Monitoring Motion

NFL Big Data Bowl
2025 Kaggle
Competition

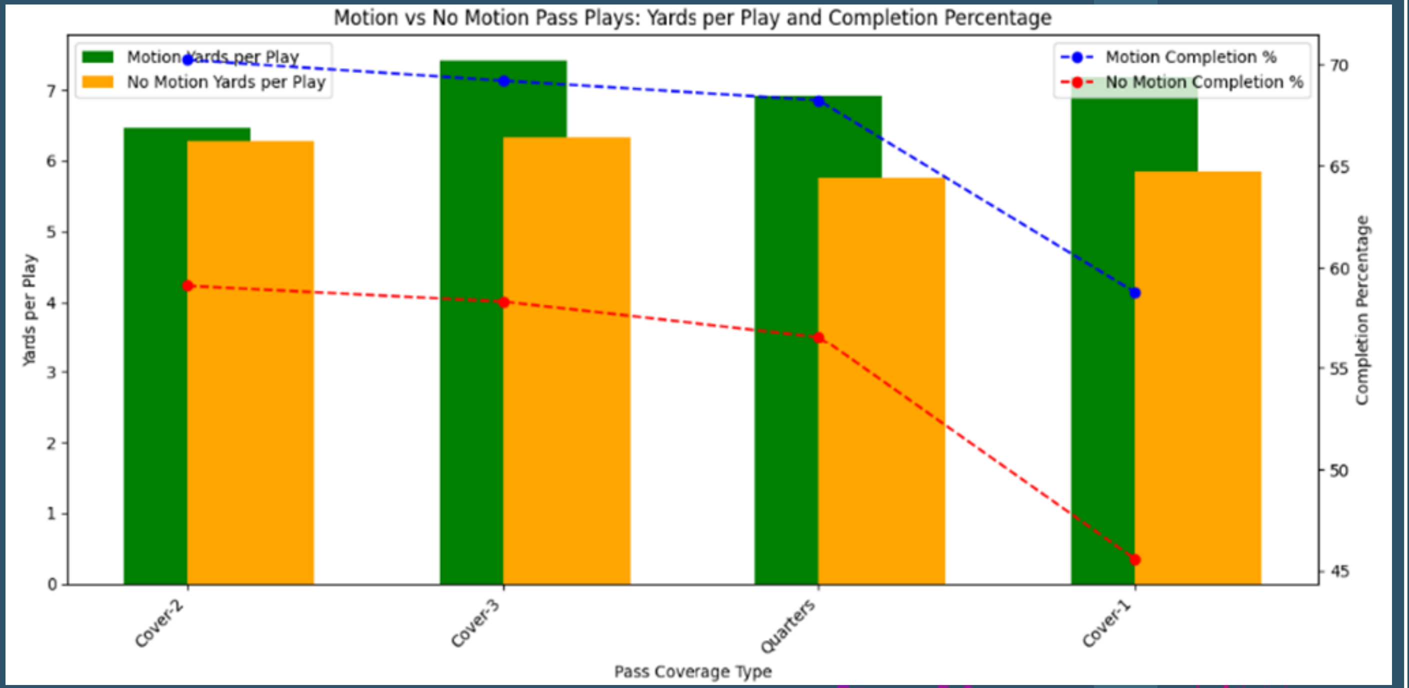
Coleton Reitan

Analysis Overview

1. Introduction & Goal
2. Receiver Labeling Process
3. Feature Engineering & Understanding
4. Clustering Motion Plays – Attributes & Clusters
5. Predicting Receiver Routes – Setting Up & Method
6. Analysis – Performance Discussion & Results
7. Summary – Discussion & Forward Thoughts

Introduction

- In today’s NFL, offenses are more versatile than ever, constantly pushing defenses to adapt and respond. Despite their adjustments, defenses continue to struggle against one particular strategy — **motion pass plays.**
- These plays have become some of the most effective tools in an offense’s playbook, creating mismatches and exploiting defensive weaknesses, as a clear discrepancy is shown between motion and non-motion pass plays in terms of completion percentage and yards gained.
- This analysis focuses on motion pass plays from weeks 1 through 9 of the 2021 NFL season against four of the common defensive coverages — Cover-1, Cover-2, Cover-3, and Quarters.

