

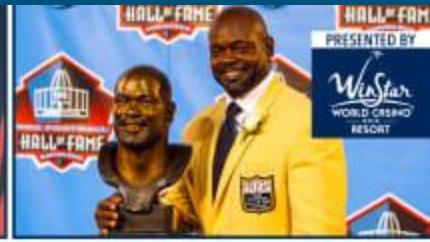
NFL Hall of Fame Predictions



John Harrigan & Evan Staffen

Hall of Fame Selection

- Eligible for nomination 5 years after retirement
- Combination of players, coaches and front office



Business Problem

Stakeholder: Agents of recently retired NFL players.

Hall of fame NFL players have marketability long after their playing days and well beyond the sphere of football.

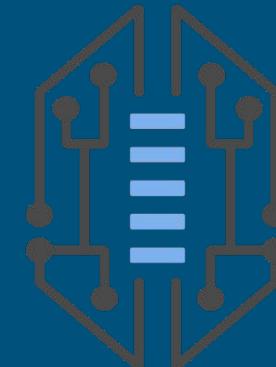
We want to create a model that can predict these Hall of Fame players based on their career statistics. This could be used by agents looking to maximize the value of the players they have signed.



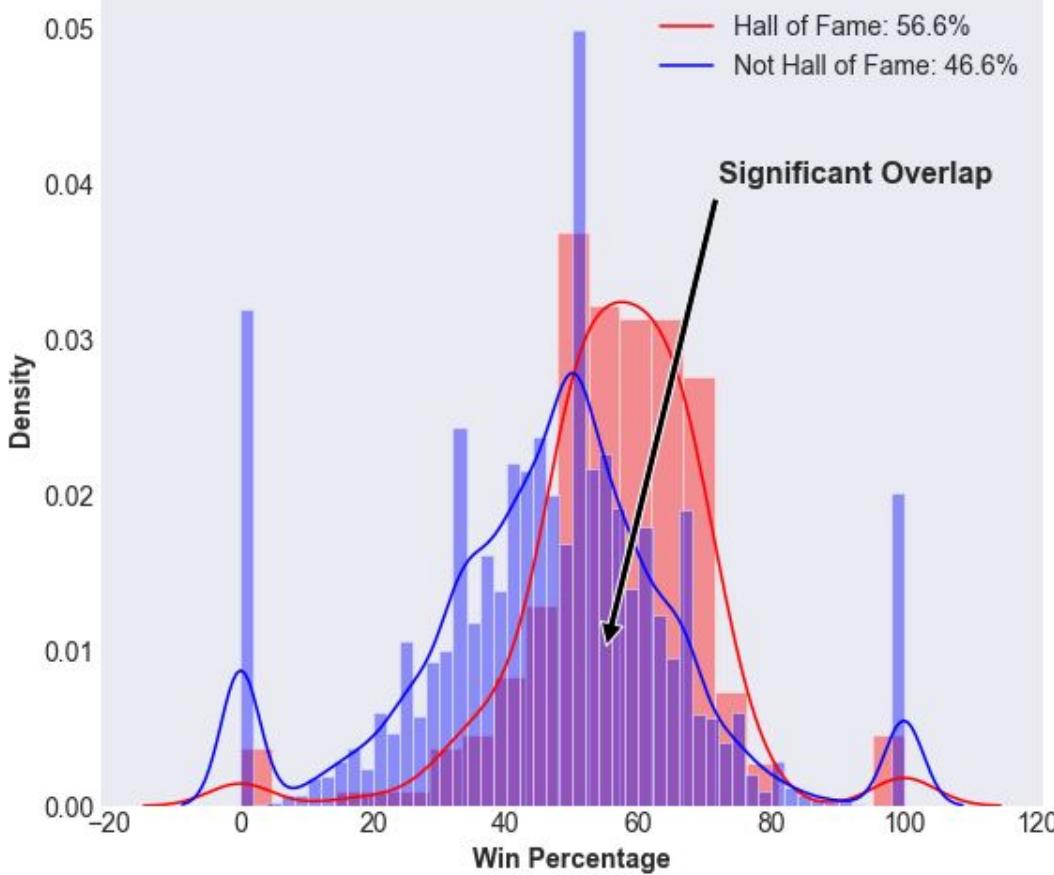
Datasets

