Gridiron Genius: Using Neural Networks to Predict College Football

Luke Boll
University of Michigan
lukeboll@umich.edu

Abstract

This report details a project that leverages neural networks to enhance the accuracy of college football game predictions as a regression task. The system utilizes a vast amount of historical game data and incorporates various factors, including team statistics and home-field advantage, to train the model. The resulting system provides highly reliable insights and predictions for college football enthusiasts. This project has the potential to revolutionize the field of college football game predictions and pave the way for more accurate and data-driven approaches in the future.

1 Credits

I would like to express my gratitude to the following individuals and organizations for their contributions to this project. Jason McCormick for his assistance and guidance as both my Capstone advisor and as my Honors advisor. The University of Michigan College of Engineering Honors College for giving me the opportunity to pursue such an exciting project in an area I am interested in and passionate about. I would also like to thank many other of my professors, mentors, friends, and family for their support and encouragement throughout the project. Finally, I acknowledge the sources of any external materials used in this report, including data and literature. Proper citations can be found in the reference section.

2 Introduction

One of the longest-standing challenges for sports enthusiasts, analysts, and bettors is predicting the outcome of a game. In particular, one of the notoriously hardest sports to predict is college football. The task of predicting college football outcomes is exceedingly difficult due to the vast number of teams, the inherent variability in team performance and quality, and the influence of a wide range of external factors. With the advent of machine learning and artificial intelligence, researchers have turned towards utilizing advanced statistical models to predict the results of games accurately. In recent years, neural networks have emerged as a powerful tool for predicting sports outcomes due to their ability to learn complex relationships between statistical inputs and game outcomes. These challenges have led to the development of various prediction models that utilize machine learning and statistical techniques to extract insights from large volumes of data, with the goal of improving the accuracy of predictions in the ever-evolving landscape of college football.

This study looked into the use of neural networks to predict college football games. Various neural network architectures were assessed based on their ability to predict game outcomes. Additionally, in order to offer a benchmark against which to assess the baseline models, more conventional machine learning algorithms were used as well. These findings show how neural networks serve as a powerful tool in nondeterministic environments like college football. The results of this research can be valuable for those looking to predict the outcomes of college football games. Furthermore, this study highlights the importance of feature selection and engineering, as well as the need for a comprehensive understanding of the underlying dynamics of the game, in order to improve the accuracy of the predictions made by machine learning models.

3 Related Work

There have been some attempts at using machine learning and neural networks to predict college football games in the past. The work done by South and Egros (South and Egros, 2020) uses a lasso regression technique to take in statistics and predict the margin of victory. However, this method differs from what is being proposed in this paper as one of the features selected is the difference in winning percentage between the two teams. This features does not allow the models to learn from the on-field statistics as is proposed in this paper. Moreover, the approach proposed in this study also considers more advanced neural network architectures which may capture trends in the data not reflected by lasso regression.

Blaikie, et al. (Blaikie et al., 2011) attempt a very similar project to this one. They attempt to rectify some of the disparity in college football by including attendance and power rankings. This again has a similar issue to what is proposed by the paper by South and Egros, where team performance is already calculated into the model through one of the input features. This paper also more focuses on the differences in modelling the NFL and NCAA, whereas this project is exclusively focused on the college game. Additionally, the current study adopts a more comprehensive approach, considering a range of input features that reflect various aspects of the game, such as offensive and defensive performance, home field advantage, and historical head-to-head matchups, to develop more accurate models for predicting college football outcomes.

4 Data

The data used in this research project was collected from a compiled database of college football gambling spreads and results. This was then combined with scraped statistics that served as input features to the model. The dataset contained information on various features such as game location, game date, game result, and various other game statistics. Preprocessing of the data was done to improve the quality and relevance of the dataset to improve model performance. Furthermore, to ensure the validity and reliability of the results, the dataset was split into training, validation, and test sets to compare the performance of different neural network architectures and machine learning algorithms in generalizing their predictions of game outcomes.

4.1 Data Collection from TeamRankings.com

In order to collect the inputs to train the model, a vast collection of statistics for each game was going to need to be created and collected. While many datasets existed, none of the sets were extensive enough for the ambitions of this project. Therefore, the decision was made to collect the data by using the BeautifulSoup library in Python to scrape TeamRankings.com (tea, 2005-2023), a website with a large collection of statistics for every team in college football. The scraped data was then processed and aggregated to create a comprehensive set of input features for the models.

