Journal of Quantitative Analysis in Sports

Volume 7, *Issue* 2	2011	*Article* 9

NFL Prediction using Committees of Artificial Neural Networks

John A. David, *The College of Wooster*
R. Drew Pasteur, *The College of Wooster*
M. Saif Ahmad, *The College of Wooster*
Michael C. Janning, *The College of Wooster*

Recommended Citation:
David, John A.; Pasteur, R. Drew; Ahmad, M. Saif; and Janning, Michael C. (2011) "NFL Prediction using Committees of Artificial Neural Networks," *Journal of Quantitative Analysis in Sports*: Vol. 7: Iss. 2, Article 9.

DOI: 10.2202/1559-0410.1327

NFL Prediction using Committees of Artificial Neural Networks

John A. David, R. Drew Pasteur, M. Saif Ahmad, and Michael C. Janning

Abstract

This paper analyzes the ability of a neural network model to predict the outcome of NFL games. This model uses only readily available statistics, such as passing yards, rushing yards, fumbles lost, and scoring. A key component of this model is the use of statistical differentials to compare teams. For example, the offensive passing yards gained by one team are compared to the defensive passing yards allowed by an opposing team to create a data set of expected values for a given matchup. By using principal component analysis and derivative based analysis, we determined which statistics influence our model the most. We assessed the performance of the model by comparing its performance to that of published prediction algorithms and the Las Vegas oddsmakers over multiple seasons. Two novel aspects of this work include the use of multiple committees of machines for prediction and the use of our model to simulate virtual round-robin tournaments to establish an objective ranking of the teams.

KEYWORDS: artificial neural networks, NFL, prediction

Author Notes: This work was supported in part by an institutional grant to the The College of Wooster from the Howard Hughes Medical Institute and was completed in part during the The College of Wooster's Applied Mathematical Research Experience (AMRE).

1 Introduction

Football is arguably the American sport with the most interest and most passionate fans. Predicting the outcome of National Football League professional games is a difficult problem. The best predictors still have an average error of 10-12 points per game as seen in `thepredictiontracker.com`. It is largely an art, with little to no underlying scientific theory. Every game has a large probabilistic element to it, and the same two teams playing a hundred times could have a wide range of outcomes. These factors make football prediction an interesting problem and suggest artificial neural networks (ANNs) as a potential tool to use in making predictions.

Applications of ANNs include system identification and control (vehicle control, process control), quantum chemistry, game-playing and decision making (backgammon, chess, racing), pattern recognition (radar systems, face identification, object recognition and more), sequence recognition (gesture, speech, handwritten text recognition), medical diagnosis, financial applications (automated trading systems), data mining, visualization and e-mail spam filtering according to Wikipedia (2010). Any problem with data and notions of connection between input and output without any underlying fundamental laws, e.g. Newton's law of gravitation or cooling, is an attractive problem for this method of generalized nonlinear regression. Also, systems for which the physical laws are known, but models built from these laws are difficult to construct, validate and use, are attractive areas for ANNs. Several applications in chemistry where ANNs have proven particularly useful include Blasco, Rueyo, Larroya, Dopazo, and Chen (1999), Shenvi, Geremia, and Rabitz (2004), and Dauber, Israel, and Taylor (2010).

In this paper we illustrate how simple statistical comparisons of teams, combined with committees of ANNs, are quite accurate in predicting the outcome of NFL games over multiple seasons as seen in several metrics including absolute error and winning percentage. The remainder of this paper is organized as follows. Section 2 contains a description of previous work in ANN prediction involving sports. We then describe the input statistics used and the basic ANN structure in Section 3. Section 4 contains information on the use of a committee of machines approach used to combine multiple networks and its prediction outcomes over multiple seasons of games. Several techniques of data and model analysis are described in Section 5. Finally, we describe the use of this model to perform weekly round-robin tournaments and "power rankings" in Section 6.