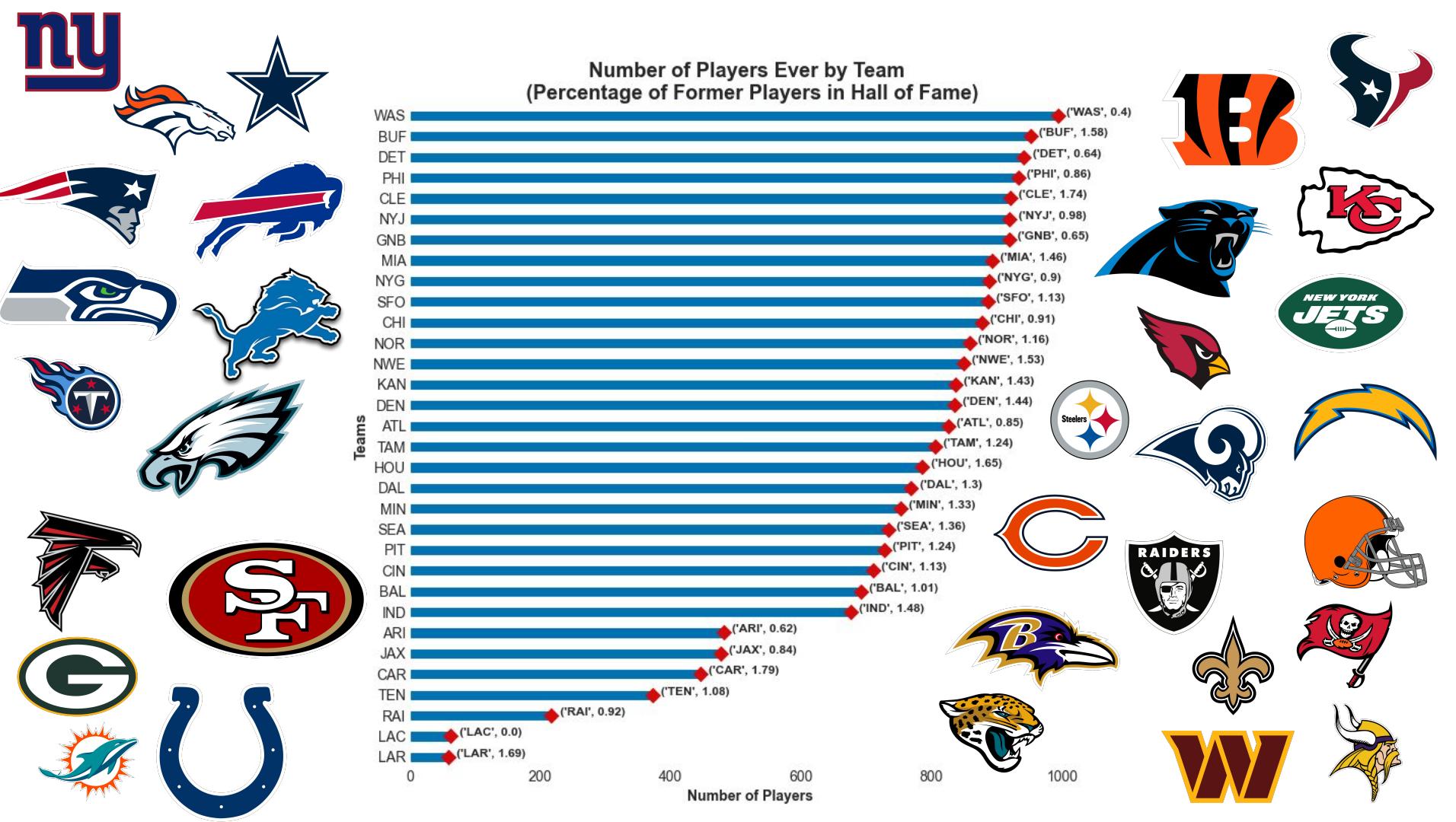


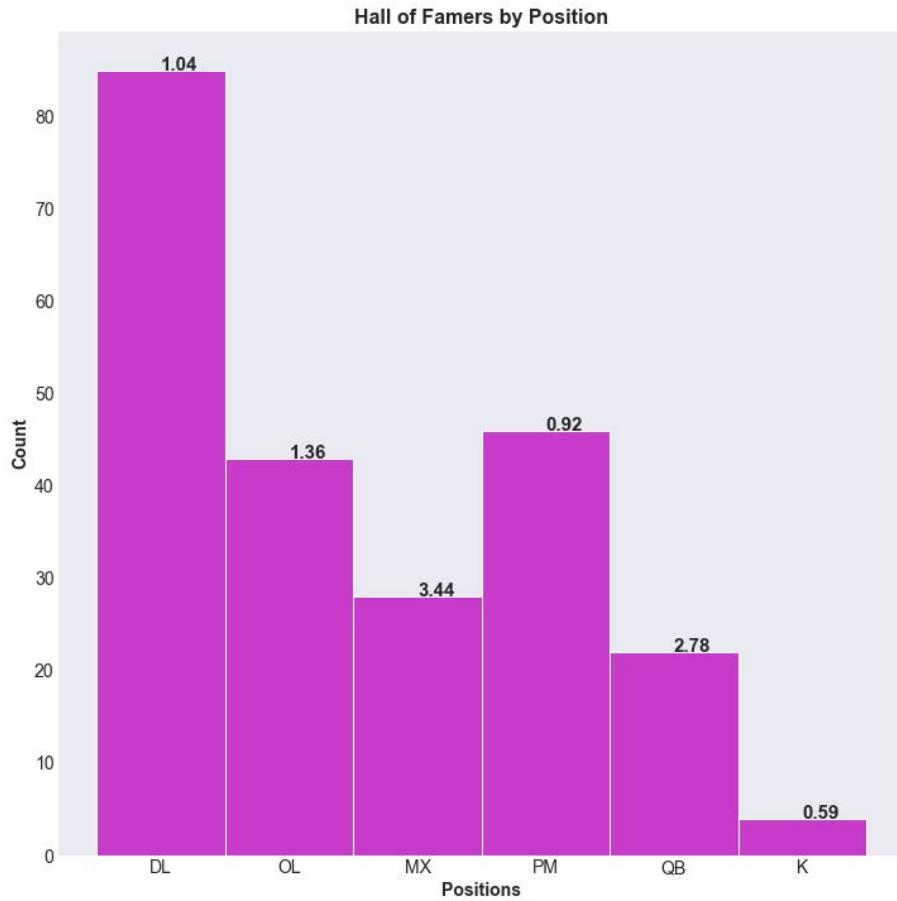
- Kaggle
 - 25K+ players
 - Over 1 million individual players game data
 - 1950-2017
- Profootballreference
 - Used to calculate awards won throughout players' career
- Gridiron AI
 - Updated database of all players past and present
 - Scraped the website to obtain the last 5 years of data



