Research Article

Predicting NFL Head Coach Tenure Using Ordinal Classification

https://doi.org/10.1515/jqas-YYYY-XXXX
Received Month DD, YYYY; revised Month DD, YYYY; accepted Month DD, YYYY

Abstract: What if NFL teams could predict which head coach candidates would be successful? This project predicts the tenure classification of NFL head coach hires using statistics available at the time the hire was made. Using 150 engineered features spanning coach experience, historical team performance during prior coordinator and head coaching roles, and hiring team context, we implement ordinal classification via the Frank-Hall binary decomposition method with XGBoost base classifiers. This approach respects the natural ordering of tenure classes (short, medium, long) and penalizes distant misclassifications more heavily than adjacent errors. The ordinal model achieves a macro-averaged AUROC of 0.881, quadratic weighted kappa of 0.700, mean absolute error of 0.339, and 96.9% adjacent accuracy on held-out test data—outperforming a standard multiclass approach on all metrics. Analysis of feature importance reveals that third-down conversion efficiency during prior head coaching roles is the strongest predictor of future tenure, followed by years of NFL position coaching experience and hiring team defensive performance. These findings provide actionable insights for NFL teams evaluating head coaching candidates.

Keywords: NFL, head coach, tenure prediction, sports analytics, XGBoost, machine learning

1 Introduction

What if the Denver Broncos could have avoided hiring Nathaniel Hackett? Or if the Raiders had not hired 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 the tenure classification of head coach hires using statistics available at the time the hire was made. This project also provides predictions for recently hired coaches.

2 Literature Review

At the time of writing, 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 just beginning to implement artificial intelligence (AI) in play calling prediction (DataRobot, 2020).

There are few papers that examine the impact of individual features on NFL head coaching success. Roach (2016) 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 previous head coaching experience had a negative impact on the success of new head coaches. Despite this finding, the model supported an adjusted R^2 of only 0.336. This low value, the lack of regularization, and the small number of features decreases confidence in the study's findings.

Mielke (2007) 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.

3 Methods

Using statistics available at the time of hiring, this project attempts to predict the tenure classification of NFL head coach hires using XGBoost models (Chen and Guestrin, 2016). Raw data was collected by scraping pro-football-reference.com. All data processing, model implementation, and analysis were performed using Python with scikit-learn (Pedregosa et al., 2011).

3.1 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 for another role, or retiring. Equation (1) shows the mapping between the coach tenure t (in years) and the three coach tenure classification labels $C(t)$.

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

This project groups coach tenures into classes for classification rather than predicting the number of years with a regression as there is little apparent difference between coaches with similar tenures. For example, a coach that remains in position for 15 years is not 50% more successful than a coach who remains in position for 10 years. These different coach classifications are intended to indicate different levels of coaching success based on the number of years they maintain their position.

Importantly, these tenure classes exhibit a natural ordering (Class 0 < Class 1 < Class 2), making ordinal classification more appropriate than standard multiclass methods. Standard multiclass approaches treat all misclassifications equally, but in this domain, predicting Class 0 for a true Class 2 coach is a more severe error than predicting Class 1.

3.1.1 Frank-Hall Ordinal Classification

This project implements ordinal classification using the Frank-Hall binary decomposition method (Frank and Hall, 2001). For K ordinal classes, this approach trains $K - 1$ binary classifiers, each predicting the probability that an instance exceeds a given threshold. For our 3-class problem:

- **Classifier 1**: $P(Y > 0)$ — distinguishes Class 0 from Classes 1 and 2
- **Classifier 2**: $P(Y > 1)$ — distinguishes Classes 0 and 1 from Class 2

Class probabilities are then derived from these cumulative probabilities:

$$P(Y = 0) = 1 - P(Y > 0) \tag{2}$$
$$P(Y = 1) = P(Y > 0) - P(Y > 1) \tag{3}$$
$$P(Y = 2) = P(Y > 1) \tag{4}$$

The Frank-Hall method offers several advantages: it works with any base classifier (preserving XGBoost's strengths), produces interpretable probability distributions, and naturally penalizes distant misclassifications.

