

# NFL Big Data Bowl 2025 - Player Performance and Prediction Analysis

## Introduction

The goal of this project is to leverage tracking data from NFL games to evaluate player performance, focusing on pre-snap player movement, and to develop a predictive model to assist in understanding key aspects of the game. This analysis explores player speed, acceleration, and position-specific behaviors, ultimately aiming to provide valuable insights that can be used by NFL teams or the league office for weekly game planning and performance evaluation.

## Data Overview

The project uses three datasets:

1. **Plays Dataset:** Contains details about each play, including game context (quarter, down, yards to go), play type, offensive and defensive formations.
2. **Players Dataset:** Includes player details such as position, height, weight, and other attributes.
3. **Tracking Week 1 Dataset:** Captures player locations and movement on the field for each play, including their speed, acceleration, and direction at different frames.

## Data Preprocessing and Feature Engineering

To effectively use the tracking data, significant preprocessing and feature engineering steps were taken:

1. **Memory Optimization:** Data types were optimized to handle large datasets, reducing memory usage and improving computational efficiency.
2. **Handling Missing Values:** Missing values were handled using a combination of **SimpleImputer** for numerical data and **fillna** for categorical data.
3. **Feature Engineering:**
  - **Speed and Acceleration Metrics:** Average, maximum, and minimum speeds and accelerations before the snap were calculated for each player.
  - **Distance from Line of Scrimmage (LOS):** Calculated to measure each player's initial positioning relative to the line of scrimmage.
  - **One-Hot Encoding:** Offensive formations and defensive coverage types were encoded to make them usable for machine learning models.

## Model Building and Evaluation

A **Logistic Regression** model was used to classify the outcomes of each play. The features included metrics like **speed, acceleration, distance from LOS**, and **formation types**. The model achieved:

- **Accuracy:** 95%
- **Precision:** 95%
- **Recall:** 95%
- **F1 Score:** 94%

These metrics indicate a well-performing model that can identify play outcomes accurately, particularly focusing on individual player contributions during pre-snap and immediately after the snap.

## Football Score Evaluation

- **Week-to-Week Usage:** The current analysis provides detailed insights into player movement, which could help teams assess individual player readiness, identify mismatches, and track consistency over a season. Incorporating more **game context variables** (e.g., down, distance, score) could improve the usefulness for weekly planning.
- **Accounting for Complexity:** Football data involves many contextual factors, such as opponent strategies and situational play calls. The current model partially addresses these complexities by including **defensive and offensive formations** but could be expanded to provide deeper situational insights.
- **Uniqueness of Ideas:** Analyzing pre-snap movement metrics like **average speed and acceleration** adds value, but further innovation could be achieved by focusing on **matchup-specific outcomes** or **player-vs-player performance** over multiple games.

## Data Science Evaluation

- **Correctness:** The workflow includes **data cleaning, imputation, feature scaling**, and appropriate model selection. Handling of missing values and standardization was performed correctly, leading to reliable model performance.
- **Backing Up Claims:** The claims are well-supported with **metrics and visualizations** showing distribution and player performance patterns. Future improvements could include **statistical tests** for hypothesis validation.

- **Statistical Model Appropriateness:** Logistic Regression was chosen for interpretability. However, given the **non-linear relationships** in football data, more advanced models like **Random Forests** or **Neural Networks** could be explored.
- **Innovation:** Using tracking data for **pre-snap analysis** is a unique aspect of the analysis, which can be further enhanced by adding predictive components for play success or player matchup efficiency.

## Report Evaluation

- **Clarity:** The report is structured with logical sections, but more narrative context could be added to explain **the motivation** behind each step of analysis and **the implications of the findings**.
- **Ease of Following:** More **in-depth explanations** of the football context for each feature, such as why certain formations matter, would improve readability for non-experts.
- **Motivation:** While player evaluation and performance tracking are implied, clearly defining objectives like **player readiness evaluation** or **team strategy insights** would provide better direction to the report.

## Data Visualization

- **Accessibility:** The charts, including **heatmaps**, **bar plots**, and **histograms**, are informative, but some would benefit from **additional annotations** or **legends** to improve accessibility.
- **Accuracy:** The visualizations are accurate and help convey the findings, particularly the distribution of player speed and position counts. Including **comparative visualizations** could add more depth.
- **Innovation:** The visualizations used are largely standard. Introducing **interactive plots** or **more dynamic visualizations** (e.g., play animations, player tracking paths) would increase engagement and provide a clearer picture of player movements.

## Conclusion and Next Steps

The current model and analysis provide a strong foundation for evaluating **player performance metrics** using NFL tracking data. To improve the project's impact, the following next steps are suggested:

- Incorporate **contextual game variables** (e.g., down, distance, quarter) to give more situational insights.

- **Expand the model** to incorporate **non-linear relationships** through **tree-based models** or **neural networks**.
- Improve **visualization quality** by adding **interactive elements** or **animated player tracking** to illustrate player paths.
- Develop predictive models for **specific game situations** or **matchups**, making the analysis more actionable for NFL teams.

The overall objective is to provide meaningful insights that NFL teams and analysts can use to improve performance and optimize game-time decisions. By iterating on the model, expanding the dataset features, and enhancing the analysis context, this project has the potential to deliver even more value for week-to-week football analysis.