Predicting the Average Two-Year Win Probability and Hire Tenure of NFL Head Coach Hires: Three Approaches

Jon C. Williamson
Dept. of Computer Science & Engineering
Texas A&M University
College Station, TX, USA
jonwilliamson@tamu.edu

I. Introduction

In the NFL, successful head coach hiring is tremendously valuable, as a differentiated head coach can bring about lasting success and divisional dominance, subsequently increasing the historical importance of a franchise and improving its top line. A machine learning model that can increase the hit probability of head coaching hires has the potential to add immense value to NFL franchises.

II. Methods

Using statistics available at the time of hiring, this project attempts to predict two outcomes of head coach hires: the average two-year winning percent and the hire tenure, using three machine learning approaches: regularized linear models, XGBoost models, and Multi-layer perceptron models. This project uses root mean squared error and macro-averaged one-versus-rest area under the receiver operating characteristic curve to measure model performance for the regressors and classifiers, respectively. The tenure of a coach hire is defined as the number of years the hired coach remains in the same position before being fired, leaving, or retiring. Equation (1) shows the mapping between the coach tenure, t, and the four coach tenure classification labels, $C(t)$.

$$C(t) = \begin{cases} 0 & t \leq 2 \\ 1 & 2 < t \leq 4 \\ 2 & 4 < t \leq 7 \\ 3 & t > 7 \end{cases} \tag{1}$$

Table I shows the 25 features used. Features 1-18 are characteristics of head coaches at time of hiring, while features 19-25 are characteristics of the hiring team. Features 10-18 and 20-23 reference average normalized team ranks in different categories. This rank normalization also allows coaches across eras to be compared, as performance is purely comparative to other teams in the same era.

Raw data was collected by scraping pro-football-reference.com. The crawling script extracted three performance tables for all head coaches in the database and two performance tables for all franchises in the database. All data is mean-imputed prior to being fed into any model.

TABLE I
MODEL FEATURE LIST

No.	Description
1	Age at hiring
2	Number of times previously hired as head coach
3	Number of years' experience as college position coach
4	Number of years' experience as college coordinator
5	Number of years' experience as college head coach
6	Number of years' experience as NFL position coach
7	Number of years' experience as NFL coordinator
8	Number of years' experience as NFL head coach
9	Demotion presence in hiring history
10	During years as NFL OC, team's avg. norm. yardage rank
11	During years as NFL OC, team's avg. norm. point rank
12	During years as NFL OC, team's avg. norm. giveaway rank
13	During years as NFL DC, team's avg. norm. yardage rank
14	During years as NFL DC, team's avg. norm. point rank
15	During years as NFL DC, team's avg. norm. turnover rank
16	During years as NFL HC, team's avg. norm. yardage differential rank
17	During years as NFL HC, team's avg. norm. point differential rank
18	During years as NFL HC, team's avg. norm. turnover ratio rank
19	Hiring team's avg. winning percent in previous two years
20	Hiring team's avg. norm. turnover ratio rank in previous two years
21	Hiring team's avg. norm. point differential rank in previous two years
22	Hiring team's avg. norm. yard differential rank in previous two years
23	Hiring team's avg. norm. divisional placement in previous two years
24	Hiring team's number of playoff appearances in previous two years
25	Hiring team's number of playoff wins in previous two years

III. Results

Table II compares the results of the three implementations for both models. All regression models showed poor RMSE performance when compared to predicting the expected value. All classification models showed better performance when compared to predicting the expected value.

TABLE II
PREDICTION RESULT COMPARISON

Winning Probability Data Set	**RMSE**		
	Reg. Lin.	**XGBR**	**MLPR**
Train	0.192	0.171	0.189
Test	0.199	0.200	0.206
Validation	0.222	0.227	0.224
Validation, Expected Outcome[1]	0.233		

Tenure Classification Data Set	**OVR AUROC**		
	Reg. Log.	**XGBC**	**MLPC**
Train	0.706	0.972	0.787
Test	0.620	0.671	0.638
Validation	0.593	0.669	0.621
Validation, Expected Outcome[1]	0.500		

[1]Not influenced by any model.

IV. CONCLUSIONS

The three winning probability prediction models showed poor performance when compared to predicting the expected value. The best RMSE value was 0.222, equivalent to predicting the number of won games in a 17 game season to within ± 3.77 wins. These findings suggest that the features in this project, largely driven by characteristics of the head coach, are not sufficient to predict a team's winning probability.

The coach tenure classification models showed significantly better performance than predicting the most prevalent class. These results suggest that the features in this project have some ability to predict the tenure of head coach hires. Additionally, the regularized logistic regression and the XGBoost classifier showed that characteristics of successful hiring teams were important in determining coaches with longer tenures, suggesting that successful franchises may be better at evaluating head coach candidates. Future iterations of these models could provide significant value to NFL franchises by increasing the likelihood of successful head coach hires.

REFERENCES

[1] Pedregosa et al., "Scikit-learn: machine learning in Python," in Journal of Machine Learning Research, vol. 12, pp.2825-2830.

[2] T. Chen, and C. Guestrin, "XGBoost: a scalable tree boosting system," in KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.785-794.