NFL Draft Prediction

Overview

- Taking information for the top 200 players that enter the combine every year seeing if we can combine that with College career stats and high school ranking to predict when they will get drafted.
- To see how important the combine really is in determining draft position.

Data

Data for the combine was scraped from Stathead.com but limited to only 200 players out of the around 300 invited every year.

Data for College season stats and high school ratings were gathered by College Football Data API

I used information for Combines and season stats from 2001 to 2015.

High school rating were only available to 2000 so I have from 2000 to 2015

32783 rows of recruit ratings

20083 rows of season stats

3000 rows of combine data

EDA

- 32783 rows of recruit ratings

- 20083 rows of season stats was boiled down from 325997 which was every player individual stat per year. These were grouped by and averaged to get per season average for all players.
    - It was noticed that the stats were for offensive skill players, kickers and interceptions.

Pos	Vertical	BenchReps	Broad Jump
C	28.463235	26.743243	101.911765
CB	36.364815	14.667969	122.106618
DE	33.507732	24.376884	115.256545
DT	29.475410	28.864865	104.692737
FB	32.363636	23.500000	112.490909
ILB	33.530435	23.438596	114.754386
K	33.500000	13.000000	116.000000
LS	NaN	15.000000	NaN
OG	28.127660	26.136986	101.170370
OLB	34.710383	22.983516	117.716578
OT	28.357143	25.348837	102.844311
P	30.500000	17.000000	120.000000
QB	31.936937	20.833333	110.990909
RB	34.819048	20.084158	118.368421
S	35.846774	17.213904	120.405405
TE	33.340278	20.827815	114.755245
WR	35.895765	15.277419	121.023729

- After merging all datasets together we were left with 936 players that we had combine, high school rating and career stats for.

player_id	team	player	stat_category	
-7111	Pittsburgh	Team	puntReturnsAVG	-3.0
			puntReturnsLONG	0.0
			puntReturnsNO	1.0
			puntReturnsTD	0.0
			puntReturnsYDS	-3.0
			rushingCAR	1.0
			rushingLONG	0.0
			rushingTD	0.0
			rushingYDS	-4.0
			rushingYPC	-4.0
-7088	Penn State	Team	rushingCAR	2.0
			rushingLONG	0.0
			rushingTD	0.0
			rushingYDS	-7.0
			rushingYPC	-3.5
-7014	Ohio State	Team	rushingCAR	1.0
			rushingLONG	0.0
			rushingTD	0.0
			rushingYDS	-1.0
			rushingYPC	-1.0

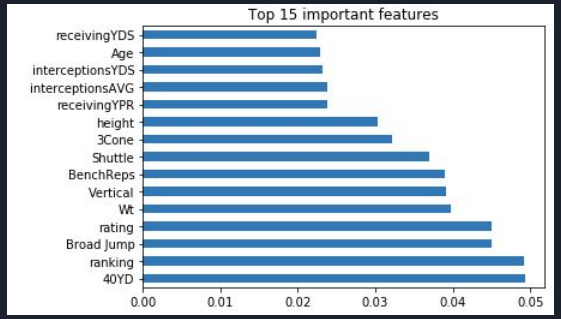
Models

I dummied position and schools to see if they could boost your chances of getting drafted higher up. The highest R^2 was .612

```
Train RMSE: 43.572865490456394
 Test RMSE: 161.64779510306542
Percent change:  270.983
Percent change (Base Model vs. Updated Model):  -40.704
```

Modeling for Round prediction through random forest

```
Testing Accuracy Random Forest Grid Bagged Classifier: 24.91%
```



Models Cont.

The best R^2 that was obtained from not binning position or schools was .395 and was not very accurate at that with

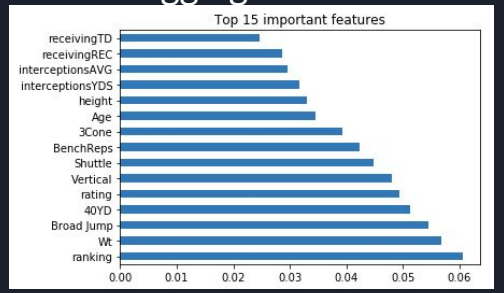
```
Train RMSE: 54.4189259886845
 Test RMSE: 85.88548704608179
Percent change:  57.823
Percent change (Base Model vs. Updated Model):  -33.507
```

Interesting interactions when comparing by logistic regression

```
Shuttle and passingYDS: 0.346
Shuttle and passingATT: 0.346
40YD and height: 0.345
Shuttle and puntingNO: 0.344
Shuttle and passingINT: 0.344
40YD and passingPCT: 0.344
Vertical and passingYDS: 0.343
Vertical and passingATT: 0.343
Broad Jump and passingYDS: 0.343
Broad Jump and passingATT: 0.343
puntReturnsAVG and puntReturnsLONG: 0.342
```

Then modeled to predict round drafted by using random forest and bagging and gridsearch

```
Testing Accuracy Random Forest Grid Bagged Classifier: 25.27%
```



Conclusion

There is some promise to predicting based on Combine as these are usually the top players getting invited.

More data is needed to obtain a more accurate prediction as with the first modeling we could predict 62 percent of the variance but not accurately.

Next Steps

Raise money to pay for the 500 dollars to obtain accurate historical college stats data.

Remove combine restriction to include all 256 players that are drafted every year.

Predict against AV which is approximate value that is how much impact a player has made at their position in the NFL (of course Tom brady has the highest)

Analyze impact of team needs and how they affect the early rounds.

Thank you

https://github.com/murra181/draft_prediction