Predicting the 2025 College Football Season with Machine Learning

Sidharth Dimri

Introduction

Having recently taken interest in College Football, I wanted to do some research into predictions of next year's seasons. After looking at some of the predictions ESPN gave on next year's potential top 25 teams, I was inspired to make some rankings and predictions myself, especially for teams I supported that were outside of the rankings. I decided that a machine learning (ML) model would be useful for this purpose.

Dataset

I created a dataset containing information of 134 Division I College Football teams, consisting of all teams in the FBS subdivision excluding Kennesaw State, University of Delaware, and Missouri State. The information I collected goes from 2023 to 2025. In it are both the 2023 and 2024 Football Power Index (FPI) ratings of each team, which measures a team's expected point margin against an average college team, calculated by ESPN.[1] Additionally collected is information of how a team did in both the 2023 and 2024 season. Teams that made a bowl game, measured by if they got six or more wins in the regular season were collected, along with the teams that made the playoffs, teams that made the national championship, and the team that won the national championship. The recruiting class of each team was also collected as a score rating how well each team managed to recruit both new freshman and transfer players. This would show how well a team improved from season to season. This information was collected by 247 Sports, who are recognized nationally for their collection of college sports recruitment data.

[1] ESPN, "2024 College Football Power Index - ESPN," ESPN, 2024, https://www.espn.com/college-football/fpi/_/season/2024.

[2] Finally, information on a team's returning production was collected, which provides information on what percentage of a team stays from one season to the next. Returning production information for 2024 and 2025 was provided on X (formerly Twitter) by the ESPN analyst Bill Connelly.[3] All other information on the spreadsheet was not used in this investigation

Process

My aim for this investigation is to create an ML model that can evaluate data from the previous year to create predictions for any college football team's success in the 2025 season.

To do this, I decided to split this investigation into two separate goals. The first would be to estimate an FPI rating for any team in 2025 using FPI information from the previous years. Using the previous FPI of a team and then evaluating how teams change based on offseason processes (returning production and recruiting class) should theoretically be able to predict next year's FPI for any team. In order to estimate a number for the FPI of a team, I would have to use some sort of regression ML model, so I narrowed down my choices for this part's model to Linear, Polynomial, and KNN regression models. I would test each of these models to evaluate the best one for the purpose of this goal.

The next step would theoretically be to use that predicted FPI to then calculate the probability of a team to achieve two major milestones in a college football season: Earning a bowl game at the end of the season (usually requiring at least 6 wins), and earning a spot in the college football playoffs (top 12 teams). However, since the FPI itself is predicted from available factors mentioned in part 1, using those original factors instead of a predicted factor such as FPI would prevent any unnecessary information loss. For this model, I would need a classification

[2] "2026 247Sports Composite Player Rankings," 247Sports, 2025, https://247sports.com/season/2025-football-20251/overallteamrankings/?OrderBy=Points.
[3] Bill Connelly, X (formerly Twitter), 2025, https://x.com/ESPN_BillC/status/1790056401382920497.

ML model, but specifically one that works with probabilities. I narrowed down my options for this part to Gaussian Naive-Bayes, Logistic Regression, and SVM.

With the possible models that could be used for the investigation, I could move on to the next step of deciding the optimal model (and hyperparameters) through cross-validation.

Set-Up

The next step of this process was to cross-validate each model, evaluating its error/accuracy in comparison to other models and choosing the optimal model to complete the investigation.

Part 1: FPI Prediction

To find the optimal model for predicting a college football team's FPI, instead of predicting 2025 FPI information, which did not exist and could not be validated, I decided to train and test my model of 2024 FPI information. Therefore, to set up this prediction, I used the features "2023 FPI," "2024 Recruiting Class," and "2024 Returning Production," all of which are features that would directly impact the 2024 FPI. Keeping the 2024 FPI as my target variable, I used K-Fold Cross Validation with four folds, using the Mean Squared Error (MSE) as an evaluator for a model's accuracy. The lower the MSE was, the closer actual data points were to the predicted, meaning that the model is more accurate and precise.[4] Predictions of a team's future FPI are not necessarily available online, meaning that I had no real world information to

[4] Jim Frost, "Mean Squared Error (MSE)," Statistics By Jim, November 12, 2021, https://statisticsbyjim.com/regression/mean-squared-error-mse/.

compare my model to. Therefore, there is no baseline for accuracy, and the lowest MSE ML model will automatically be accepted.

