual value by $+/-$ 18.

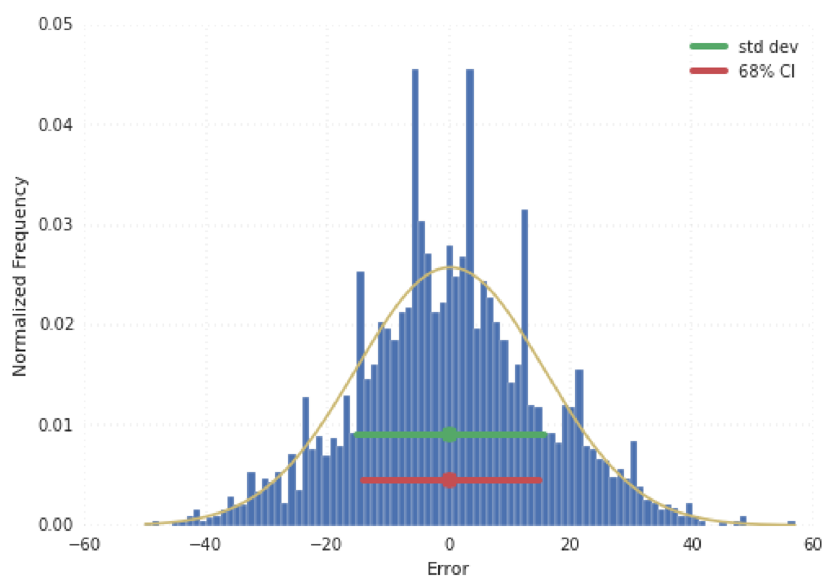


Fig. 7: Error distribution of the predicted point difference for NFL.

The best linear regression model predicts 62.6% of the NFL games and 60.42% of the College Football games correctly.

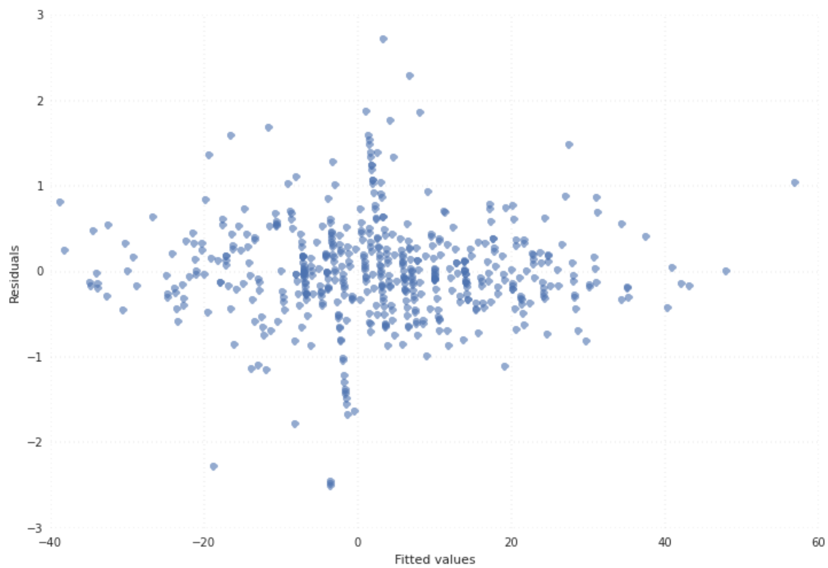


Fig. 6: Residual plot for linear regression on NFL. Residuals are spread randomly, which suggests the model fits the data well. Here the intensity of the colour represents the density of values.

We would expect linear regression model to produce more accurate predictions, but it is not giving us results which are significantly better than the baseline models. The reason for this is explained below using the following tables.

Figure 8(a) shows the top 10 features with highest weight that the model uses to predict score difference.

Figure 8(b) shows a column containing the features that we have estimated for one of the coming games based on averaging twelve previous games' features (we have described this technique in the Predicting procedure section of this report) and a column containing the actual game features (we get this after the game gets played).

We can see that the estimated feature vector that we are generating by averaging previous games differs a lot (in terms of absolute value difference and sign difference) when compared to the actual values. This is because, averaging technique won't generate a good feature vector as it will not be able to capture the variations in every feature of the game based on few historical games. We found this to be one of the major reasons for the mediocre accuracy of the model.