2 Previous Research

Neural network algorithms are an advanced field of study with a significant body of literature. However as with most mathematical algorithms, application to real problems and data sets reveals challenging issues. The work by Purucker (1996) introduced the idea of using ANNs to predict the outcome of NFL games. This work offers an excellent introduction to basic ANNs in the context of NFL prediction. Purucker applied several types of supervised and unsupervised networks to predicting weeks 15 and 16 of the 1994 NFL season. Work by Kahn (2003) used 13 weeks of game data, including total yardage differential, rushing yardage differential, time of possession differential, turnover differential and home field advantage to predict weeks 14 and 15. He achieved 62.5% and 75% accuracy in these two weeks. Considering the ESPN experts in 2009 had between 61% and 67% accuracy predictions, this methodology shows promise. Loeffelholz *et. al.* performed a more extensive study of neural network prediction of NBA games in Loeffelholz, Bednar, and Bauer (2009). The work in this paper incorporated ideas of these past studies, with several key additions including season-to-date statistics combined with previous seasons, a simple method for incorporating team statistics, a robust method of combining multiple ANNs and analysis of the performance of this methodology over several seasons.

3 Basic Model

The fundamentals of the model involve two key components: the statistics used to describe the teams and the details of the ANNs used.

3.1 Statistical Data

There are a seemingly-infinite number of statistics we could use as we try to predict the point differential in these games. In this first attempt at using ANNs to predict sports outcomes, we attempted to identify the key statistics that may predict how two teams will perform when matched against each other. Choosing which statistical categories to use is unquestionably the most subjective part of this modeling

DOI: 10.2202/1559-0410.1327

process. We identified scoring, passing and rushing yardage, fumbles, and interceptions as key statistics, and track the numbers for and against each team (e.g. rushing yards gained offensively and allowed defensively) in each category. Additionally, we also include win-loss record and a home-field advantage parameter. We then used the season-to-date average of these statistics to describe each team, except in the first five weeks of the season where we used a weighted average between the current season and the previous season statistics. The ratio was 100% previous season in week 1, then 80% previous season and 20% current season in week 2. We then incremented the process by 20% each week until week 6, when only the current season's statistics were used.

We used one final preprocessing step before applying the ANNs to the data. We averaged the statistics between the two units that are on the field at the same time. For example, if team A scores 20 points per game and team B gives up 10 points per game then we average the two numbers to get an estimate that team A will score 15 points in the game. This number is then used as an input to the ANNs. We repeat this process for every statistic except for record and home-field advantage. We do not include home-field advantage as an explicit variable, rather this is accounted for by always subtracting the home team's win percentage from that of the road team. This approach has shown a home-field advantage of approximately three points. In regular-season games from 2007 through 2009, home teams had an average scoring margin of +2.5 points, so our estimate seems reasonable. This gives us 11 total inputs into the ANNs for each game.

3.2 ANN Structure

We used feed-forward ANNs in this project. Each ANN had four nodes in the first layer, and six nodes in the second layer. Between the input layer and the first layer and between the first and second layers we used hyperbolic tangent sigmoid transfer functions. Between the second layer and the output layer we used a linear transfer function. We randomly divided each data set into training, testing and validation data sets. We used mean square error (MSE) as the goal function and used the Levenberg-Marquardt optimization routine to train the networks.

4 Committees and Committees of Committees

Our basic ANNs, while useful, still had a good deal of variability in predictive ability, e.g. two networks with the same structure trained on different data sets may make substantially different predictions in future seasons. In order to create an algorithm that will give more consistent and robust results, we took an approach generally referred to as a committee of machines (CoM). For a survey of methods for combining networks into committees, see Sharkey (1999). In this approach many networks are trained against different random partitionings of the data set. Then based on the mean square error against the testing data, the top models are chosen to use predictively. These models are then combined using a measure of central tendency (i.e. mean, median, etc.) to make a final prediction. In order to understand the performance of our models, we have compared them to the predictions listed on `thepredictiontracker.com`, a website that tracks the results of approximately 60 computer-based NFL prediction systems, including the consensus "line" of the Las Vegas bookmakers.

We want an approach that will consistently give accurate results, not an approach that occasionally is the best, but generally performs poorly. Even committees with a hundred members showed some variation in their performance. To get the most robust approach to NFL game prediction, we used a committee of committees approach where many committees' predictions are combined to form our final prediction. Using this approach with 500 ANNs in the training stage, the best 100 were used in each committee, then 50 such committees were used to achieve the results below. The mean was used to form each committee vote and to combine the committees' predictions.