Linear Regression

No hyperparameters were required for this model, meaning that the tuning phase of the model could be skipped.

The resulting MSE of this model was 42.79.

Polynomial Regression

The sole significant hyperparameter was the degree of the polynomial regression line, which I hypothesized could either be valued at 2 or 3, since we only have three features to predict FPI. However, to be safe, I cross-validated every polynomial degree value from 2 to 10, and chose the lowest MSE out of every degree to ensure the optimal polynomial model for the investigation.

Sure enough, the best degree for the polynomial regression line was at 2, resulting in an MSE of 41.27, slightly better than the linear regression model.

KNN Regression

KNN Regression had two notable hyperparameters for the purposes of our investigation. The first was the k value. Going off of the rule of thumb formula for deciding an optimal k value:

$$k = \sqrt{n}$$

with k being the optimal k value and n being the total number of samples in the dataset, I concluded an optimal k value of 11 given a dataset of 134 data points. However, since this formula is not always accurate, I decided to cross validate every odd k value from 1 to 23, comparing each k value to find the optimal k value.

The second hyperparameter is the distance metric used for KNN. Since many metrics typically used for classification purposes, such as Cosine and Hamming distance, did not particularly work for this investigation, I kept my options for the distance metric to Minkowski distance metrics, which are simply calculated as:

$$D = (\sum_{i=1}^{n} |a_i - b_i|^p)^{1/p}$$

where n is the number of dimensions used, a_i and b_i are data points and their values with a certain feature, and p is the power that they are raised to.[5] I decided to try three different exponential values, testing Manhattan distance (p=1), Euclidean distance (p=2), and p=3 distance.

This means that concerning all hyperparameters, every combination has to be evaluated to find the optimal version of the model.

The best set of hyperparameters for this model was with a k value of 5 and a p value of 2, yielding an MSE of 40.83. Being the lowest out of all three tested models, this model will therefore be used to predict the FPI of each team.

Dimensionality Reduction

Before we move on to the next part, I decided to quickly test out a dimensionality reduction of the dataset to see if removing any of the features would improve the dataset's quality. The algorithm to handle this is simple. First, consider the original dataset as the starting

---

[5] Pulkit Sharma, "Distance Metrics | Different Distance Metrics in Machine Learning," Analytics Vidhya, February 25, 2020, https://www.analyticsvidhya.com/blog/2020/02/4-types-of-distance-metrics-in-machine-learning/.

point. Then, remove each feature while adding back any previously removed features from the starting dataset, evaluating the cross validation score of the new dataset afterwards. Out of each dataset with a removed feature, consider the dataset with the lowest MSE and compare it to the MSE of the starting dataset. If the starting dataset has the lower MSE out of the two, that dataset is the optimal dataset for the investigation. Otherwise, the dataset with a feature removed becomes the new starting dataset, and the process repeats until the optimal dataset is found.

After testing out the dimensionality reduction of the dataset for this part, I realized the MSE score was equal between the original dataset, and the dataset that removed the “Returning Production” feature. I decided to remove this feature in hopes that it would make the model more efficient. Note that this change does not necessarily apply to part 2’s model, and dimensionality reduction will be separate for that model.