7.2 PageRank Model

PageRank is a ranking method developed to rank nodes of a graph based on its link structure. PageRank was developed at Stanford University by Larry Page and Sergey Brin in 1996 as part of a research project about a new kind of search

	feature_description	coeff_value
RTD_diff	Punts/Kickoff Touchdowns Difference	2.828946
TDP_diff	Touchdowns-Passing Difference	2.588013
TDR_diff	Touchdowns-Rushing Difference	2.574888
TDT_diff	Touchdowns via Turnovers Difference	1.559245
FGM_diff	Field Goals Made Difference	1.215837
3QP_diff	3rd Quarter Points Difference	0.610539
1QP_diff	1st Quarter Points Difference	0.609781
4QP_diff	4th Quarter Points Difference	0.607957
2QP_diff	2nd Quarter Points Difference	0.598245
P1A_diff	Pass Attempts-1st Down Difference	0.424798

((a)) Top 10 linear regression features with highest weight and their weight values.

feature	estimated_feature	actual_feature
RTD_diff	0	0
TDP_diff	0	1
TDR_diff	0	-1
TDT_diff	0	0
FGM_diff	1	-1
3QP_diff	-2	-7
1QP_diff	4	-3
4QP_diff	0	7
2QP_diff	1	0
P1A_diff	-2	6

((b)) Estimated vs. actual feature values of a sample game taken from the data-set.

Fig. 8: Reason for mediocre performance of linear regression model.

engine [12]. It was later used by the company that they formed, Google, to rank websites based on their incoming and outgoing links.

PageRank works by counting the number and quality of links to a page to determine a rough estimate of how important the website is. The underlying assumption is that more important websites are likely to receive more links from other websites. PageRank can be computed either iteratively or algebraically. The iterative method can be viewed as the power iteration method or the power method.

This method can be used to rank football teams if we can model the problem as a graph ranking problem. We do so by defining the football teams as the nodes of the graph and every encounter between two teams adds an edge between them. The weight of the edge defines how well the ranking works. Here we use point differential as edge weights.

$$edge\ weight = \frac{abs(home\ score - away\ score)}{(home\ score + away\ score)} \tag{8}$$

$$r = M \times r \tag{9}$$

where,
r - rank
M - stochastic adjacency matrix

We use the above formula to calculate the PageRank by power iteration. The PageRank model predicts 62.5% of the NFL games and 62.4% of the College Football games correctly.

You can see that PageRank model is slightly better than the linear regression model and is more consistent in predicting the winners. In the next section, we'll describe how we have used the probabilities from these models to predict winners using Monte Carlo simulation.

7.3 Crowd sourcing Prediction for NFL

Crowd sourcing methods have been known to be good at predicting events. Though we did not get a chance to get data from a large crowd, the data obtained by interviewing 13 people gave us following prediction as shown in table 7.

Table 7: Crowd's view

Team	Votes
New England Patriots	7
Arizona Cardinals	3
Green Bay Packers	2
San Francisco 49ers	1

7.4 Evaluation of all the models

Table 8 and table 9 shows the accuracy for different models on train and test data set for both NFL and College football. We can infer that in general college football has higher accuracy than NFL because of its large data-set.

Table 8: Accuracy over train data-set

Models	Train Accuracy(%)	
	NFL	College Football
Baseline 1	60.34	67.49
Baseline 2	60.70	68.35
Unconstrained Regression (Separate Features)	52.34	55.64
Ridge Regression (Separate Features)	50.47	57.86
Lasso Regression (Separate Features)	62.62	61.08
Unconstrained Regression (Paired Features)	52.34	54.03
Ridge Regression (Paired Features)	55.14	55.44
Lasso Regression (Paired Features)	62.62	61.08
PageRank	63.55	65.33
Point Spread	67.18	68.59

Table 9: Accuracy over test data-set

Models	Test Accuracy(%)	
	NFL	College Football
Baseline 1	57.76	67.44
Baseline 2	59.70	68.40
Unconstrained Regression (Separate Features)	48.13	53.17
Ridge Regression (Separate Features)	46.77	58.00
Lasso Regression (Separate Features)	62.62	60.42
Unconstrained Regression (Paired Features)	46.44	54.32
Ridge Regression (Paired Features)	45.11	55.44
Lasso Regression (Paired Features)	57.50	61.08
PageRank	62.50	62.39
Point Spread	67.18	68.59

8 Final Prediction and Conclusions

We use Monte Carlo Simulation to predict the winners of the tournament. To do so we use probabilities obtained by the above mentioned advanced models.

