Recruiting Quality Analysis

Intro

The purpose of this analyis is to look at HS recruiting analytics & individual college performance to understand how this information translates to a player getting drafted in the NFL. We will evaluate some marquee programs in their recruiting success and build an initial model to predict Running back's NFL draft stock. Additionally, we want to understand if HS Ratings are indicative of someone being drafted or not.

Top Recruiting Programs

The chart below shows top 20 Recruiting Programs by Avg. Recruiting Rating since 2000. Since 2000, Alabama, Ohio State, and USC have dominated the recruiting trail with an average rating over that timeframe of 92 (out of 100). 16% of USC's recruits wree "5 Stars", which is the highest rate amongst any school. Alabama and Ohio State recruits have converted to becoming draft picks at a higher rate than any other school at 21%. 8% of Alabama's recruits end up being drafted in the first round, which is +2% the next closest school of Ohio State at 6% and nearly +7% higher than their rival in Auburn. That's some clear selling point for current HS Superstars.

Top 20 Recruiting Programs by Avg. Rating

committedTo	commits	meanStars	meanRating	FiveStars	FourStars	ThreeStars	FiveStarPct	Drafted	DraftedPct	First	FirstPct
Alabama	392	3.78	0.92	52	207	127	0.13	84	0.21	30	0.08
Ohio State	311	3.80	0.92	34	187	83	0.11	66	0.21	20	0.06
USC	315	3.78	0.92	49	156	102	0.16	54	0.17	10	0.03
Florida	353	3.67	0.91	30	183	133	0.08	59	0.17	14	0.04
Georgia	364	3.71	0.91	45	177	134	0.12	60	0.16	12	0.03
LSU	367	3.68	0.91	28	199	133	0.08	69	0.19	13	0.04
Texas	349	3.72	0.91	24	206	117	0.07	34	0.10	3	0.01
Florida State	351	3.63	0.90	37	155	152	0.11	53	0.15	12	0.03
Notre Dame	340	3.61	0.90	14	181	142	0.04	47	0.14	9	0.03
Oklahoma	339	3.54	0.90	15	162	153	0.04	51	0.15	6	0.02
Auburn	367	3.46	0.89	13	161	177	0.04	33	0.09	5	0.01
Clemson	336	3.48	0.89	25	128	167	0.07	58	0.17	13	0.04
Miami	333	3.45	0.89	13	134	177	0.04	46	0.14	4	0.01
Michigan	371	3.50	0.89	12	176	169	0.03	42	0.11	7	0.02
Tennessee	356	3.38	0.89	11	126	207	0.03	21	0.06	3	0.01
Penn State	320	3.42	0.88	9	130	166	0.03	43	0.13	3	0.01
Texas A&M	357	3.39	0.88	16	128	194	0.04	29	0.08	9	0.03
UCLA	331	3.32	0.88	9	114	182	0.03	34	0.10	5	0.02
Nebraska	325	3.19	0.87	1	81	225	0.00	19	0.06	2	0.01
Oregon	318	3.28	0.87	10	99	178	0.03	32	0.10	6	0.02

Note:
Filtered for data from 2005 - 2020; 10 or more commits

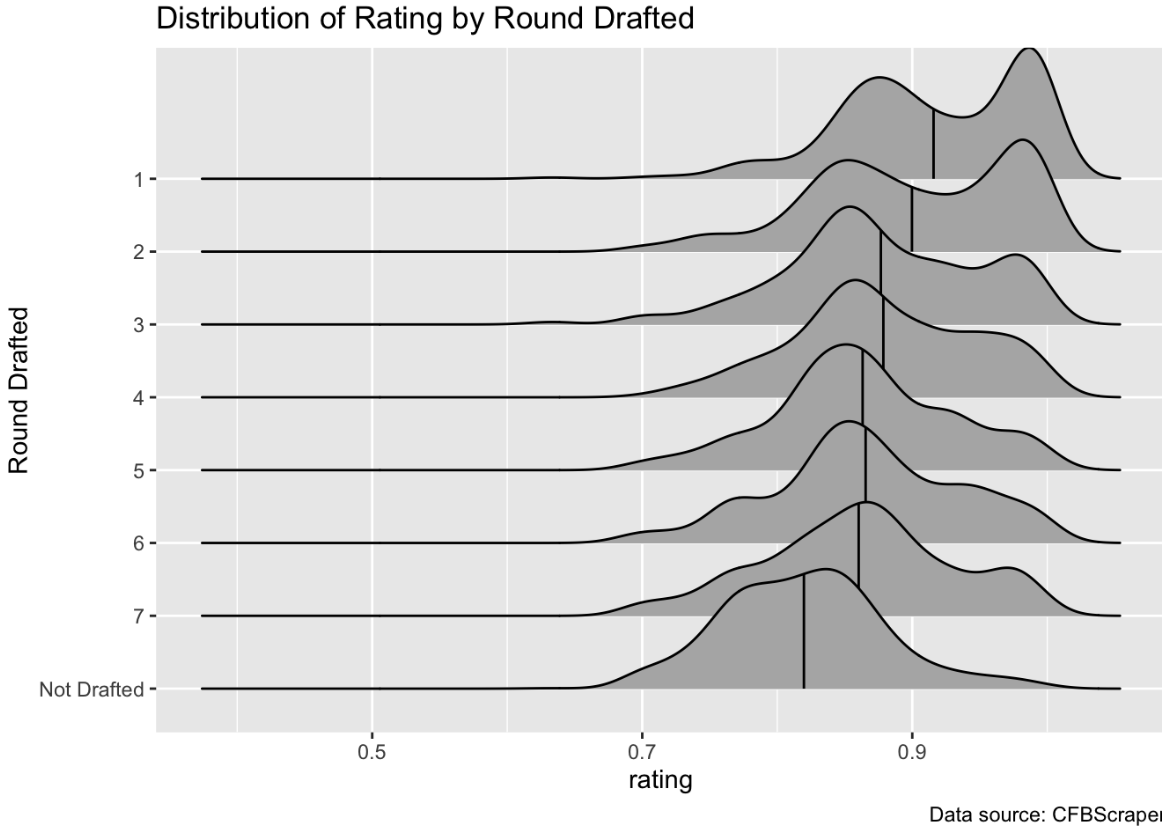
The best classes ever, in terms of mean rating were the 2008 Ohio State class and 2013 USC class, which had an average rating 0.96. I would through the 2014 Alabama class in there as an equal given they 23 recruits with an average rating 0.95, which is pretty impressive. What also makes this Alabama class impressive was that 43% of the recruits were drafted into the NFL and 2 were drafted in the first round (which wasn't even the best Alabama First Round Draft Year). Florida in 2007 had 13% of their recruits drafted in the first round with a high 26% of their recruits being 5 stars. The USC classes on the board although high 5 star and average rating, the recruits haven't necessarily panned out out in terms of a high percentage being drafted, with the exception of the 2005 USC class, which had 62% of their recruits drafted.

Top 20 Recruiting Classes All-Time in Avg. Rating

committedTo	year	commits	meanStars	meanRating	FiveStars	FourStars	ThreeStars	FiveStarPct	Drafted	DraftedPct	First	FirstPct
Ohio State	2008	14	4.21	0.96	4	9	1	0.29	2	0.14	0	0.00
USC	2013	12	4.33	0.96	4	8	0	0.33	1	0.08	0	0.00
Alabama	2014	23	4.13	0.95	6	14	3	0.26	10	0.43	2	0.09
USC	2007	18	4.17	0.95	7	7	4	0.39	5	0.28	0	0.00
USC	2010	17	4.18	0.95	5	10	2	0.29	3	0.18	0	0.00
Alabama	2012	25	3.84	0.94	3	15	7	0.12	8	0.32	1	0.04
Alabama	2015	24	4.08	0.94	6	14	4	0.25	7	0.29	3	0.12
Alabama	2017	28	4.00	0.94	6	17	4	0.21	4	0.14	3	0.11
Alabama	2019	28	4.07	0.94	3	24	1	0.11	0	0.00	0	0.00
Alabama	2020	24	4.00	0.94	4	16	4	0.17	0	0.00	0	0.00
Florida	2007	23	4.04	0.94	6	12	5	0.26	8	0.35	3	0.13
Florida	2008	16	4.06	0.94	2	13	1	0.12	0	0.00	0	0.00
Florida	2010	28	3.96	0.94	5	17	6	0.18	7	0.25	3	0.11
Florida State	2012	18	4.22	0.94	7	8	3	0.39	7	0.39	1	0.06
Georgia	2018	26	4.04	0.94	7	14	4	0.27	0	0.00	0	0.00
Georgia	2019	21	4.14	0.94	5	14	2	0.24	0	0.00	0	0.00
Ohio State	2017	22	4.09	0.94	6	13	2	0.27	3	0.14	2	0.09
Texas	2010	24	4.08	0.94	5	16	3	0.21	1	0.04	0	0.00
USC	2005	16	4.00	0.94	5	6	5	0.31	10	0.62	2	0.12
USC	2008	19	3.84	0.94	2	12	5	0.11	5	0.26	3	0.16

Note:
Filtered for data from 2005 - 2020; 10 or more commits

When takinga look at the distribution of HS Rating to drafted, it is apparent that the higher the rating the more likely to be drafted in higher rounds. However, The distribution for 1st/2nd rounds appear to be much more bi-modal, which shows that teams are willing to take a chance on lower ratings (maybe proved themselves more in college) and higher ratings with raw skillset.



Running Backs Progression to the Draft

This next chart shows rushers with high average points per attempt, whihch helps to shows explosiveness and effectiveness of runers in college. As we can see w/ Georgia Tech & Army ranked high in this list that Triple Option programs are likely to have higher mean points per attempt. A couple hidden gems in this list are both Kenyan Drake and Kareem Hunt. Both drafted in the 3rd round with less overall rushes and therefore film to be evaluated from. We can hypothesize that their high mean points per attempt could gave contributed to their success in the NFL.