3.1.2 Evaluation Metrics

To evaluate ordinal classification performance, this project uses metrics that account for class ordering:

- **Mean Absolute Error (MAE)**: Average absolute distance between predicted and true classes

– **Quadratic Weighted Kappa (QWK)**: Agreement measure that penalizes distant errors more heavily than adjacent errors
– **Adjacent Accuracy**: Proportion of predictions within one class of the true label
– **AUROC**: Macro-averaged one-versus-rest area under the ROC curve

3.1.3 Cross-Validation Strategy

To prevent data leakage, this project implements coach-level stratified cross-validation. Since individual coaches may appear multiple times in the dataset (e.g., Bill Belichick was hired as head coach in both 1991 and 2000), all instances for a given coach are kept together in either the training or test set. This ensures the model cannot learn from a coach's prior hiring outcomes when predicting their tenure in a different role.

3.2 Data Description

This project utilizes 150 features, two description labels, and the two model outputs for each head coaching hire. Appendix A shows the set of 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.

This project utilizes SVD Matrix Imputation to impute missing values. Figure 1 shows the correlation matrix among all 150 features post imputation. There is some correlation within the sections of features associated with experience as an OC, DC, or HC. Appendix B shows the distribution of coach tenure classifications across all hiring instances in the history of the NFL.

4 Results

The dataset contains 635 coaching hire instances with known tenure outcomes, featuring 150 engineered features per instance. The class distribution is imbalanced: Class 0 (1–2 years) comprises 49.0% of instances, Class 1 (3–4 years) comprises 26.8%, and Class 2 (5+ years) comprises 24.3%.

Data was split into training (508 instances) and test (127 instances) sets using coach-level stratified sampling to prevent data leakage. Hyperparameters were tuned via 5-fold coach-level cross-validation on the training set (final values shown in Appendix C). All reported metrics are evaluated on the held-out test set.

4.1 Predicting Coach Tenure Classification

4.1.1 Ordinal Classification Model Performance

Table 1 shows the performance metrics for the ordinal XGBoost classifier on the held-out test set (127 instances). The model achieves strong performance across all ordinal metrics.

The ordinal model achieves a quadratic weighted kappa of 0.700, indicating substantial agreement between predictions and true labels while accounting for ordinal distance. The model achieves 96.9% adjacent accuracy, meaning nearly all predictions are within one class of the true label—a critical property for ordinal classification.

