

Hockey Equipment Marketing Strategy Using NLP and Subreddit Classification

by Charles Ramey, Hockey For Everyone

Background

- Hockey For Everyone
 - Affordable
 - High-quality
- Reddit
 - r/hockey
 - r/hockeyplayers
- Natural Language Processing
- Machine Learning



Problem Statement

Which subreddit should HFE target for its initial marketing campaign, and can we design a classification model to help HFE's marketing team draft advertisements with appropriately targeted language?



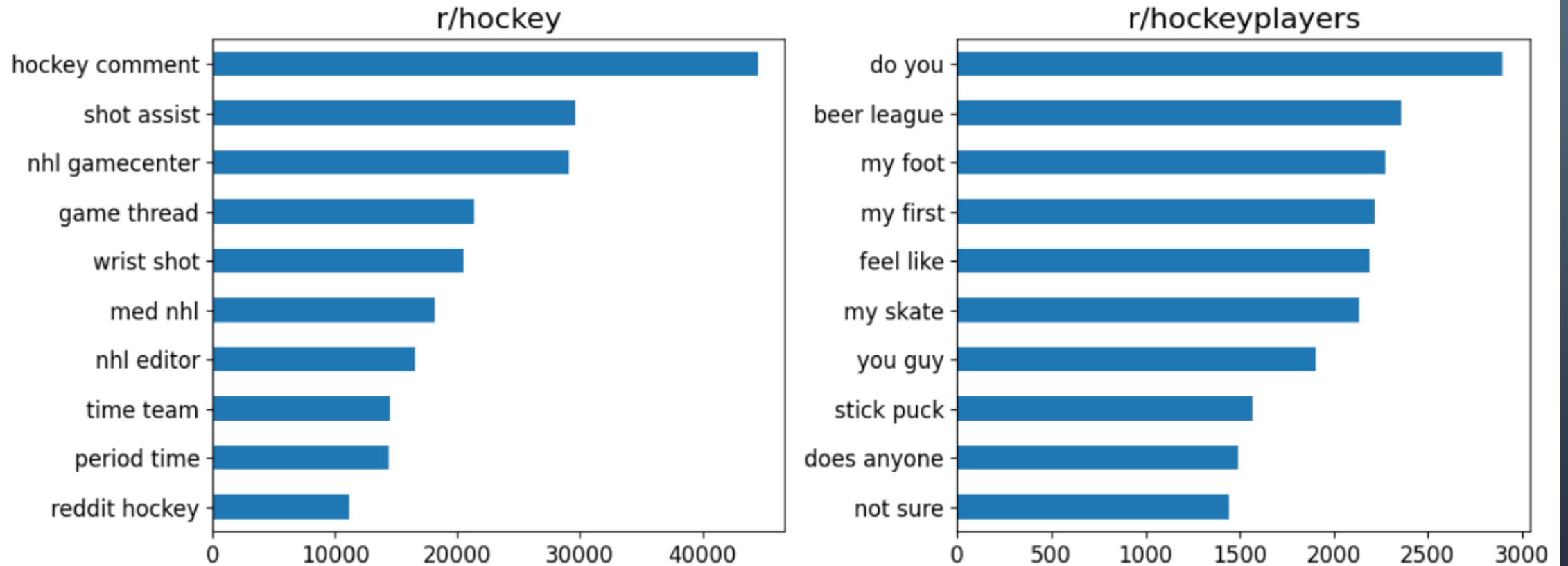
Data Collection

- Pushshift API
- 200,000 total submissions
 - r/hockey 150,000
 - r/hockeyplayers 50,000
- 63,000 total submissions after cleaning
 - r/hockey 32,000
 - r/hockeyplayers 31,000
- Text = Title + Body

	subreddit	Text	text
0	hockeyplayers	Warrior Goalie Pants	Has anyone tried the new ...
1	hockeyplayers	G Title	ny tips for a playe Body n net fo...
2	hockeyplayers	Huge Hockey Facility up for Auction	Lots and I...
4	hockeyplayers	Skates for kids with wide feet	Hope I'm in the...
6	hockeyplayers	Gear for sled hockey	Posting here for a friend...

