DraftKings NFL Predictor

Using Data Modeling to predict NFL player performance

By Zach Fuller

Inspiration

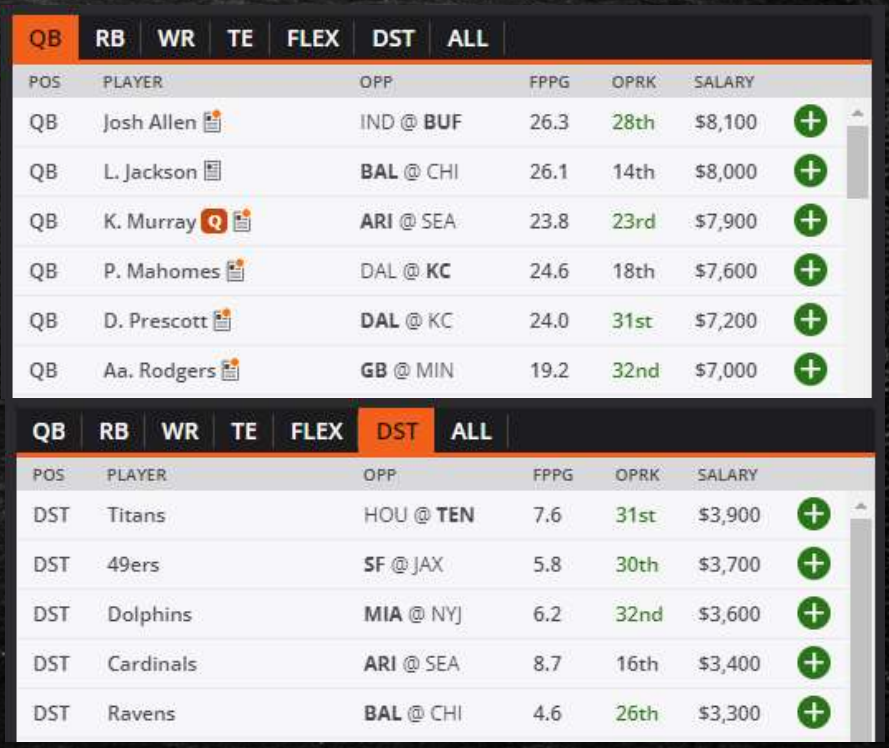
- David Bergman – UConn professor who won the Millionaire Maker last year using data science. (https://www.espn.com/fantasy/football/story/_/id/32166378/in-business-winning-how-dfs-champ-used-analytics-win-25m)
- David's area of expertise and interest: Large-scale automated decision making, decision diagrams, discrete optimization, integer programming, machine learning, integration of optimization techniques
- David's current classes: OPIM 5272 – Business Process Modeling and Data Management and OPIM 5603 – Statistics in Business Analytics

What is DraftKings? Daily fantasy sports?

- Fantasy sports take live-action sports and assign performance points to various performance statistics accrued in the live-action game.
- Season long fantasy sports: draft your initial roster of players at the beginning of the season and manage that roster for the duration of the season while competing weekly against other league members. Typically, league members are friends, family, or co-workers and range from 8 – 14 members per league
- Daily fantasy sports: draft your roster of players related to only one day or week of the season often competing against other unknown members. Individual contests can range from head-to-head all the way to 500,000 members.
- DraftKings has gained in popularity after agreeing to co-exclusive rights with ESPN in 2020.
- DraftKings' most popular contest is the Millionaire Maker where ~200,000 players compete for the top prize of $1 million.

$1 Million?!?! How does that work?

- Each contestant is given $50,000 to spend as a salary cap on a player pool designated by the players from teams that are considered for the contest. This contest only considers games played on Sunday for 12pm and 3-3:30pm CST start times.
- Must allocate the $50,000 between 9 roster spots consisting of 1 QB, 2 RBs, 3 WRs, 1 TE, 1 FLEX (any RB, WR, or TE), and 1 DST. This means that there is an average of ~$5,555 to spend on each roster spot.
- Highest performing players from a season perspective have the highest salary. However, week to week adjustments are made to player salary depending on performance, injuries, and matchups.