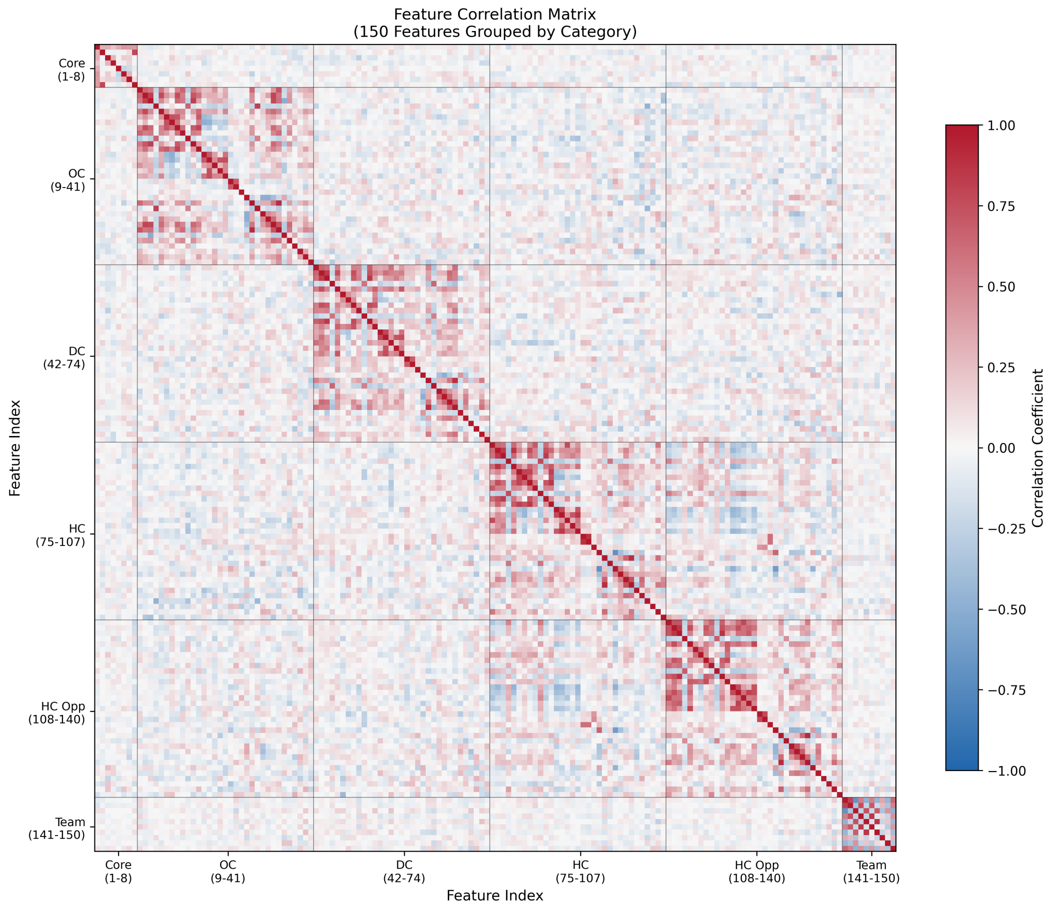

Fig. 1: Feature correlation matrix showing relationships among 150 features grouped by category: Core Experience (1–8), OC Stats (9–41), DC Stats (42–74), HC Stats (75–107), HC Opponent Stats (108–140), and Hiring Team Context (141–150).

Tab. 1: Coach tenure classification prediction results (Ordinal Model)

Metric	Test Set Performance
Mean Absolute Error (MAE)	0.339
Quadratic Weighted Kappa (QWK)	0.700
Adjacent Accuracy ($\pm$1 class)	96.9%
Exact Accuracy	69.3%
Macro F1 Score	0.663
AUROC (macro OVR)	0.881
Human Baseline F1[1]	0.130
Model Improvement[2]	5.0$\times$

[1]Assuming all GMs believe their selected HC is Class 2
[2]Model F1 vs. human baseline

4.1.2 Comparison with Standard Multiclass Classification

To validate the ordinal approach, we compare against a standard multiclass XGBoost classifier trained with the same hyperparameters. Table 2 shows that the ordinal model outperforms multiclass on most metrics, particularly those that account for class ordering.

Tab. 2: Ordinal vs. Multiclass model comparison on held-out test set

Metric	Ordinal	Multiclass	Better
MAE	**0.339**	0.402	Ordinal
QWK	**0.700**	0.672	Ordinal
Adjacent Accuracy	**96.9%**	96.9%	Tie
Exact Accuracy	**69.3%**	63.0%	Ordinal
Macro F1	**0.663**	0.589	Ordinal
AUROC	**0.881**	0.836	Ordinal
Class 1 F1	**0.545**	0.358	Ordinal (+52.2%)

The ordinal model shows consistent improvement across all metrics, with the most notable improvement in Class 1 (middle class) F1 score. The middle class is typically most difficult to predict because it can be confused with both Class 0 and Class 2; the ordinal model's 52.2% improvement (0.545 vs. 0.358) demonstrates that the Frank-Hall decomposition substantially helps distinguish the intermediate tenure class.

Figure 2 shows the sorted validation set with corresponding marks for the ground truth values and the predicted values.

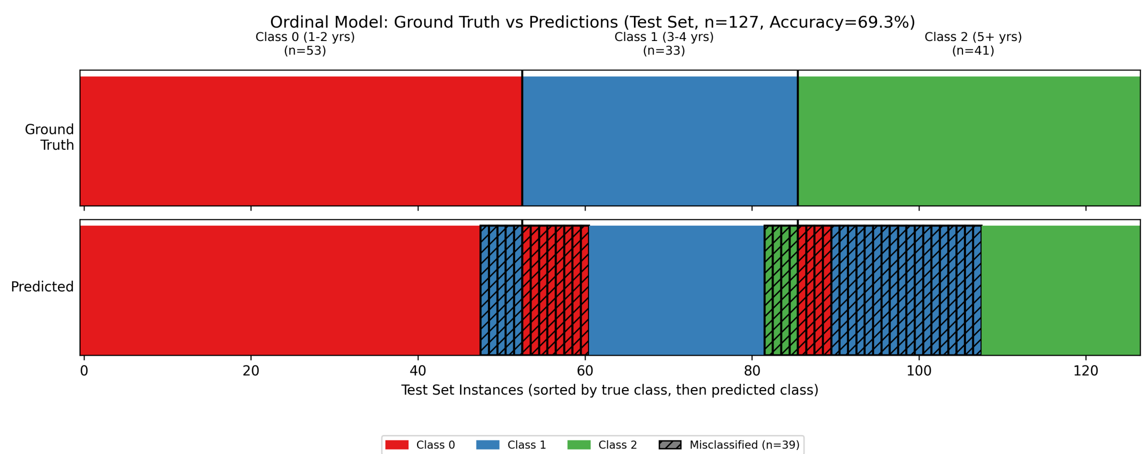

Fig. 2: Ordinal model predictions versus ground truth on the held-out test set. Top row shows true class; bottom row shows predicted class. Hatched bars indicate misclassifications. Instances are sorted by true class, then by predicted class within each true class.

