Bumble vs. Tinder

A NLP analysis of Reddit.com posts

Carly Chamberlain Sharma
sharma.carly@gmail.com

Warning: This presentation includes data that contains profane language.

Problem

- Single people who want a serious relationship have limited time/resources
- According to a 2019 study 39% of couples meet online*
- Hundreds of apps and dating sites, which to choose?

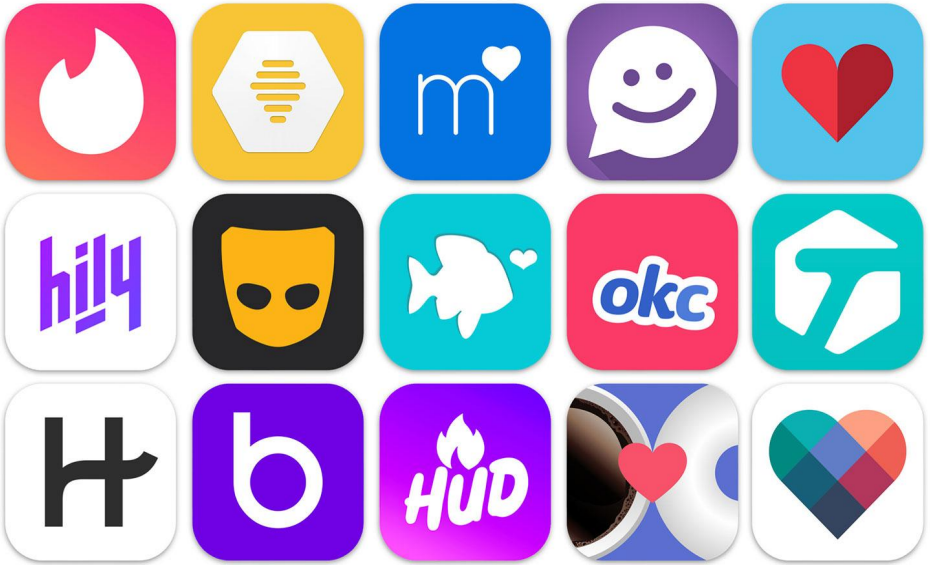


Image credit: 2019 article

* Only heterosexual couples included in study

Data Gathering

- Friend recommendations are best but hard to get at scale
- Access Reddit communities through API
- Compare top 2 dating app subreddits from Reddit.com (Bumble and Tinder)
- 1000 posts for each subreddit

Project

- Step 1: Prove a distinction between subreddits based on language in posts
- Step 2: Analyze language - does one community seem more serious than the other?
- Step 3: Ideal outcome - recommend one app over the other

Data Cleaning

- Pulled post text (‘self_text’) and title through Reddit API
- Created “combined” column of post text + title
- Created “text_word_count” for post text
- Lemmatized text

subreddit	selftext	title	combined	text_word_count	lem_text
0	Things that make me swipe left on you without ...	Things that will make me swipe left on you wit...	[Things, that, make, me, swipe, left, on, you,...	212	Things that make me swipe left on you without ...
0	If you’ve hit it off with a girl on Bumble, wo...	Men: Do you ever intentionally wait a couple d...	[If, you, ’ve, hit, it, off, with, a, girl, on...	60	If you ’ve hit it off with a girl on Bumble , ...
0	I haven’t gotten a single match since I downlo...	Would love some profile help!	[I, haven, ’t, gotten, a, single, match, since...	30	I haven ’t gotten a single match since I downl...

Model Preprocessing

- Countvectorized dataframe

- Constantly updated list of stop words and combined with English stop words

STOP WORDS =

- tinder
- bumble
- https
- reddit
- www
- just
- ve
- feel
- like
- com
- jfe
- ca
- mailto
- flava

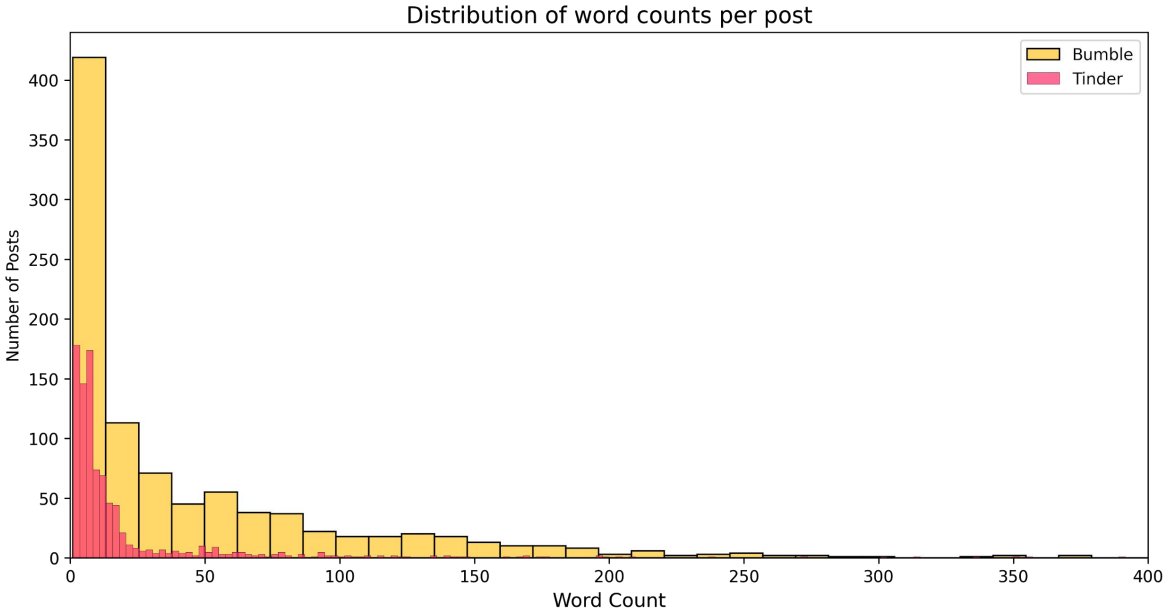
Model Scores (Score to beat! = 50%)

