

Projecting Wide Receiver Stats With Machine Learning

By Duncan Park and Sameer Nathawat





Overview

01

**Background and
Overview**

02

**Gaussian Mixture
Model & Random
Forest**

03

**Data Collection &
Preparation**

04

**Results &
Conclusion**



Background



- Predicting WR performance is challenging due to variability in their usage and many other details. While basic statistics provide insight, they don't fully capture how a player's performance evolves over time.
- Our goal was to generate projected statistics & rankings for WRs in 2026 using a **Random Forest** for the stats and a **Gaussian Mixture Model** to help find hidden correlation within tiers of WRs, and rank them via a composite score.

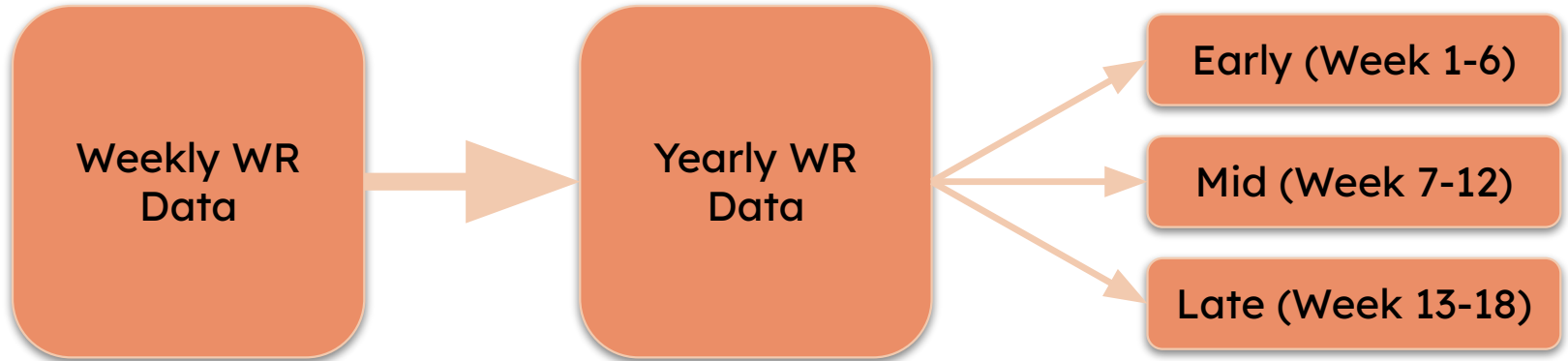


Collecting and Cleaning Data

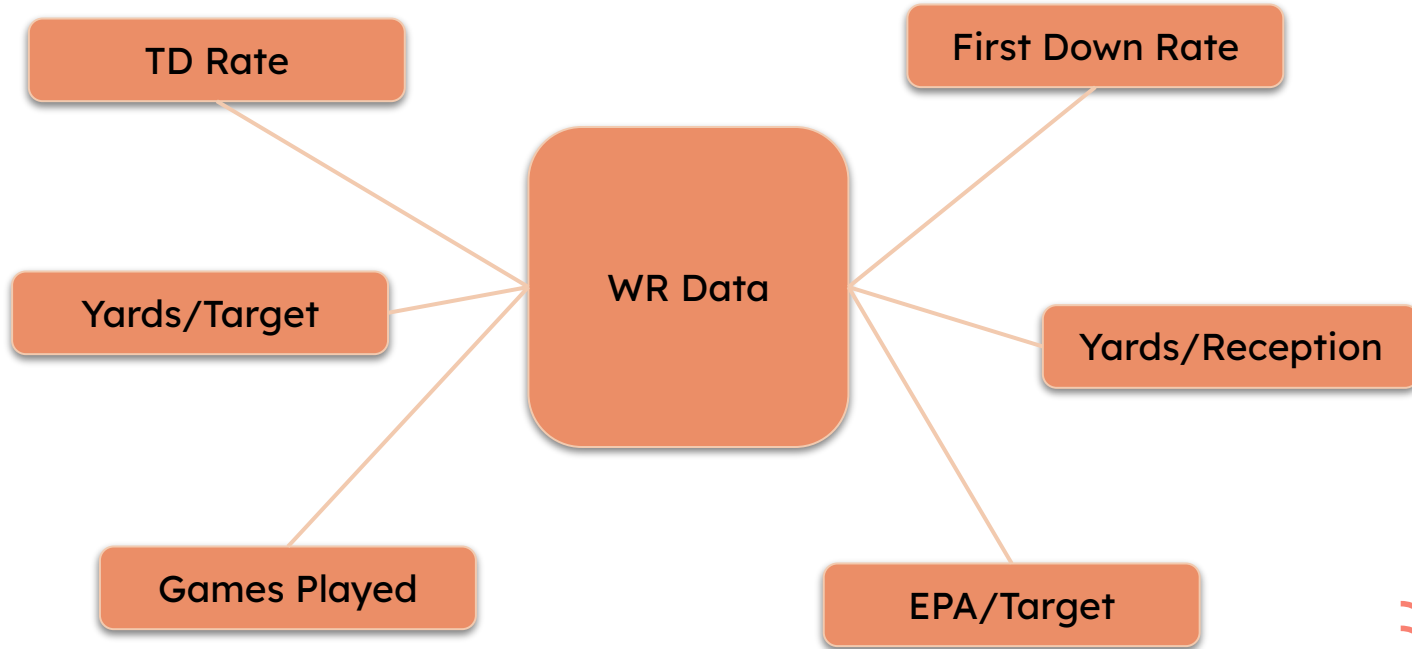
- Weekly wide receiver data from 2012-2025 using the **nflreadpy** Python package
- Data included:
 - Yards, targets, TDs
 - Snaps played, snaps %
 - Fantasy points
- Only wide receivers who **played more than 20% of their teams offense snaps** were included