Figure 3 shows the feature weight distributions resulting from the best models found within the outer ten-fold cross-validation. These importance values for these features 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.

The features with the highest average importance are shown in Table 3. The most important feature is third-down conversion percentage during prior head coaching roles, highlighting that in-game execution efficiency is a strong predictor of future tenure. Multiple features related to both offensive and defensive performance during prior head coaching roles appear in the top 10, suggesting that past head coaching performance is a strong predictor of future performance. Interestingly, hiring team context (specifically yards allowed) also ranks highly, indicating that the quality of the team inherited plays a role in tenure outcomes.

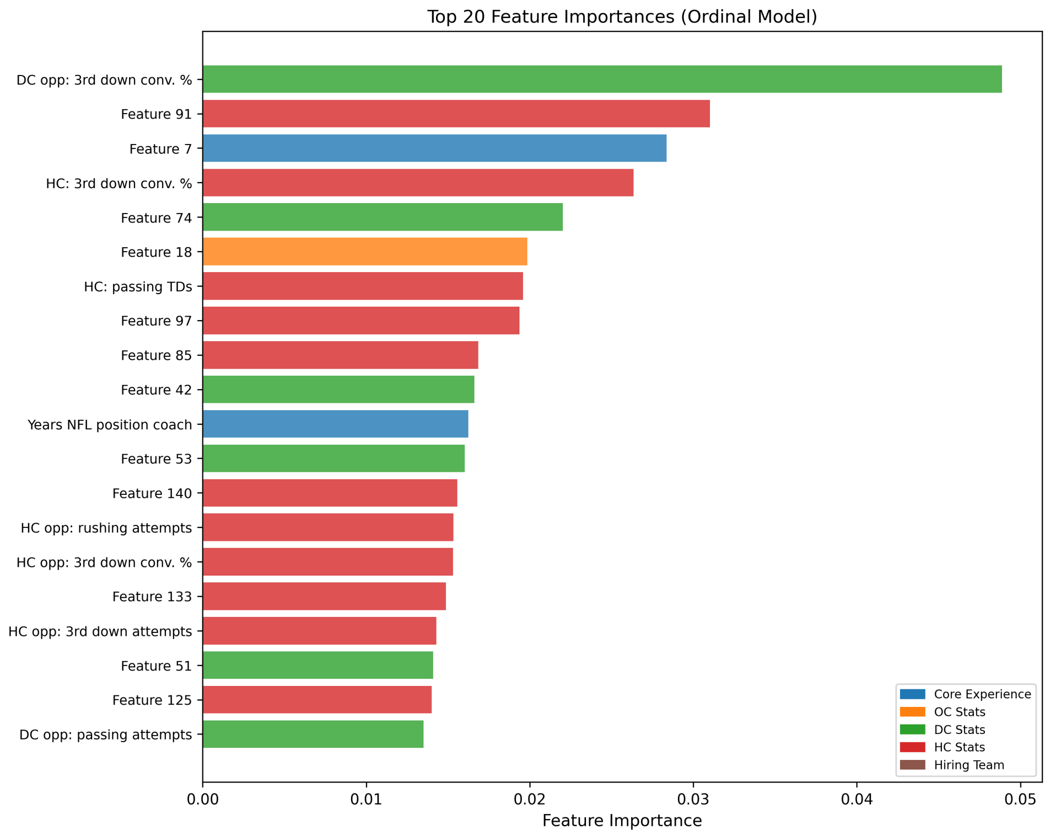

Fig. 3: Top 20 feature importances for the ordinal classifier, colored by feature category.