Top 15 Mean Points Per Attempt by Rating

rusher_player_name	rating	position	pos_team	Rnd1	mean_ppa	success_rate	ypc	rushes	rush_yds	yps	TDs
Orwin Smith	0.8764	RB	Georgia Tech	Not Drafted	0.40	0.65	9.29	197	1831	14.01	21
Jerrod Johnson	0.8795	ATH	Texas A&M	Not Drafted	0.39	0.59	5.97	245	1462	8.74	17
Terry Baggett	0.7667	RB	Army	Not Drafted	0.35	0.62	8.32	160	1332	12.34	10
Embry Peeples	0.8479	RB	Georgia Tech	Not Drafted	0.33	0.57	7.51	136	1021	11.87	4
Roddy Jones	0.8497	RB	Georgia Tech	Not Drafted	0.32	0.54	7.37	237	1746	12.55	14
Malcolm Brown	0.7222	ATH	Army	Not Drafted	0.30	0.56	6.32	218	1377	9.70	11
Tavon Austin	0.9188	WR	West Virginia	1	0.30	0.57	9.26	117	1084	15.42	7
Percy Harvin	0.9989	WR	Florida	1	0.29	0.67	9.56	194	1855	13.78	19
Melvin Gordon	0.8983	RB	Wisconsin	1	0.25	0.57	8.14	287	2335	12.96	16
Kenyan Drake	0.9454	APB	Alabama	3	0.24	0.61	7.28	134	975	10.63	13
James Rodgers	0.7889	WR	Oregon State	Not Drafted	0.23	0.58	8.10	165	1336	13.57	8
Kain Colter	0.8733	ATH	Northwestern	Not Drafted	0.22	0.54	5.81	402	2335	9.49	28
Kareem Hunt	0.8032	RB	Toledo	3	0.22	0.58	6.32	137	866	9.72	6
Waymon James	0.9108	RB	TCU	Not Drafted	0.22	0.53	6.27	281	1763	9.90	15

Note:
Filtered for data from 2000-2020
[1] Filtered for 100 plays or more
[a] Not including QBs

rusher_player_name	rating	position	pos_team	Rnd1	mean_ppa	success_rate	ypc	rushes	rush_yds	yps	TDs
Raymond Maples	0.7444	RB	Army	Not Drafted	0.20	0.52	5.93	439	2603	9.60	7

Note:
Filtered for data from 2000-2020
[1] Filtered for 100 plays or more
[a] Not including QBs

A successful run is one that is more than more than 50% of distance to 1st down in rushing yards on 1st down (i.e. First down with 10 yards to go, a rush of over 5 yards is successful run), more than 70% on 2nd down and distance to go, or if the 3rd or 4th down run resulted in a first down. A success rate would yield the proportion of successful runs to total rushes. A high success rate would indicate a high big play ability and efficiency. Not surprisingly, Percy Harvin's 67% success rate and a 0.99 rating out of HS translated to his appeal as a first round pick. Again, Kenyan Drake and Kareem Hunt's high success rate with a lower sample size translated well to the NFL.

Top 15 Success Rate by Rating

rusher_player_name	rating	position	pos_team	Rnd1	mean_ppa	success_rate	ypc	rushes	rush_yds	successes	yps	TDs
Percy Harvin	0.9989	WR	Florida	1	0.29	0.67	9.56	194	1855	130	13.78	19
Orwin Smith	0.8764	RB	Georgia Tech	Not Drafted	0.40	0.65	9.29	197	1831	128	14.01	21
Terry Baggett	0.7667	RB	Army	Not Drafted	0.35	0.62	8.32	160	1332	99	12.34	10
Kenyan Drake	0.9454	APB	Alabama	3	0.24	0.61	7.28	134	975	82	10.63	13
Imani Cross	0.9065	RB	Nebraska	Not Drafted	0.13	0.59	5.51	140	772	83	8.22	18
Jerrod Johnson	0.8795	ATH	Texas A&M	Not Drafted	0.39	0.59	5.97	245	1462	145	8.74	17
Henry Melton	0.9727	SDE	Texas	4	-0.02	0.58	4.63	115	532	67	6.73	14
James Rodgers	0.7889	WR	Oregon State	Not Drafted	0.23	0.58	8.10	165	1336	95	13.57	8
Kareem Hunt	0.8032	RB	Toledo	3	0.22	0.58	6.32	137	866	80	9.72	6
Thomas Tyner	0.9899	RB	Oregon	Not Drafted	0.16	0.58	5.85	114	667	66	9.86	9
D.J. Adams	0.9269	RB	Maryland	Not Drafted	-0.03	0.57	3.87	105	406	60	6.03	15
Embry Peeples	0.8479	RB	Georgia Tech	Not Drafted	0.33	0.57	7.51	136	1021	78	11.87	4
Melvin Gordon	0.8983	RB	Wisconsin	1	0.25	0.57	8.14	287	2335	164	12.96	16
Tavon Austin	0.9188	WR	West Virginia	1	0.30	0.57	9.26	117	1084	67	15.42	7
Bryce Beall	0.8181	S	Houston	Not Drafted	0.16	0.56	5.84	540	3155	301	9.14	39

Note:
Filtered for data from 2000-2020
[1] Filtered for 100 plays or more
[a] Not including QBs

Yards Per Carry is a lot more well-known statistic than success rate and not surprisingly we see similar names on this list. Jahvid Best's 7.37 YPC on 362 rushes is pretty incredible. Especially when you consider he had a 47% success rate.

Top 15 Yards Per Carry by Rating

rusher_player_name	rating	position	pos_team	Rnd1	mean_ppa	success_rate	ypc	rushes	rush_yds	successes	yps	TDs
Percy Harvin	0.9989	WR	Florida	1	0.29	0.67	9.56	194	1855	130	13.78	19
Orwin Smith	0.8764	RB	Georgia Tech	Not Drafted	0.40	0.65	9.29	197	1831	128	14.01	21
Tavon Austin	0.9188	WR	West Virginia	1	0.30	0.57	9.26	117	1084	67	15.42	7
Terry Baggett	0.7667	RB	Army	Not Drafted	0.35	0.62	8.32	160	1332	99	12.34	10

Note:
Filtered for data from 2000-2020
[1] Filtered for 100 plays or more
[a] Not including QBs

rusher_player_name	rating	position	pos_team	Rnd1	mean_ppa	success_rate	ypc	rushes	rush_yds	successes	yps	TDs
Elijah McGuire	0.8415	RB	Louisiana	6	0.15	0.50	8.23	104	856	52	14.40	8
Melvin Gordon	0.8983	RB	Wisconsin	1	0.25	0.57	8.14	287	2335	164	12.96	16
James Rodgers	0.7889	WR	Oregon State	Not Drafted	0.23	0.58	8.10	165	1336	95	13.57	8
De'Anthony Thomas	0.9926	APB	Oregon	4	0.16	0.54	7.74	242	1874	131	13.08	26
Embry Peeples	0.8479	RB	Georgia Tech	Not Drafted	0.33	0.57	7.51	136	1021	78	11.87	4
Dri Archer	0.7000	APB	Kent State	3	-0.02	0.43	7.41	316	2343	136	15.47	25
Jahvid Best	0.9220	RB	California	1	0.02	0.47	7.37	362	2668	169	14.67	29
Roddy Jones	0.8497	RB	Georgia Tech	Not Drafted	0.32	0.54	7.37	237	1746	128	12.55	14
Felix Jones	0.9000	ATH	Arkansas	1	0.10	0.47	7.35	341	2506	161	14.12	15
Kenyan Drake	0.9454	APB	Alabama	3	0.24	0.61	7.28	134	975	82	10.63	13
Alex Green	0.8444	APB	Hawai'i	Not Drafted	0.01	0.52	7.18	208	1493	108	12.88	19

Note:
Filtered for data from 2000-2020
[1] Filtered for 100 plays or more
[a] Not including QBs

The workhorses in terms of career rushing yards are very familiar names for college footballl fans. Montee Ball arguably had the best rushing career in the past 20 years and at a 50% success rate on 900+ rushes is truly incredible. Branden Oliver ended up having a short stint with the Chargers and marks a lower valued rusher that can be effective at the next level.

Top 15 Rush Attempts by Rating

rusher_player_name	rating	position	pos_team	Rnd1	mean_ppa	success_rate	ypc	rushes	rush_yds	successes	yps	TDs
Damion Fletcher	0.7896	APB	Southern Mississippi	Not Drafted	0.07	0.48	5.25	934	4901	446	9.22	40
Montee Ball	0.8844	RB	Wisconsin	2	0.03	0.50	5.58	922	5142	465	9.54	77
Robbie Rouse	0.8052	RB	Fresno State	Not Drafted	0.00	0.43	5.11	883	4512	382	9.72	36
Branden Oliver	0.7469	RB	Buffalo	Not Drafted	-0.05	0.40	4.65	868	4040	348	9.47	33
Stepfan Taylor	0.8961	RB	Stanford	5	0.00	0.45	5.09	848	4318	380	9.13	40
Rodney Stewart	0.7667	APB	Colorado	Not Drafted	-0.02	0.43	4.42	816	3605	350	8.61	24
Kevin Smith	0.7667	RB	UCF	3	-0.05	0.47	4.86	811	3940	380	8.97	38
Chris Polk	0.9267	RB	Washington	Not Drafted	0.00	0.42	5.06	798	4036	337	9.89	26
James Sims	0.8733	RB	Kansas	Not Drafted	-0.06	0.41	4.49	798	3586	328	8.55	34
Jacquizz Rodgers	0.8794	APB	Oregon State	5	-0.05	0.45	4.90	796	3903	357	9.31	46
Johnathan Franklin	0.8943	RB	UCLA	4	-0.01	0.44	5.55	794	4410	352	10.73	32
Montel Harris	0.8101	RB	Boston College	Not Drafted	-0.02	0.40	4.77	769	3667	311	9.62	27
DeMarco Murray	0.9901	RB	Oklahoma	3	-0.05	0.46	4.83	759	3667	352	8.67	49
Tim Cornett	0.8333	RB	UNLV	Not Drafted	-0.07	0.41	4.88	755	3683	313	9.65	35
Ka'Deem Carey	0.8799	RB	Arizona	4	0.05	0.49	5.67	747	4235	367	9.72	48

Note:
Filtered for data from 2000-2020
[1] Filtered for 100 plays or more
[a] Not including QBs
[*] Filtered out Mike Davis from South Carolina given duplicate names at the school