Win Percentage vs Hall of Fame Status

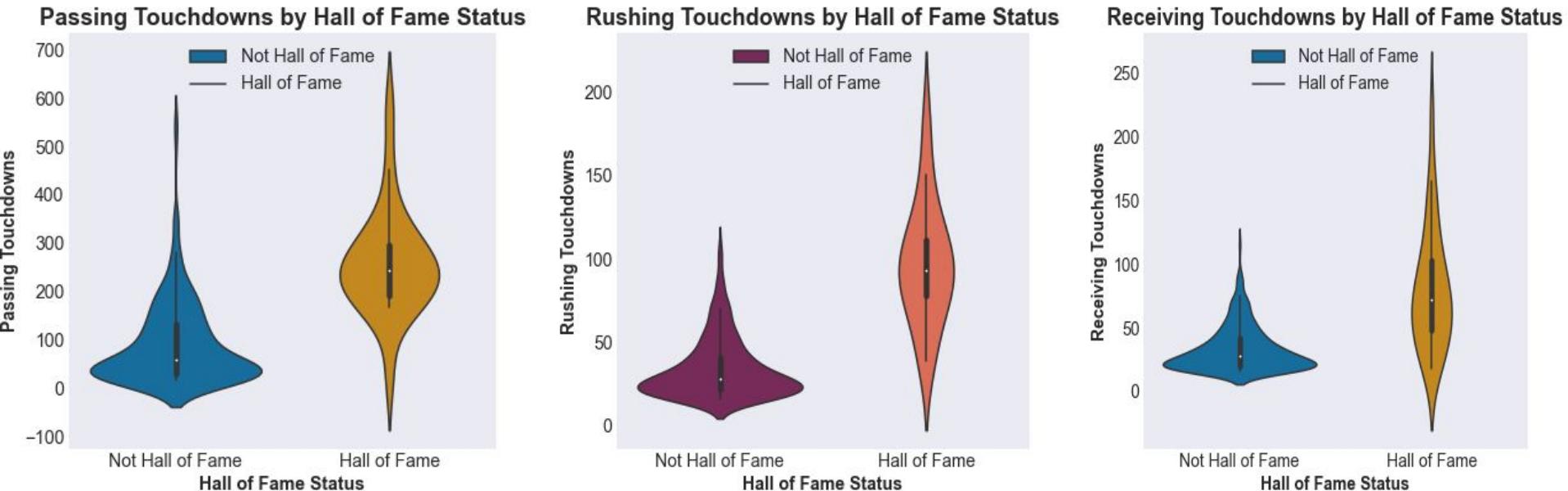






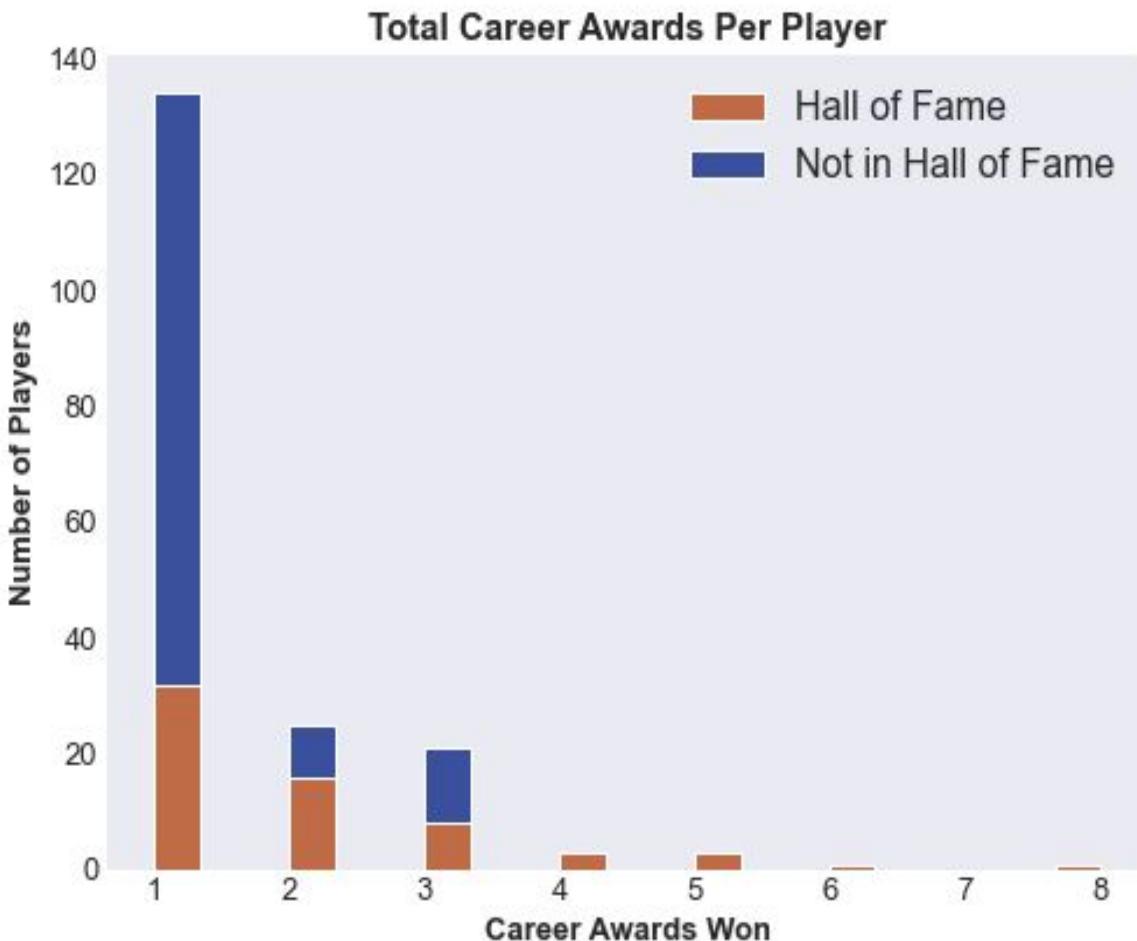
Positional Categories

- DL = Defensive Linemen
- OL = Offensive Linemen
- MX = Mixed Positions (2+)
- PM = Running backs, wide receivers and tight ends
- QB = Quarterbacks
- K = Kickers and Punters



Awards Used

- Most Valuable Player (MVP)
- Super bowl MVP
- Offensive Player of the Year
- Offensive Rookie of the Year
- Defensive Player of the Year
- Defensive Rookie of the Year



Modeling

1. Focus on improving false negative rate
 - a. **False negative** = players who are predicted to make the Hall of Fame when they do not
 - b. **False positive** = players who are not predicted to make the Hall of Fame when they do
2. 7 Models Explored
 - a. Basic Logistic Regression
 - b. Logistic Regression with SMOTE
 - c. Hyperparameter-tuned Logistic Regression
 - d. Random Forest Classifier
 - e. XGBoost Classifier
 - f. SVC Classifier
 - g. KNN (K-Nearest Neighbors)

Best Model

- Parameters

- C: 0.1
- Gamma: scale
- Kernel: rbf

Metrics	Training	Testing
Accuracy	0.927	0.918
Precision	0.138	0.104
Recall	0.963	0.780
F1 Score	0.241	0.184

Top 10 Coefficients

Coefficients	Column
0.827881	receiving_receptions
0.830011	passing_rating
0.831294	defense_tackle_assists