Monte Carlo Simulation

Out of the league's 32 teams, six (four division winners and two wild-card teams) from each conference compete in the NFL playoffs, a single-elimination tournament culminating in the Super Bowl, played between the champions of the NFC and AFC.

After regular season, we have knowledge of which 12 teams would appear in playoff and which will play against which. By putting all those possible team combinations into our model, we obtain the probability p that a team T1 beats the other team T2. Then for a certain game played between T1 and T2, T1 will win with the probability of p. Follow the rule of winning team goes to the next round, we can figure out the probability for every 12 team to win the Super Bowl.

Based on 250000 simulations of 2014 playoff games, final champion prediction and probabilities are shown as below.

We have used 14 weeks regular season data to generate the probabilities that any team T1 wins against T2. With this we found out team standings in their respective conferences and divisions. And then, using the playoff selection criteria, we selected 12 teams and used Monte Carlo method to generate the probability of these teams win the Super Bowl championship. The resulting probabilities are as shown in table 10.

Table 10: Winning probability of teams playing NFL calculated using Monte Carlo method

Team Name	Win Probability
SEA	0.257668
NE	0.194108
DEN	0.139708
DAL	0.092936
GB	0.083004
BAL	0.044044
PIT	0.043560
ARI	0.042300
IND	0.036980
CIN	0.035992
CAR	0.011068
DET	0.018632

The 25 teams selected by the college football selection committee is out and we selected the top 4 teams and used the Monte Carlo method to simulate college football games and the winning probabilities. The final winning probabilities are shown in table 11.

Table 11: Winning probability of college football teams calculated using Monte Carlo method

Team Name	Win Probability
Oregon	0.4685
Alabama	0.3241
Florida State	0.1186
Ohio State	0.0887

Final Prediction:

NFL Champion		**Seattle Seahawks**
College Football Champion		**Oregon**

Suggestions:

1. We have used point differential as the edge weight in our PageRank model. The edge weight can be modelled in a better way by changing the edge weights and including self loops. This would make the model much more robust.
2. Trying out ELO Rating, Bradley-Terry-Luce Model, Pythagorean Wins and comparing it with existing models.
3. Incorporating crowd sentiments which can be crawled from news sites, social networking sites such as twitter and facebook.

Acknowledgments.

1. Thanks to Armchair Analysis [8] for the NFL data-set they have provided.
2. Thanks to Sunshine Forecast Enterprises [9] and Dr. Wags Blog [10] for providing vast data-set for college football.

9 Bibliography

References

1. http://en.wikipedia.org/wiki/National_Football_League
2. Y. Hong, S. Skiena, C.: The Wisdom of Bookies? Sentiment Analysis Versus the NFL Point Spread. In:Fourth International AAAI Conference on Weblogs and Social Media (2010)
3. Boulier, B. L., Stekler, H. O.: Predicting the outcomes of National Football League games. International Journal of Forecasting (2003)
4. Harville, D.: Predictions for National Football League games via linear-model methodology. In:Journal of the American Statistical Association 75.371 (1980): 516-524.
5. David Harville: The Use of Linear-Model Methodology to Rate High School or College Football Teams, Journal of the American Statistical Association, 1977
6. Andrew D. Blaikie et. al: NFL & NCAA Football Prediction using Artificial Neural NetworksJournal of the American Statistical Association, 1977

7. Soren P. Sorensen: An Overview of Some Methods for Ranking Sports Teams, 2012.
8. http://www.armchairanalysis.com/
9. http://www.repole.com/sun4cast/index2014.html
10. http://www.drwagpicks.com/
11. http://espn.go.com/nfl/lines
12. http://en.wikipedia.org/wiki/PageRank