Box and whisker plots show that relative to the other prediction algorithms at `thepredictiontracker.com` this methodology is a consistently good way to predict the outcome of NFL games. On each figure the median of the prediction algorithms is indicated by the bar, with the upper and lower quartiles being indicated by the box. In Figure 1, the lower bar indicates the performance of the best model, while the upper bar indicates either the worst model or 1.5 times the interquartile range plus the upper quartile, whichever is less. In Figure 2, the approach is similar, but the best models are at the top of the graph, predicting a higher percentage of games correctly. The results for 2008-2010 are shown in the following figures. Figure 1 shows that we are generally in the best quartile of predictors, in terms of

DOI: 10.2202/1559-0410.1327

average error. Figure 2 indicates the performance at picking winners, with an ‘x’ to indicate the win percentage of the Las Vegas oddsmakers according to Repole (2010), which is relatively comparable to our approach over the last three years.

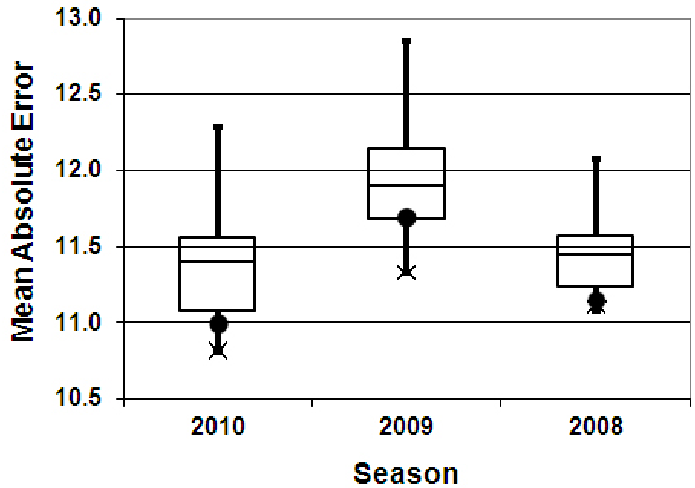

Figure 1: Absolute Error Performance in 2008-2010

5 More Analysis

With a model that has shown reasonable predictive capability, we applied additional analysis techniques to further understand the relationship between an NFL team’s statistical performance and their likelihood of winning.

The first technique we considered was principal component analysis (PCA). This technique is unsupervised, meaning that it does not use the game results and only uses the input statistics. It attempts to determine the most important ways in which games can look statistically different. It looks for combinations of data which describe the maximum variability in the data as described in Hand, Mannila, and Smyth (2001). For example, if all teams had the same rushing yardage, but a wide range of passing offenses and defenses, this technique would reveal this and

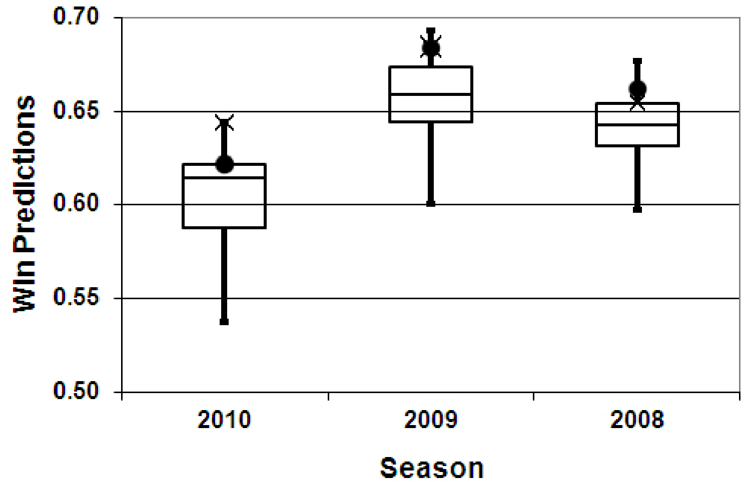

Figure 2: Win Percentage Performance in 2008-2010

quantify this variability. Figure 3 summarizes the information gained from the PCA analysis. When we analyzed the season-to-date statistical comparison between two teams most of the variability in how two teams compare is in how many yards we expected them to gain either on the ground or in the air. 99.7% of the variance in the statistics was the passing and rushing differentials. The interesting thing is that going into a game, we often expected a substantial difference in how (and even if) teams will move up and down the field, but we generally expected far less difference in how many points they would score and how often they would turn the ball over. As a note, the amount of variance described by turnovers was virtually negligible.