0.893769	passing_sacks_yards_lost
1.105871	passing_completions
1.116944	kick_return_yards
1.174337	kick_return_attempts
1.191905	position_PM
1.724438	weight
2.121725	draft_year

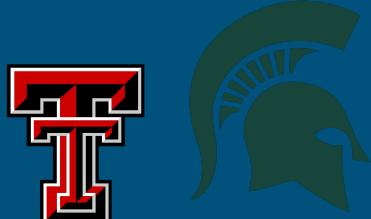
Conclusions

- Agents should be focusing on players in the realms of these average statistics based on their positions

Position	Yards	Touchdowns	Completions	Games	Win %	Awards	Superbowl wins
QB	37763	258	5090	194	59.5	1.8	1.18
PM	6284	44.5	Nan	178	56.3	0.89	0.86
Position	Sacks	Tackles	Tackle Assists	Interceptions	Win %	Awards	Superbowl Wins
DL	28.4	26.7	7.8	22.9	59.3	0.42	0.91
Position	Point After Attempts	Point After Makes	Field Goal Makes	Win%	Awards	Superbowl Wins	
K	443	453	375	57.7	0	1	

Next Steps

- Incorporate NCAA Football data
- Data on player characteristics/popularity
- Different metrics and models by positions
- Incorporate GridironAI more scraped data



Thank You!

Evan's contacts:

- Email: evan.staffen@gmail.com
- Linkedin: <https://www.linkedin.com/in/evan-staffen-a74045207/>
- Github: <https://github.com/evanstaffen/NFL-Hall-of-Fame-Predictions>

Jack's Contacts

- Email: ooharrigan@gmail.com
- Linkedin: <https://www.linkedin.com/in/john-laramee-harrigan/>
- Github: <https://github.com/jlharrigan/NFL-hall-of-fame>



Questions?
