Fantasy Predictions

Fort Hunter, Vincent Broda, Austin Smith, Dillon Frankenstein

Objective

Make a model to predict Fantasy Football scores.

We want to be able to predict and gain insight to the top overall performers for every position each week.

Motivation

- Selecting fantasy football players each week is quite risky and players might feel overwhelmed by the sheer amount of choices.
- Enabling players to be more confident and secure in their choices will increase enjoyment.
- There is often prizes or punishments associated with placement in a fantasy league.

Technologies Used

- Scikit-Learn
  - Linear regression, training/testing pipeline
- Pandas & NumPy
  - Data processing and management
- Seaborn & Matplotlib
  - Visualization of data and results
- NFL Data-Py
  - Up-to-date NFL statistics (2022–2024)
- PyTorch
  - Player and team embeddings

Dataset

- Data Source
  - nfl-data-py: Python library for NFL statistics spanning the 2022–2024 seasons.
- Key Features Gathered
  - Skill Positions: Quarterbacks, running backs, wide receivers, and tight ends.
  - Offensive Stats: Both basic and advanced statistics.
  - Game Context: Weekly schedules, home/away teams, and final scores.
  - Coaching: Names of home/away head coaches.
  - Fantasy Scoring: Player-level stats such as PPR fantasy points.

Weekly Player Dataset

```
team,player_id,player_name,position,season,week,game_type,pass_attempts,complete_pass,incomplete_pa
TEN,00-0035676,A.J. Brown,WR,2019,1,REG,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,4.0,3.0,100.0,54.0,56.0,0.0
TEN,00-0035676,A.J. Brown,WR,2019,2,REG,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,5.0,3.0,25.0,25.0,8.0,0.0,
TEN,00-0035676,A.J. Brown,WR,2019,3,REG,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,5.0,1.0,4.0,69.0,-2.0,0.0,
TEN,00-0035676,A.J. Brown,WR,2019,4,REG,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,3.0,3.0,94.0,43.0,51.0,2.0,
TEN,00-0035676,A.J. Brown,WR,2019,5,REG,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,2.0,2.0,27.0,25.0,2.0,0.0,
TEN,00-0035676,A.J. Brown,WR,2019,7,REG,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,8.0,6.0,64.0,56.0,17.0,0.0,
TEN,00-0035676,A.J. Brown,WR,2019,8,REG,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,3.0,2.0,11.0,32.0,1.0,1.0,
TEN,00-0035676,A.J. Brown,WR,2019,9,REG,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,7.0,4.0,81.0,102.0,25.0,0.0
TEN,00-0035676,A.J. Brown,WR,2019,10,REG,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,4.0,1.0,17.0,89.0,12.0,0.0
TEN,00-0035676,A.J. Brown,WR,2019,12,REG,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,5.0,4.0,135.0,47.0,95.0,1.
TEN,00-0035676,A.J. Brown,WR,2019,13,REG,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,4.0,3.0,45.0,38.0,17.0,0.0
TEN,00-0035676,A.J. Brown,WR,2019,14,REG,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,7.0,5.0,153.0,115.0,84.0,2
TEN,00-0035676,A.J. Brown,WR,2019,15,REG,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,13.0,8.0,114.0,178.0,36.0,
TEN,00-0035676,A.J. Brown,WR,2019,16,REG,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,2.0,1.0,34.0,50.0,5.0,0.0,
TEN,00-0035676,A.J. Brown,WR,2019,17,REG,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,8.0,4.0,124.0,145.0,53.0,1
TEN,00-0035676,A.J. Brown,WR,2020,1,REG,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,8.0,5.0,39.0,51.0,14.0,0.0,
TEN,00-0035676,A.J. Brown,WR,2020,5,REG,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,9.0,7.0,82.0,87.0,23.0,1.0,
TEN,00-0035676,A.J. Brown,WR,2020,6,REG,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,7.0,5.0,58.0,93.0,14.0,2.0,
TEN,00-0035676,A.J. Brown,WR,2020,7,REG,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,8.0,6.0,153.0,90.0,96.0,1.0
TEN,00-0035676,A.J. Brown,WR,2020,8,REG,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,7.0,4.0,24.0,29.0,19.0,1.0,
TEN,00-0035676,A.J. Brown,WR,2020,9,REG,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,9.0,4.0,101.0,122.0,45.0,1.
TEN,00-0035676,A.J. Brown,WR,2020,10,REG,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,4.0,1.0,21.0,57.0,10.0,0.0
TEN,00-0035676,A.J. Brown,WR,2020,11,REG,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,7.0,4.0,62.0,55.0,38.0,1.0
TEN,00-0035676,A.J. Brown,WR,2020,12,REG,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,6.0,4.0,98.0,57.0,59.0,1.0
TEN,00-0035676,A.J. Brown,WR,2020,13,REG,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,7.0,4.0,87.0,157.0,26.0,1.
TEN,00-0035676,A.J. Brown,WR,2020,14,REG,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,9.0,7.0,112.0,109.0,20.0,1
TEN,00-0035676,A.J. Brown,WR,2020,15,REG,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,6.0,5.0,44.0,37.0,17.0,1.0
```

