

Reading Reddit

When Football Meets
Fantasy

Chad Richter
DSI 907

Problem Statement

Can we use natural language processing to determine whether fans speak differently about real football vs fantasy football on **Reddit**? If so, can we use this data to make more informed decisions about which **new feature** to prioritize adding to our fantasy football platform?

A Challenging Problem

- 'Team'
- 'Points'
- Player / Team Names
- Win / Loss



<https://www.wkbn.com/wp-content/uploads/sites/48/2021/06/football-and-football-field-2.jpg?w=1280>

The Data

- **1,929** Reddit post titles
- r/NFL and r/fantasyfootball
- **10/01/2021** through **10/10/2021**
- **993** fantasy football posts
- **936** NFL posts



<https://operations.nfl.com/media/1553/ap420659576066.jpg?mode=max&width=995>

The Subreddits – Usage Statistics

	r/fantasyfootball	r/NFL
Subscribers:	755,000	2,000,000
Posts Per Day:	230	410
Comments Per Day:	15,000	58,000
Average Title Word Count	12.4	19.2
Subreddit Sharing Most Users	r/NFL	r/NBA

Data Cleaning

- Punctuation stripped
- Common **stop words** eliminated (The, I, you, etc..)
- Text **lemmatized** (ex: passed, passes, and passing all become 'pass')
- **'Fantasy', 'NFL', and 'Morning'** removed



https://www.q4industries.com/wp-content/uploads/2014/06/DSC_0017_0105_crop.jpg

Modeling - Logistic Regression

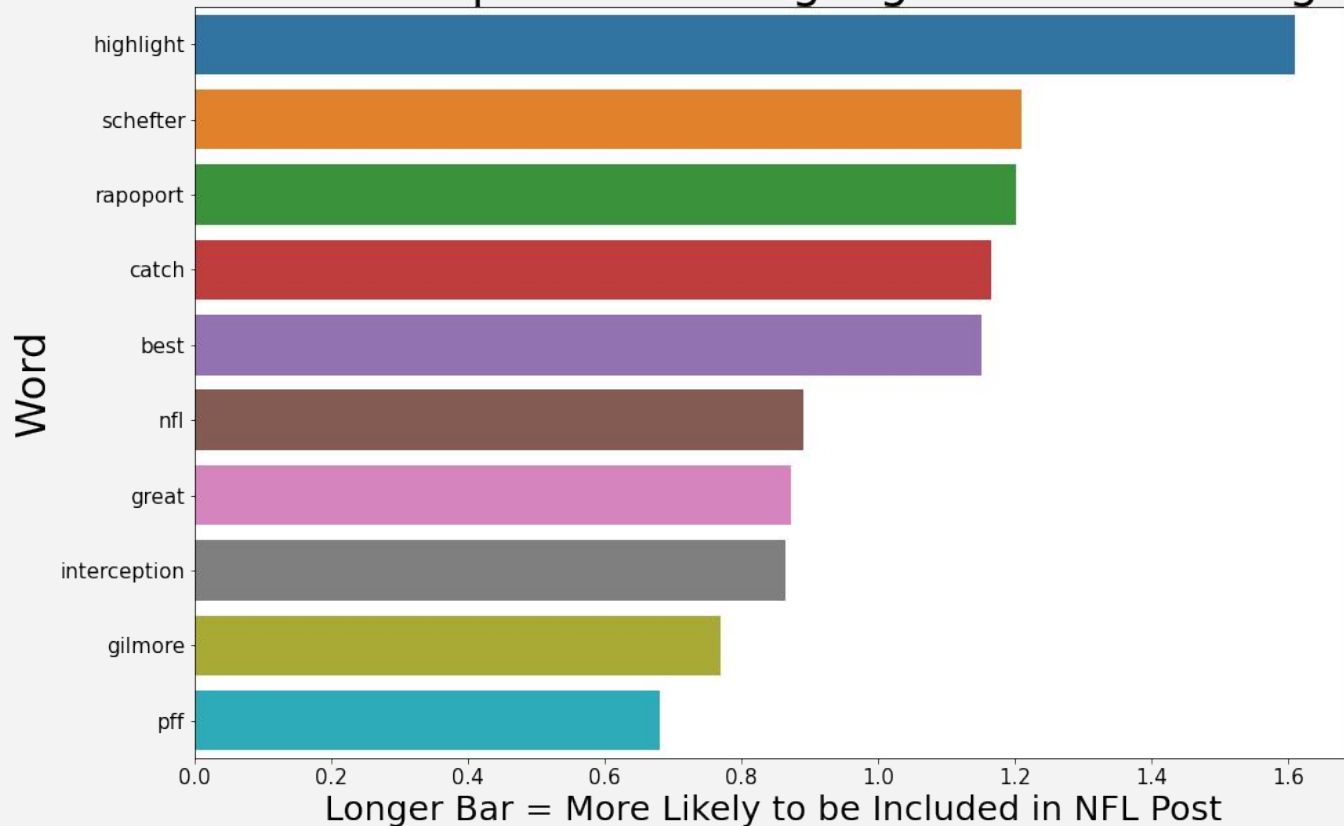


<https://www.profootballhof.com/assets/1/6/FGPost600TW.jpg>

- 77.6% accuracy vs 51.5% baseline
- CountVectorizer
- Less accurate, but more easily interpretable
- 500 max features
- Analyzed combinations of up to three words