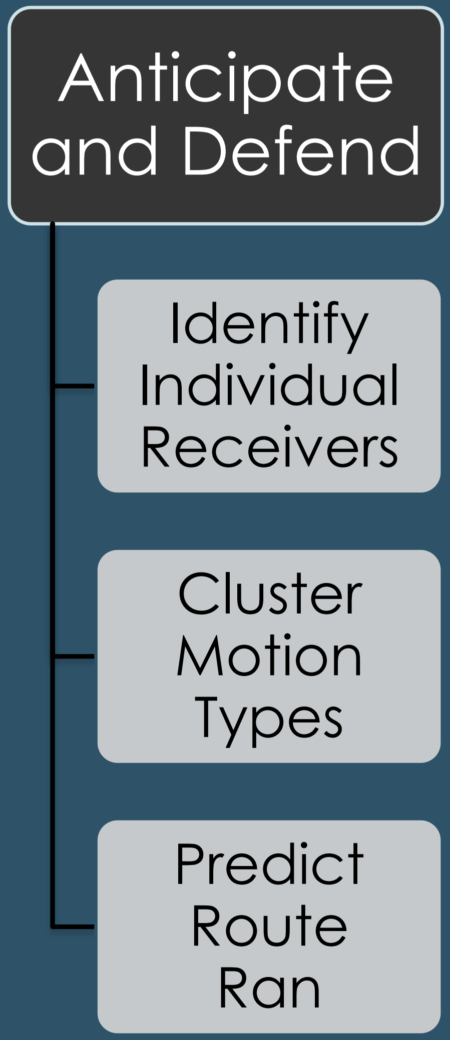


Introduction - Goal

- The goal of this analysis is to **predict eligible receiver routes on motion pass plays** to help defenses improve their ability to anticipate and defend against these plays effectively.

- To achieve this, motion pass plays are grouped into 5 distinct motion types using clustering techniques. Within each cluster, the focus shifts to predicting individual receiver routes based on patterns identified in the data.

- This approach provides defenses with actionable insights into offensive tendencies, giving them a better chance to counter these high-success plays.



Receiver Labeling Process

- To ensure consistent receiver labeling, an offensive strong side label was created. Based on this label, each eligible receiver on the field was assigned a number (1–5) relative to the strong side.
- This approach maintains consistency even if the formation is mirrored (flipped)—the receiver labels will mirror accordingly.
- Additionally, variables were created for both line set and ball snap events to capture key moments prior to each play

Strong Side

Variables:

- Line_set_strong_side
- Ball_snap_strong_side

- Determines the side of the formation with the most eligible receivers
- Eligible receivers 4 or more yards behind the LOS are not used to determine strong side (such as RB or FB)
- If each side has the same number of receivers, the side where the TE is present gets labeled the strong side

Receiver Label

Variables:

- Line_set_receiver_label
- Ball_snap_receiver_label

- Numeric labeling of eligible receivers starting from 1
- Count begins on strong side and increments with next closest receiver
- Eligible receivers 4 or more yards behind the LOS are counted last

