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

Football

193982

1. **Introduction**

What if the Denver Broncos could have avoided hiring Nathaniel Hackett? Or if the Raiders could have avoided hiring Josh McDaniels? Certainly, these teams would be in a different position today if they had hired different candidates. But more broadly, what if NFL teams could predict which head coach candidates would be successful? That is the aim of this project. Specifically, this project attempts to predict two outcomes of head coach hires: the average two-year winning probability and the hire tenure, using statistics available at the time the hire was made. This project also provides predictions of these metrics for recently hired coaches.

1. **Literature Review**

There have been no journal publications that attempt to predict the success of NFL coaching hires through statistical learning techniques. Currently, the NFL is only beginning to implement artificial intelligence (AI) in play calling prediction [1]. Additionally, there are few papers that examine the impact of individual features on NFL head coaching success. Reference [2] used a linear regression with seven features to attempt to predict the number of wins of head coaches in their first three years to understand if prior NFL head coaching experience impacts success in position. This paper found that a previous head coaching experience had a negative impact on the success of new head coaches. Despite this finding, the model supported an adjusted of only . This low value, the lack of regularization, and the small number of features decreases confidence in the study's findings. Reference [3] reviews research in sports economics and suggests that hiring decisions made solely on playing success are unlikely to be optimal given financial (resource) inequality among sports franchises.

1. **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 probability and the hire tenure, using three machine learning approaches: regularized linear models, XGBoost models, and Multi-layer perceptron models [4-5]. 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. Raw data was collected by scraping pro-football-reference.com.

* 1. **Predicting Coach Tenure Classification**

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* (in years), and the three coach tenure classification labels, *C(t)*. Different coach classifications are intended to indicate different levels of coaching success based on the number of years they maintain their position.

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

This model seeks to predict the coaching tenure classification of head coach hires based on statistics available at the time of hiring. This project utilizes three implementations of this model:

1. Logistic Regression with Lasso Regularization [4]
2. XGBoost Classifier [5]
3. Multi-layer Perceptron (MLP) Classifier [4]

This project utilizes extensive cross-validation to determine hyperparameter values for each implementation. This project uses macro-averaged one-versus-rest (OVR) area under the receiver operating characteristic curve (AUROC) to measure model performance. This performance metric accounts for class imbalance.

* 1. **Predicting Average Two-Year Winning Probability**

Equation (2) defines the calculation of average winning probability, , as a function of the number of wins, , the number of losses, , and the number of ties, , of a head coach over any interval.

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

This winning probability is bound within . This project implements three regressors to attempt this prediction:

1. Linear Regression with Lasso Regularization [4]
2. XGBoost Regressor [5]
3. Multi-layer Perceptron (MLP) Regressor [4]

The first implementation of this model is a simple linear model with norm regularization. This project uses regularization due to its tendency to remove features from the model. The second implementation of the model is an XGBoost regressor. This regressor uses gradient boosting to build a single predication model through the aggregation of weak learners. This project uses trees as the universal model for the XGBoost weak learners. The third and final implementation of this model is through a multi-layer perceptron regressor. This regressor is a basic neural network with a final regression node. It extracts features without supervision. This project utilizes extensive cross-validation to determine the optimal values for hyperparameters for each model implementation.

This project uses root mean squared error (RMSE) as the evaluation metric for these regression models. This metric was chosen for two reasons. Firstly, its dimensions are identical to the dimensions of the prediction variable. Secondly, it punishes outliers greater than absolute error. It is important to note that the thresholds that constitute “good” and “bad” RMSE are impacted by the scale of the prediction. As a result, this project compares model RMSE against the RMSE for predicting the expected outcome to understand model performance.

* 1. **Data Description**

This project utilizes 150 features, two description labels, and the two model outputs for each head coaching hire. Appendix A shows the 150 features used. Abbreviations included in feature descriptions include offensive coordinator (OC), defensive coordinator (DC), and head coach (HC). Features 1-140 are characteristics of head coaches at time of hiring, while features 141-150 are characteristics of the hiring team. Features 9-140 and 141-150 reference average normalized metrics, utilizing a traditional z-score distance from league average. This normalization allows coaches across eras to be compared, as team performance is measured relative to other teams in the same year. Figure 1 shows the correlation matrix among all 150 features. There is some correlation within the sections of features associated with experience as an OC, DC, or HC. Appendix B shows the distribution of two-year win probabilities and the distribution of coach tenure classifications across all hiring instances in the history of the NFL.