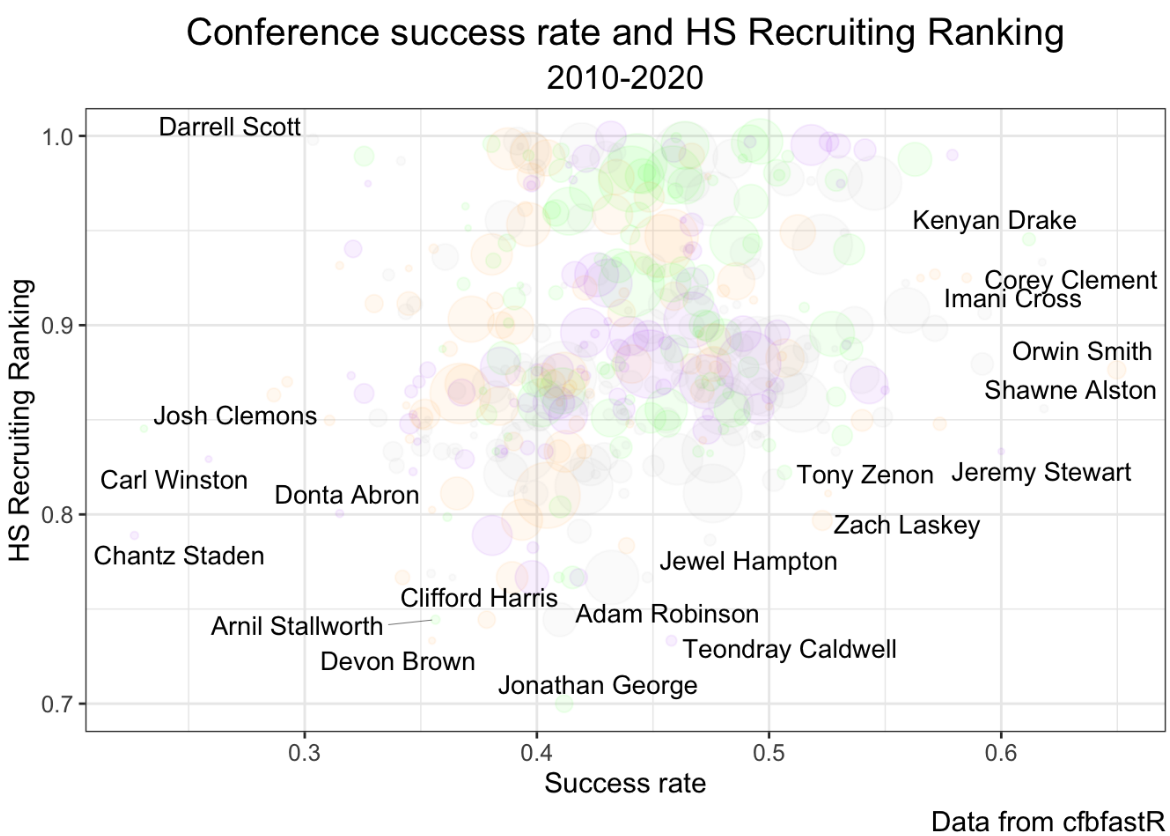
Bringing it altogether, we can see that the SEC is the preferred destination for running backs with nearly 28% of RBs getting drafted at an average draft position of ~95 +/- 67 draft position. What that means is that nearly a 1/3 of running backs in the SEC project to be drafted by 3 round, which is nearly a whole round sooner on average than the ACC. The 2nd highest draft position belongs to the PAC 12, which had less rusher overall. It may be that the longevity of a PAC12 rushers career made it less desirable of a conference to commit to compared to other conferences. The MAC & the WAC had a few prolific running bucks drafted in the 3rd round, which goes to show there are hidden gems in some of the smaller conferences.

Top RB Conferences by Avg Rating

offense_conference	position	players	mean_rating	sd_rating	drafted	drafted_pct	mean_draft_pick	sd_draft_pick	mean_ppa	success_rate	ypc	plays_mean
SEC	RB	201	0.92	0.06	56	0.28	94.6	66.9	0.00	0.45	5.26	131.69
ACC	RB	193	0.89	0.06	29	0.15	123.7	68.0	-0.05	0.42	4.87	123.47
Big Ten	RB	171	0.87	0.05	21	0.12	104.4	65.4	0.00	0.45	5.12	131.26
Pac-12	RB	148	0.89	0.06	33	0.22	114.5	57.6	-0.02	0.44	5.11	143.04
Big 12	RB	145	0.89	0.06	22	0.15	136.7	62.1	-0.02	0.45	5.00	142.13
Conference USA	RB	140	0.81	0.05	6	0.04	118.0	54.8	-0.03	0.44	4.87	110.15
Mid-American	RB	134	0.78	0.04	4	0.03	88.1	5.8	-0.06	0.42	4.76	129.09
Mountain West	RB	98	0.80	0.05	11	0.11	115.6	59.6	-0.01	0.44	4.99	122.71
Big East	RB	95	0.85	0.06	11	0.12	121.8	61.0	-0.06	0.42	4.81	123.44
Sun Belt	RB	84	0.80	0.04	3	0.04	131.3	52.4	-0.04	0.43	4.91	117.13
Western Athletic	RB	62	0.78	0.05	9	0.15	93.3	75.6	-0.03	0.44	5.23	127.34
FBS Independents	RB	28	0.86	0.09	3	0.11	159.6	31.8	0.04	0.45	5.30	131.71

Note:
Filtered for data from 2000-2020
[1] Avg above 100 rushes per player
[a] Big East was disbanded in 2012
[*] Combined All-Purpose Back, Fullback and General Athletes that have high rushing attempts into Runningback

The chart below shows Success rate compared to HS Recruiting rankings of all individual running backs. The names represent some outliers in the dataset. I think one thing is interesting call-out is the consistently lower success rate of rushers in ACC compared to both SEC & PAC 12. That may be what has hurted ACC RB's draft capital.

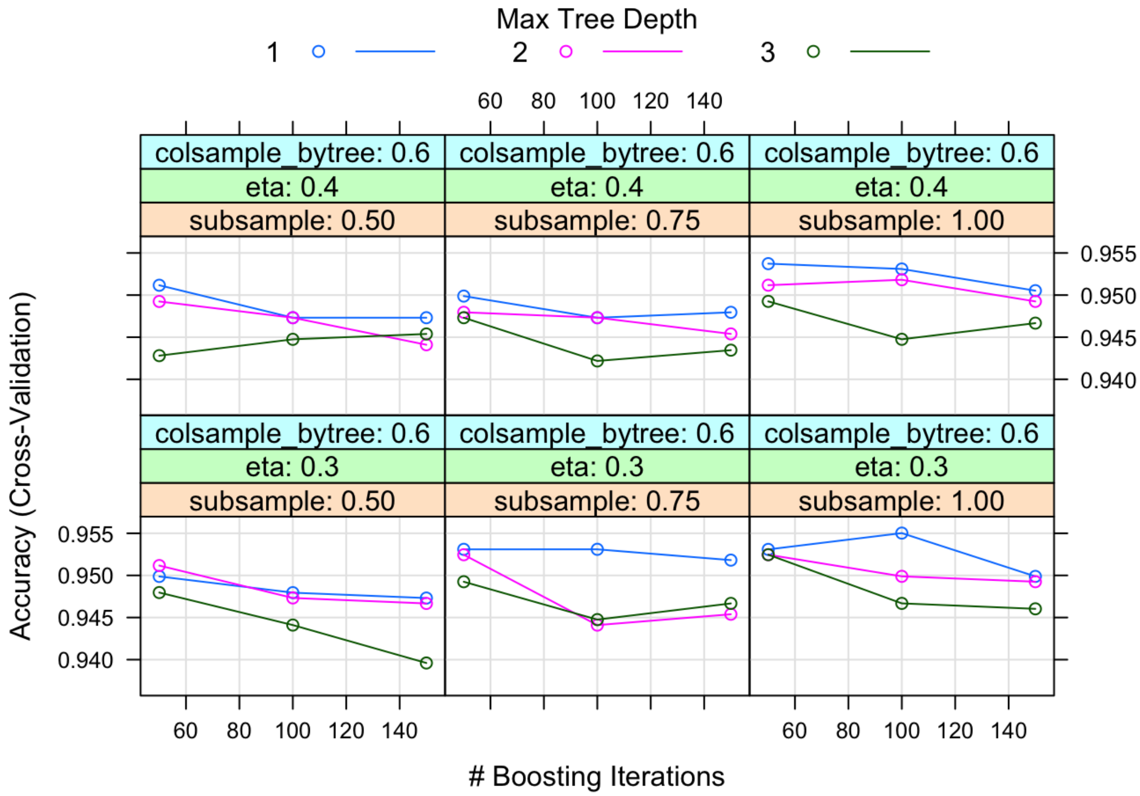


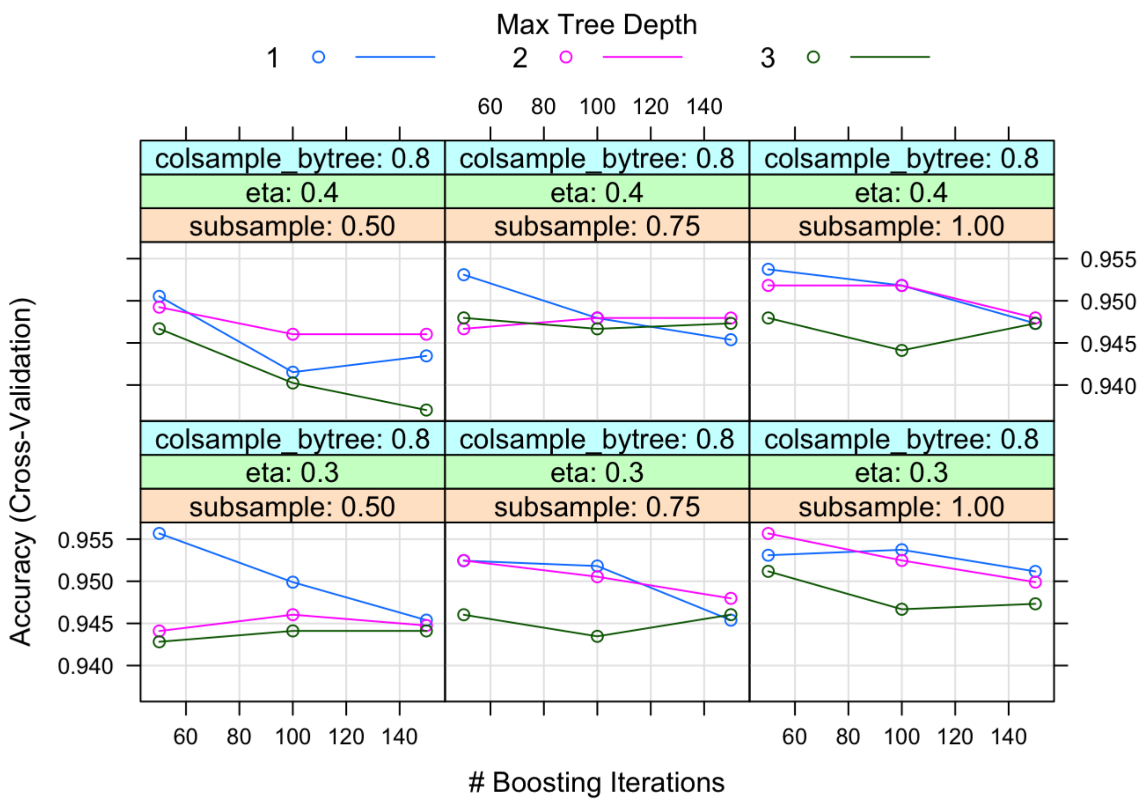
Predicted Individuals Probability of Being Drafted