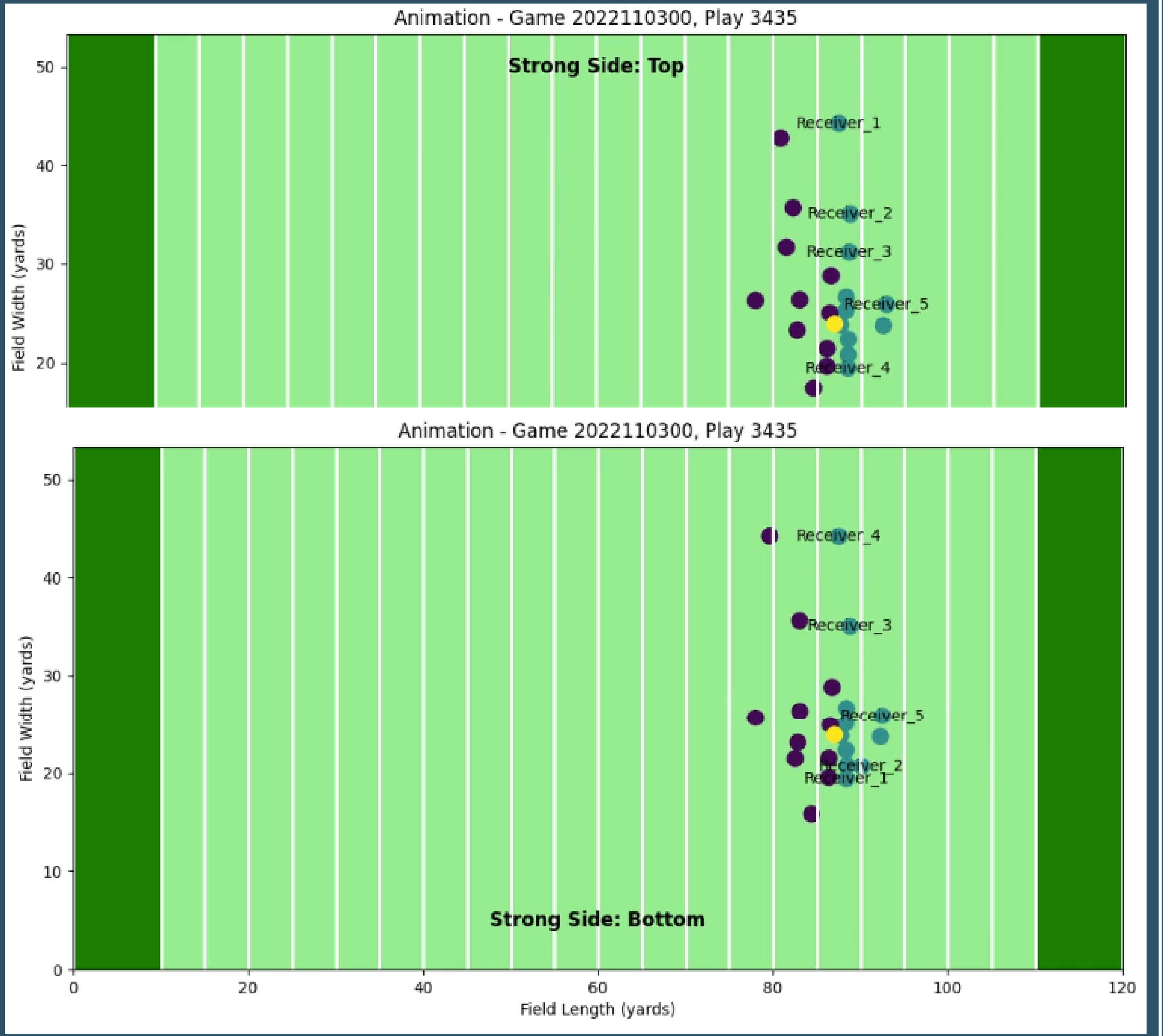
Receiver Labeling Visualization

Line Set Event

The Strong side is the top, and receiver count starts top to bottom.

Ball Snap Event

The strong side is now the bottom (after the motion) and receiver count starts bottom to top.



Feature Engineering

- The goal of feature engineering in this project was to create **clear and easily recognizable variables** that defensive players could quickly interpret.
- These variables were then considered for inputs during **motion clustering**, enabling the identification of patterns and tendencies within motion pass plays to support defensive strategies.

Motion Distance

- Measures how many yards a motion receiver moves from their original Y position during motion.

Motion Direction

- Indicates the direction a motion receiver is moving relative to the ball's position and distance to the sideline (short side or long side).

Motion Speed

- Captures the maximum speed a motion receiver reaches during motion.

Crosses Ball

- Identifies whether a motion receiver crosses the Y position of the ball during motion.

Feature Understanding

- In addition to the engineered variables, the original dataset also included some clear and easily recognizable variables that defensive players could quickly interpret.
- These pre-existing variables were also reviewed and considered as potential inputs for motion clustering to enhance defensive strategies.

Yards To Go

- Distance needed for a first down

Offensive Formation

- Formation used by offensive team

In Motion at Ball Snap

- Indicates whether the player was in motion at ball snap

Receiver Alignment

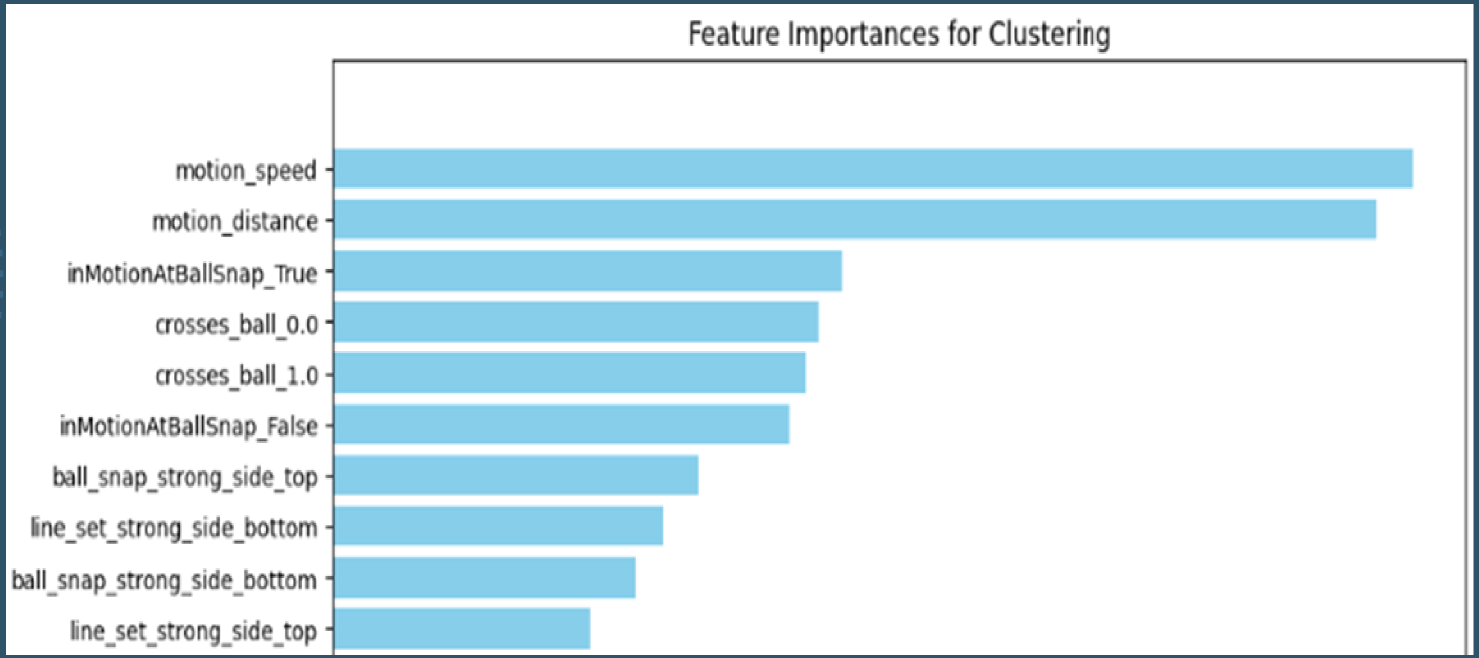
- Alignment of receivers in offensive formation

