NFL Big Data Bowl 2021

Ty Hammond, Blake Hammarstrom, Jaret Stickrod, and Luke Spencer

Motivation and Design

Main Point: Understanding how defensive strategies impact NFL play outcomes.

- **Focus:** Investigate if more defenders in the box reduce offensive yards gained.
- **Significance:**
  - Helps coaches balance between rushing and coverage plays.
  - Guides decisions on optimizing defensive play-calling.
- **Inspiration:**
  - Shared passion for football and analytics.
  - Contribute to the growing field of sports analytics.

EDA

Data Quality

Methodology

- Data Preparation
  - Filtered passing plays; cleaned data to 17,330 observations. Response variable: offensive yards gained (excluding penalties). Predictors: number of defenders in the box and pass rushers. Exploratory Data Analysis (EDA):
  - Visual tools: boxplots, heatmaps, correlation matrix. Pass rushers showed slightly stronger correlation with play result (0.032) than defenders in the box (0.006).
- **Modeling:**
  - Linear Regression: Captures direct linear relationships.
  - Generalized Additive Model (GAM): Accounts for non-linear relationships with smooth terms.
  - Train-test split (80/20) and 5-fold cross-validation for performance evaluation.
  - Metrics: RMSE, MAE, and R-squared.

Results

Main Point: Defensive predictors have weak explanatory power for play outcomes.

- **Linear Regression:**
  - Adjusted R-squared = 0.0005 (predictors explain 0.05% variance).
  - Significant predictor: Pass rushers (estimate = 0.30, p = 0.002).
    Test RMSE: 9.892.
- **GAM Results:**
  - Adjusted R-squared = 0.0014 (slight improvement).
  - Significant non-linear relationships for both predictors ($p < 0.001$) &Test RMSE: 9.879.

Conclusion

References

The National Football League. 2021. NFL Big Data Bowl 2021. Retrieved Oct 29, 2024, from https://www.kaggle.com/competitions/nfl-big-data-bowl-2021/data

Kaplan, Andee. Statistical Learning, Class Notes. PDF File. 2024. https://dsci445-csu.github.io/notes/2_stat_learning/20240829_2_stat_learning.pdf

Kaplan, Andee. Regression, Class Notes. PDF File. 2024. https://dsci445-csu.github.io/notes/3_regression/20240910_3_regression.pdf

Kaplan, Andee. Classification, Class Notes. PDF File. 2024. https://dsci445-csu.github.io/notes/4_classification/20240919_4_classification.pdf

Kaplan, Andee. Regularization, Class Notes. PDF File. 2024. https://dsci445-csu.github.io/notes/6_regularization/20241015_6_regularization.pdf