The statistics collected were in the following categories:

- Scoring Offense
- Scoring Defense
- Offense by Quarter
- Defense by Quarter
- Total Offense
- Total Defense
- Passing Offense
- Passing Defense
- Rushing Offense
- Rushing Defense
- Special Teams Offense
- Special Teams Defense
- Turnovers
- Penalties

Additionally, these statistics were taken as averages for the following splits:

- Current Season
- Previous Season
- Previous Three Games
- Previous Game
- Home
- Away

From this web scraping, a dataset of 1704 input features was created as an attempt to encapsulate all aspects of the chaotic, nondeterministic environment of college football.

4.2 Vegas Point Spreads and Game Results

Utilizing a compiled database of college football gambling spreads and results from Sportsbook Reviews Online (spo, 2023) is an additional component in this research on how neural networks predict college football outcomes. Using the point spreads from Vegas as a baseline, the predictive accuracy of the neural network model can be compared against the industry standard. Additionally, this data from Vegas projections may provide a wealth of information that can be used to identify trends and patterns in college football results. The use of the Vegas point spreads also provides an opportunity to study the behavior of betting markets, as the spread reflects not only the perceived strength of each team but also public sentiment and market demand. Overall, this database allows for more efficient and effective research, leading to a better understanding of how neural networks can predict college football outcomes. By combining this dataset with the team statistics outlined previously, the total available data becomes 1708 features for 7369 inputs (i.e., games).

4.3 Preprocessing

Preprocessing is a crucial step in preparing a dataset for use in a neural network model. Preprocessing refers to the steps taken to clean, transform, and normalize raw data before it is used in a machine learning or deep learning model. This can include tasks such as removing missing values, scaling or normalizing data, feature selection, and other transformations to ensure the data is properly formatted and free of noise, which can improve the performance and accuracy of the model. The main reason for this is that neural networks are highly sensitive to the quality and formatting of the input data. By properly preprocessing the data, the performance of the neural network model can be improved, leading to more accurate predictions and better generalization to new data. For this project, several preprocessing steps were used to decrease the noise of the dataset and remove features that would be problematic and unnecessary to the final predictions.

4.3.1 Expert Knowledge

One of the initial preprocessing steps taken in the research project was the removal of college football games played before the month of October. This decision was made based on expert knowledge within the field of college football with the aim of improving the accuracy of the neural network model. The rationale for this step was that games played before October are typically out of conference, and as a result, there may not be enough information available to accurately project the strengths and weaknesses of the teams involved. This decision to remove games played before October is an example of how domain knowledge can be utilized in the preprocessing stage to improve the accuracy of the model. Removing these games from the dataset allowed the network to learn from games played within a more specific timeframe, enabling more accurate modeling of the performance of teams in conference play. This reduced the dataset to 5097 games. This preprocessing step was an important aspect of the research project, ensuring that the data used to train the neural network model was of the highest quality and relevance.

4.3.2 Removal of Correlated Input Features

Another critical preprocessing step taken in the research project was the removal of highly correlated features from the dataset. Specifically, all features that had a correlation coefficient of 0.8 or higher with another feature were removed, keeping only the feature with the highest correlation coefficient with the outcome variable (i.e., game result measuring the difference between home and away final score). This step was taken to prevent the model from becoming overfit and not learning effectively, as highly correlated input features can cause the model to become biased towards certain features and overlook other important features. By removing these highly correlated features, the dataset was streamlined to ensure that the model was only trained on the most relevant and informative features, leading to a more effective neural network model. This reduced the feature space to 966 features. Removing highly correlated features is an important preprocessing step as it reduces the redundancy of the dataset, improves the efficiency of the model, and prevents the overfitting of the model. This preprocessing step was crucial in improving the performance of the model and was an important contribution to the overall success of the research project.

4.4 Removal of the Vegas Spread

Another important experiment in this research project was to remove the Vegas spread as an input feature. This was done because some of the machine learning algorithms used in the analysis were found to be overly reliant on the Vegas spread as the main feature for making predictions. By removing the Vegas spread as an input feature, the models were forced to learn from the statistics of the game rather than just selecting the Vegas spread as the prediction. By removing the Vegas spread as an input feature, the analysis was able to focus on other features that could be more informative for predicting the outcomes of college football games. Additionally, this has the potential to lower the chances of the models finding a local minima based on the Vegas spreads, and instead find their own solutions to the minimization of the loss function. Overall, this allows for better comparision and learning from the models.