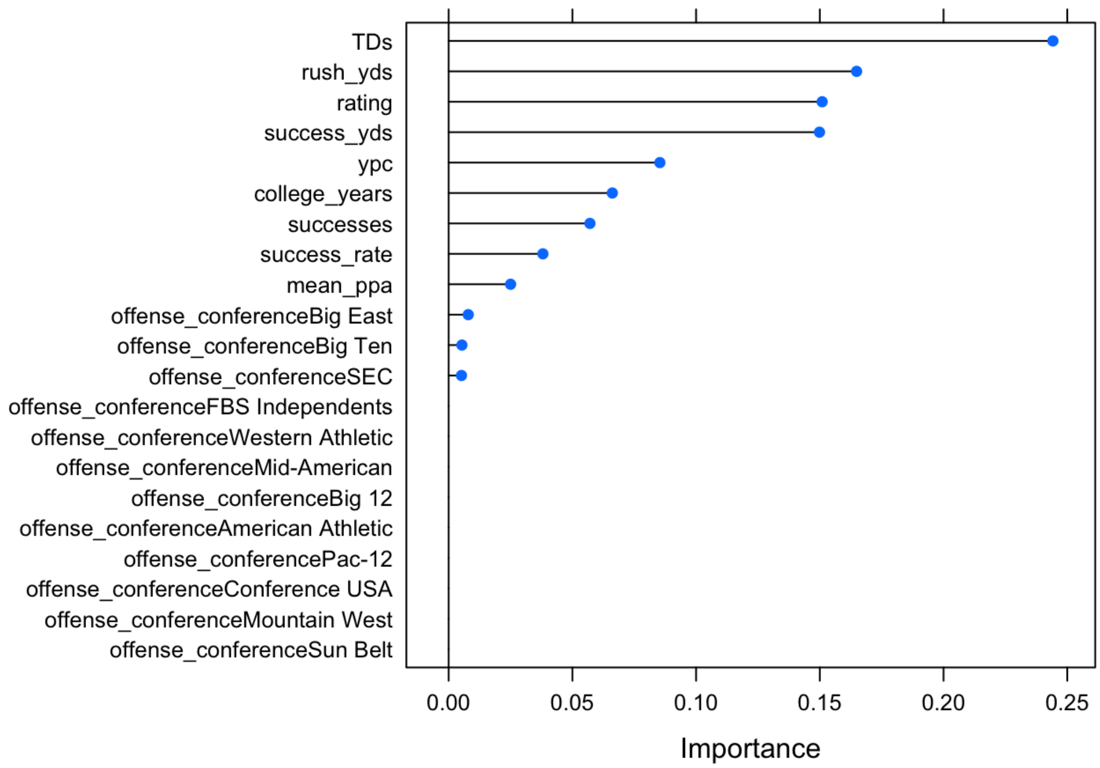
We will be utilizing a classification model with 5 cross-folds (or resampling of data for training & testing the model) to build a prediction model. As we previewed the variables we intend to add to our model consist of the following: HS Rating (on a 0 to 1 scale), conference the individual played in, mean points per attempt (mean_ppa), success rate, yards per carry (ypc), career rush yards, successes, success yards, career Touchdowns, and years of college playing experience. Our first model will be to build an

XGBoost Tree model that uses gradient boosted decision tree to rank and order important variables. XGBoost models work well in cases where there are a mixture of categorical (such as conference) and numeric field. The concept of gradient booting is an ensemble of decision trees designed to enhance the learning of the decision trees. The goal is to spend much more time on weak learners to optimize or boot its learning capability and reduce information loss. It does so in batch or ensemble learning.

The next two charts show the iteration of learning from the XGBoost model and shows that through several iterations we are getting an ~94-95% accuracy in our resampling. The third importance chart show which variables had the most predictive importance in determining someone's ability to be drafted. We can see that someone's Success Yards has over 0.3 importance, which said different has over 30% of importance in weights of model. This tells us that someone career explosive in being able to gain big chunks of yardage per rush. This intuitively makes sense as frequent big plays would stand out on the tape. Career Touchdowns (TDs), Rush Yards and HS Rating all have between 0.15 to 0.20 average importance in rating. It's interesting to see how well-known recruits translates to ability to be drafted, this makes sense given the exposure of well-known recruiting programs and the conversion ability of these programs to get high-profile running backs drafted. SEC & PAC 12 as shown prior have a relative boost in probability of being drafted relative to all other conferences.


Max Tree Depth
1
2
3
colsample_bytree: 0.6
eta: 0.4
subsample: 0.50
subsample: 0.75
subsample: 1.00
eta: 0.3
0.955
0.950
0.945
0.940
60
80
100
120
140
Accuracy (Cross-Validation)
Boosting Iterations


Max Tree Depth
1
2
3
colsample_bytree: 0.8
eta: 0.4
subsample: 0.50
subsample: 0.75
subsample: 1.00
eta: 0.3
0.955
0.950
0.945
0.940
60
80
100
120
140
Accuracy (Cross-Validation)
Boosting Iterations



Our next model is to fit a generalized linear model or logistic regression to fit to our dataset to see if there is any improvement of the model. The GLM model will also help us to understand the coefficients of the model to see how they affect the output. What's also helpful with this output is the p-value, which helps to see statistical significance in the model's output. This biggest coefficient in value and signficance is college years. This model suggest that an increase in college years actually decreases their chances of being drafted. That's an interesting takeaway that long career running backs may be more banged up going into the NFL. Similar to the previous model, HS Rating still has high weight of draftability. Being an SEC rusher highly increases your odds of being drafted according to this model too. TDs and Successes are also contributing variables in the models too.

```
## + Fold1: parameter=none
## - Fold1: parameter=none
## + Fold2: parameter=none
## - Fold2: parameter=none
## + Fold3: parameter=none
## - Fold3: parameter=none
## + Fold4: parameter=none
## - Fold4: parameter=none
## + Fold5: parameter=none
## - Fold5: parameter=none
## Aggregating results
## Fitting final model on full training set
```

```
##
## Call:
## NULL
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.7405  -0.2519  -0.1659  -0.0766   3.2391
##
## Coefficients:
##                                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)                           -8.909e+00  2.352e+00  -3.787 0.000152
## rating                                 6.605e+00  2.455e+00   2.691 0.007132
## `offense_conferenceAmerican Athletic` -1.410e+01  7.084e+02  -0.020 0.984119
## `offense_conferenceBig 12`            -4.266e-01  5.331e-01  -0.800 0.423557
## `offense_conferenceBig East`           2.423e-02  6.435e-01   0.038 0.969963
## `offense_conferenceBig Ten`           -1.026e-01  5.266e-01  -0.195 0.845495
## `offense_conferenceConference USA`    -1.811e+00  8.633e-01  -2.097 0.035971
## `offense_conferenceFBS Independents`  -4.228e-01  1.180e+00  -0.358 0.720073
## `offense_conferenceMid-American`      -1.778e+00  9.596e-01  -1.853 0.063907
## `offense_conferenceMountain West`     -1.345e+00  9.287e-01  -1.448 0.147608
## `offense_conferencePac-12`             2.522e-01  4.802e-01   0.525 0.599492
## offense_conferenceSEC                  4.373e-01  4.682e-01   0.934 0.350281
## `offense_conferenceSun Belt`          -2.505e+00  1.274e+00  -1.966 0.049349
## `offense_conferenceWestern Athletic`   4.081e-01  8.189e-01   0.498 0.618249
## mean_ppa                              -5.140e-01  6.359e-01  -0.808 0.418887
## success_rate                          -6.660e-02  1.367e+00  -0.049 0.961158
## ypc                                    2.286e-02  6.818e-02   0.335 0.737419
## rush_yds                              -1.568e-04  2.248e-03  -0.070 0.944377
## successes                             -8.822e-03  1.051e-02  -0.839 0.401262
## success_yds                            2.284e-03  2.205e-03   1.036 0.300210
## TDs                                    8.475e-02  3.144e-02   2.696 0.007016
## college_years                         -5.171e-01  1.594e-01  -3.244 0.001180
##
## (Intercept)                           ***
## rating                                **
## `offense_conferenceAmerican Athletic`
## `offense_conferenceBig 12`
## `offense_conferenceBig East`
## `offense_conferenceBig Ten`
## `offense_conferenceConference USA`    *
## `offense_conferenceFBS Independents`
## `offense_conferenceMid-American`      .
## `offense_conferenceMountain West`
## `offense_conferencePac-12`
## offense_conferenceSEC
## `offense_conferenceSun Belt`          *
## `offense_conferenceWestern Athletic`
## mean_ppa
## success_rate
## ypc
## rush_yds
## successes
## success_yds
## TDs                                   **
## college_years                         **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 731.62  on 1555  degrees of freedom
```