A blue and yellow grid with a black border

Description automatically generated with medium confidence

Figure 1. Feature correlation matrix

1. **Results**

For all models and all implementations, the data was split into training and validation sets via an 80/20 ratio. Each model was created using multiple levels of internal cross-validation in order to tune hyperparameters (final hyperparameter values are shown in Appendix C). This section reports performance metrics on the training set, the testing set, and the validation set. The testing set is the set of data set aside within internal cross-validation during hyperparameter tuning. The validation set is 20% of the entire data set that was not used in any portion of training. Final claims of model performance are made on performance with the validation set using a single model with tuned hyperparameters.

* 1. **Predicting Coach Tenure Classification**
     1. **Model Performance Comparison**

Table 1 shows the macro-averaged one-versus-rest (OVR) area under the receiver operating characteristic curve (AUROC) values for the three implementations across the train, test, and validation sets.

Table 1. Coach tenure classification prediction result comparison

A close-up of a number

Description automatically generated

This table shows that the XGBoost classifier had the best performance on the validation set. All models showed significantly better performance when compared to predicting the expected value. These findings suggest that these features, largely driven by characteristics of the head coach, do have value in predicting the tenure of head coaches. In other words, these models have the potential to help NFL teams find candidates that are predicted to remain a head coach longer using information available at the time of hire.

* + 1. **Best model: XGBoost Classifier**

This subsection provides an overview of the best classifying model. For details on the other models, see Appendix D. The XGBoost classifier had an OVR AUROC value of , indicating that the model has predictive utility. Fig. 2 shows the sorted validation set with corresponding marks for the ground truth values and the predicted values.

A screenshot of a computer

Description automatically generated

Figure 2. Validation set, XGBoost classifier prediction versus ground truth

Fig. 3 shows the feature weight distributions resulting from the best models found within the outer ten-fold cross-validation. These feature importances do not infer a monotonic relationship between feature value and predicted value. Rather, these importance values result from feature prevalence in the model's weak estimators. A feature with higher importance is present in more estimators than a feature with low importance.

A black and white image of lines and dots

Description automatically generated

Figure 3. XGBoost classifier feature importance distributions

The features with the highest average importance are shown below in Table 2. 10 of the top 20 most important features are attributes of the coach’s team during a previous head coaching role—suggesting that past performance is a predictor of future performance. One notable finding is that the number of years as an NFL coordinator is less important than the number of years of experience as an NFL position coach. Additionally, three of the top 20 features are attributes of the coach’s team during their role as a defensive coordinator, but none of the top 20 features are attributes of the coach’s team during their role as an offensive coordinator.

Table 2. XGBoost Classifier feature importance for top 20 most important features

|  |  |  |  |
| --- | --- | --- | --- |
| Rank | Feature No. | Feature Description | Avg. Importance |
| 1 | 6 | Number of years' experience as NFL position coach | 0.0308 |
| 2 | 84 | During years as NFL HC, team's average passing interceptions | 0.0212 |
| 3 | 1 | Age at hiring | 0.0200 |
| 4 | 124 | During years as NFL HC, opponent team's average rushing 1st downs | 0.0188 |
| 5 | 85 | During years as NFL HC, team's average NY/A | 0.0180 |
| 6 | 79 | During years as NFL HC, team's average 1st downs | 0.0179 |
| 7 | 108 | During years as NFL HC, opponent team's average points scored | 0.0175 |
| 8 | 112 | During years as NFL HC, opponent team's average 1st downs | 0.0156 |
| 9 | 50 | During years as NFL DC, opponent team's average passing touchdowns | 0.0155 |
| 10 | 53 | During years as NFL DC, opponent team's average passing first downs | 0.0148 |
| 11 | 148 | Hiring team's average turnovers forced in previous two years | 0.0148 |
| 12 | 49 | During years as NFL DC, opponent team's average passing yards | 0.0143 |
| 13 | 7 | Number of years' experience as NFL coordinator | 0.0142 |
| 14 | 116 | During years as NFL HC, opponent team's average passing touchdowns | 0.0141 |
| 15 | 47 | During years as NFL DC, opponent team's average passing completions | 0.0140 |
| 16 | 60 | During years as NFL DC, opponent team's average penalty yards | 0.0137 |
| 17 | 123 | During years as NFL HC, opponent team's average rush yards per play | 0.0136 |
| 18 | 78 | During years as NFL HC, team's average turnovers | 0.0135 |
| 19 | 119 | During years as NFL HC, opponent team's average passing first downs | 0.0135 |
| 20 | 146 | Hiring team's average yards / play in previous two years | 0.0134 |