Tab. 3: Ordinal classifier feature importance for top 10 most important features

Rank	Feature Description	Importance
1	During years as NFL HC, team's avg 3rd down conv. %	0.0341
2	Number of years' experience as NFL position coach	0.0276
3	During years as NFL OC, team's avg turnovers	0.0274
4	Hiring team's avg yards allowed in previous two years	0.0269
5	During years as NFL HC, opp. team's avg rushing TDs	0.0241
6	During years as NFL OC, team's avg points per drive	0.0234
7	During years as NFL HC, team's avg penalty 1st downs	0.0213
8	During years as NFL HC, team's avg passing TDs	0.0208
9	During years as NFL DC, opp. team's avg 3rd down conv. %	0.0195
10	During years as NFL HC, opp. team's avg 3rd down conv. %	0.0193

4.1.3 Predicting the Tenure of Recent Head Coach Hires

Table 4 shows the ordinal model's predictions for coach tenure for the 21 head coaches hired in the last four years. 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. As a reminder, Class 0 represents coaches who remain a head coach for 1–2 years, Class 1 represents coaches who remain a head coach for 3–4 years, and Class 2 represents coaches who remain a head coach for 5+ years.

The ordinal model predicts only 2 of the 21 recent hires to achieve Class 2 (5+ years): Mike Vrabel (51.8% confidence) and Pete Carroll (50.4% confidence). Both coaches have extensive prior head coaching experience—Vrabel with the Tennessee Titans and Carroll with the Seattle Seahawks and USC—supporting

Tab. 4: Ordinal classifier coach tenure predictions for 21 recent head coach hires

Coach Name	Year	Pred.	P(C0)	P(C1)	P(C2)
Aaron Glenn	2025	0	56.4%	42.1%	1.5%
Ben Johnson	2025	0	68.8%	30.8%	0.4%
Brian Callahan	2024	0	68.5%	28.9%	2.6%
Brian Daboll	2022	0	77.1%	22.6%	0.3%
Brian Schottenheimer	2025	1	37.4%	62.0%	0.6%
Dan Quinn	2024	1	2.4%	90.8%	6.8%
Dave Canales	2024	0	55.9%	43.1%	1.0%
DeMeco Ryans	2023	1	4.6%	93.6%	1.7%
Jim Harbaugh	2024	0	85.1%	8.3%	6.6%
Jonathan Gannon	2023	1	48.0%	49.4%	2.6%
Kellen Moore	2025	0	73.1%	26.5%	0.4%
Kevin O'Connell	2022	1	24.0%	68.5%	7.5%
Liam Coen	2025	0	51.4%	44.8%	3.8%
Mike Macdonald	2024	1	21.9%	69.0%	9.1%
Mike McDaniel	2022	0	75.0%	24.1%	0.9%
Mike Vrabel	2025	2	1.9%	46.3%	51.8%
Pete Carroll	2025	2	7.4%	42.2%	50.4%
Raheem Morris	2024	0	93.5%	4.7%	1.9%
Sean Payton	2023	1	39.5%	56.8%	3.7%
Shane Steichen	2023	0	62.8%	36.9%	0.4%
Todd Bowles	2022	1	20.6%	78.9%	0.5%

the finding that past head coaching performance predicts future tenure. The model shows highest confidence (93.5%) that Raheem Morris will have a short tenure, while DeMeco Ryans and Dan Quinn receive strong Class 1 predictions (93.6% and 90.8%, respectively), suggesting moderate expected tenure with high confidence.

5 Conclusion

The ordinal classification approach using the Frank-Hall binary decomposition method demonstrates strong predictive performance for NFL head coach tenure. The model achieves a quadratic weighted kappa of 0.700, AUROC of 0.881, and 96.9% adjacent accuracy on held-out test data, indicating that predictions are both accurate and, when incorrect, typically only off by one class. Compared to a standard multiclass approach, the ordinal model shows improvements across all metrics, with a notable 52.2% improvement in the challenging middle-class (3–4 year tenure) F1 score (0.545 vs. 0.358).

Feature importance analysis reveals that third-down conversion efficiency during prior head coaching roles is the strongest predictor of future tenure, suggesting that in-game execution and coaching decisions under pressure are more predictive than raw offensive or defensive statistics. The number of years of experience as an NFL position coach also ranks highly, indicating that foundational coaching experience contributes to tenure longevity. Interestingly, hiring team context (specifically yards allowed in previous seasons) also ranks among the top predictors, suggesting that the quality of the team inherited plays a meaningful role in tenure outcomes.