Clustering Motion Plays
- Attributes -

- A **random forest model** was used to identify the most important features for defining **motion type clusters**.
  - Note; the clustering data is based on the motion receiver.

- The following variables were included in the clustering process:

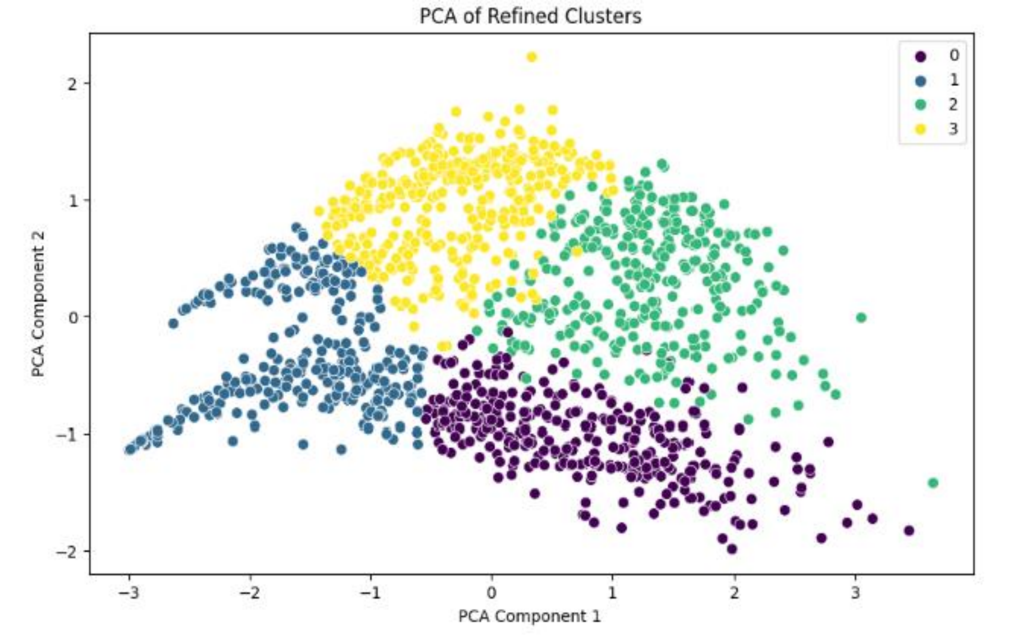
1. Motion Speed
2. Motion Distance
3. Crosses Ball
4. In Motion at Ball Snap
5. Ball Snap Strong Side



Clustering Motion Plays
- Clusters -

An elbow and silhouette plot were used to determine the optimal number of clusters within the most important motion play attributes.

- The results indicate that both 4 and 6 clusters provide similar performance.
- **4 clusters may be more practical for defensive personnel to recognize** and develop actionable strategies against.
- The PCA plot further illustrates the grouping and fundamental separation of these clusters, highlighting distinct patterns in the data.