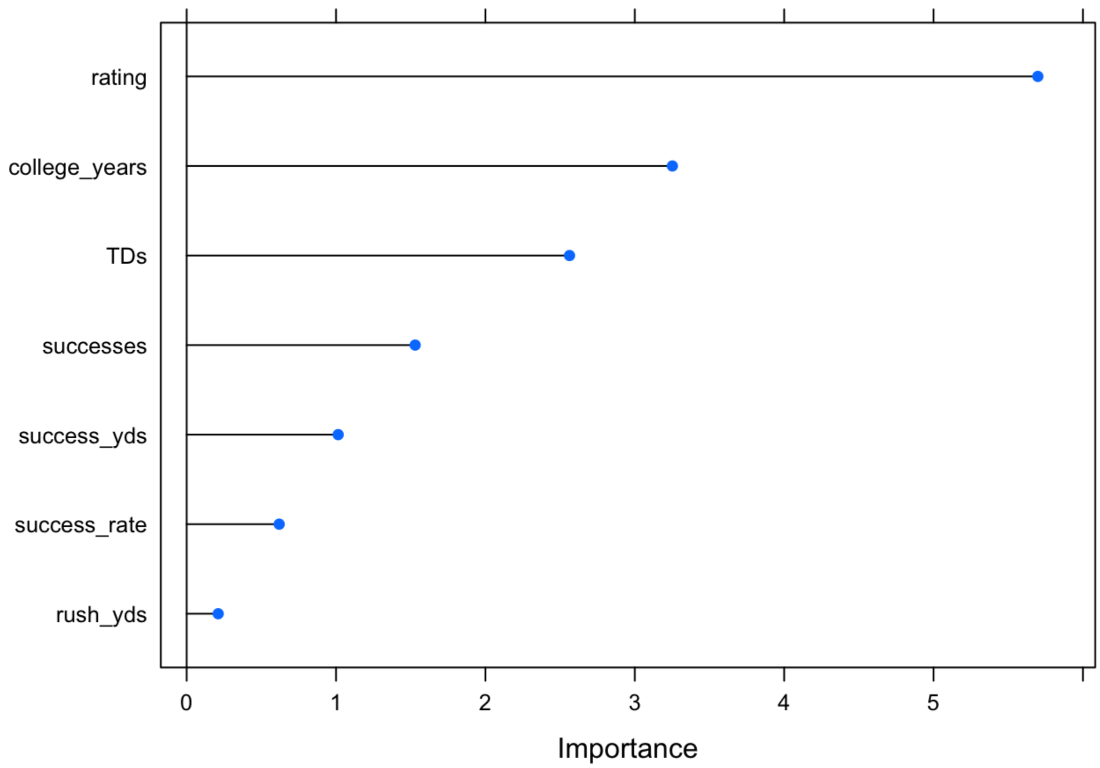
```
## Residual deviance: 420.64  on 1534  degrees of freedom
## AIC: 464.64
##
## Number of Fisher Scoring iterations: 16
```

The last model will take the same GLM approach but with a lighter and insignificant amount of variables. For this lighter model, we solely included HS rating, success rate, career rush yards, total career successes, career success yards, career TDs, and college years.

Improved AIC by 2pts; From this we tested the train and test set and returned a 96% and 95% accuracy, which is a slight improvement from the XGBTree model. Given the simplicity of the model, this makes sense. To confirm, statistical significance we ran a Hosmer and Lemeshow Goodness of Fit and see our p-value is statistically significant. In this model, HS Rating is the most important followed by years of experience (which we confirmed has a slight negative coefficient). Given the simplicity and accuracy, we will choose this model.

```
## + Fold1: parameter=none
## - Fold1: parameter=none
## + Fold2: parameter=none
## - Fold2: parameter=none
## + Fold3: parameter=none
## - Fold3: parameter=none
## + Fold4: parameter=none
## - Fold4: parameter=none
## + Fold5: parameter=none
## - Fold5: parameter=none
## Aggregating results
## Fitting final model on full training set
```

```
##
## Call:
## NULL
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.9115  -0.2483  -0.1673  -0.1138   3.4379
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.278e+01  1.856e+00  -6.887 5.71e-12 ***
## rating         1.131e+01  1.986e+00   5.698 1.21e-08 ***
## success_rate  -6.184e-01  9.981e-01  -0.620  0.53555
## rush_yds       4.473e-04  2.118e-03   0.211  0.83277
## successes     -1.476e-02  9.647e-03  -1.530  0.12607
## success_yds    2.129e-03  2.099e-03   1.014  0.31042
## TDs            7.724e-02  3.014e-02   2.563  0.01038 *
## college_years -5.046e-01  1.552e-01  -3.252  0.00115 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 731.62  on 1555  degrees of freedom
## Residual deviance: 442.44  on 1548  degrees of freedom
## AIC: 458.44
##
## Number of Fisher Scoring iterations: 7
```




```
##
##  Hosmer and Lemeshow goodness of fit (GOF) test
##
## data:  dfc.train$rush_drafted, fitted(pts_glm1)
## X-squared = 1556, df = 8, p-value < 2.2e-16
```

```
##
##      FALSE       TRUE
## 0.03675345 0.96324655
```

```
##
##      FALSE       TRUE
## 0.04562982 0.95437018
```

Using this model, we can predict based off RBs career stats, the probability of that runningback getting drafted. The table below shows the Top 20 in probability of getting drafted as a running back in the NFL.

We can see the model works well for the most part with the exception of Noel Devine, Denard Robinson and John Clay. Noel Devine with Pat White were one of the winningest duos in CFB history in the past 15 years. Unfortunately, Noel Devine was hampered by ankle injuries in his senior season (this kept him out of draft combine). Noel Devine ended up being undrafted and signed with the Philadelphia Eagles. He spent 4 days with the team before retiring from the NFL. He spent some time in the Canadian Football League and American Arena League. Denard Robinson was actually drafted as a Runningback in the NFL but given he was mostly a QB in college, he didn't quite qualify as runningback drafted. This shows a shortcoming to the model of people switching positions. John Clay was a slight anomaly. He forgoed his senior season to enter the 2011 NFL Draft and ended up being undrafted. He was signed by Steelers and ended up only recording 41 career rushing yards. Outside of those anomalies, we would say the model is fairly accurate for those looking at making the leap, which may have the application for RBs deciding to enter the draft or not. All of the running backs were fairly household names by the time of the draft.

Top 20 Draft Probability

rusher_player_name	rating	rush_drafted	Draft_Prob	position	pos_team	rush_Rnd	mean_ppa	success_rate	ypc	rush_yds	TDs	college_years

rusher_player_name	rating	rush_drafted	Draft_Prob	position	pos_team	rush_Rnd	mean_ppa	success_rate	ypc	rush_yds	TDs	college_years
Montee Ball	0.8844	1	0.9967448	RB	Wisconsin	2	0.03	0.50	5.58	5142	77	4
Denard Robinson	0.9427	0	0.9890405	ATH	Michigan	0	0.18	0.52	6.76	4708	42	4
Noel Devine	0.9957	0	0.9855680	RB	West Virginia	0	-0.04	0.40	5.79	4315	30	4
Steve Slaton	0.8333	1	0.9755604	ATH	West Virginia	3	-0.06	0.45	5.77	3820	52	3
Darren McFadden	0.9855	1	0.9668772	RB	Arkansas	1	-0.01	0.44	5.65	3884	32	3
Jamaal Charles	0.9775	1	0.9592610	RB	Texas	3	0.03	0.53	6.42	3389	37	3
Ka'Deem Carey	0.8799	1	0.9584731	RB	Arizona	4	0.05	0.49	5.67	4235	48	3
James White	0.8733	1	0.9523361	RB	Wisconsin	4	0.01	0.48	6.22	3999	45	4
DeMarco Murray	0.9901	1	0.9451119	RB	Oklahoma	3	-0.05	0.46	4.83	3667	49	4
Jonathan Dwyer	0.9532	1	0.9353460	RB	Georgia Tech	6	-0.03	0.40	6.12	3071	33	3
Mark Ingram	0.9444	1	0.9281054	RB	Alabama	1	0.06	0.49	5.69	3259	42	3
Trent Richardson	0.9971	1	0.9249266	RB	Alabama	1	0.03	0.50	5.92	3091	33	3
Jacquizz Rodgers	0.8794	1	0.9225044	APB	Oregon State	5	-0.05	0.45	4.90	3903	46	3
Knowshon Moreno	0.9717	1	0.9085503	RB	Georgia	1	-0.01	0.42	5.56	2757	30	2
John Clay	0.9751	0	0.9038886	RB	Wisconsin	0	0.10	0.55	5.42	3399	41	3
Jahvid Best	0.9220	1	0.8872554	RB	California	1	0.02	0.47	7.37	2668	29	3
James Davis	0.9418	1	0.8848187	RB	Clemson	6	-0.04	0.45	4.86	3561	43	4
Lance Dunbar	0.8274	0	0.8820313	APB	North Texas	0	-0.06	0.42	5.53	3951	39	4
Andre Ellington	0.9471	1	0.8810040	APB	Clemson	6	-0.02	0.46	5.55	3443	34	4
Bishop Sankey	0.9028	1	0.8555069	RB	Washington	2	0.08	0.47	5.42	3497	37	3

Note:
Filtered for data from 2005-2019

In looking undrafted RB with highest probability, the top 4 are fairly big anomalies in the model. For instance, Lance Dunbar was a regular backup for the Cowboys with over 400+ career rushing yards. Michael Dyer had a spectacular career at Louisville/Auburn playing with both Cam Newton and Lamar Jackson but his transfer raised some off field questions to be undrafted. Malcom Brown, a undrafted signee with the LA Rams ended up being pivotal with the Rams run to the Superbowl when Todd Gurley went down and has as of late proved to be an effective NFL runningback. Byron Marshall was used as a hybrid RB/WR in his role at Oregon proving to be more of hyrid athlete on the field in the right scheme. Taylor Martinez, is still active in college and a QB putting him in a similar situation as Denard Robinson.

Top 15 Draft Probability for Undrafted RBs

rusher_player_name	rating	rush_drafted	Draft_Prob	position	pos_team	rush_Rnd	mean_ppa	success_rate	ypc	rush_yds	TDs	college_years
Denard Robinson	0.9427	0	0.9890405	ATH	Michigan	0	0.18	0.52	6.76	4708	42	4
Noel Devine	0.9957	0	0.9855680	RB	West Virginia	0	-0.04	0.40	5.79	4315	30	4
John Clay	0.9751	0	0.9038886	RB	Wisconsin	0	0.10	0.55	5.42	3399	41	3
Lance Dunbar	0.8274	0	0.8820313	APB	North Texas	0	-0.06	0.42	5.53	3951	39	4
Damion Fletcher	0.7896	0	0.7839627	APB	Southern Mississippi	0	0.07	0.48	5.25	4901	40	4
Percy Harvin	0.9989	0	0.6861610	WR	Florida	0	0.29	0.67	9.56	1855	19	3
Adam Muema	0.8372	0	0.6746392	RB	San Diego State	0	0.01	0.46	5.50	2957	34	3
Tyrell Sutton	0.8333	0	0.6614800	RB	Northwestern	0	0.04	0.48	5.37	3914	33	4
Michael Dyer	0.9960	0	0.6584740	RB	Auburn	0	-0.02	0.49	5.53	2340	15	2
Tim Cornett	0.8333	0	0.6313628	RB	UNLV	0	-0.07	0.41	4.88	3683	35	4
Antone Smith	0.9932	0	0.6195528	RB	Florida State	0	-0.17	0.39	4.49	2154	26	4

Note:
Filtered for data from 2005-2020
[1] Filtered for those with 1000 career yards

rusher_player_name	rating	rush_drafted	Draft_Prob	position	pos_team	rush_Rnd	mean_ppa	success_rate	ypc	rush_yds	TDs	college_years
Zach Line	0.7444	0	0.5700005	ILB	SMU	0	0.04	0.45	5.32	3823	44	4
Cierre Wood	0.9694	0	0.5674237	RB	Notre Dame	0	0.02	0.44	5.45	2460	16	3
Onterio McCalebb	0.8938	0	0.5634013	APB	Auburn	0	0.11	0.48	6.37	2597	24	4
P.J. Hill	0.8111	0	0.5493958	FB	Wisconsin	0	0.03	0.48	5.05	3428	37	3
James Sims	0.8733	0	0.5460228	RB	Kansas	0	-0.06	0.41	4.49	3586	34	4
David Fluellen	0.8528	0	0.5304531	RB	Toledo	0	0.10	0.50	5.97	3365	28	4
Antonio Andrews	0.8188	0	0.5106627	APB	Western Kentucky	0	0.13	0.51	5.94	3641	29	4
Bryce Beall	0.8181	0	0.4871150	S	Houston	0	0.16	0.56	5.84	3155	39	4
Kapri Bibbs	0.8211	0	0.4769954	RB	Colorado State	0	0.12	0.50	6.19	1746	30	1

Note:
Filtered for data from 2005-2020
[1] Filtered for those with 1000 career yards

Below lists (in order) the top draft probability of RBs from non Power 5 conferences. You'll notice some recognizable names below. What you may notice is that a lot of these runningbacks have late round appeal, which goes to show you can still find some value in later rounds. Our next goal would be to project what round we think runningback's would go in or their average draft position would be.