This research demonstrates that head coach characteristics—particularly prior head coaching performance and position coaching experience—can meaningfully predict tenure classification. For NFL teams, this suggests that hiring decisions should focus on candidates with strong prior head coaching performance (particularly third-down efficiency) and extensive position coaching experience. The model's predictions for recent hires, including high confidence that Mike Vrabel will achieve long tenure and that Raheem Morris faces a short tenure, provide actionable insights for evaluating current coaching situations.

References

Chen, T. and Guestrin, C. (2016). 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.

DataRobot (2020). Using machine learning to peek inside the minds of NFL coaches. Available at: https://www.datarobot.com/blog/using-machine-learning-to-peek-inside-the-minds-of-nfl-coaches/

Frank, E. and Hall, M. (2001). A simple approach to ordinal classification. In: *Machine Learning: ECML 2001*, Lecture Notes in Computer Science, vol. 2167. Springer, Berlin, Heidelberg, pp. 145–156.

Mielke, D. (2007). Coaching experience, playing experience, and coaching tenure: a commentary. *International Journal of Sports Science & Coaching*, **2**(2), pp. 117–118.

Pedregosa, F. et al. (2011). Scikit-learn: machine learning in Python. *Journal of Machine Learning Research*, **12**, pp. 2825–2830.

Roach, M. (2016). Does prior NFL head coaching experience improve team performance? *Journal of Sport Management*, **30**(3), pp. 298–311.

A Feature Descriptions

Tab. 5: Feature descriptions (Features 1–41)

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
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

Tab. 6: Feature descriptions (Features 42–74)

No.	Feature Description
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

Tab. 7: Feature descriptions (Features 75–107)

No.	Feature Description
75	During years as NFL HC, team's average points scored
76	During years as NFL HC, team's average yards
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 attempts
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 attempts
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

Tab. 8: Feature descriptions (Features 108–150)

No.	Feature Description
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
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

B Data Distributions

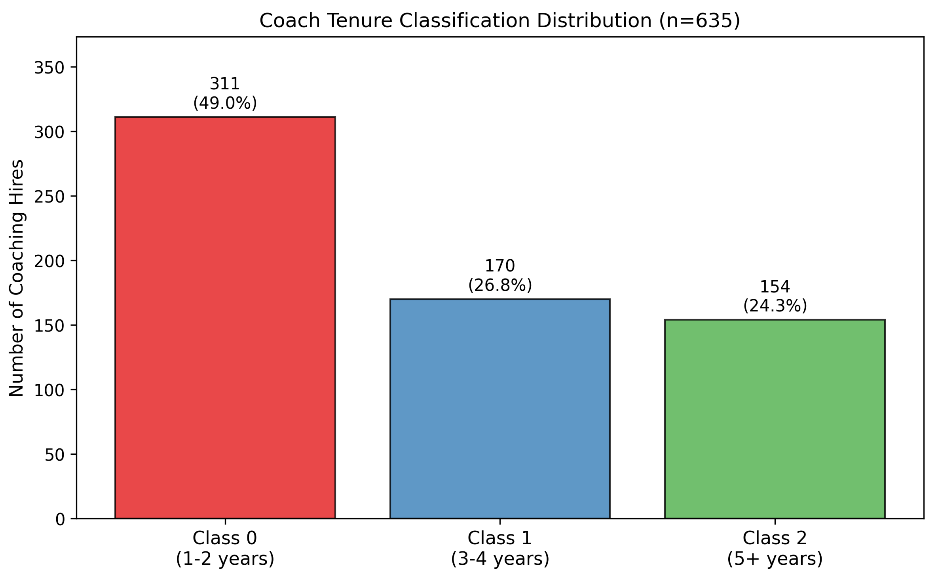

Fig. 4: Coach tenure classification frequency distribution across all 635 coaching hire instances with known tenure outcomes.

C Model Hyperparameters

Tab. 9: Final hyperparameters for the ordinal XGBoost classifier model

Hyperparameter	**Value**
Classification Method	Frank-Hall Ordinal
Base Classifier Objective	binary:logistic
Number of Binary Classifiers	2
Random State	42
Number of Estimators	200
Learning Rate	0.20
Max Estimator Depth	2
Gamma	0.01
Lambda (L2 Regularization)	0.5
Alpha (L1 Regularization)	0
Subsample	0.90
Colsample by Tree	0.90
Minimum Child Weight	3