Another technique that was useful in understanding the model was a derivative-based technique. For small perturbations of each input, we calculated how much the output prediction changed. A similar methodology is used to analyze the most important factors of a deterministic model of HIV dynamics in David, Tran, and Banks (2009). Figure 4 gives a visual cartoon of how this technique works in the context of an ANN.

DOI: 10.2202/1559-0410.1327

Amount each column matters

Amount each statistic matters for the column

35%	32%	23%	10%	
0.03	-0.8	0.06	0.08	Road Scoring
-0.25	-0.13	0.64	0.71	Road Passing Yards
0.57	-0.80	-0.10	0.13	Road Rushing
0.06	0.06	-0.05	0.07	Home Scoring
-0.08	0.14	-0.72	0.66	Home Passing Yards
0.78	0.56	0.20	0.19	Home Rushing Yards

Figure 3: Summary of PCA analysis of 2007-2009 season to date statistics.

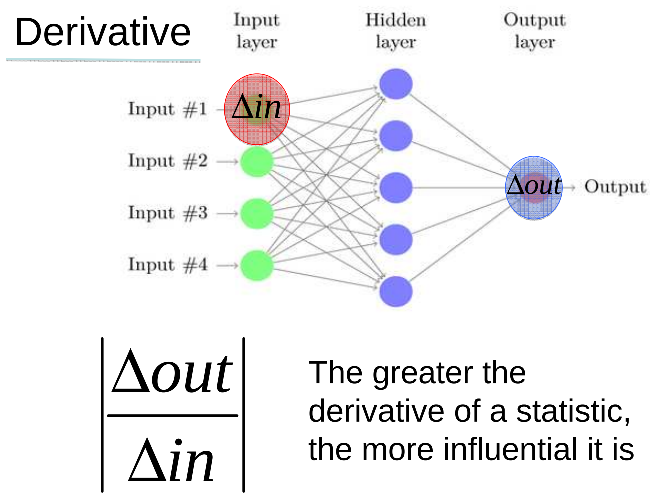

Figure 4: Summary of derivative based analysis technique.

Analyzing a model in this fashion is common, however there were several challenges with respect to using this technique on our CoM of CoMs approach. First, as ANNs are nonlinear functions, their derivative will depend on the input value, e.g. the derivative with respect to a given parameter may be different in a close game than in a blowout. Second, our full model has 100 CoMs, each with 50 ANNs. In order to get a sense of the relationship between input and output in this model we looked at the derivative of our 5000 ANNs when two teams are evenly matched, *i.e.* all stats are equal and at league average levels, except one team does have home field advantage. We then normalized the derivative by the output value, which is generally referred to as semi-normalization. We were unable to normalize by the input value as the win percentage differential being equal gives us a value of zero and such a normalization would require us to divide by zero. We then calculated the total derivative information, by summing the absolute value of all derivative values over all networks, and then calculating how much each input contributed to that value. Table 1 has a summary of these values. Predicted points account for 44.5% of the derivative information, making it by far the most influential input in our model. This intuitively makes a lot of sense; if we are trying to predict a point differential in a game, then a lot of information is gained by knowing how many points we expect each team to score. One other interesting aspect of this chart is the fact that the derivative with respect to team record is zero. This comes from the fact that the difference in win percentage takes on discrete input values and the difference increment used to compute the derivative and capture the local behavior does not pass that first discrete record differential value. The effect of record differential can be seen in the analysis where we vary the input over a range of plus and minus 50% of the league average values.