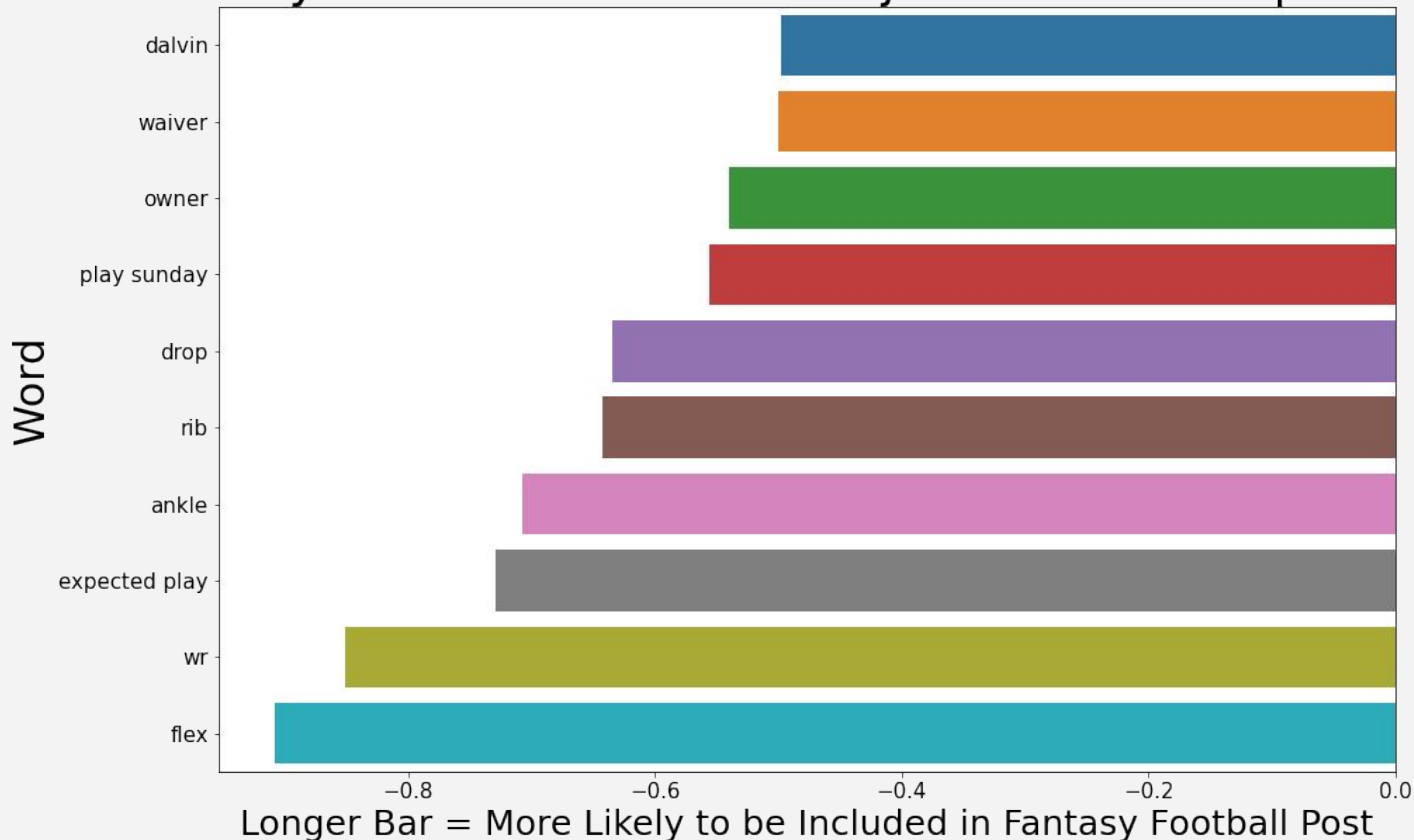
Interpretation - Logistic Regression

NFL Fans Love Spectacular Highlights and Breaking News



Interpretation - Logistic Regression

Fantasy Football is All About Injuries and Lineup Advice



Modeling - Gradient Boosting

- 80.12% accuracy vs 51.5% baseline
- CountVectorizer
- Most accurate model, but less easily interpretable
- 500 max features
- Analyzed combinations of up to two words



https://static.clubs.nfl.com/image/private/t_editorial_landscape_12_desktop/giants/oxyayee7khxkrtn3iko

Interpretation - Gradient Boosting

Title of Post	Prediction	Probability	Correct?
Bengals' RB Joe Mixon was diagnosed with a low-grade ankle sprain that will leave him week to week, per source.	Fantasy	95%	Yes
Bears RB Damien Williams suffered a thigh bruise in Week 4 but "should be OK," a source confirmed. With David Montgomery out 4-5 weeks, here's a look at the Bears' running back depth chart	Fantasy	91%	Yes
If Tom Brady, and the Bucs defeat the New England Patriots tomorrow, Brady will become the fourth Quarterback of all time to get a Quarterback Win against every current NFL Team	NFL	99%	Yes
Is Urban Meyer the most disastrous Head Coach hire in NFL history?	NFL	97%	Yes

Interpretation - Gradient Boosting

Title of Post	Prediction	Probability	Correct?
Can we talk about how disappointing some of the early round picks have been so far?	Fantasy	50%	Yes
How many of Mahomes' TDs are actually shovel passes?	Fantasy	58%	No
[Yates] Ravens RB Ty'Son Williams is inactive today against the Broncos	Fantasy	83%	No
This years NFL sleepers	NFL	89%	No
Sam Darnold is the only player in NFL history to have 5 rushing and 5 passing TDs after the first 4 games of the season.	NFL	84%	No

Conclusion – Injuries Are the Key!

Fantasy players consistently turn to Reddit for information on player **status updates** and **injury reports**

- **Real-time** injury updates and news
- Backup **suggestions** for sidelined players



<https://i.insider.com/5a0620203dbef484008b56eb?width=1000&format=jpeg&auto=webp>

Thank you

The floor is open for Q&A