Top Draft Probability for RB Drafted from Non Power 5 Conferences

rusher_player_name	rating	Draft_Prob	position	pos_team	offense_conference	rush_Rnd	mean_ppa	success_rate	ypc	rush_yds	TDs	college_years
Steve Slaton	0.8333	0.9755604	ATH	West Virginia	Big East	3	-0.06	0.45	5.77	3820	52	3
Donald Brown	0.8333	0.8496733	RB	Connecticut	Big East	1	-0.02	0.45	5.48	3713	34	3
Ronnie Hillman	0.8479	0.7984916	ATH	San Diego State	Mountain West	3	-0.02	0.47	5.52	3105	34	2
Ryan Mathews	0.8198	0.7533292	RB	Fresno State	Western Athletic	1	0.04	0.46	6.09	3146	38	3
Bernard Pierce	0.7667	0.7525464	RB	Temple	Mid-American	3	-0.03	0.45	5.24	3257	49	3
Dion Lewis	0.8733	0.7197777	APB	Pittsburgh	Big East	5	-0.04	0.42	5.23	2839	30	2
Jordan Todman	0.8306	0.6465469	RB	Connecticut	Big East	6	0.00	0.42	5.19	3201	32	3
Kevin Smith	0.7667	0.6067585	RB	UCF	Conference USA	3	-0.05	0.47	4.86	3940	38	3
Robert Turbin	0.7674	0.5295358	RB	Utah State	Western Athletic	4	0.02	0.46	5.87	3299	39	4
Jay Ajayi	0.8528	0.3554248	RB	Boise State	Mountain West	5	0.05	0.48	5.98	1990	22	2
Latavius Murray	0.8333	0.3540728	RB	UCF	Conference USA	6	-0.04	0.48	5.31	2368	36	4
Kenneth Dixon	0.8472	0.3007806	RB	Louisiana Tech	Western Athletic	4	0.12	0.59	6.12	1212	27	1
Ray Rice	0.8333	0.2921127	APB	Rutgers	Big East	2	-0.02	0.47	5.17	1789	21	1
Kerwynn Williams	0.7667	0.2856189	APB	Utah State	Western Athletic	7	0.03	0.48	6.66	2436	22	3
Jamaal Williams	0.8335	0.2458454	APB	BYU	FBS Independents	4	0.00	0.41	5.22	1983	19	2
Jawan Jamison	0.8622	0.1737401	APB	Rutgers	Big East	7	-0.14	0.35	4.04	1973	12	2
Delone Carter	0.8333	0.1586102	RB	Syracuse	Big East	4	-0.01	0.40	4.62	2874	20	4
Bilal Powell	0.8111	0.1522220	RB	Louisville	Big East	4	-0.06	0.45	5.39	2338	20	4
Elijah McGuire	0.8415	0.0955185	RB	Louisiana	Sun Belt	6	0.15	0.50	8.23	856	8	1
Aaron Brown	0.7667	0.0886455	RB	TCU	Mountain West	6	0.04	0.45	5.55	2479	20	4
Kenneth Dixon	0.8472	0.0669166	RB	Louisiana Tech	Conference USA	4	0.12	0.47	6.10	915	3	1
Donnel Pumphrey	0.8042	0.0506131	APB	San Diego State	Mountain West	4	-0.06	0.44	5.94	749	8	1

Note:
Filtered for data from 2005-2020
[1] Filtered for those with 1000 career yards

rusher_player_name	rating	Draft_Prob	position	pos_team	offense_conference	rush_Rnd	mean_ppa	success_rate	ypc	rush_yds	TDs	college_years
Kareem Hunt	0.8032	0.0379660	RB	Toledo	Mid-American	3	0.22	0.58	6.32	866	6	1
Theo Riddick	0.8939	0.0245440	APB	Notre Dame	FBS Independents	6	0.01	0.47	4.81	1187	5	4
Anthony Sherman	0.8243	0.0039550	ILB	Connecticut	Big East	5	-0.30	0.18	3.59	61	0	4
Eddie Williams	0.7000	0.0027138	RB	Idaho	Western Athletic	7	0.67	0.71	9.71	136	3	2

Note:
Filtered for data from 2005-2020
[1] Filtered for those with 1000 career yards

#Predict Average Draft Position # For Prediciting Average Position, we will building a regression model. Given a majority of running backs will not be drafted we will have a fair amount of zeros inflating our model. We will building a regression model specifically handling zero inflation to correct our sampling for. To illustrate, we've constructed 3 models. The first, includes the following formula:

formula = rush_pick ~ rating + mean_ppa + success_rate + ypc + success_yds + TDs + college_years + ACC + MountainWest + WesternAthletic + SEC + PAC12 + BigTen + Big12,data=dfc.train)

The second formula is a lighter model not accounting for the conferences:
m2<-zeroinfl(formula = rush_pick ~ rating + mean_ppa + success_rate + ypc + success_yds + TDs + college_years,data=dfc.train)

The Third formula is the lightest model only accounting for rating, success yards and college years, which we've shown to be major inflences.
m3<-zeroinfl(formula = rush_pick ~ rating + success_yds + college_years,data=dfc.train)

We can see that regardless of the variables added or subtracted, the best R-squared we can get is 0.19, which means that 19% of predicted variables can be explained of the current dataset. To improve, we likely need other variables to add to improve.