5 Methods and Algorithms

This report seeks to investigate the power of neural networks in nondeterministic environments with high levels of variance. College football serves as an excellent example of this and shows how machine learning, particularly neural networks, can be used to make accurate predictions in complex and unpredictable environments. This demonstrates the potential for machine learning to revolutionize the way we approach data analysis in a wide range of fields. This section details how those neural networks are created and how they are compared to baseline models built with more traditional and simpler modelling methods.

5.1 Baseline Models

The aim of this research project was to predict the outcomes of college football games using neural networks. To establish a baseline, several different traditional machine learning algorithms were used:

- Elastic Net Regression
- Lasso Regression
- Linear Regression
- Ridge Regression
- Support Vector Regression (SVR)

All of these models, including the neural networks, were given the same training, validation, and training split. Additionally, these baseline models were all created using the `sklearn` package in Python. This was done in order to ensure accurate comparisons between the different algorithms.

Grid search was used to identify the best hyperparameters for each algorithm based on the training dataset. This involved a systematic search with cross-validation over a specified range of hyperparameters, with the aim of finding the hyperparameters that led to the best mean performance on the validation set during cross-validation. Once the best hyperparameters had been identified for each algorithm, the models were trained on the full training set. The performance of each algorithm was then evaluated on both the validation and testing sets. The loss and R^2 scores for the testing, validation, and testing sets were recorded for each of the algorithms at this point. To ensure consistency in the comparison of these models, the evaluation metrics were calculated using the same loss functions and scoring metrics across all the algorithms.

5.2 Neural Network

In order to create the neural network model, the `pytorch` package in Python was used. The neural network model consisted of several fully connected layers with a ReLU activation function. Similarly to the traditional baseline models, a hyperparameter search was conducted by using all the possible combinations of hyperparameters and finding the one that produced the lowest validation loss. The values of hyperparameters studied can be seen in Table 1.

Parameter	Values
Hidden Layers	1, 2, 3
Neurons	32, 64, 128, 256, 512
Dropout Rate	0.2, 0.3, 0.4, 0.5

Table 1: Hyperparameters studied for the neural networks

During the training process, the weights of the neural network were updated using the Adam backpropagation algorithm. Adam is a popular optimization algorithm used for updating the weights of a neural network during the training process. One of the main advantages of Adam is that it adapts the learning rate for each weight based on the gradient history, allowing for faster convergence and better optimization. It also takes into account the momentum of the gradient updates, which can help it escape from local minima.

After finding the optimal network structure for the dataset, the architecture would then be trained on the training set with the validation set being used for early stopping. Early stopping is a technique used in neural networks to prevent overfitting by stopping the training process when the performance on the validation set stops improving. This step is performed after every epoch, or pass through the data. The training would stop when the validation loss had not improved for ten epochs. An example of the training plot can be seen in Figure 1 where R^2 score is on the left and mean squared error is on the left. Additionally, the training set values can be seen in blue, with the validation set seen in red.

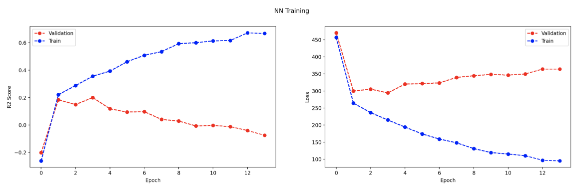

Figure 1: Example of the training plot for the Neural Networks

Finally, the loss and R^2 scores for the testing, validation, and testing sets were recorded for each of the algorithms at this point.

5.2.1 Lasso Regression

Additionally, one experiment done with the neural networks was changing the data that was used in the process. Building off the work of South and Egros and the usage of lasso regression to predict college football, the full range of input features was cut down by first running lasso regression on the training set and selecting only the features with non-zero coefficients from the resulting solution. Running Lasso regression before a neural network can be a good thing because it can help with feature selection, identifying and removing irrelevant or redundant input features that may negatively affect the performance of the neural network. This idea is corroborated by Sun, et al. (Sun et al., 2016). Overall, this gives another point of comparison for the usage of different approaches to machine learning for the regression task of predicting college football. Using Lasso regression before a neural network can also help with reducing overfitting and improving the generalization performance of the neural network, as the reduced feature set is less likely to cause the model to memorize noise in the training data.