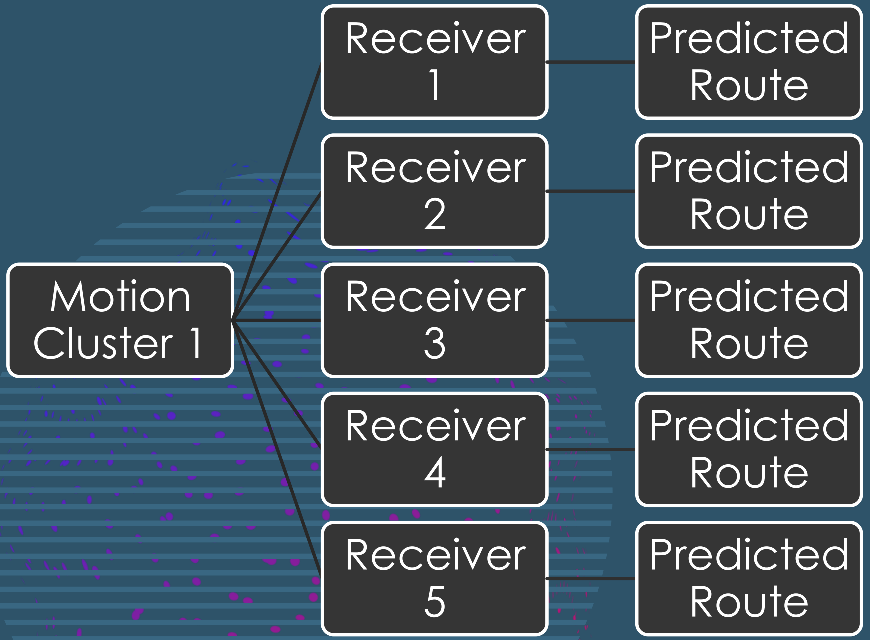
Motion Type Cluster Table

	Motion Speed (Yds/sec)	Motion Distance (Yards)	Crosses Ball	In Motion at Ball Snap	Ball Snap Strong Side	Motion Description
Cluster 1	5.10	12.18	Yes	No	Top	A medium-speed, high-distance motion that crosses the ball, creates a top strong side, and gets set (Calm motion across the field)
Cluster 2	2.51	2.65	No	No	Bottom	A low-speed, low-distance motion that does not cross the ball, holds a bottom strong side, and gets set (Slight motion, often from TE or RB)
Cluster 3	6.36	11.21	Yes	Yes	Top	A high-speed, high-distance motion that crosses the ball, creates a top strong side, and does not get set (Jet Motion Type)
Cluster 4	5.56	4.20	No	Yes	Top	A high-speed, medium-distance motion that does not cross the ball, holds a top strong side, and does not get set (Jet Motion Type, but ball snap before crossing)

- Speed and Distance shows the Average values,
- Crosses, In Motion, and Strong side show Mode values

Predicting Receiver Routes – Setting Up

- To begin the process of predicting receiver routes, eligible **receivers were first grouped based on their respective clusters**. Remember, each receiver is assigned a label (1-5) for each play, ensuring consistency in how they are identified and tracked throughout the dataset. This labeling system is crucial for **maintaining uniformity** across the analysis.



(This representation is repeated for each cluster)

- Within each cluster, receivers are isolated into individual data frames that **contain only other receivers from the same cluster**. This approach enables a more focused analysis, where each receiver is considered in the context of their cluster, capturing the nuances of how they interact with teammates and their specific play characteristics.

- Next, all the variables that contributed to the clustering process, along with additional relevant features pertaining to the play pre-snap, such as offensive formation, were utilized. By leveraging these variables, advanced techniques can be applied to predict the routes each receiver will run based on the historical data and patterns within their cluster.

Predicting Receiver Routes – Method

- The **Random Forest Classifier**, a multi-class classification model, was utilized to predict the route ran by each receiver.
  - Before training the model, a feature importance analysis was performed to identify the most influential variables for each receiver. This process was conducted individually for each receiver within each cluster.
  - Based on the results of the feature importance test, the top 5 most significant features were selected and used as input for training the model, ensuring that only the most relevant variables were considered for each receiver in their respective clusters.

Accuracy Report Table	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Receiver 1	0.15	0.00	0.38	0.29
Receiver 2	0.07	0.25	0.29	0.62
Receiver 3	0.25	0.14	0.43	0.53
Receiver 4	0.33	0.20	0.40	0.16
Receiver 5	0.66	0.10	0.43	0.75

- While the individual accuracy scores may seem modest at first glance, it's crucial to remember that the true value lies in the predictions as a whole.