QB | RB | WR | TE | FLEX | DST | ALL

POS	PLAYER	OPP	FPPG	OPRK	SALARY
QB	Josh Allen	IND @ **BUF**	26.3	28th	$8,100
QB	L. Jackson	**BAL** @ CHI	26.1	14th	$8,000
QB	K. Murray Q	**ARI** @ SEA	23.8	23rd	$7,900
QB	P. Mahomes	DAL @ **KC**	24.6	18th	$7,600
QB	D. Prescott	**DAL** @ KC	24.0	31st	$7,200
QB	Aa. Rodgers	**GB** @ MIN	19.2	32nd	$7,000

QB | RB | WR | TE | FLEX | DST | ALL

POS	PLAYER	OPP	FPPG	OPRK	SALARY
DST	Titans	HOU @ **TEN**	7.6	31st	$3,900
DST	49ers	**SF** @ JAX	5.8	30th	$3,700
DST	Dolphins	**MIA** @ NYJ	6.2	32nd	$3,600
DST	Cardinals	**ARI** @ SEA	8.7	16th	$3,400
DST	Ravens	**BAL** @ CHI	4.6	26th	$3,300

And the scoring system?

Offense		DST (Defense and Special Teams)	
Passing TD	+4 Pts	Sack	+1 Pt
25 Passing Yards	+1 Pt (+0.04 Pts/Yards)	Interception	+2 Pts
300+ Yard Passing Game	+3 Pts	Fumble Recovery	+2 Pts
Interception	-1 Pt	Punt/Kickoff/FG Return for TD	+6 Pts
Rushing TD	+6 Pts	Interception Return TD	+6 Pts
10 Rushing Yards	+1 Pt (+0.1 Pts/Yard)	Fumble Recovery TD	+6 Pts
100+ Yard Rushing Game	+3 Pts	Blocked Punt or FG Return TD	+6 Pts
Receiving TD	+6 Pts	Safety	+2 Pts
10 Receiving Yards	=1 Pt (+0.1 Pts/Yard)	Blocked Kick	+2 Pts
100+ Receiving Yard Game	+3 Pts	2 Pt Conversion/Extra Point Return	+2 Pts
Reception	+1 Pt	0 Points Allowed	+10 Pts
Punt/Kickoff/FG Return of TD	+6 Pts	1 – 6 Points Allowed	+7 Pts
Fumble Lost	-1 Pt	7 – 13 Points Allowed	+4 Pts
2 Pt Conversion (Pass, Run, or Catch)	+2 Pts	14 – 20 Points Allowed	+1 Pt
Offensive Fumble Recovery TD	+6 Pts	21 – 27 Points Allowed	+0 Pts
		28 – 34 Points Allowed	-1 Pts
		35+ Points Allowed	-4 Pts

So the goal is to win, right?

- My goal was to win the Millionaire Maker by creating a predictive model that generates highest performing player by position on a given week (week 10) based on opposing team. There are 4 positions that will be predicted (QB, RB, TE, and WR). DST will be selected after other positions have been filled on the roster.
- Data was collected from Pro-Football-Reference using their Stathead tool for most of the data sets. Stathead tool allows for easy filtering, selecting of data, and exporting to various file types.
- My target variable will be predicting DK fantasy points and my features will be a combination of general and advance statistics (more on this). Data will cover NFL games played from week 1 to week 9.

You said a lot of data… Exactly how much?

- 52 datasets were collected to study and consider for modeling (12 of these just for predictions for week 10). Datasets include stats by team offense, team defense, and player offense and consider standard recorded metrics (attempts, yards, touchdown) in addition to advance metrics (3rd down efficiency, time of possession, and red zone efficiency).
- Stats collected can also be divided into season long accumulative stats and game by game recorded stats (more on this).
- DK salary and weekly results were also tracked and used for analysis.