6 Results

This section of the report summarizes the findings of this study. The results of the different machine learning methods were compared to determine which modelling technique works best for the regression task of college football. Additionally, this process was done with both the Vegas spread as an input feature and with it removed from the input feature space.

6.1 Comparison of Methods

To predict college football outcomes, various models were developed and trained for their accuracy. To compare the performance of these models, their mean squared error values between the predicted and actual difference between the home and away were calculated using Equation 1. Moreover, the R^2 scores were also computed to evaluate the goodness of fit of the models.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2 \tag{1}$$

In Equation 1, y_i is the actual result, and $\hat{y_i}$ is the predicted result, where n is the number of games in the dataset. Based on this metric, the results for the different algorithms can be seen in Figure 2:

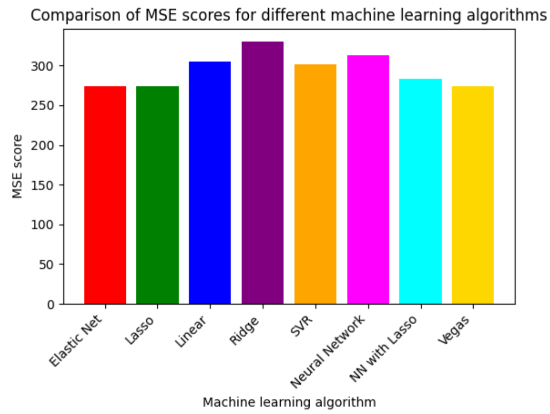

Figure 2: Comparison of mean squared error for the different models

In this figure, it can be seen that Vegas, lasso regression, and elastic net regression have the lowest testing mean squared error of all the algorithms included in this study. Both lasso and elastic net are regression techniques that penalize the magnitude of coefficients in the model. However, the difference is that while lasso regression uses a L1 penalty, elastic net combines a L1 and L2 penalty. In this way, elastic net regression effectively combines lasso and ridge regression in one regression technique. In this study, elastic net takes on a similar value to lasso regression because lasso was much more effective than ridge regression at modeling the data. While initially, these results from lasso regression seemed promising in that they approached the industry standard in Vegas point spreads, it was found that the reason for this was that Vegas is one of only two terms retained in lasso regression. This suggests that the betting lines set by Vegas are a strong predictor of college football outcomes and that models may be overfitting to the Vegas spread as an input feature. This is problematic as the goal is to have the models learn from the data features and approach or exceed the Vegas point spread in performance.

Another approach that was explored in this study was to use a neural network with lasso regression ahead of time for feature selection. This approach outperformed the standard neural network approach in terms of testing mean squared error, suggesting that the feature selection step improved the model's ability to generalize to new data. This finding highlights the fact that neural networks can be highly sensitive to the data being fed into them and that preprocessing steps such as feature selection can have a significant impact on their performance. It also underscores the importance of careful data exploration and preprocessing when working with neural networks, as small changes in the input data can have a large impact on the output. Overall, the results of this study suggest that incorporating techniques such as lasso regression and feature selection can improve the accuracy and robustness of neural network models for predicting college football outcomes.

For complete results of this experiment, see Appendix A.

6.2 Removal of the Vegas Spread as Input Feature

After observing that the Vegas spread was one of the only terms retained in the lasso regression model, further investigation was carried out to assess the impact of removing this feature from the input space. This was motivated by the concern that relying solely on the Vegas spread could lead to overfitting and limit the generalization ability of the model to new data. To investigate this, models were trained using the same set of predictors as before, but with the Vegas spread excluded. The results of this investigation in regards to testing mean squared error can be seen in Figure 3.