* + 1. **Predicting the tenure of recent head coach hires**

Including Josh McDaniels, who was only fired in the last week, there were 13 head coaches hired in the NFL in the past two offseasons who are still with their team. Table 3 below shows the final model's predictions for coach tenure for these 13 hires. This table also shows the probabilities associated with each class prediction; these probabilities sum to 1, and the class with the greatest probability is the final predicted class.

Table 3. XGBoost Classifier coach tenure predictions for the 13 ongoing head coach hires made in the last two years

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  | **Class Probability** | | |
| **Coach Name** | **Hire Year** | **Predicted Class** | **Class 0** | **Class 1** | **Class 2** |
| Brian Daboll | 2022 | 0 | 35.9% | 31.3% | 32.8% |
| DeMeco Ryans | 2023 | 2 | 27.1% | 24.7% | 48.3% |
| Dennis Allen | 2022 | 0 | 53.3% | 26.0% | 20.7% |
| Doug Pederson | 2022 | 1 | 28.6% | 53.9% | 17.5% |
| Frank Reich | 2023 | 1 | 30.7% | 49.2% | 20.1% |
| Jonathan Gannon | 2023 | 1 | 31.4% | 35.5% | 33.1% |
| Josh McDaniels | 2022 | 0 | 38.8% | 32.4% | 28.8% |
| Kevin O'Connell | 2022 | 2 | 31.8% | 33.0% | 35.2% |
| Matt Eberflus | 2022 | 2 | 26.0% | 28.4% | 45.6% |
| Mike McDaniel | 2022 | 0 | 34.7% | 31.6% | 33.6% |
| Sean Payton | 2023 | 1 | 29.3% | 49.6% | 21.1% |
| Shane Steichen | 2023 | 0 | 39.2% | 29.2% | 31.6% |
| Todd Bowles | 2022 | 1 | 35.5% | 36.3% | 28.2% |

Many of these predictions seem to track with the coaches’ trajectories. The clearest example is Josh McDaniels, who was fired last week, ending his tenure at less than 2 years (Class 0). The model predicted Class 0 for Josh McDaniels. The model predicted 3 of the hires to be Class 2: DeMeco Ryans, Kevin O’Connell, and Matt Eberflus. Although Eberflus seems to be trending towards being fired this year, DeMeco Ryans and Kevin O’Connell have both had strong starts to their careers. Also note that Miami Dolphins coach Mike McDaniel was within one percentage point of also being categorized as Class 2.

* 1. **Predicting Average Two-Year Winning Probability**
     1. **Model Performance Comparison**

Table 4 shows the RMSE values for the three implementations across the train, test, and validation sets.

Table 4. Winning probability prediction result comparison

A white sheet with black text

Description automatically generated

This table shows that the XGBoost regressor had the best performance on the validation set. However, all models showed poor RMSE performance when compared to predicting the expected value. These findings suggest that these features, largely driven by characteristics of the head coach, are not sufficient to predict a team's winning probability. In other words, it appears that a head coach hire will not determine win probability based on these features.

* + 1. **Best model: XGBoost Regressor**

This subsection provides an overview of the best regression model. For details on the other models, see Appendix D. The XGBoost regressor had a RMSE value of , indicating that the model has poor predictive utility. Fig. 4 shows the sorted validation set with corresponding marks for the ground truth values and the predicted values.

A graph of a graph

Description automatically generated with medium confidence

Figure 4. Validation set, XGBoost regressor prediction versus ground truth

Fig. 5 shows the feature weight distributions resulting from the best models found within the outer ten-fold cross-validation.