One other thing we wanted to look at with respect to the derivative analysis is whether the ANNs we are building are relatively consistent in the importance they assign to each input variable. For example, if the relative importance of the expected scoring value was driven by just a few networks with extremely large derivatives with respect to this input, we would have to question the robustness of our model's predictive capability. To analyze this, we looked at the percentage of time each parameter was ranked first through eleventh. This methodology supported the idea that the same factors were generally given similar importance across all 5000 models. Road team scoring was first or second 81% of the time and home scoring 71% of the time. By comparison, a variable such as road team interceptions taken, which accounts for 6% of the derivative information is only first or second 3% of the time and is usually one of the less important factors for a given ANN. This makes intuitive sense as to its importance in prediction.

DOI: 10.2202/1559-0410.1327

Table 1: Summary of derivative based analysis results in percent of total derivative information.

	Percent
Home Scoring	24.5
Road Scoring	20.0
Home Rushing Yards	12.9
Road Rushing Yards	10.7
Road Passing Yards	8.5
Home Passing Yards	6.9
Road Int. Taken	6.1
Home Int. Taken	5.4
Road Fumbles Taken	2.6
Home Fumbles Taken	2.4
Record	0

One final way we chose to analyze this model is to look at how the prediction changes over a range of values if all other factors are even for both teams and at league average levels. We have looked at levels plus or minus 50% of the league average. As before, expected scoring is the biggest indicator and changes the output the most. Other factors have relative expected levels of change, with home teams' rushing performance having the largest range of predicted output values, besides scoring. This could potentially be related to the idea that a home team that does not run the ball well will have trouble protecting a lead and will be more prone to upsets. Table 2 has a summary of these values.

6 Round Robin Power Rankings

While far less important than in major college football, where a hybrid ranking generally determines which teams play for the unofficial championship, there is still interest in a league-wide ranking of NFL teams. Most rankings of this sort are determined by expert voting, for example, `www.espn.com` uses a poll involving four of their staff members at espn.com (2009). There are many ways to

Table 2: Summary predicted scoring margin based on varying input values plus and minus 50% of the league average.

	50% below avg.	50% above avg.
Home Scoring	-11.6	6.1
Road Scoring	-10.2	6.5
Home Rushing Yards	-6.6	2.3
Road Rushing Yards	-3.3	-1.6
Road Passing Yards	-5.2	-0.5
Home Passing Yards	-4.9	-0.4
Road Int. Taken	-4.4	-0.6
Home Int. Taken	-3.3	-1.3
Road Fumbles Taken	-4.0	-1.9
Home Fumbles Taken	-3.6	-2.3
Record	-4.9	-1.4

rank teams; predictive models attempt to forecast future games, while retrodictive models are designed to accurately reflect past results, as discussed in Pasteur (2010). One significant challenge in ranking football teams is that a season constitutes an *incomplete tournament*, meaning that not all teams play one another. In an NFL regular season, a team plays two games against each of the other three teams in their division, but then plays only ten of the other 28 teams during the regular season. Also, due to the small number of games, record and performance data are often "noisy", filled with statistical oddities that are possible with small sample sizes `http://www.advancednflstats.com/2010/11/randomness-of-win-loss-records.html`.

One of the advantages of having a model is that each week we can play virtual games between each pair of teams, one at each team's home site. In order to test this methodology, we have performed such a simulation using the final 2009 statistics. We did this twice, using different assumptions to perform this analysis. Our first assumption is that the favored team would win each time. Under this assumption, we find that six teams win over 80% of their games in our simulation, led by the New York Jets, with a record of 56-6. Using the actual 2010 schedules, we have nine teams with predicted records of 12-4 or better, twice as many as the league average over the last few seasons.

DOI: 10.2202/1559-0410.1327

Our second assumption corresponds to the idea that a team favored by a half-point in all 16 games should hardly be expected to go 16-0. In fact a record closer to 8-8 should probably be expected. Based on historical data, a very accurate model (the consensus Las Vegas oddsmakers' line as seen in Repole (2010)) has a mean absolute error of about 11 points per game, over several seasons. Considering each game as a normal random variable, with standard deviation 11 points, a team favored by a half-point has only a 51.8% chance of winning the game. Under 16 identical trials, we would expect a team in this situation to win only 8.29 games. Under this assumption, we can calculate the expected 2010 win total for each team, and find the New Orleans Saints to be the leader, with 10.5 expected wins. Our previous favorite, the Jets, are fifth-best, at 9.77 expected wins.

