Telemetry Sports: Win Probability

Chloe Click, Michael Cruz, Denton Paul, Adam Quinn, Daniel Castro

INTRODUCTION

Telemetry Sports integrates and provides data for NFL and College football teams to help their clients have more scouting efficiency, as well as aiding in better decision making.

Our goal: Create a 4th-down situational book to provide to clients, giving coaches a tool to help aid in 4th-down decision making. In order to recommend the optimal choice, we must first model the expected results for each of the three 4th-down options:

Model 1 (Field Goal):
- Type: Classification
- Description: Predict whether a field goal was made or missed

Model 2 (Punt):
- Type: Regression
- Description: Predict the net distance on the punt

Model 3 (Go-For-It):
- Type: Regression
- Description: Predict the yards gained on the attempt. This will directly determine if the conversion was successful or not.

RESEARCH WORKFLOW

Data Collection:
- Our source of data was the College Football Data API accessed using Python.
- We used the SQLite database engine to store our play-by-play, team, and player information

Data Cleaning + Analysis:
- Dropping null and inconsistent data
- Scraping player names and important values from play description text
- Analyzing distribution and trends of occurrences (seen in Fig. 5).

Modeling:
- (1) Baseline Modeling
- (2) Feature Engineering (i.e., Player Stats)
- (3) Feature Selection
- (4) Hyperparameter Tuning
- (5) Build more Complex Model
- (6) Evaluate

CONCLUSION + FUTURE GOALS

Our team has developed three accurate models for the prediction of each 4th-down option. These expected outcomes can be paired with a third-party win percentage model to produce the results in Figure 4. We are very proud of our work and have learned how to apply the data science workflow to a sports setting. We hope to improve these models and create our own win percentage model in the future.

ACKNOWLEDGEMENTS

We would like to thank our Telemetry Sports Corporate Partner Mentors Nate Jahn and Tyler Fuelling for providing us with this opportunity and guiding us along the way. We would also like to thank Dr. Mark Daniel Ward, Maggie Ann Betz, Heather Goodwin, and David Glass for providing the resources and support needed for a successful completion of our project.

Up 14, 4th & 3, 40 yards from opponent end zone
Qtr 1, 02:59 | Timeouts: Off 3, Def 3

	Win %	Success %[1]	Win % if Fail	Win % if Succeed
Go for it	91	56	88	93
Punt	90	NA	NA	NA
Field goal attempt	89	51	87	91

[1] Likelihood of converting on 4th down or of making field goal
Source: @ben_bot_baldwin

Figure 4: Twitter Bot Output (Ben 2021)
(This is an example of what we are trying to replicate and improve)

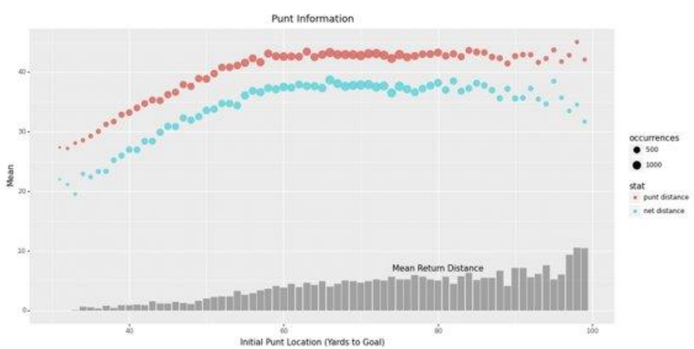

Figure 5: Punt Occurrences; Punt, Net, and Return Distance

Field Goal

- Final Model: XGBoost
- Optimized on ROC_AUC scoring function
- Optimal Features: Distance, Kicker's Success Rate, Kicker's Longest Made, Quarter, and Score Difference
- Optimal hyperparameters: default parameters with max tree depth of 5
- Evaluation: Model consistently runs with **84% accuracy**

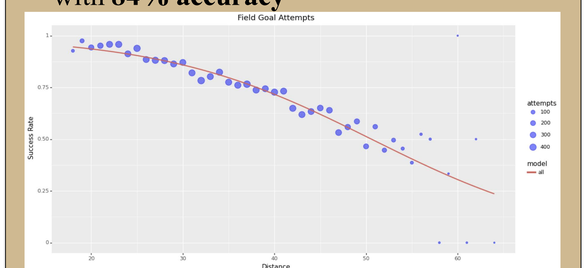

Figure 1: Field Goal Success Rate vs Distance

Punt

- Final Model: KNN Regressor
- Optimized on distance to neighbor
- Optimal Features: Yards-to-Goal, and Punter's Yards-per-Punt
- Optimal hyperparameters: K = 150
- Evaluation: MAE of **7. 3** and RMSE of **10.3**

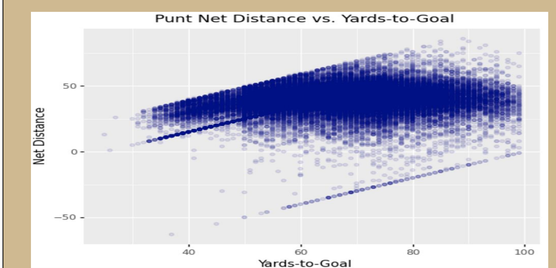

Figure 2: Punt Net Distance vs. Yards-to-Goal

Go-For-It

- Final Model: XGBoost
- Optimized on negative mean absolute error scoring metric
- Optimal Features: Yards-to-Goal, Distance, and Play Type
- Optimal hyperparameters: default parameters with max tree depth of 10
- Evaluation: Model consistently runs with MAE of **5.5** and RMSE of **8.2**

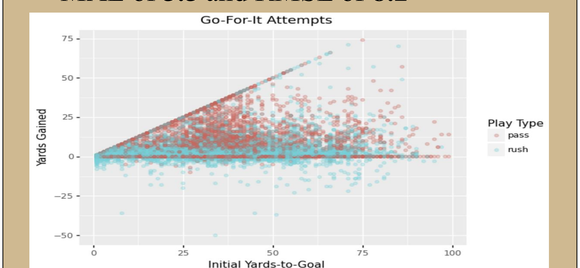

Figure 3: 4th-down Yards Gained vs. Init Yards-to-Goal