A black and white image of lines and dots

Description automatically generated

Figure 5. XGBoost regressor feature importance distributions

The features with the highest average importance are shown below in Table 5. The most important predictor of a head coach’s two-year winning percentage is the previous two-year winning percentage of the hiring team. This feature is more than 20% more important than the 2nd ranked feature. Moreover, of the top 20 features, 8 are attributes of the hiring team. These findings suggest that the hiring team itself, not the head coach, drive win probability. Equivalently, the features of the coach themselves, including whether or not they have been a successful coordinator in the past, are not sufficient to predict two-year winning percentage.

Table 5. XGBoost regressor feature importance for top 20 most important features

|  |  |  |  |
| --- | --- | --- | --- |
| Rank | Feature No. | Feature Description | Avg. Importance |
| 1 | 141 | Hiring team's average winning percent in previous two years | 0.0472 |
| 2 | 7 | Number of years' experience as NFL coordinator | 0.0372 |
| 3 | 142 | Hiring team's average points scored in previous two years | 0.0369 |
| 4 | 145 | Hiring team's average yards of offense allowed in previous two years | 0.0343 |
| 5 | 146 | Hiring team's average yards / play in previous two years | 0.0340 |
| 6 | 56 | During years as NFL DC, opponent team's average rushing touchdowns | 0.0314 |
| 7 | 108 | During years as NFL HC, opponent team's average points scored | 0.0295 |
| 8 | 144 | Hiring team's average yards of offense in previous two years | 0.0283 |
| 9 | 143 | Hiring team's average points allowed in previous two years | 0.0230 |
| 10 | 147 | Hiring team's average yards / play allowed in previous two years | 0.0227 |
| 11 | 38 | During years as NFL OC, team's average number of 4th down attempts | 0.0217 |
| 12 | 43 | During years as NFL DC, opponent team's average yards | 0.0216 |
| 13 | 75 | During years as NFL HC, team's average points scored | 0.0215 |
| 14 | 149 | Hiring team's average turnovers in previous two years | 0.0191 |
| 15 | 84 | During years as NFL HC, team's average passing interceptions | 0.0186 |
| 16 | 116 | During years as NFL HC, opponent team's average passing touchdowns | 0.0184 |
| 17 | 21 | During years as NFL OC, team's average rushing attempts | 0.0168 |
| 18 | 42 | During years as NFL DC, opponent team's average points scored | 0.0167 |
| 19 | 10 | During years as NFL OC, team's average yards | 0.0165 |
| 20 | 5 | Number of years' experience as college head coach | 0.0147 |

* + 1. **Predicting the two-year win probability of recent head coach hires**

Table 6 below shows the model’s prediction for average two-year win probability for the 13 head coaches hired in the NFL in the past two offseasons who are still with their team.

Table 6. XGBoost regressor average two-year win probability predictions for the 13 ongoing head coach hires made in the last two years

|  |  |  |
| --- | --- | --- |
| **Coach Name** | **Hire Year** | **Predicted 2-Year Win Prob.** |
| Brian Daboll | 2022 | 0.422 |
| DeMeco Ryans | 2023 | 0.457 |
| Dennis Allen | 2022 | 0.378 |
| Doug Pederson | 2022 | 0.415 |
| Frank Reich | 2023 | 0.451 |
| Jonathan Gannon | 2023 | 0.454 |
| Josh McDaniels | 2022 | 0.410 |
| Kevin O'Connell | 2022 | 0.551 |
| Matt Eberflus | 2022 | 0.382 |
| Mike McDaniel | 2022 | 0.491 |
| Sean Payton | 2023 | 0.395 |
| Shane Steichen | 2023 | 0.408 |
| Todd Bowles | 2022 | 0.471 |

The model’s prediction for average two-year win probability among these coaches ranges from 0.382 (Matt Eberflus) to 0.551 (Kevin O’Connell). As shown in section 4.2.2., the model tends to predict winning probabilities near the average. This set of coaches is no exception to this tendency.

1. **Conclusion**

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. This finding poses immense potential value for NFL teams, as enabling the hiring of a differentiated NFL head coach could bring about lasting success and divisional dominance, subsequently increasing the historical importance of a franchise and improving its bottom line.