Part 2: Success Probability

To find the optimal model for a college football team’s chance to either make a bowl game or make the college football playoffs, I decided to use the same three features used in part 1, as my idea was that the 2024 FPI should correlate somewhat with 2024’s performance. I again used K-Fold cross validation to cross validate the model, but this time, I used the generic classification accuracy score to determine the quality of a classification model. To establish a baseline accuracy required for my model, I decided to view ESPN’s predictions for the now-complete 2024-2025 season[6] to compare their accuracy to mines. Since their predictions are only for the top 25 teams, I will only be able to compare their playoff predictions to my own, so

[6] ESPN, “AP Top 25 Reaction - 2024-25 College Football Predictions - ESPN,” ESPN.com (ESPN, August 12, 2024), https://www.espn.com/college-football/story/_/id/40751760/ap-top-25-reaction-2024-25-college-football-predictions.

for bowl game accuracy, the higher that accuracy is, the better. Meanwhile, for ESPN's playoff predictions, they predicted 6 out of the 12 playoff teams correctly, meaning that in order to have a better model than them, I would need to predict playoff contenders for 2024 at an accuracy better than 91.04%. This is because since ESPN predicted 6 incorrect playoff teams, they missed the other 6 teams that did make the playoffs, giving them an accuracy across the 134 FBS teams (excluding Kennesaw State) of 122 teams out of 134, or 91.04%

Gaussian Naive Bayes

There are no significant parameters for this model.

The resultant accuracy of the model predicting 2024 bowl game teams was 63.99% 91.76% for 2024 playoff predictions. This guarantees me a model more accurate than ESPN's playoff predictions.

SVM

There were two hyperparameters that were important to keep track of with SVMs. The first was making sure probability was set to true, which would set the SVMs to calculate the probability of each team hitting a certain seasonal milestone. The second was the choice of kernel for the SVM. Although I do not understand much about these kernels, I was able to narrow down possible kernel choices to two: Linear and RBF kernels. From what I could tell, linear kernels were simpler while RBF kernels were a general case for SVM.

The better kernel for the model was the linear kernel, with a 67.71% accuracy for bowl games and a 90.24% accuracy for playoff games. However, after running this model through a confusion matrix, I noticed that the SVM model was solely predicting every team to win 6 or more games, making a bowl game, while predicting every team to miss the playoffs. This allowed them to get a “good enough” accuracy to seem to be the best model for the investigation. Since then, I have disregarded this model for the investigation.

Logistic Regression

There are no significant hyperparameters for the purposes of this investigation. However, due to a misunderstanding on my own part on what the class weight hyperparameter of logistic regression represented, my output for this model was less accurate than it would have been without this hyperparameter. This model was in fact the most accurate out of all models for part 2, but I had continued with Gaussian Naive Bayes instead. I have since replaced Naive Bayes with Logistic regression.

Logistic Regression had a bowl game accuracy of 70.7% and a playoff accuracy of 92.51%. With both the highest bowl game and playoff prediction accuracy, I decided to make this my model of choice for the predictions.