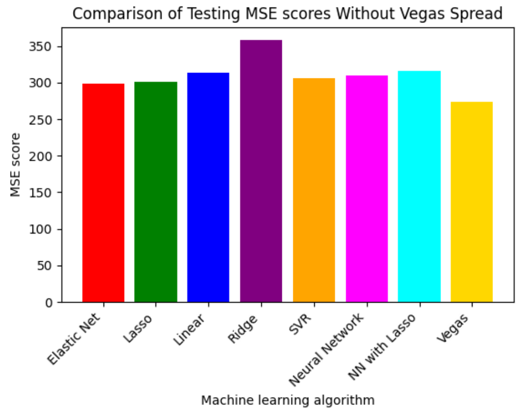

Figure 3: Comparison of mean squared error for the different models without the Vegas spread as an input feature

After removing the Vegas spread from the input feature space, it can be seen that all of the regression methods used here perform worse regarding the testing mean squared error. This demonstrates how powerful of a predictor the Vegas point spread is and explains why it is the industry standard.

Another finding that supports this claim is that the performance of the neural network with lasso regression as a feature selector reverts to the performance of the standard neural network. This shows how powerful the information provided by the Vegas point spread was in driving predictions for the model.

Additionally, another interesting finding is that neural network with lasso regression as a feature selector does not perform as well as lasso regression by itself. This is a strange finding as theoretically the network should set its weights and biases to the same values found by lasso regression and mirror its performance if lasso regression has found a global minimum from the data. This reflects that the process of determining the architecture and hyperparameters of the neural network may be flawed as the model does not effectively capture the relationships between the input features and the target variable. This demonstrates how neural networks are not an all-powerful technique and should be the discussion of future research.

Lasso and elastic net regression still exceed the rest of the algorithms in regards to testing mean squared error. This is most likely due to the effectiveness of lasso regression in environments where the relationship between the input features and the response variable is complex and there is a high level of variance. This finding suggests that lasso regression may be a good start to build upon for future methods of exploring college football datasets, and may transfer to other sports. For a complete list of input features retained by lasso regression see Appendix C.

For complete results of this experiment, see Appendix B.

7 Ethical Considerations

A major issue with developing a sports prediction model is that it can promote unhealthy behavior, including a gambling addiction. Although actual betting was not a part of this study, it is vital to understand that these uses are possible and that there is a chance that the models could be abused or exploited in this way.

Assuring that the developed models are used responsibly and ethically is crucial to addressing these concerns. This could involve measures such as not publicizing the predictions and results in a manner that encourages sports betting. Additionally, it is also wise to ensure that individuals who seek to use these models in the context of gambling are aware of the potential risks and are using the models in a responsible and educated way.

Overall, even if using predictive modeling for sports betting can be an effective tool for improving outcomes and making well-informed judgments, it is crucial to proceed cautiously and make sure that all pertinent ethical issues are taken into consideration. By doing this, the hazards that could arise can be reduced. Thus, these models and this project as a whole are being used in a moral and responsible manner.

8 Conclusion

The goal of this study was to investigate how well neural networks would compare to traditional machine learning algorithms in predicting the results of college football games. The results showed that lasso regression was the most effective algorithm in terms of predicting game outcomes. Furthermore, the Vegas point spread was found to be a powerful predictor of game outcomes, as removing it from the input feature space resulted in worse performance for all the regression methods used. Additionally, the neural network with lasso regression as a feature selector performed worse than lasso regression by itself, suggesting that further research is needed to improve the process of determining the architecture and hyperparameters of the neural network.

These findings highlight the potential of machine learning models in predicting sports outcomes and provide insights for future research in this field. Overall, this study contributes to the body of literature on using machine learning techniques for predicting sports outcomes. It can hopefully serve as a resource for those looking to improve their predictions in the challenging domain of college football.

9 References

References

2005-2023. Team rankings.

2023. Ncaa football odds archives.

Andrew Blaikie, Gabriel Abud, and John A. David. 2011. Nfl & ncaa football prediction using artificial neural networks.

Charles South and Edward Egros. 2020. Forecasting college football game outcomes using modern modeling techniques. *Journal of Sports Analytics*.

Kai Sun, Shao-Hsuan Huang, David Shan-Hill Wong, and Shi-Shang Jang. 2016. Design and application of a variable selection method for multilayer perceptron neural network with lasso. *IEEE Transactions on Neural Networks and Learning Systems*, 28(6):1386–1396.

Appendix A: Full Results from the Models with Vegas Spread

This appendix of the report contains the results with the Vegas spread as an input feature.