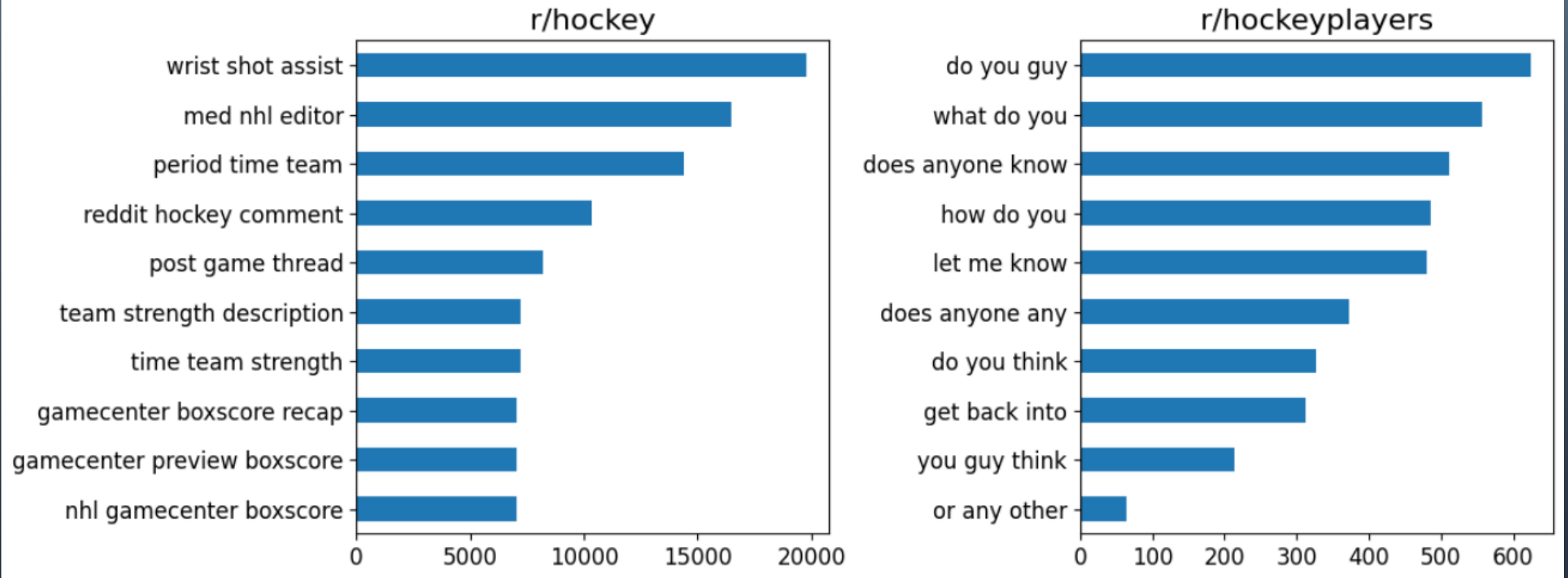
Common Language

Bigram Frequencies

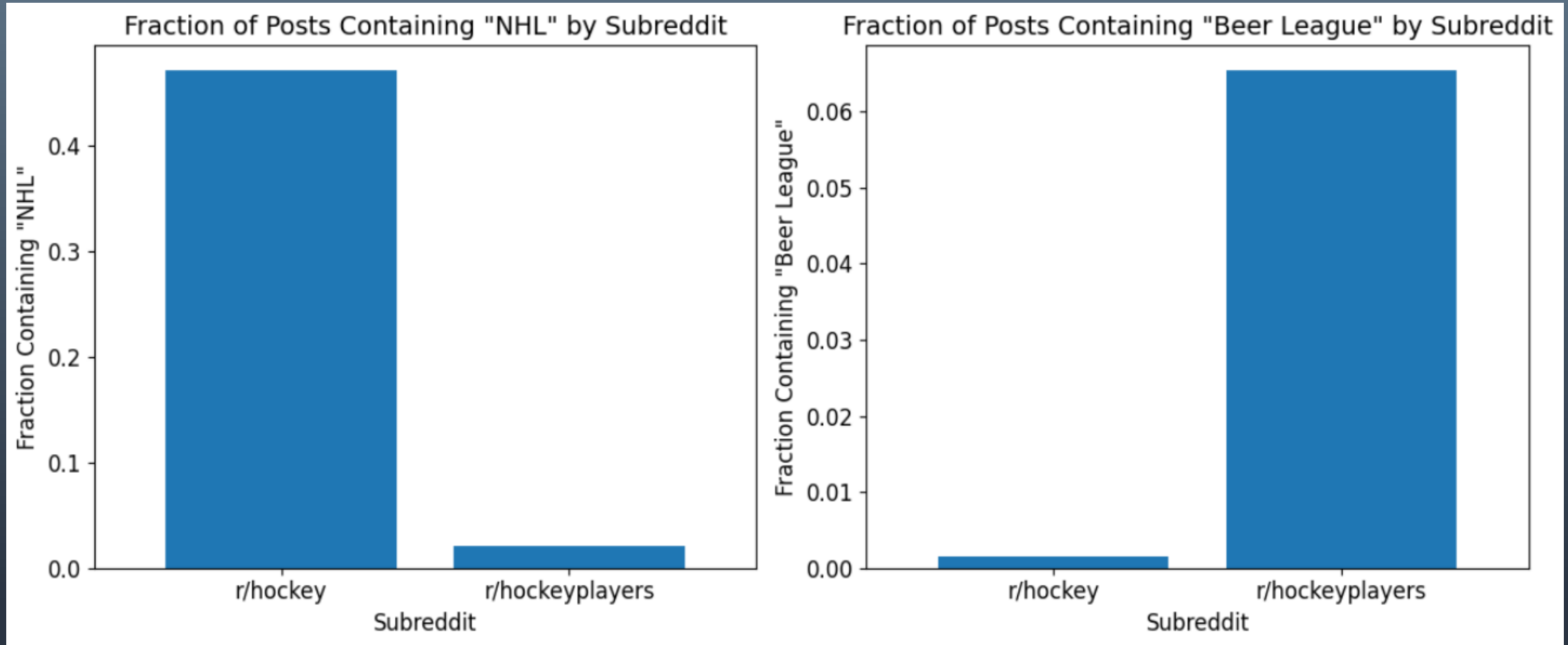


Common Language

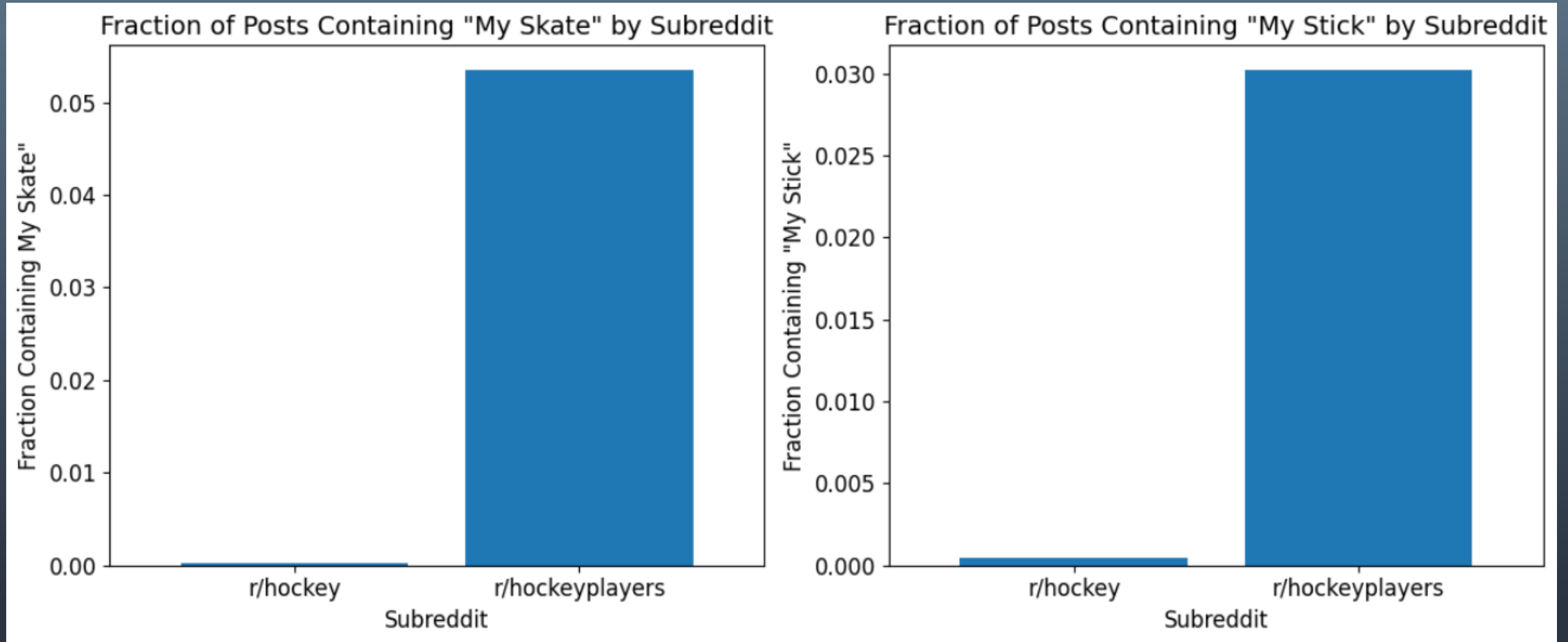
Trigram Frequencies



NHL vs. Beer League



My “Equipment”



Classification Modeling

- Tested 7 different classification methods
- Stacked all models for optimization

Model	Training Accuracy	Test Accuracy
Logistic Regression	0.96	0.95
Adaptive Boost Classifier	0.94	0.94
Stacked Logistic Regression	0.95	0.95

Conclusions

- Target r/hockeyplayers
- Successful modeling

Recommendations

- r/hockeygoalies
- Youth hockey
- Other platforms and forums
 - Facebook, Twitter



Questions?