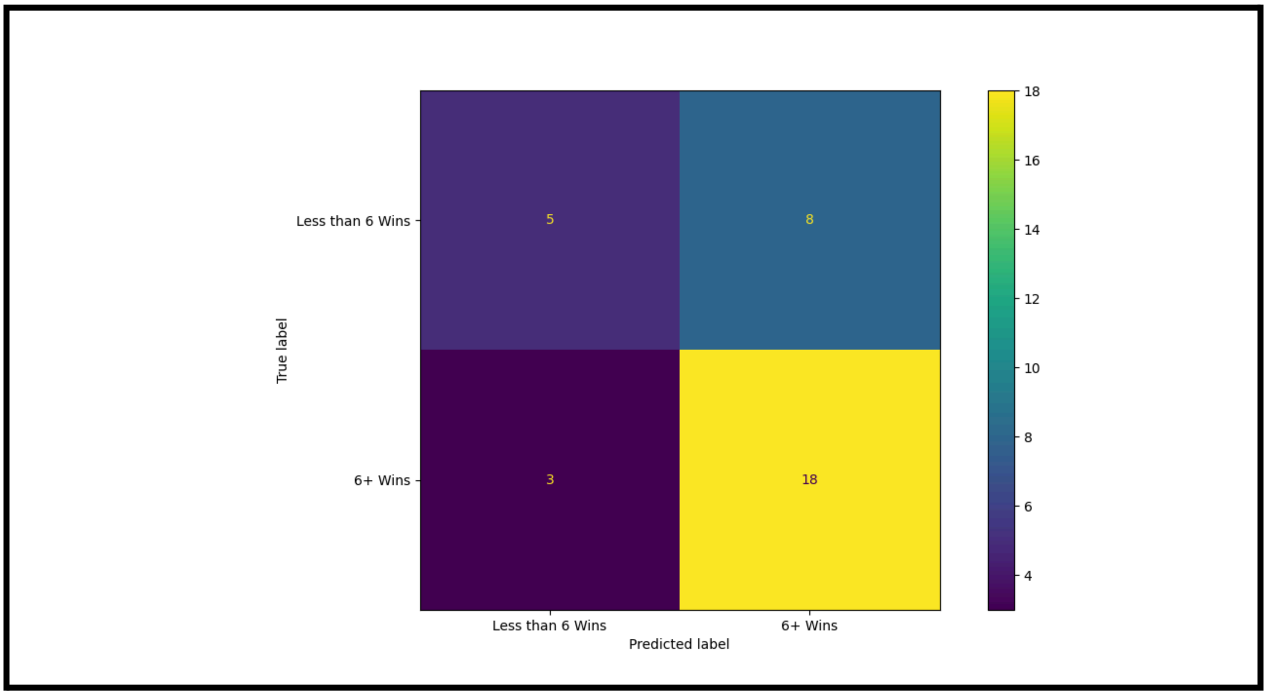
Dimensionality Reduction

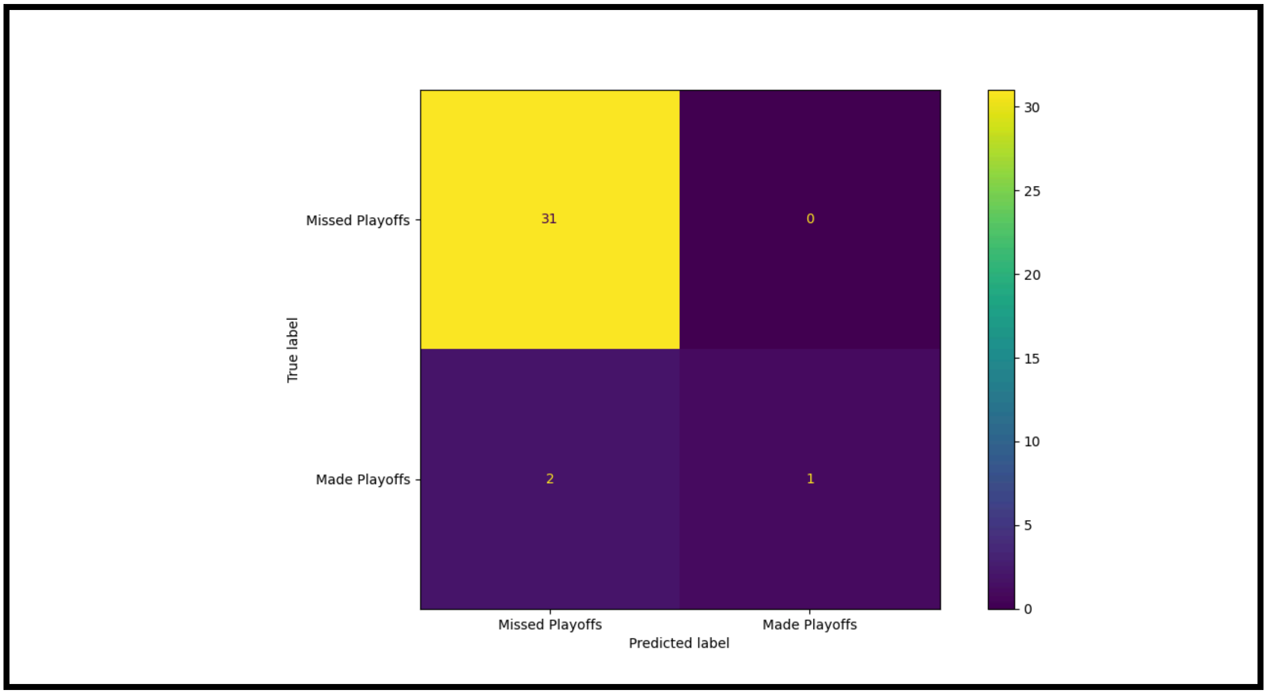
After a quick investigation, I concluded that the optimal dataset for part 2 was the original dataset itself. This means that the datasets for part 1 and part 2 are different, since part 2 utilizes returning production.