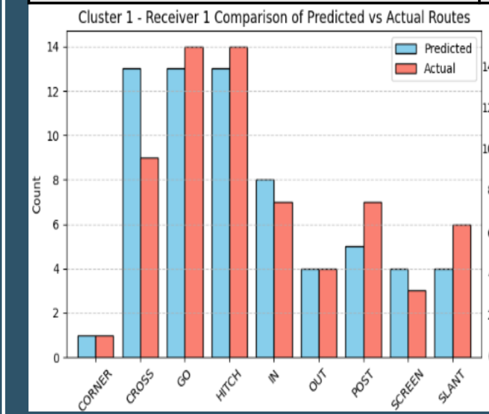
Analysis – Performance Discussion

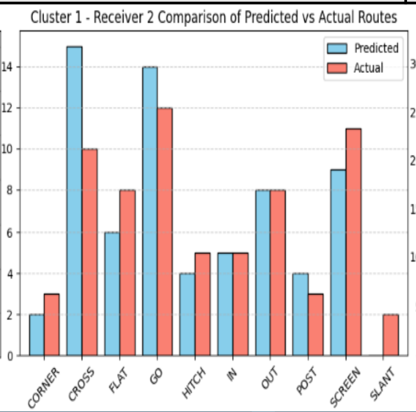
- The primary drivers of the predicted routes are the **clusters themselves**, as they encapsulate essential patterns and relationships that influence the outcome.
- Therefore, even if the accuracy for each individual receiver within a cluster isn't perfect, **the collective predictions** — driven by the clustering — are key to understanding and predicting the route behavior effectively.
- Investigating the MAE (Mean Absolute Error) is more important in this situation as it directly measures how close the model's predicted counts are to the actual counts of each route, which is essential for assessing the accuracy of the model's predictions on a continuous scale. This helps ensure the model is making predictions that closely reflect the actual distribution of routes.
  - The MAE measures how close the model's predictions are to the actual values, with a lower MAE indicating better performance. Simply put, it reflects the average difference between the predicted and actual route counts for each cluster and receiver, so an MAE of 1.5 means the model was off by approximately 1.5 routes on average.

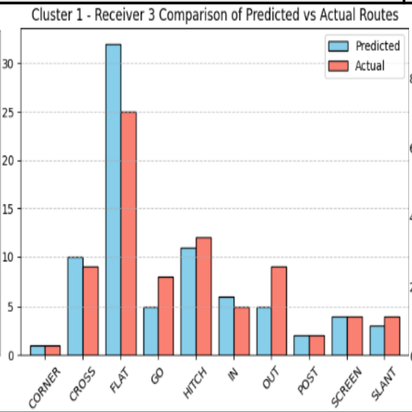
Analysis - Results

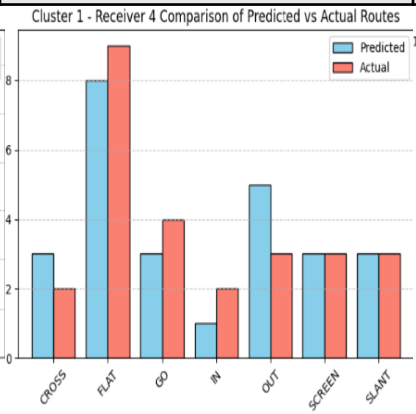
MAE Cluster 1

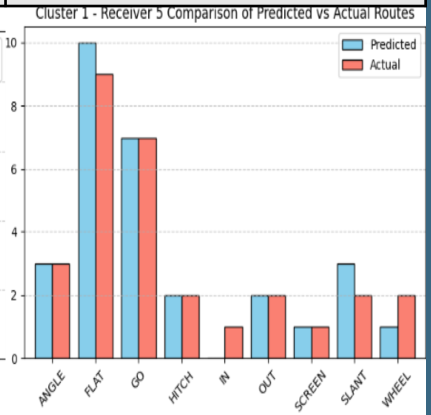
Receiver 1: 1.33	Receiver 2: 1.56	Receiver 3: 1.80	Receiver 4: 0.86	Receiver 5: 0.38