The three winning probability prediction models showed poor performance when compared to predicting the expected value. The best RMSE value was 0.192, equivalent to predicting the number of won games in a 17-game season to within 3.26 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. In other words, NFL teams will have to look at factors other than a coach’s career history (e.g., team personnel, coordinators, etc.) to drive their winning probability.

**References**

[1] “Using machine learning to peek inside the minds of NFL coaches,” DataRobot. https://www.datarobot.com/blog/using-machine-learning-topeek-inside-the-minds-of-nfl-coaches/ (accessed Oct. 26, 2020)

[2] M. Roach, “Does prior NFL head coaching experience improve team performance?,” in Journal of Sport Management, vol. 30, no. 3, pp.298-311.

[3] D. Mielke, “Coaching experience, playing experience, and coaching tenure: a commentary,” in International Journal of Sports Science & Coaching, vol. 2, no. 2, pp. 117-118.

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

[5] 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.

[6] M. Ribeiro, S. Singh, and C. Guestrin, ““Why should I trust you?”: Explaining the predictions of any classifier,” in KDD ’16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.1135-1144.

**Appendix A**

|  |  |
| --- | --- |
| Feature No. | Feature 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 | During years as NFL OC, team's average points scored |
| 10 | During years as NFL OC, team's average yards |
| 11 | During years as NFL OC, team's average yards/play |
| 12 | During years as NFL OC, team's average turnovers |
| 13 | During years as NFL OC, team's average 1st downs |
| 14 | During years as NFL OC, team's average passing completions |
| 15 | During years as NFL OC, team's average passing attempts |
| 16 | During years as NFL OC, team's average passing yards |
| 17 | During years as NFL OC, team's average passing touchdowns |
| 18 | During years as NFL OC, team's average passing interceptions |
| 19 | During years as NFL OC, team's average NY/A |
| 20 | During years as NFL OC, team's average passing first downs |
| 21 | During years as NFL OC, team's average rushing attempts |
| 22 | During years as NFL OC, team's average rushing yards |
| 23 | During years as NFL OC, team's average rushing touchdowns |
| 24 | During years as NFL OC, team's average rush yards per play |
| 25 | During years as NFL OC, team's average rushing 1st downs |
| 26 | During years as NFL OC, team's average number of penalties |
| 27 | During years as NFL OC, team's average penalty yards |
| 28 | During years as NFL OC, team's average penalty 1st downs |
| 29 | During years as NFL OC, team's average number of drives |
| 30 | During years as NFL OC, team's average scoring percentage |
| 31 | During years as NFL OC, team's average turnover percentage |
| 32 | During years as NFL OC, team's average drive duration |
| 33 | During years as NFL OC, team's average plays per drive |
| 34 | During years as NFL OC, team's average yards per drive |
| 35 | During years as NFL OC, team's average points per drive |
| 36 | During years as NFL OC, team's average number of 3rd down attempts |
| 37 | During years as NFL OC, team's average third down conversion percentage |
| 38 | During years as NFL OC, team's average number of 4th down attempts |
| Feature No. | Feature Description |
| 39 | During years as NFL OC, team's average 4th down conversion percentage |
| 40 | During years as NFL OC, team's average red zone attempts |
| 41 | During years as NFL OC, team's average red zone percentage |
| 42 | During years as NFL DC, opponent team's average points scored |
| 43 | During years as NFL DC, opponent team's average yards |
| 44 | During years as NFL DC, opponent team's average yards/play |
| 45 | During years as NFL DC, opponent team's average turnovers |
| 46 | During years as NFL DC, opponent team's average 1st downs |
| 47 | During years as NFL DC, opponent team's average passing completions |
| 48 | During years as NFL DC, opponent team's average passing attempts |
| 49 | During years as NFL DC, opponent team's average passing yards |
| 50 | During years as NFL DC, opponent team's average passing touchdowns |
| 51 | During years as NFL DC, opponent team's average passing interceptions |
| 52 | During years as NFL DC, opponent team's average NY/A |
| 53 | During years as NFL DC, opponent team's average passing first downs |
| 54 | During years as NFL DC, opponent team's average rushing attempts |
| 55 | During years as NFL DC, opponent team's average rushing yards |
| 56 | During years as NFL DC, opponent team's average rushing touchdowns |
| 57 | During years as NFL DC, opponent team's average rush yards per play |
| 58 | During years as NFL DC, opponent team's average rushing 1st downs |
| 59 | During years as NFL DC, opponent team's average number of penalties |
| 60 | During years as NFL DC, opponent team's average penalty yards |
| 61 | During years as NFL DC, opponent team's average penalty 1st downs |
| 62 | During years as NFL DC, opponent team's average number of drives |
| 63 | During years as NFL DC, opponent team's average scoring percentage |
| 64 | During years as NFL DC, opponent team's average turnover percentage |
| 65 | During years as NFL DC, opponent team's average drive duration |
| 66 | During years as NFL DC, opponent team's average plays per drive |
| 67 | During years as NFL DC, opponent team's average yards per drive |
| 68 | During years as NFL DC, opponent team's average points per drive |
| 69 | During years as NFL DC, opponent team's average number of 3rd down attempts |
| 70 | During years as NFL DC, opponent team's average third down conversion pct. |