Data Processing

- Condensed Statistics - To handle all the information collected, we created a condensed statistic.
    - Each position (QB, WR, RB, TE) has its unique formula combining relevant features.
    - Weights were assigned to each feature based on its impact on fantasy performance and general intuition.
        - Example: Passing Yards are Weighted at +0.04, Interceptions are Penalized at -2.0.
    - This results in a single score summarizing a player's contribution per game.

Data Processing

Embedding Generation

- Player Embeddings:
  - Generated unique embeddings for each player using their player_id.
  - Captures player-specific performance trends and context.
- Team Embeddings:
  - Helps represent team-level strategies and performance nuances.
  - Head Coach and Matchups: Incorporates coaching influence on team dynamics and provides insights into how coaching strategies might impact gameplay outcomes.

Why Two Embeddings?

- Separating player and team embeddings allows the model to distinguish between individual player performance and the situational context provided by team dynamics.

Model

Random Forest Regressor

- Chosen for its robustness and ability to model high-dimensional, non-linear relationships in data.
- Number of Trees (Estimators): 100
  - Selected to effectively capture diverse feature interactions.
- Leverages embedding-based input to learn complex relationships between players, teams, and coaching dynamics.

.

Results

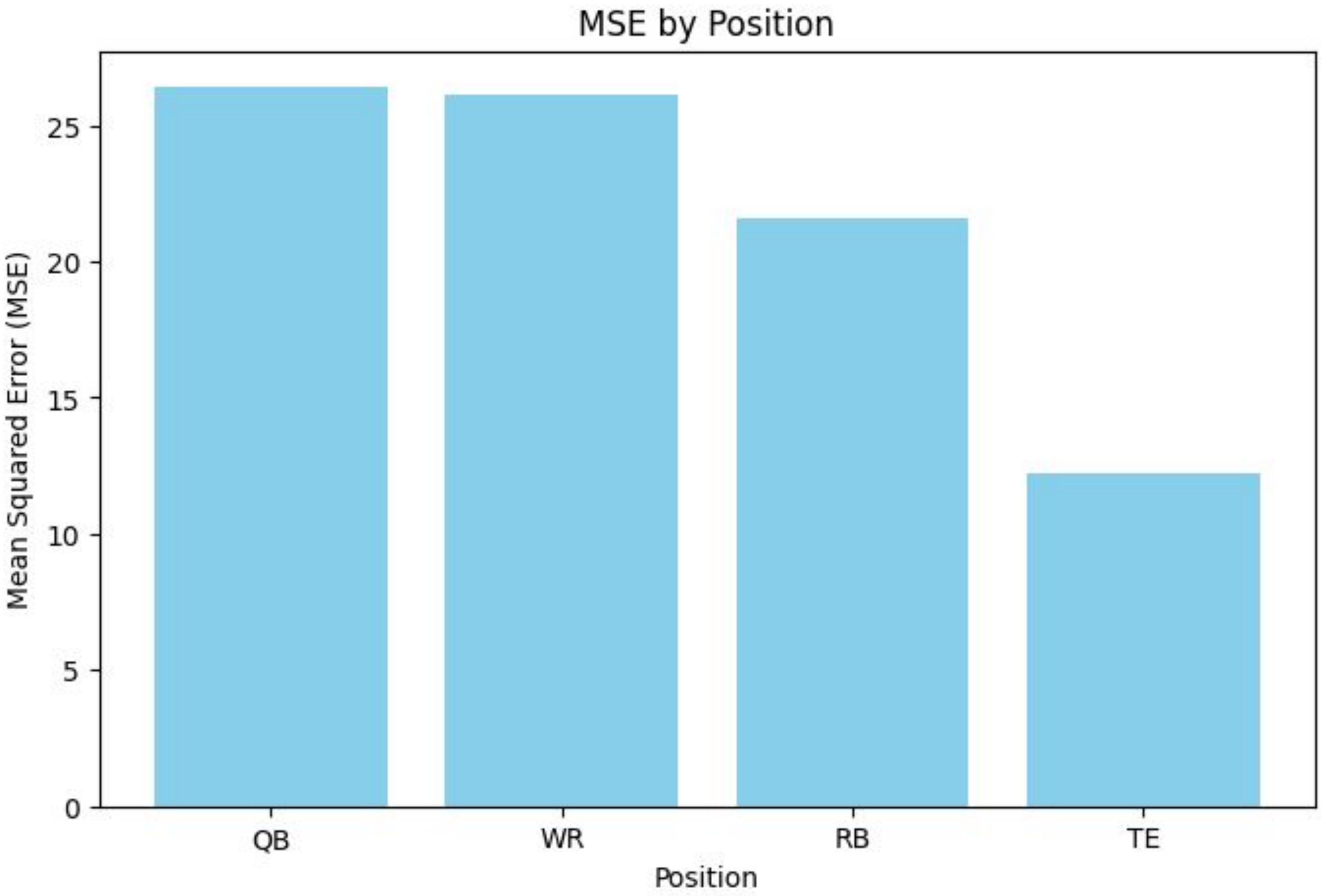
Model looks at season performance and considers weekly factors like matchups to predict a PPR (Points Per Reception) score.

```
Testing on Season 2024, Weeks 1 to 11

Overall Test MSE: 22.33
Overall Test MAE: 3.27
Overall Mean Error: 0.21
Overall Standard Deviation of Errors: 4.72
```

```
Best Overall Prediction:
player_name      Kalif Raymond
team                       DET
position                    WR
predicted_ppr         5.697389
actual_ppr                 5.7
error                -0.002611
Name: 282, dtype: object

Worst Overall Prediction:
player_name      Ja'Marr Chase
team                       CIN
position                    WR
predicted_ppr        20.250524
actual_ppr                55.4
error               -35.149476
Name: 2436, dtype: object
```