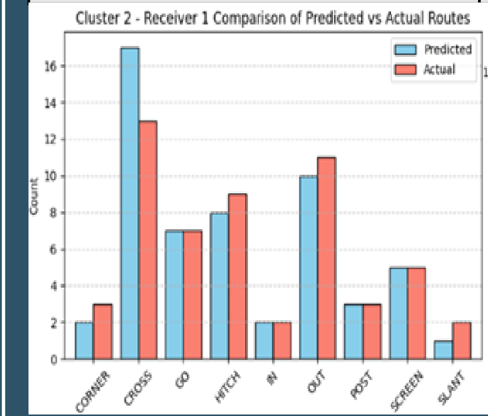


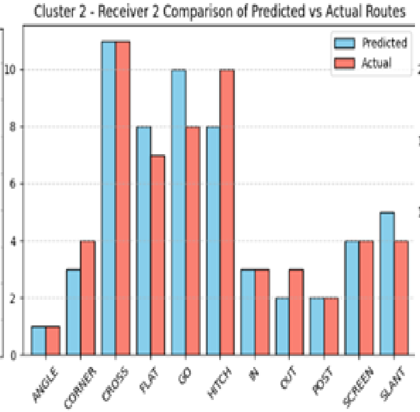


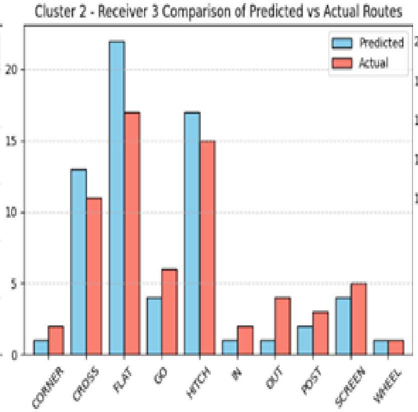


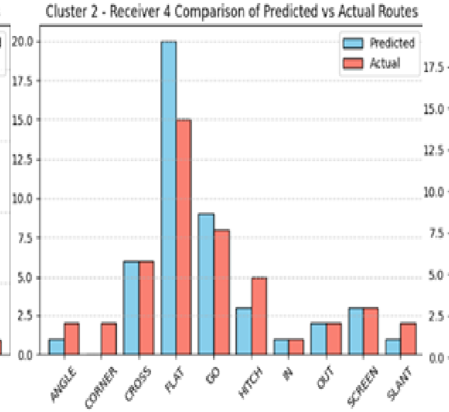
MAE Cluster 2

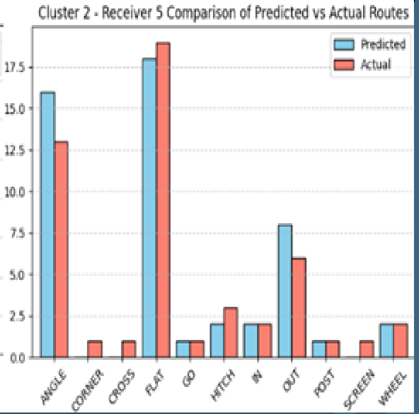
Receiver 1: 0.89	Receiver 2: 0.73	Receiver 3: 1.80	Receiver 4: 1.11	Receiver 5: 0.88