| 71 | During years as NFL DC, opponent team's average number of 4th down attempts |
| 72 | During years as NFL DC, opponent team's average 4th down conversion pct. |
| 73 | During years as NFL DC, opponent team's average red zone attempts |
| 74 | During years as NFL DC, opponent team's average red zone percentage |
| 74 | During years as NFL DC, opponent team's average red zone percentage |
| 75 | During years as NFL HC, team's average points scored |
| 76 | During years as NFL HC, team's average yards |
| Feature No. | Feature Description |
| 77 | During years as NFL HC, team's average yards/play |
| 78 | During years as NFL HC, team's average turnovers |
| 79 | During years as NFL HC, team's average 1st downs |
| 80 | During years as NFL HC, team's average passing completions |
| 81 | During years as NFL HC, team's average passing attempts |
| 82 | During years as NFL HC, team's average passing yards |
| 83 | During years as NFL HC, team's average passing touchdowns |
| 84 | During years as NFL HC, team's average passing interceptions |
| 85 | During years as NFL HC, team's average NY/A |
| 86 | During years as NFL HC, team's average passing first downs |
| 87 | During years as NFL HC, team's average rushing attempts |
| 88 | During years as NFL HC, team's average rushing yards |
| 89 | During years as NFL HC, team's average rushing touchdowns |
| 90 | During years as NFL HC, team's average rush yards per play |
| 91 | During years as NFL HC, team's average rushing 1st downs |
| 92 | During years as NFL HC, team's average number of penalties |
| 93 | During years as NFL HC, team's average penalty yards |
| 94 | During years as NFL HC, team's average penalty 1st downs |
| 95 | During years as NFL HC, team's average number of drives |
| 96 | During years as NFL HC, team's average scoring percentage |
| 97 | During years as NFL HC, team's average turnover percentage |
| 98 | During years as NFL HC, team's average drive duration |
| 99 | During years as NFL HC, team's average plays per drive |
| 100 | During years as NFL HC, team's average yards per drive |
| 101 | During years as NFL HC, team's average points per drive |
| 102 | During years as NFL HC, team's average number of 3rd down attemps |
| 103 | During years as NFL HC, team's average third down conversion percentage |
| 104 | During years as NFL HC, team's average number of 4th down attemps |
| 105 | During years as NFL HC, team's average 4th down conversion percentage |
| 106 | During years as NFL HC, team's average red zone attempts |
| 107 | During years as NFL HC, team's average red zone percentage |
| 108 | During years as NFL HC, opponent team's average points scored |
| 109 | During years as NFL HC, opponent team's average yards |
| 110 | During years as NFL HC, opponent team's average yards/play |
| 111 | During years as NFL HC, opponent team's average turnovers |
| 112 | During years as NFL HC, opponent team's average 1st downs |
| 113 | During years as NFL HC, opponent team's average passing completions |
| 114 | During years as NFL HC, opponent team's average passing attempts |
| 115 | During years as NFL HC, opponent team's average passing yards |
| Feature No. | Feature Description |
| 116 | During years as NFL HC, opponent team's average passing touchdowns |
| 117 | During years as NFL HC, opponent team's average passing interceptions |
| 118 | During years as NFL HC, opponent team's average NY/A |
| 119 | During years as NFL HC, opponent team's average passing first downs |
| 120 | During years as NFL HC, opponent team's average rushing attempts |
| 121 | During years as NFL HC, opponent team's average rushing yards |
| 122 | During years as NFL HC, opponent team's average rushing touchdowns |
| 123 | During years as NFL HC, opponent team's average rush yards per play |
| 124 | During years as NFL HC, opponent team's average rushing 1st downs |
| 125 | During years as NFL HC, opponent team's average number of penalties |
| 126 | During years as NFL HC, opponent team's average penalty yards |
| 127 | During years as NFL HC, opponent team's average penalty 1st downs |
| 128 | During years as NFL HC, opponent team's average number of drives |
| 129 | During years as NFL HC, opponent team's average scoring percentage |
| 130 | During years as NFL HC, opponent team's average turnover percentage |
| 131 | During years as NFL HC, opponent team's average drive duration |
| 132 | During years as NFL HC, opponent team's average plays per drive |
| 133 | During years as NFL HC, opponent team's average yards per drive |
| 134 | During years as NFL HC, opponent team's average points per drive |
| 135 | During years as NFL HC, opponent team's average number of 3rd down attempts |
| 136 | During years as NFL HC, opponent team's average third down conversion pct. |
| 137 | During years as NFL HC, opponent team's average number of 4th down attempts |
| 138 | During years as NFL HC, opponent team's average 4th down conversion pct. |
| 139 | During years as NFL HC, opponent team's average red zone attempts |
| 140 | During years as NFL HC, opponent team's average red zone percentage |
| 141 | Hiring team's average winning percent in previous two years |
| 142 | Hiring team's average points scored in previous two years |
| 143 | Hiring team's average points allowed in previous two years |
| 144 | Hiring team's average yards of offense in previous two years |
| 145 | Hiring team's average yards of offense allowed in previous two years |
| 146 | Hiring team's average yards / play in previous two years |
| 147 | Hiring team's average yards / play allowed in previous two years |
| 148 | Hiring team's average turnovers forced in previous two years |
| 149 | Hiring team's average turnovers in previous two years |
| 150 | Hiring team's number of playoff appearances in previous two years |