Model	Train MSE	Train R2
LinReg	222.0500	0.3107
Ridge	183.6900	0.5864
Lasso	257.5176	0.4202
ElasticNet	258.0649	0.4190
SVR	161.7125	0.6359
Neural Network	217.8364	0.3429
NN with Lasso	259.7003	0.0867
Vegas	256.9844	0.4214

Table 2: Mean squared error and R^2 for training set with Vegas spread as an input feature

Model	Val MSE	Val R2
LinReg	297.1206	0.1696
Ridge	331.9119	0.2701
Lasso	251.0294	0.4479
ElasticNet	254.6290	0.4400
SVR	292.3897	0.3570
Neural Network	300.6448	0.1739
NN with Lasso	251.8629	0.3024
Vegas	249.9001	0.4504

Table 3: Mean squared error and R^2 for validation set with Vegas spread as an input feature

Model	Test MSE	Test R2
LinReg	305.3909	0.0282
Ridge	330.2338	0.2071
Lasso	273.7758	0.3426
ElasticNet	273.7599	0.3427
SVR	301.3672	0.2764
Neural Network	312.8432	0.0190
NN with Lasso	283.2719	0.0867
Vegas	274.4007	0.3411

Table 4: Mean squared error and R^2 for testing set with Vegas spread as an input feature

Appendix B: Full Results from the Models without Vegas Spread

This appendix of the report contains the results from removing the Vegas spread as an input feature. This was a significant shift that demonstrates reduced accuracy in almost all of the models. However, it does reflect the models' ability to learn instead of simply overfitting to the Vegas value.

Model	Train MSE	Train R2
LinReg	261.711	0.2222
Ridge	203.3245	0.5422
Lasso	292.7595	0.3409
ElasticNet	280.5701	0.3683
SVR	166.2968	0.6256
Neural Network	267.8622	0.1886
NN with Lasso	284.0472	0.1477
Vegas	256.9844	0.4214

Table 5: Mean squared error and R^2 for training set with Vegas spread removed as an input feature

Model	Val MSE	Val R2
LinReg	309.5968	0.1414
Ridge	379.1533	0.1662
Lasso	299.2335	0.3419
ElasticNet	298.2011	0.3442
SVR	299.5841	0.3412
Neural Network	300.2645	0.1865
NN with Lasso	309.9510	0.1588
Vegas	249.9001	0.4504

Table 6: Mean squared error and R^2 for validation set with Vegas spread removed as an input feature

Model	Test MSE	Test R2
LinReg	313.1907	0.0069
Ridge	358.4697	0.1393
Lasso	301.1578	0.2769
ElasticNet	299.1767	0.2817
SVR	306.1294	0.2650
Neural Network	309.3097	0.0368
NN with Lasso	316.7257	0.0028
Vegas	274.4007	0.3411

Table 7: Mean squared error and R^2 for testing set with Vegas spread removed as an input feature

Appendix C: Input Features Retained by Lasso Regression

This appendix of the report includes a list of features that are retained by the lasso regression. This serves a good starting point for the statistics that have the most relevancy to the on-field results of a college football game. This should serve as a good starting point for further research in this area.

- Away Average Scoring Margin
- Away Average Scoring Margin Previous Season
- Away Red Zone Scoring Percentage (TDs and FGs) Previous Season
- Home Yards per Game Last 3
- Home Third Downs per Game Last 3
- Away Third Downs per Game Last 3
- Away Punts per Play Last 3
- Home Completion Percentage Last 3
- Home Completion Percentage Away
- Home QB Sacked per Game Last 1
- Home Punt Attempts per Game
- Away Gross Punt Yards per Game Away
- Away Opp Yards per Point Previous Season
- Home Opponent Average Scoring Margin
- Home Opponent Average Scoring Margin Last 1
- Home Opponent Average Scoring Margin Previous Season
- Away Opponent Average Scoring Margin Last 3
- Away Opponent Average Scoring Margin Home
- Away Opponent Average Scoring Margin Previous Season
- Away Opponent Punts per Play Previous Season
- Home Opponent Rushing Yards per Game
- Home Opponent Rushing Yards per Game Home
- Away Opponent Rushing First Downs per Game Last 3

- Away Opponent Yards per Rush Attempt Last 3
- Away Opponent Yards per Rush Attempt Home
- Home Opponent Passing Yards Percentage Last 1
- Away Sacks per Game
- Home Opponent Gross Punt Yards per Game Previous Season
- Away Interceptions per Game Previous Season