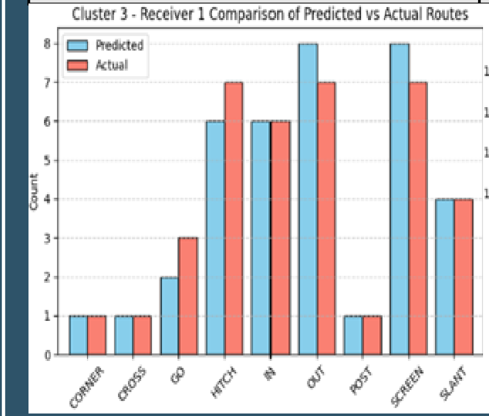




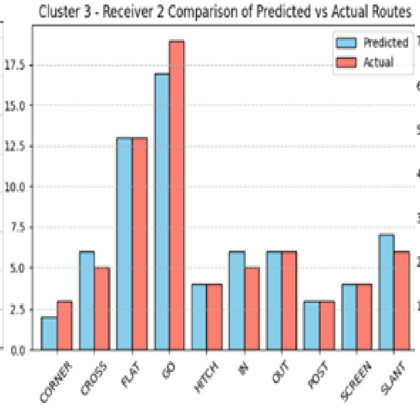
Analysis - Results

MAE Cluster 3

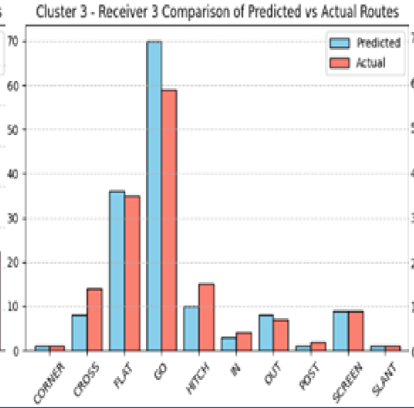
Receiver 1: 0.44



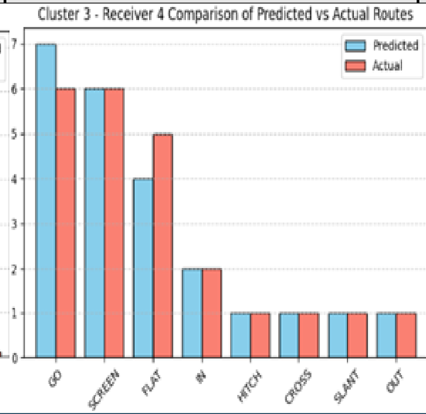
Receiver 2: 0.60



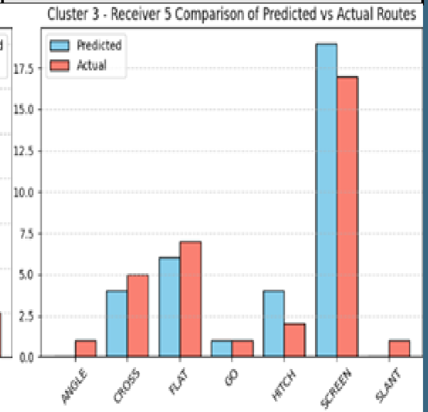
Receiver 3: 2.60



Receiver 4: 0.25

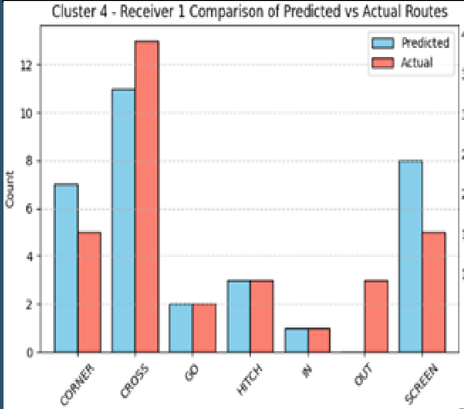


Receiver 5: 1.20

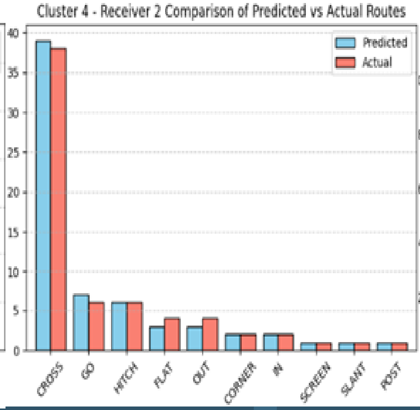


MAE Cluster 4

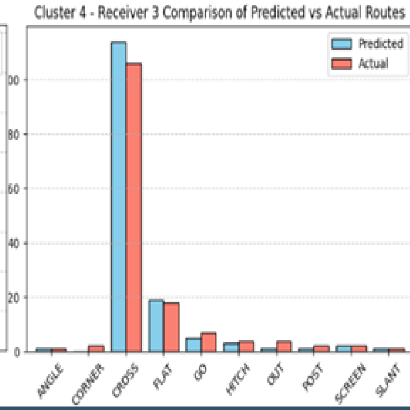
Receiver 1: 1.17



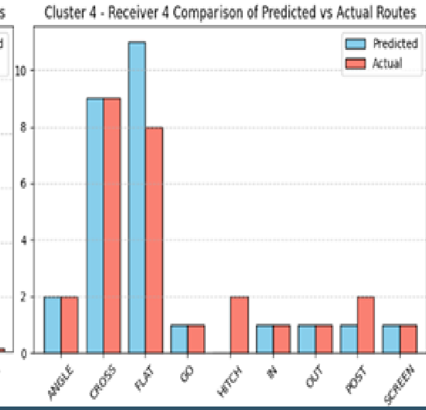
Receiver 2: 0.40



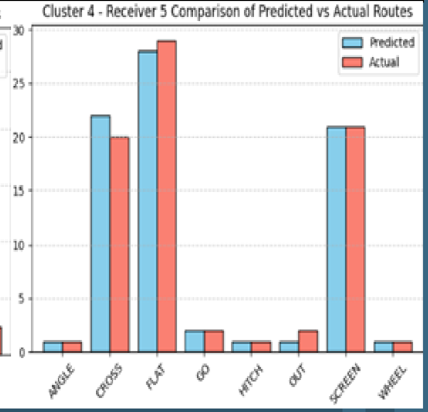
Receiver 3: 1.78



Receiver 4: 0.50



Receiver 5: 0.50



Summary - Discussion

- The MAE highlighted a key insight in the analysis, revealing that the predictions were more consistent and reliable when assessed within their respective clusters, compared to when evaluated individually

- Shown is a representation of the most common predicted routes ran by each receiver within each cluster

Summary – Forward Thoughts

- For teams seeking to enhance their defense against offenses through predictive analytics, this analysis provides a fresh perspective on motion pass plays and could mark the beginning of a shift away from motion-heavy offenses.
- To draw more definitive conclusions, the study would benefit from focusing on individual teams rather than the league as a whole.
- Further analysis could take into consideration the play results, combined with in game play tendencies, which in turn could lead to a deeper understanding of the motion play types and their objectives.

In total, this project aimed to predict route types in motion pass plays for NFL teams, using a range of features such as offensive formation, motion characteristics, and receiver alignment. By applying machine learning models like Random Forest and analyzing feature importance, the project explored how different factors influenced the prediction of pass routes. The key focus was on predicting routes accurately across different clusters of data representing various team formations and motion strategies. The analysis used Mean Absolute Error (MAE) to assess model performance, highlighting the models' ability to predict routes more accurately when considering the data as a whole within each cluster rather than on an individual basis. The findings could potentially offer teams valuable insights for refining their defensive strategies against motion-heavy offenses.

Thank You!

NFL Big Data Bowl
2025 Kaggle
Competition

Coleton Reitan
coletonreitan@gmail.com

Code Link:
https://www.kaggle.com/code/coletonreitan21/motion-analysis-3