Pro Bowl: Finding Snubs Using Regression

Matthew Kraus

Introduction

It's that time of year again. The Pro Bowl (Games) rosters have been announced and the NFL world was once again quick to judge who should and should not have been selected. Like many others, I saw a few players on the roster that, in my opinion, did not deserve to make the trip to Vegas. However, as I delve deeper into my sport analytics journey, I wanted to determine whether or not each player should have made the final roster according to the statistics each player recorded.

Notes

1. Data includes Pro Bowl selections from 2012 to 2022.
2. Positions such as offensive linemen and special teams players are not included.
3. Only the original Pro Bowl selections are used. Replacements are labeled as not selected.
4. All data is from Pro Football Reference.

Example

Each player in the data set is assigned either a 0 or 1 which determines whether or not they made the Pro Bowl during that specific season. If you wanted to know the original six quarterback selections for the 2014 season, the output would look like this.

```
players %>% filter(Pos == 'QB' & Year == 2014 & ProBowler == 1)
```

```
##                 Player  Tm Pos  G GS Cmp PassAtt PassYds PassTD PassInt Year
## 1:        Aaron Rodgers GNB  QB 16 16 341     520    4381     38       5 2014
## 2:            Tom Brady NWE  QB 16 16 373     582    4109     33       9 2014
## 3:            Tony Romo DAL  QB 15 15 304     435    3705     34       9 2014
## 4: Ben Roethlisberger PIT  QB 16 16 408     608    4952     32       9 2014
## 5:          Andrew Luck IND  QB 16 16 380     616    4761     40      16 2014
## 6:        Peyton Manning DEN  QB 16 16 395     597    4727     39      15 2014
##    RushAtt RushYds RushTD Rec RecYds RecTD DefInt Sacks Solo ProBowler
## 1:      43     269      2   0      0     0      0     0    0         1
## 2:      36      57      0   0      0     0      0     0    0         1
## 3:      26      61      0   0      0     0      0     0    0         1
## 4:      33      27      0   0      0     0      0     0    0         1
## 5:      64     273      3   0      0     0      0     0    0         1
## 6:      24     -24      0   0      0     0      0     0    0         1
##    probowl_prob
## 1:        0.977
## 2:        0.205
## 3:        0.468
## 4:        0.748
## 5:        0.752
## 6:        0.447
```

As you can see above, Aaron Rodgers was given a 97.7% chance to make the Pro Bowl based on the model. However, Tom Brady was selected with only a 20.5% chance to make it.

We can see who the model thought was more deserving of a Pro Bowl bid below.

```
players %>% filter(Pos == 'QB' & Year == 2014) %>% arrange(desc(probowl_prob)) %>% head(6)
```

```
##                 Player  Tm Pos  G GS Cmp PassAtt PassYds PassTD PassInt Year
## 1:        Aaron Rodgers GNB  QB 16 16 341     520    4381     38       5 2014
## 2:       Russell Wilson SEA  QB 16 16 285     452    3475     20       7 2014
## 3:          Andrew Luck IND  QB 16 16 380     616    4761     40      16 2014
## 4: Ben Roethlisberger PIT  QB 16 16 408     608    4952     32       9 2014
## 5:            Tony Romo DAL  QB 15 15 304     435    3705     34       9 2014
## 6:        Peyton Manning DEN  QB 16 16 395     597    4727     39      15 2014
##    RushAtt RushYds RushTD Rec RecYds RecTD DefInt Sacks Solo ProBowler
## 1:      43     269      2   0      0     0      0     0    0         1
## 2:     118     849      6   1     17     0      0     0    0         0
## 3:      64     273      3   0      0     0      0     0    0         1
## 4:      33      27      0   0      0     0      0     0    0         1
## 5:      26      61      0   0      0     0      0     0    0         1
## 6:      24     -24      0   0      0     0      0     0    0         1
##    probowl_prob
## 1:        0.977
## 2:        0.825
## 3:        0.752
## 4:        0.748
## 5:        0.468
## 6:        0.447
```

Russell Wilson, clearly helped out by his stellar rushing stats on the year, was given a 82.5% chance to be selected. The model correctly predicted the other five Pro Bowl quarterbacks, but thought that Tom Brady should've been swapped for Wilson this year.

Limitations

Unfortunately, this model comes with a few notable drawbacks that are hard to ignore.

1. The model does a very poor job of predicting Pro Bowl berths for defensive backs. Because interceptions are the only true statistic that the model has at its disposal, it essentially makes its decision of whether a cornerback or safety should be in the Pro Bowl based off of that. Another variable like pass breakups, QBR allowed, or completion percentage when targeted would most likely be beneficial.

2. Conferences are not listed as variables. Excluding 2012-2015 (when players were drafted by NFL legends), the Pro Bowl teams are always separated by AFC and NFC. Just because the top six players in a certain position have the highest probability to be selected doesn't mean there will be room for them. Because the same number of players from the same position are chosen from each conference, some deserving players might be left out.

3. The probabilities for each player making the Pro Bowl are calculated with every position in mind. For more accurate probabilities, running separate regressions for each individual position would be more beneficial.

Conclusion

Overall, the model predicted the Pro Bowl chances for some positions better than others. A position like wide receiver can have their impact on the game fit into a box score while defensive backs influence the game in ways that can't be easily as quantified. I look forward to refining this model to potentially get more accurate predictions in the future. But for the time being, social media will continue to be a battleground for debates about certain players selections into the Pro Bowl.

The code for this project can be found on my GitHub, mkraus24, in the Pro Bowl Regression repository.