MSE by Position
Mean Squared Error (MSE)
0
5
10
15
20
25
QB
WR
RB
TE
Position

Prediction Results

Week 11 Predictions compared with actual results

	player_name	team	position	opponent_team	predicted_ppr	Actual
207	Amon-Ra St. Brown	DET	WR	JAX	30.007105	38.7
55	Taysom Hill	NO	TE	CLE	28.206170	42.5
19	Davante Adams	NYJ	WR	IND	27.278648	13.2
46	Jared Goff	DET	QB	JAX	26.554370	34.6
43	Tyreek Hill	MIA	WR	LV	26.431625	19.1
105	Saquon Barkley	PHI	RB	WAS	26.311640	33.8
270	Breece Hall	NYJ	RB	IND	26.186217	31.1
174	Joe Burrow	CIN	QB	LAC	25.682229	29
199	Ja'Marr Chase	CIN	WR	LAC	24.236432	26.5
167	Jalen Hurts	PHI	QB	WAS	23.737582	18.7
351	Bo Nix	DEN	QB	ATL	23.659200	29
169	Tee Higgins	CIN	WR	LAC	23.560333	29.8
107	Josh Allen	BUF	QB	KC	23.525223	25
333	Brock Bowers	LV	TE	MIA	23.472000	31.3
76	Cooper Kupp	LA	WR	NE	22.542795	28.6
137	David Montgomery	DET	RB	JAX	22.076000	24.5
324	Puka Nacua	LA	WR	NE	22.014033	25.3
201	DeVonta Smith	PHI	WR	WAS	21.973222	6.9
4	Matthew Stafford	LA	QB	NE	21.793731	27.8
50	Christian McCaffrey	SF	RB	SEA	21.464377	14.6

Week 12 predictions

Position: WR

	player_name	team	opponent_team	predicted_ppr
8	Keenan Allen	CHI	MIN	32.687064
143	Justin Jefferson	MIN	CHI	26.164850
146	CeeDee Lamb	DAL	WAS	22.404945
97	Jakobi Meyers	LV	DEN	20.985893
72	Cooper Kupp	LA	PHI	18.850612

Position: QB

	player_name	team	opponent_team	predicted_ppr
148	Jalen Hurts	PHI	LA	25.789552
327	Bo Nix	DEN	LV	21.121433
90	Lamar Jackson	BAL	LAC	19.915757
67	Patrick Mahomes	KC	CAR	18.158727
105	Kyler Murray	ARI	SEA	17.343212