However, the second model allows for fractional wins and losses. Even if every game were decided by a coin flip (i.e. each team is equally likely to win), we would statistically expect one of the 32 teams to finish with a record of 12-4 or better. Given that all teams are not equal, this model clearly overestimates the parity of the NFL.

To rank teams in a sensible manner, at any point during the season, we simulate a double-round-robin tournament (in which each team plays all of the others both at home and away) each week, based on the updated season-to-date statistics. We compute each team's rating as its expected average point differential (i.e. aggregate net points divided by games played) and then rank the teams in descending order. In practice, we add 100 to the ratings, to avoid causing confusion with negative team ratings. Using these ratings, we can obtain a quick prediction for a neutral-site game between any pair of teams by subtracting their ratings. To account for home-field advantage, we calculate the average scoring margin of all home teams in our simulation, and use this number as a universal home-field advantage constant, to be added to the home team's rating. Using this simplified system generally gives a predicted result within 1-2 points of that obtained from the full model.

7 Summary and Future Work

In this paper we have described a basic methodology for building an ANN to predict point differential outcomes in NFL games based on a relatively simple set of past data. We have described a Committee of Committee approach of using these

ANNs to make a single prediction for future games and presented the performance in seasons that were not used to train the model. The results are comparable to the Las Vegas line and better than most of the predictive models presented on `thepredictiontracker.com`. We have presented unsupervised and supervised ways to analyze our model, including principal component analysis, derivative based, and parameter sweep methods. This model allows us to perform experiments that cannot be done in a real season, including weekly round robin tournaments and subsequent rankings.

There are several avenues of experimentation we would like to pursue with this research, including the following:

- Use other types of ANN architectures, e.g. radial basis functions
- Include additional statistics, such as time of possession, strength-of-schedule, kicking game measures, and injuries
- Look at how to best predict early season games, before season-to-date averages have necessarily become a good indicator of a team's ability
- Try this on other levels of football (e.g. NCAA) and with other sports

References

Blasco, J., N. Rueyo, J. Larroya, C. Dopazo, and Y. Chen (1999): "A single-step time-integrator of a methane-air chemical system using artificial neural networks," *Computers and Chemical Engineering*, 23, 1127–1133.

Dauber, H., N. Israel, and R. Taylor (2010): "Modeling the tan δ dynamic mechanical spectrum," Technical report, The College of Wooster.

David, J. A., H. T. Tran, and H. T. Banks (2009): "HIV model analysis and estimation implementation under optimal control based treatment strategies," *International Journal of Pure and Applied Mathematics*, 57, 357–392.

espn.com (2009): "2009 NFL power rankings: Week 18," `http://espn.go.com/nfl/powerrankings/_/year/2009/week/18`.

Hand, D., H. Mannila, and P. Smyth (2001): *Principles of Data Mining*, MIT Press.

Kahn, J. (2003): "Neural network prediction of NFL football games," World Wide Web electronic publication, URL `homepages.cae.wisc.edu/~ece539/project/f03/kahn.pdf`, .

Loeffelholz, B., E. Bednar, and K. W. Bauer (2009): "Predicting NBA games using neural networks," *Journal of Quantitative Analysis in Sports*, 5.

Pasteur, R. D. (2010): *Mathematics and Sports*, MAA, chapter When Perfect Isn't Good Enough: Retrodictive Rankings in College Football.

Purucker, M. (1996): "Neural network quarterbacking," *IEEE Potentials*.

DOI: 10.2202/1559-0410.1327

Repole, W. (2010): “Historical NFL scores and lines,” http://www.repole.com/sun4cast/data.html

Sharkey, A. J. (1999): *Combining Artificial Neural Nets: Ensemble and Modular Multi-Net Systems*, Springer-Verlag.

Shenvi, N., J. Geremia, and H. Rabitz (2004): “Efficient chemical kinetic modeling through neural network maps,” *Journal of Chemical Physics*, 120, 9942–9951.

Wikipedia (2010): “Aritficial neural network,” World Wide Web electronic publication, URL http://en.wikipedia.org/wiki/Artificialneuralnetwork.