**Appendix B**

A blue and black graph

Description automatically generated

Figure B1. Two-year winning probability frequency distribution

A blue and white graph

Description automatically generated

Figure B2. Coach tenure classification frequency distribution

**Appendix C**

Table C1. Final hyperparameters for the logistic regression with lasso regularization model

A black text on a white background

Description automatically generated

Table C2. Final hyperparameters for the XGBoost classifier model

A table of measurements with black text

Description automatically generated

Table C3. Final hyperparameters for the multi-layer perceptron classifier model

A table with black text

Description automatically generated

Table C4. Final hyperparameters for the regularized linear regression model

A close-up of a sign

Description automatically generated

Table C5. Final hyperparameters for the XGBoost regression model

A table of measurements with black text

Description automatically generated

Table C6. Final hyperparameters for the multi-layer perceptron regression model

A table with text and numbers

Description automatically generated

**Appendix D**

**D.1. Additional two-year win probability prediction models**

**D.1.1.** **Linear Regression with Lasso Regularization**

A line of dots with black dots

Description automatically generated with medium confidence

Figure D1. Validation set, regularized linear model prediction versus ground truth

A white background with black lines and dots

Description automatically generated

Figure D2. Regularized linear model feature importance distributions

**D.1.2. Multi-layer Perceptron Regressor**

A line of dots with a white background

Description automatically generated

Figure D3. Validation set, MLP regressor prediction versus ground truth

A screenshot of a computer screen

Description automatically generated

Figure D4. MLP regressor feature importance distributions from LIME analysis [6]

**D.2. Additional coach tenure prediction models**

**D.2.1.** **Logistic Regression with Lasso Regularization**

A screenshot of a computer

Description automatically generated

Figure D5. Validation set, regularized logistic model prediction versus ground truth

A graph of a person

Description automatically generated

Figure D6. Regularized logistic model feature importance graph

**D.2.2. Multi-layer Perceptron Classifier**

A screenshot of a computer

Description automatically generated

Figure D7. Validation set, MLP classifier prediction versus ground truth

A screenshot of a computer screen

Description automatically generated

Figure D8. MLP classifier feature importance distributions from LIME analysis [6]