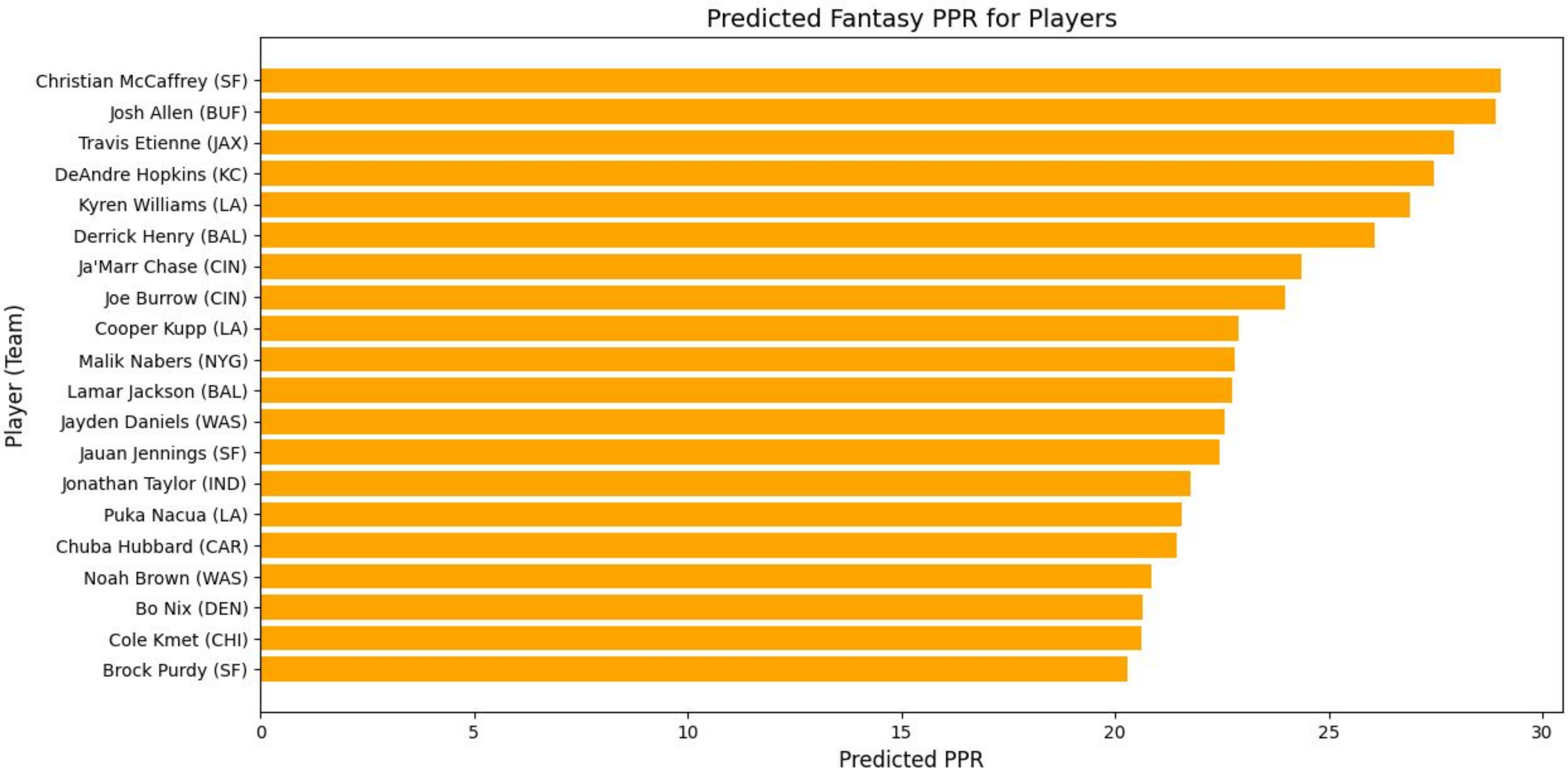
Position: RB

	player_name	team	opponent_team	predicted_ppr
58	James Conner	ARI	SEA	25.263900
36	Derrick Henry	BAL	LAC	21.847417
50	Aaron Jones	MIN	CHI	21.813655
199	Jerome Ford	CLE	PIT	20.074600
122	Josh Jacobs	GB	SF	19.859817

Position: TE

	player_name	team	opponent_team	predicted_ppr
307	Brock Bowers	LV	DEN	22.443000
80	Dallas Goedert	PHI	LA	18.608708
10	Travis Kelce	KC	CAR	15.900848
68	David Njoku	CLE	PIT	14.304217
83	Dalton Schultz	HOU	TEN	13.175133

Predicted Top Performers, Every Position, 2024, Week 3



Issues

- Data Cleaning and Feature Selection
  - Identifying the most relevant data points for accurate predictions.
- Weighing Defensive Strength
  - Determining how to account for the impact of opposing defenses on predicted player performance.
- Unpredictability of Sports
  - The inherently random nature of sports events makes precise predictions challenging, even with comprehensive data.
- Balancing Recent and Historical Data
  - Combining recent player performance with historical trends to generate accurate predictions.
- Real-World Testing
  - Conducting thorough testing to evaluate the model under realistic conditions.

Future work

- Incorporate Additional Data Sources
  - Integrate factors like weather conditions, teammate injuries, officiating, and venue-specific trends to improve prediction accuracy.
- Model and Data Processing Refinement
  - Fine-tune the model's parameters and enhance data processing pipelines for better predictions.
- Expand Features and Capabilities
  - Enable predictions for full fantasy teams to support strategic decisions.
  - Add player comparison tools for easier analysis of matchups and trade evaluations.
- Explore Advanced Model Architectures
  - Experiment with ensemble methods or neural network models to handle complex feature interactions.
- Real-Time Updates
  - Implement real-time data integration for up-to-the-minute predictions during live games.