Analysis

To review, the chosen models for this investigation is a KNN regression model with a Euclidean distance metric and k value of 5, and a Logistic Regression (for classification) model.

To test the KNN regression model, I trained a KNN regression model to fit the 2024 FPI dataset in reference to the two features, afterwards comparing the prediction of each team to its actual 2024 FPI. I discovered that for the large part, the regression model was accurate, and wasn't often far off from the actual model, only being off by a few points most of the time. However, several team's predicted FPIs were far off from their original FPI. To test the logistic regression model, I created a confusion matrix to analyze the results, with a 75% training size and 25% testing size.





As seen from the confusion matrices above, the model was generally pretty accurate in predicting the success rate of each college football team.

Conclusion and Predictions

Now that the model is complete and valid, it is time to make predictions based on the 2025 season. I made my training set on the 2024 season, using 2024 FPI, 2025 Recruiting Class, and 2025 Returning Production to predict the 2025 FPI, 2025 bowl game chance, and 2025 playoff chance of a college football team.

Results for individual teams can be seen with the project code provided. Generally, however, no team has over a 70% chance of making the playoffs, and only a few teams have a 90% chance of winning 6 or more games next year. I feel that the model provided is very conservative with its prediction, however, I believe that this is a good thing for the model. The model is not overly committed to predicting certain teams to make the playoffs. I had originally conducted these predictions with Gaussian Naive Bayes, and although the two models' accuracies are only about 3% off, Gaussian Naive Bayes had more extreme predictions for teams, marking some teams with a 0% chance to make the playoffs and others with over 80% of a chance. The Logistic Regression model brings into consideration more teams for the playoffs and bowl games, keeping an essence of the unpredictability of college football.

Evaluation

There were some issues with this project. The dataset was somewhat small, with only 134 data points. Additionally, the model cannot take into account newer FBS teams such as Kennesaw State and the two new teams entering in the 2025 season. This may switch up things prediction-wise, and the new FBS teams should be added as soon as enough information is available on them. Another issue of the dataset was the variability of FPI predictions, as some teams were way off prediction wise compared to their actual performance. This can be chalked up to unpredictability in the college football world, as there are always overperformers and underperformers in a season. The model has too few features to accurately predict such sudden changes, making it really only a general indicator for the next season. In the future, more features should be added to help improve the model's accuracy. One potential feature could be the strength of schedule for a team, as that is often a deciding factor for a team's quality, performance, and record. For the most part, however, the model works and is equally accurate to (and possibly even more accurate than) other popular predictions, such as ESPN's playoff predictions.

Bibliography

247Sports. "2026 247Sports Composite Player Rankings," 2025. https://247sports.com/season/2025-football-20251/overallteamrankings/?OrderBy=Points.

Connelly, Bill. X (formerly Twitter), 2025. https://x.com/ESPN_BillC/status/1790056401382920497.

ESPN. "2024 College Football Power Index - ESPN." ESPN, 2024. https://www.espn.com/college-football/fpi/_/season/2024.

———. "AP Top 25 Reaction - 2024-25 College Football Predictions - ESPN." ESPN.com. ESPN, August 12, 2024. https://www.espn.com/college-football/story/_/id/40751760/ap-top-25-reaction-2024-25-college-football-predictions.

Frost, Jim. "Mean Squared Error (MSE)." Statistics By Jim, November 12, 2021. https://statisticsbyjim.com/regression/mean-squared-error-mse/.

Sharma, Pulkit. "Distance Metrics | Different Distance Metrics in Machine Learning." Analytics Vidhya, February 25, 2020. https://www.analyticsvidhya.com/blog/2020/02/4-types-of-distance-metrics-in-machine-learning/.