Feature Engineering



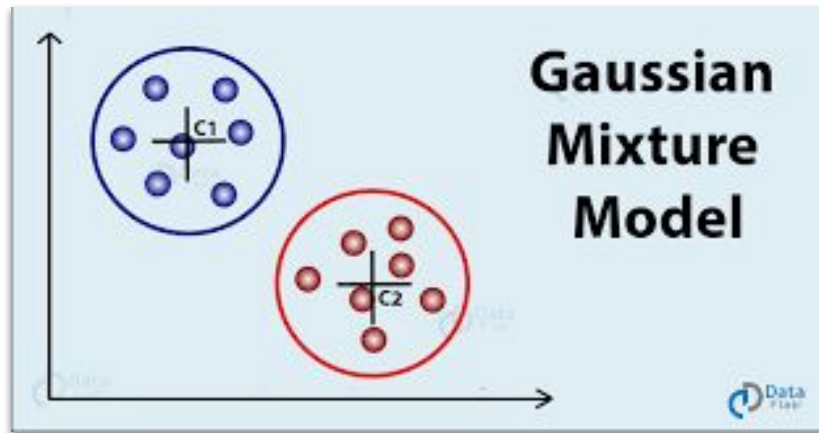
Feature Engineering



Gaussian Mixture Model & Random Forest

What is GMM?

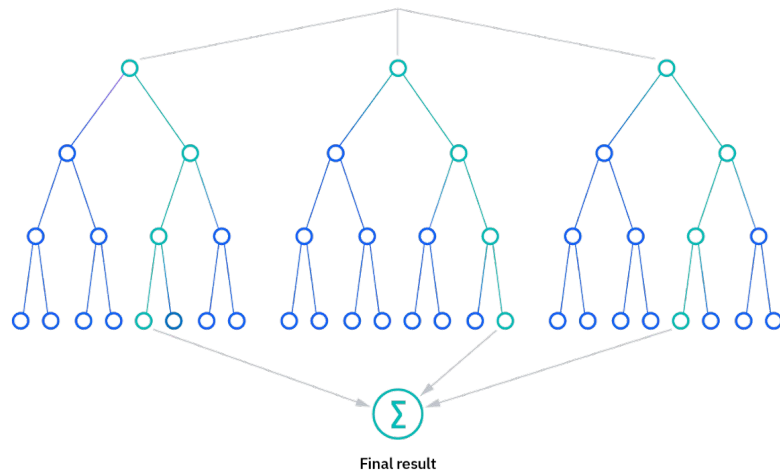
- Probabilistic clustering method; assumes data is generated from a combination of several underlying distributions called Gaussians.
- Estimates the probability that a point belongs to each cluster
- Allows for more flexible and realistic modeling of complex data.
- Each cluster is defined by its own mean and variance.
- Makes GMM useful when clusters overlap/not clearly separated



Gaussian Mixture Model & Random Forest

What is Random Forest?

- ML model that combines many decision trees to make predictions.
- Each tree is trained on a random subset of data & features → helps reduce overfitting.
- Makes a final prediction by averaging the outputs of all the trees
- Captures complex, nonlinear relationships between variables.



How We Calculated Rankings



Trained on historical data from 2012-2024:

- Created year-to-year pairs so the model learns how a player's current season predicts their next season.

Projected 2026 stats using Random Forest:

- Built three models to predict **receptions, yards, and touchdowns**
- Took in 2025 player stats, outputted projected 2026 production

Identified player archetypes using GMM:

- Clustered historical seasons into **5 tiers** based on performance profiles
- Each player receives a **probability of belonging to each cluster**

Converted probabilities into a GMM score:

- Clusters ranked by average receiving yards
- Best cluster = highest score; worst = lowest
- Each player gets a **GMM score = weighted average of cluster probabilities**

Normalized all metrics (0-1 scale)

- Ensured GMM score and projected stats are comparable

Calculated a composite score:

$$\text{COMPOSITE SCORE} = 0.5 \cdot \text{GMM Score} + 0.3 \cdot \text{Proj. Yds} + 0.1 \cdot \text{Proj Rec.} + 0.2 \cdot \text{Proj TDs}$$



Results

[Table of Results](#)

Some things to note:

- Forecasted less outstandingly good seasons (like Puka/JSN's 2025 seasons) and also less extremely bad seasons (anyone with under 200 yards was projected to have over 200 yards)
- Big jumps from 1-2 to 3-7 to 8-10 and the rest
 - It makes sense given what we know about these players
- We chose not to round the projections, even though you can't have 0.5 of a TD
- Doesn't consider injury, QB change, or playtime



Future Projects

- **Developing a breakout prediction model** by building a classification model to predict which receivers will significantly outperform expectations the following year.
- **Expand to other positions** with the same modeling framework and analyze running backs or tight ends to compare positional trends.
- **Add uncertainty and ranges** to generate ceiling/floor projections instead of simply single-point estimates for each player.





Thank you for listening!

Questions, concerns,
comments?