# Predicting Fourth Down Conversions

STAT 451

Madeline Olson, Aiden Carney, Alec Stuedemann, Mathew Kuyper

**Introduction:**

The focus of this study examines factors of fourth-down plays and using them to predict whether a fourth down attempt is converted. Fourth-down plays are crucial moments in a game where teams must decide whether to advance the ball to continue their drive, making it an ideal point of study. The data originates from nflastr.com, containing NFL play-by-play data from 1999 to 2023. Within each year, there are 372 variables. The machine learning models used were support vector machine (SVM), logistic regression, and decision tree. Grid search found optimal models and hyperparameters. The best model performance was logistic regression, with an accuracy score of 0.848.

**Data:**

2020 was an outlier year because of COVID-19, so to ensure relevancy, the years 2021 through 2024 were selected, totaling 2,427 fourth down attempts. The initial selection of columns was guided by what might predict fourth down conversions, starting off with a total of 12 attributes. An initial look at the data is shown in Figure 1.

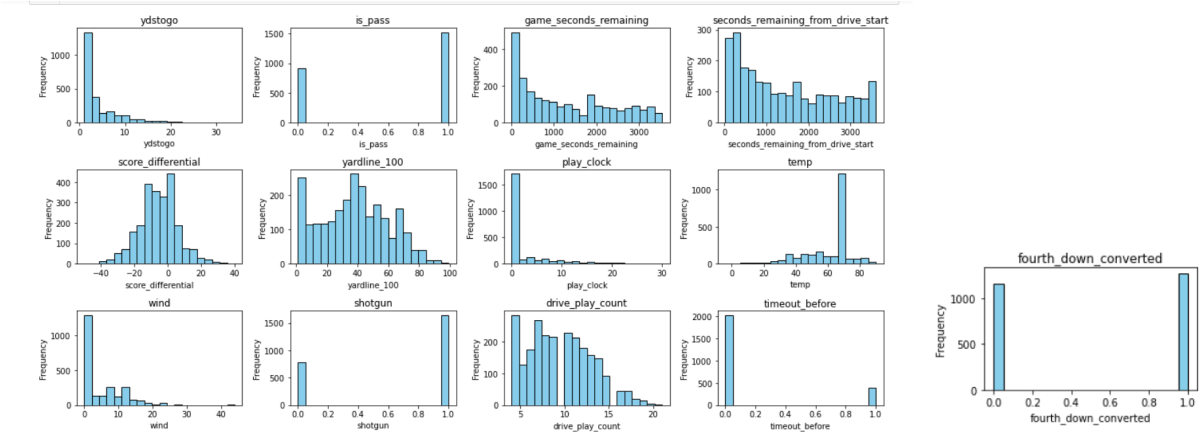


Figure 1 - Initial Features

While several variables (play_clock, temp, wind) have rough skews, they were removed later on during feature selection. Additionally, scaling data did not impact results.

**Machine Learning Analysis:**

The models chosen to test were support vector machine, logistic regression, and decision tree classifier. These models are best suited to binary classification and fit the context of our problem. Of the models chosen, the decision tree classifier has the best explainability but is prone to overfitting. The other two can be good for avoiding overfitting due to regularization terms but

have lower explainability with high-dimension data. Additionally, the SVM struggles to predict data that is not separable if the proper kernel is not used.

An 80%-10%-10% train-validate-test split of the data was used for model selection and hyperparameter tuning. Grid search was utilized to find the optimal model and hyperparameters. See Figure 2 for the hyperparameters tried for each model. The SVM and logistic regression models needed a max iteration parameter because the models failed to converge. Scaling the data did not solve this deficiency.

A second grid search with several features removed did not significantly impact the model to reduce overfitting. The final features utilized by the models were yards to go, pass or run, seconds remaining in the game, seconds remaining at the start of the drive, yard line, and number of plays in the drive. See Figure 3.

| Model | Parameters |
|---|---|
| SVM | **Kernel**: linear, rbf<br>**C:** 0.01, 1, 10, 100, 1000<br>**max_iter**: 10000 |
| Logistic Regression | **C:** 0.001, 0.01, 0.05, 0.1, 1, 10, 100, 1000<br>**max_iter:** 5000 |
| Decision Tree Classifier | **criterion:** entropy<br>**max_depth:** 1, 3, 5, …, 15 |

Figure 2 - Hyperparameter Tuning

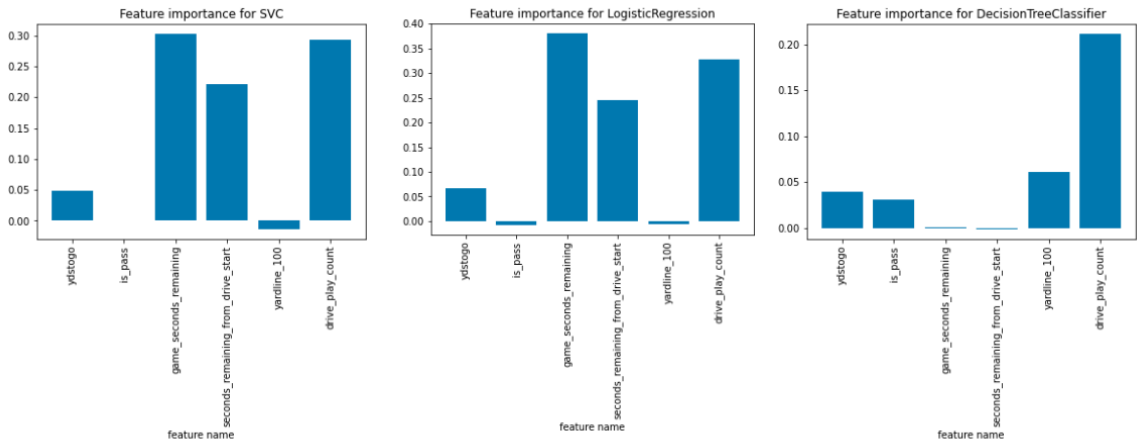


Figure 3 - Feature Importance

**Results:**

The logistic regression model with inverse regularization term 0.01 performed the best on the validation data and was selected for the model. See Figure 4 for a comparison of the other models. The small value for the inverse regularization term is likely used to account for larger values in the dataset because scaling was not used. The model had an accuracy score of 0.848 on the validation data and 0.835 on the unseen test data. The similarity of these scores implies that the model did not have a high level of overfitting.

| Model | Best Parameters | Validation Score |
|---|---|---|
| SVM | **kernel:** rbf<br>**C:** 1000 | 0.815 |
| Logistic Regression | **C:** 0.01 | 0.848 |
| Decision Tree Classifier | **criterion:** entropy<br>**max_depth:** 5 | 0.782 |

Figure 4 - Grid Search Outcome

**Conclusion:**

The goal was to identify key factors contributing to successful conversions on fourth downs. The models indicated that the longer a drive goes on, the more easily 1st downs are converted, as drive_play_count was the largest factor. Going for fourth downs earlier in the game was also exceptionally relevant. This holds logically, as teams, who skew conservative, will attempt only the most advantageous fourth downs earlier in the game. The initial anticipation was the yardage factor (ydstogo) would be the most relevant, but it was a distant fourth compared to these other features. A weakness of this model is that it does not predict future fourth down success, but whether one converted a fourth down or not after the fact.

| Member | Proposal | Coding | Presentation | Report |
|---|---|---|---|---|
| Aiden | 1 | 1 | 1 | 1 |
| Alec | 1 | 1 | 1 | 1 |
| Madeline | 1 | 1 | 1 | 1 |
| Matthew | 1 | 1 | 1 | 1 |