How it started…

- Analysis started with identifying what my null model was. Four data sets were collected that detail the fantasy points accumulated by season and by position (QB, RB, TE, and WR).
- The culmination of the data is the DK fantasy points per game metric which tracks defensive performance against the four respective positions.
- At the bare minimum, you could always play the null model in your lineup. However, you would find that this rarely works. There are many factors involved an outcome of which we will explore.

Picture the null model

QB		RB		TE		WR	
DEF	DK pt avg	DEF	DK pt avg	DEF	DK pt avg	DEF	DK pt avg
WAS	26.8	NYJ	40.6	PHI	19.6	TEN	48
KAN	24	DET	31.2	BAL	19.2	WAS	43.8
IND	22.9	SFO	31	LAC	17.1	MIA	43.7
DAL	22.6	CIN	29.3	LAR	17.1	IND	43.1
MIA	22.1	PHI	29.3	IND	16.9	MIN	42.6

Versus how it went…

- Started out looking at the team offensive data per week and season long. The first modeling was performed attempting to predict which teams will have the highest offensive output in passing and rushing points.
- What I found is that regression models perform extremely well when predicting against very linear data. Who would have guessed? In total, Linear Regression, Regularization models Ridge and LASSO, K Nearest Neighbors, Random Forests, and AdaBoost.
- At this point I realized that I will only be able to make predictions against data that is known in week 10… which essentially removes everything. I will only be able to use fantasy points scored (my original target), player team, opponent team, location, and game week.
- This smashed my model performance for Random Forest and shrunk my outlook. My R2 scores went from being nearly perfect (0.995, 0.970) for training and testing to (0.866, 0.054). My model is now extremely overfit and does respond well to unseen data.
- I applied this method of only considering player team, opponent team, location, and game week as my features while still using player DK fantasy points as my target. The majority of the models resulted in negative testing R2 scores.

Versus how it went…

	QB		RB		TE		WR	
Linear Regression	0.347	-0.362	0.157	-0.131	0.233	-0.085	0.089	-0.032
Ridge	0.333	-0.202	0.155	-0.110	0.231	-0.065	0.088	-0.024
Ridge CV	0.124	0.037	0.059	-0.016	0.138	0.008	0.001	0.001
LASSO	0.006	0.008	0.001	-0.021	0	-0.005	0.003	0.004
LASSO CV	0.007	0.015	0.058	-0.030	0.111	0.003	0	0
KNN	0.205	-0.362	0.210	-0.131	0.204	-0.085	0.199	-0.032
KNN CV GS	0.006	0.019	0.033	-0.050	0.027	-0.016	0.029	0.006
Random Forest	0.768	-0.133	0.494	-0.513	0.598	-0.254	0.341	-0.636
Random Forest CV GS	0.172	-0.009	0.122	-0.049	0.133	-0.022	0.066	0.007

You call that a prediction?

QB				RB				TE				WR			
Ridge CV		Random For CV GS		KNN CV GS		Random Forest		KNN CV GS		Random Forest		KNN CV GS		Random Forest	
Pred Opp	Actual Ranking	Pred Opp	Actual Ranking	Pred Opp	Actual Ranking	Pred Opp	Actual Ranking	Pred Opp	Actual Ranking	Pred Opp	Actual Ranking	Pred Opp	Actual Ranking	Pred Opp	Actual Ranking
BAL – 18.5	8th – 18.9	BAL – 22.5	8th – 18.9	MIN – 11.5	19th – 14.9	IND – 16.7	14th – 18.4	JAX – 7.7	20th – 6.1	ATL – 18.8	27th – 3.5	DAL – 9.3	63rd – 4.2	MIN – 12.1	8th – 17.8
LAC – 17.4	10th – 18.5	MIN – 19.0	15th – 13.0	CAR – 11.1	18th – 15.4	TEN – 15.0	9th – 20.8	DAL – 7.5	10th – 10.0	CLE – 11.0	2nd – 19.7	TEN – 9.1	18th – 14.4	LAC – 11.9	4th – 25.9
MIA – 17.0	11th – 16.4	LAC – 18.8	10th – 18.5	ATL – 11.1	8th – 21.8	JAX – 14.1	3rd – 27.6	LVR – 7.2	1st – 22.9	TEN – 10.6	12th – 8.2	MIA – 9.0	21st – 14.0	BAL – 11.7	17th – 14.6
MIN – 16.7	15th – 13.0	MIA – 17.7	11th – 16.4	DET – 11.0	11th – 20.3	GNB – 13.1	39th – 5.9	PHI – 7.1	8th – 10.9	MIA – 10.5	3rd – 18.3	GNB – 9.0	49th – 5.6	PIT – 11.5	36th – 10.1
SEA – 16.6	18th – 11.5	NYJ – 17.6	3rd – 24.9	JAX – 10.7	3rd – 27.6	DET – 11.9	11th – 20.3	WAS – 6.9	13th – 7.6	DAL – 10.3	10th – 10.0	LAC – 8.7	4th – 25.9	GNB – 10.9	49th – 5.6
KAN – 39.2	**@**	**LVR – 18.1**	**25th**	**KAN – 32.4**	**@**	**LVR – 24.5**	**17th**	**KAN – 22.9**	**@**	**LVR – 14.4**	**11th**	**BUF – 33.2**	**@**	**NYJ – 32.1**	**27th**