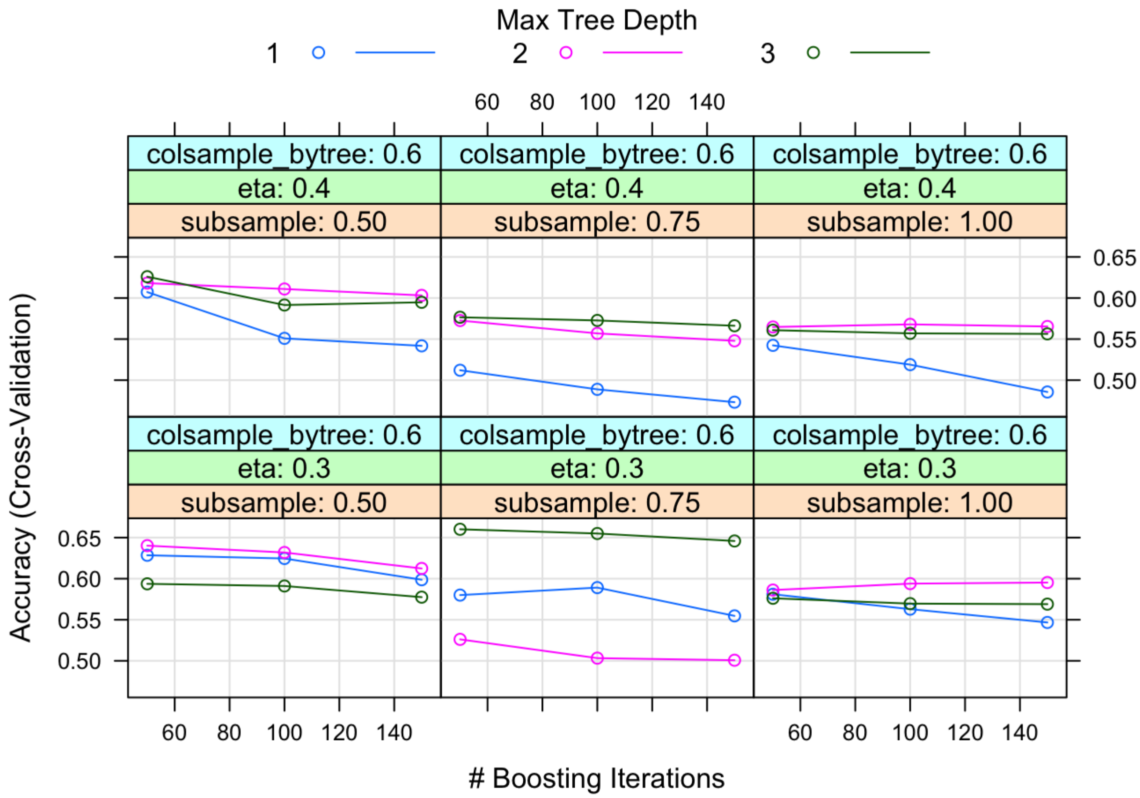
Model Comparison for ADP

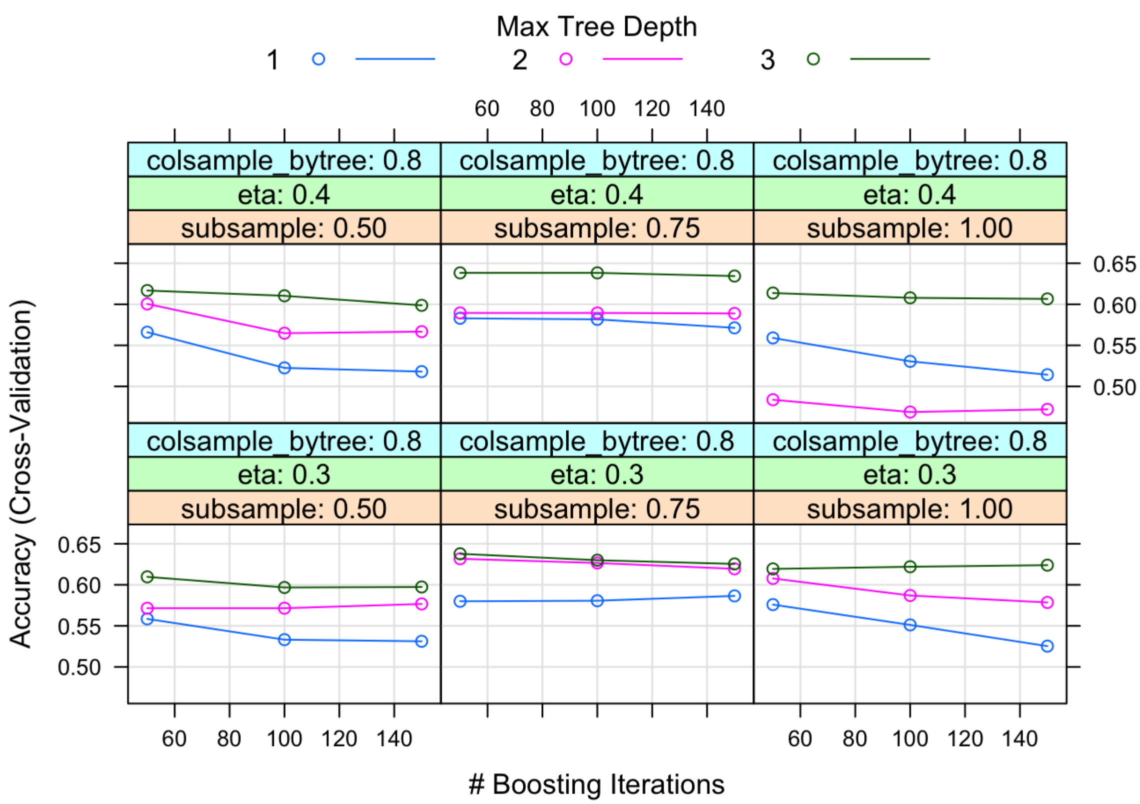
formula	R-squared
Large Model	0.0753144
No Conference Model	0.0623920
Light Model	0.0768169

Note:
Filtered for data from 2000-2020
[1] ADP = Average Draft Position

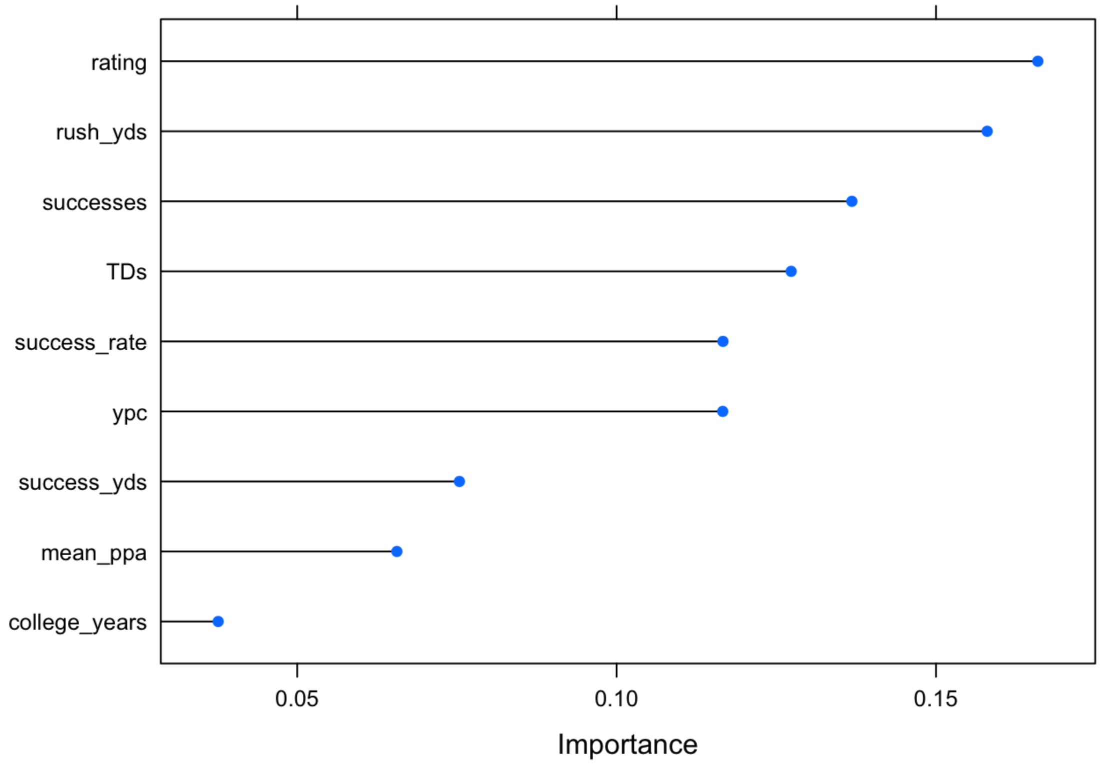
Predict Round Drafted

Since we can't accurate predict the average draft position with our current dataset, let's see if we can predicted Round drafted with our current data set. We will be building a classification model that aims to predict 8 classifications (7 rounds plus not drafted probability). To account for a same issue of an imbalance of non-drafted, we can use a "down sampling" technique to treat the prediction of each classifications separately. This helps to make sure that the probability doesn't over-generalize for zero imbalance. In using an XGBTree, we can see with our adjust imbalance classification model that we were able to achieve roughly ~55% accuracy. The 55% although not as high as 95% accuracy of a binary classification model is fairly decently given we are assigning a probability to 8 different classifcation at once.





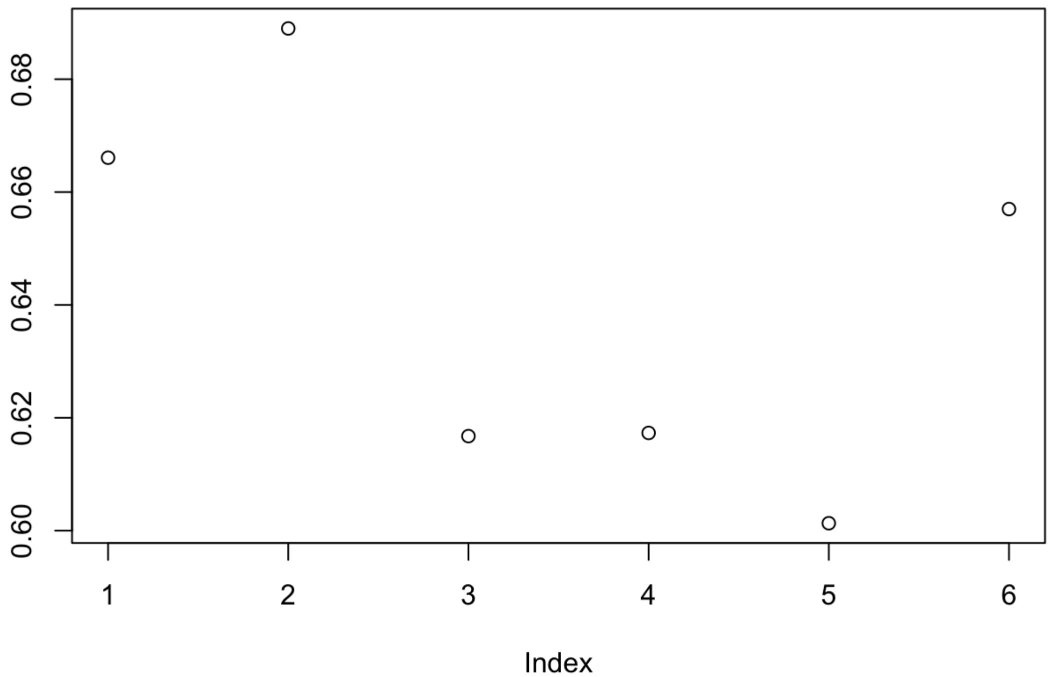
In observing the important factors of the model, we observe that HS Rating has high importance in the average round a runningback is drafted. What's interest around the importance of this model is that college years has low importance overall in what round a RB is drafted in. One area of concern is that the accuracy on the test set is 49% and 57.6% accuracy on the train, which may indicate some overfitting of the XGBTree model.




```
##
##     FALSE      TRUE
## 0.4125561 0.5874439
```

```
##
##    FALSE     TRUE
## 0.374026 0.625974
```

Given our high range of values between 49% accuracy and 57.6% accuracy, we seek to fit another type of predictive algorithm to see if we can find any improvement. With that, we have chosen to fit a random forest model to see if we will get different results and a less overfit model. Our accuracy is 75.8% on the training set and ~67% accuracy on the test set. We still have a nature of overfitting given the different but the improvement along allows us a barebone model to work to use for analytics/predictions.




```
##
##     FALSE      TRUE
## 0.3168909 0.6831091
```

```
##
##     FALSE      TRUE
## 0.2454545 0.7545455
```

With the Random Forest Model, we seek to observe how accurate the predictions are through historical values. The first is to look at those predict in the correct timeframe. The first table shows accurate 1st round predictions and the 2nd table shows accurate 2nd day drafted runningbacks.

First Round Probability & Drafted in 1st Round

rusher_player_name	rating	pred	position	pos_team	rush_Rnd1	mean_ppa	success_rate	ypc	rush_yds	TDs	college_years
Trent Richardson	0.9971	1	RB	Alabama	1	0.03	0.50	5.92	3091	33	3
Jonathan Stewart	0.9952	1	RB	Oregon	1	0.05	0.52	5.79	2706	22	3
C.J. Spiller	0.9909	1	RB	Clemson	1	-0.07	0.40	5.38	2547	22	3
Darren McFadden	0.9855	1	RB	Arkansas	1	-0.01	0.44	5.65	3884	32	3
Rashard Mendenhall	0.9775	1	RB	Illinois	1	0.15	0.51	6.49	2545	22	3
Knowshon Moreno	0.9717	1	RB	Georgia	1	-0.01	0.42	5.56	2757	30	2
Ezekiel Elliott	0.9693	1	APB	Ohio State	1	0.44	0.63	8.73	262	2	1
Todd Gurley	0.9654	1	RB	Georgia	1	0.11	0.49	6.16	2378	27	2
David Wilson	0.9585	1	RB	Virginia Tech	1	0.00	0.43	5.71	2628	18	3
Mark Ingram	0.9444	1	RB	Alabama	1	0.06	0.49	5.69	3259	42	3
Jahvid Best	0.9220	1	RB	California	1	0.02	0.47	7.37	2668	29	3

Note:
Filtered for data from 2000-2020

rusher_player_name	rating	pred	position	pos_team	rush_Rnd1	mean_ppa	success_rate	ypc	rush_yds	TDs	college_years
Felix Jones	0.9000	1	ATH	Arkansas	1	0.10	0.47	7.35	2506	15	3
Melvin Gordon	0.8983	1	RB	Wisconsin	1	0.25	0.57	8.14	2335	16	3
Ryan Mathews	0.8198	1	RB	Fresno State	1	0.04	0.46	6.09	3146	38	3

Note:
Filtered for data from 2000-2020

Second Day Probability & Drafted 2nd Day

rusher_player_name	rating	pred	position	pos_team	rush_Rnd1	mean_ppa	success_rate	ypc	rush_yds	TDs	college_years
Derrick Henry	0.9928	2	ATH	Alabama	2	0.30	0.67	10.61	382	3	1
Christine Michael	0.9880	2	RB	Texas A&M	2	0.01	0.48	5.34	2370	23	3
Christine Michael	0.9880	2	RB	Texas A&M	2	-0.09	0.49	4.74	417	12	1
Ben Tate	0.9793	2	RB	Auburn	2	-0.03	0.46	4.89	3303	24	4
Montario Hardesty	0.9602	2	RB	Tennessee	2	-0.15	0.41	4.15	2285	26	5
Giovani Bernard	0.9232	2	RB	North Carolina	2	0.05	0.49	5.89	2499	25	2
Shane Vereen	0.9222	2	APB	California	2	-0.02	0.43	5.03	2792	25	3
Bishop Sankey	0.9028	2	RB	Washington	2	0.08	0.47	5.42	3497	37	3
Jeremy Hill	0.8994	2	RB	LSU	2	0.05	0.47	6.28	2172	28	2
Toby Gerhart	0.8701	2	FB	Stanford	2	0.06	0.47	5.22	3277	41	4
Ameer Abdullah	0.8653	2	ATH	Nebraska	2	0.09	0.49	5.45	2991	20	3
Ray Rice	0.8333	2	APB	Rutgers	2	-0.02	0.47	5.17	1789	21	1
DeMarco Murray	0.9901	3	RB	Oklahoma	3	-0.05	0.46	4.83	3667	49	4
Duke Johnson	0.9894	3	APB	Miami	3	0.08	0.47	6.56	1870	16	2
Jamaal Charles	0.9775	3	RB	Texas	3	0.03	0.53	6.42	3389	37	3
Kenyan Drake	0.9454	3	APB	Alabama	3	0.24	0.61	7.28	975	13	2
Matt Jones	0.9347	3	ATH	Florida	3	-0.02	0.43	4.73	620	5	2
Knile Davis	0.9127	3	RB	Arkansas	3	-0.01	0.47	5.34	1844	20	3
Carlos Hyde	0.9078	3	FB	Ohio State	2	0.16	0.56	6.16	3216	37	4
Tre Mason	0.8954	3	RB	Auburn	3	0.10	0.53	5.78	2980	33	3
Tevin Coleman	0.8686	3	RB	Indiana	3	0.02	0.45	6.48	1167	13	2
Glen Coffee	0.8556	3	RB	Alabama	3	0.04	0.48	5.33	2169	15	3
James Conner	0.8370	3	OLB	Pittsburgh	3	0.01	0.52	5.21	786	8	3
Steve Slaton	0.8333	3	ATH	West Virginia	3	-0.06	0.45	5.77	3820	52	3
Kareem Hunt	0.8032	3	RB	Toledo	3	0.22	0.58	6.32	866	6	1
Bernard Pierce	0.7667	3	RB	Temple	3	-0.03	0.45	5.24	3257	49	3
Kevin Smith	0.7667	3	RB	UCF	3	-0.05	0.47	4.86	3940	38	3

Note:
Filtered for data from 2005-2020

This table shows where the RBs where predicted a lot higher than they were drafted. In some cases, the prediction was dead-on in evaluating late round talent (Devonta Freeman, Jay Ajayi, Le'Veon Bell), however in many other cases where the RB was drafted anecdotally makes sense with their performance in the NFL.

RBs Who Slid in the Draft (according to the Prediction)

rusher_player_name	rating	pred	position	pos_team	rush_Rnd	mean_ppa	success_rate	ypc	rush_yds	TDs	college_years

Note:
Filtered for data from 2000-2020
[1] Actual Round greater than predicted round for first 4 round talent

rusher_player_name	rating	pred	position	pos_team	rush_Rnd	mean_ppa	success_rate	ypc	rush_yds	TDs	college_years
Eddie Lacy	0.9400	1	RB	Alabama	2	0.14	0.53	6.71	2335	30	3
T.J. Yeldon	0.9876	1	RB	Alabama	2	0.19	0.56	6.14	2347	26	2
De'Anthony Thomas	0.9926	1	APB	Oregon	4	0.16	0.54	7.74	1874	26	3
Chris Rainey	0.9787	1	RB	Florida	5	0.03	0.45	6.21	2452	13	5
Jonathan Dwyer	0.9532	1	RB	Georgia Tech	6	-0.03	0.40	6.12	3071	33	3
Kenjon Barner	0.8684	1	RB	Oregon	6	0.16	0.54	6.29	2699	32	2
Kerwynn Williams	0.7667	1	APB	Utah State	7	0.03	0.48	6.66	2436	22	3
Devonta Freeman	0.9490	2	RB	Florida State	4	0.04	0.51	5.55	2241	31	3
Marcus Lattimore	0.9950	2	APB	South Carolina	4	-0.03	0.46	4.78	2639	38	3
Joseph Randle	0.8917	2	ATH	Oklahoma State	5	0.06	0.49	5.43	3049	39	3
Cyrus Gray	0.9668	2	RB	Texas A&M	6	-0.04	0.48	5.13	3236	30	4
Allen Bradford	0.9925	3	RB	USC	6	0.04	0.46	5.93	1588	16	5
James Davis	0.9418	3	RB	Clemson	6	-0.04	0.45	4.86	3561	43	4
Josh Robinson	0.8366	3	RB	Mississippi State	6	0.11	0.50	5.86	785	4	3
Chris Thompson	0.9197	4	RB	Florida State	5	-0.11	0.42	6.37	1727	14	4
Zac Stacy	0.8590	4	RB	Vanderbilt	5	-0.05	0.40	5.28	2933	29	4
Spencer Ware	0.9593	4	ATH	LSU	6	-0.08	0.41	4.22	1240	10	3
Jay Finley	0.8260	4	RB	Baylor	7	-0.03	0.42	5.57	2607	23	4

Note:
Filtered for data from 2000-2020
[1] Actual Round greater than predicted round for first 4 round talent

This table shows RBs that were picked ahead of the round projected. Again, in some cases the prediction was accurate in the assessment (Felix Jones, Rashard Mendenhall, Toby Gerhart, Shane Vereen) but in a lot of other cases, the model incorrectly predicted the round to where the runningback was drafted, which make show where some teams reached to get a player. This is likely due to other NFL draft statistics that was not included in the model.

Early Picks of Draft

rusher_player_name	rating	pred	position	pos_team	rush_Rnd	mean_ppa	success_rate	ypc	rush_yds	TDs	college_years
Donald Brown	0.8333	3	RB	Connecticut	1	-0.02	0.45	5.48	3713	34	3
Doug Martin	0.7771	3	RB	Boise State	1	0.03	0.52	6.03	2160	28	3
Doug Martin	0.7771	4	RB	Boise State	1	-0.11	0.44	4.89	1291	16	1
Carlos Hyde	0.9078	3	FB	Ohio State	2	0.16	0.56	6.16	3216	37	4
Montee Ball	0.8844	4	RB	Wisconsin	2	0.03	0.50	5.58	5142	77	4
Le'Veon Bell	0.8149	6	RB	Michigan State	2	0.02	0.42	4.98	3343	33	3
Ryan Williams	0.9782	6	RB	Virginia Tech	2	-0.04	0.43	5.36	2162	31	2
Ronnie Hillman	0.8479	5	ATH	San Diego State	3	-0.02	0.47	5.52	3105	34	2
Stevan Ridley	0.8875	5	ATH	LSU	3	-0.01	0.46	4.67	1433	19	3
Roy Helu	0.8385	0	RB	Nebraska	4	0.05	0.43	4.65	214	0	1
Javorius Allen	0.8905	5	RB	USC	4	-0.10	0.46	5.85	784	15	1
Jeremy Langford	0.8434	5	ATH	Michigan State	4	-0.07	0.49	4.80	1444	18	2
Johnathan Franklin	0.8943	5	RB	UCLA	4	-0.06	0.45	4.92	1701	14	2
Kendall Hunter	0.8590	6	RB	Oklahoma State	4	0.04	0.51	5.79	3937	34	4
Kenneth Dixon	0.8472	6	RB	Louisiana Tech	4	0.12	0.47	6.10	915	3	1

Note:
Filtered for data from 2000-2020
[1] Actual Round greater than predicted round for first 4 round talent

rusher_player_name	rating	pred	position	pos_team	rush_Rnd	mean_ppa	success_rate	ypc	rush_yds	TDs	college_years
Robert Turbin	0.7674	6	RB	Utah State	4	0.02	0.46	5.87	3299	39	4
Bilal Powell	0.8111	7	RB	Louisville	4	-0.06	0.45	5.39	2338	20	4
Wendell Smallwood	0.8524	0	APB	West Virginia	5	0.02	0.44	5.67	221	1	1
T.J. Logan	0.9250	7	APB	North Carolina	5	0.09	0.59	5.74	540	4	1
Aaron Brown	0.7667	7	RB	TCU	6	0.04	0.45	5.55	2479	20	4
Eddie Williams	0.7000	0	RB	Idaho	7	0.67	0.71	9.71	136	3	2

Note:
Filtered for data from 2000-2020
[1] Actual Round greater than predicted round for first 4 round talent

Applications for Operations & Learnings

We have learned that HS Rating is a big contributing factor to the probability of a RB getting drafted. Our initial model was being able to predict likelihood of being drafted or not helps current RBs on the fence give themselves a probability of coming back for another year or not. Our secondary models help to predict in the draft a runningback will be selected. We learned ADP is difficult to predict with Recruiting and on-field performance, which may suggest that other on-field performance and/or NFL Draft prep data may be needed. Also, we didn’t take into consideration athletic size and fit for the role, which we know is important in the evaluation process. Other data that may enhance the accuracy of the models would be injury history and potentially player “clout”. Understand these items more will help to evolve these initial models.