So who won?

- Eppy99 did…
- We see that Eppy99 also played the maximum 150 lineups (150 x $20 = $3000).
- QB: drafted the 2nd ranked
- RB: drafted the 1st, 5th, and 9th ranked
- TE: drafted the 1st ranked
- WR: drafted the 1st, 2nd, and 7th ranked
- In order to win you mostly need to draft the top ranked player by position or players in the top 5.

eppy99 (34) RANK 1 WINNING $1,021,052.63 YTP 0 PMR 0 View H2H

POS	PLAYER	DRAFT %	GAME	SCORING	FPTS
QB	J. Allen $7,900	9.7%	BUF 45 @ NYJ 17 Final	2 PaTD, 366 PaYds, 3 RuYds, 1 INT, 1 300+Pass	24.94
RB	D. Johnson $4,700	48.4%	CLE 7 @ NE 45 Final	58 RecYds, 99 RuYds, 7 REC	22.70
RB	M. Ingram II $4,500	26.2%	NO 21 @ TEN 23 Final	1 RuTD, 61 RecYds, 47 RuYds, 4 REC	20.80
WR	K. Allen $7,000	8.5%	MIN 27 @ LAC 20 Final	98 RecYds, 8 REC	17.80
WR	S. Diggs $7,500	8.0%	BUF 45 @ NYJ 17 Final	1 RecTD, 162 RecYds, 8 REC, 1 100+Rec	33.20
WR	C. Lamb Q $7,000	10.4%	ATL 3 @ DAL 43 Final	2 RecTD, 94 RecYds, 12 RuYds, 6 REC	28.60
TE	H. Henry $4,100	4.0%	CLE 7 @ NE 45 Final	2 RecTD, 37 RecYds, 4 REC	19.70
FLEX	R. Stevenson $4,500	3.3%	CLE 7 @ NE 45 Final	2 RuTD, 14 RecYds, 100 RuYds, 4 REC, 1 100+Rush	30.40
DST	Eagles $2,700	3.3%	PHI 30 @ DEN 13 Final	1 SACK, 1 DFR, 1 DefTD, 1 7-13 PA, 1 BLK	15.00

FANTASY POINTS 213.14

What did we learn?

- If making predictions was easy, everyone would do it! Also, it depends…
- Despite poor modeling performance from a traditional standpoint, we did still find some predictive success. There is a natural randomness and unpredictability about NFL games that just cannot be modeled.
- Was a player dealing with an injury? Did a player leave a game with injury? Did a player wake up on the wrong side of the bed? There are many unknown and unpredictable metrics that can never be predicted.
- Given known information what else can we do? Use and attempt to improve on 538's ELO metric for strength of game matchup. Consider if a team is going into a bye week or coming out of a bye week. Did a team take a bad loss the week before? A good team is more likely to make up for it in the next game.
- Could try using season long averages for player performance to predict week 10 outcomes. 267 entries to create data for.
- Try again next week!

What did we learn?