TRANSFORMER	MODEL	TRAINING SCORE	TESTING SCORE
CountVectorizer	Multinomial Naive Bayes	0.81	0.66
TF-IDF	Multinomial Naive Bayes	0.83	0.66
TF-IDF	KNN	0.5	0.5
CountVectorizer	SVC	0.85	0.72
CountVectorizer	**RandomForest**	**0.74**	**0.73**
TF-IDF	RandomForest	0.79	0.73
CountVectorizer	LogisticRegression	0.92	0.73
CountVectorizer*	LogisticRegression	0.93	0.67

Best Parameters: CountVectorizer + RandomForest

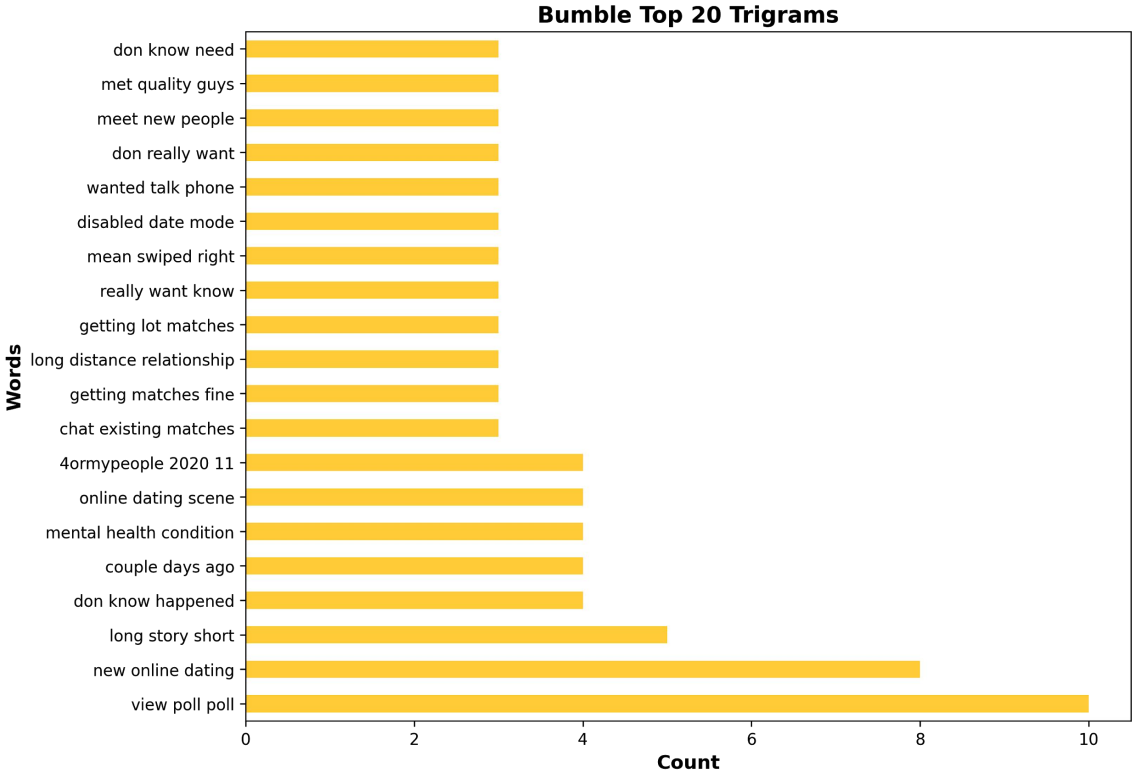
CV Param	CV Value	RF Param	RF Value
Max df	0.9	Max depth	5
Max features	3000	Max features	Auto
Min df	2	N estimators	75
Ngram range	(1, 1)		
Stop words	English		

Which subreddit is more “serious”?

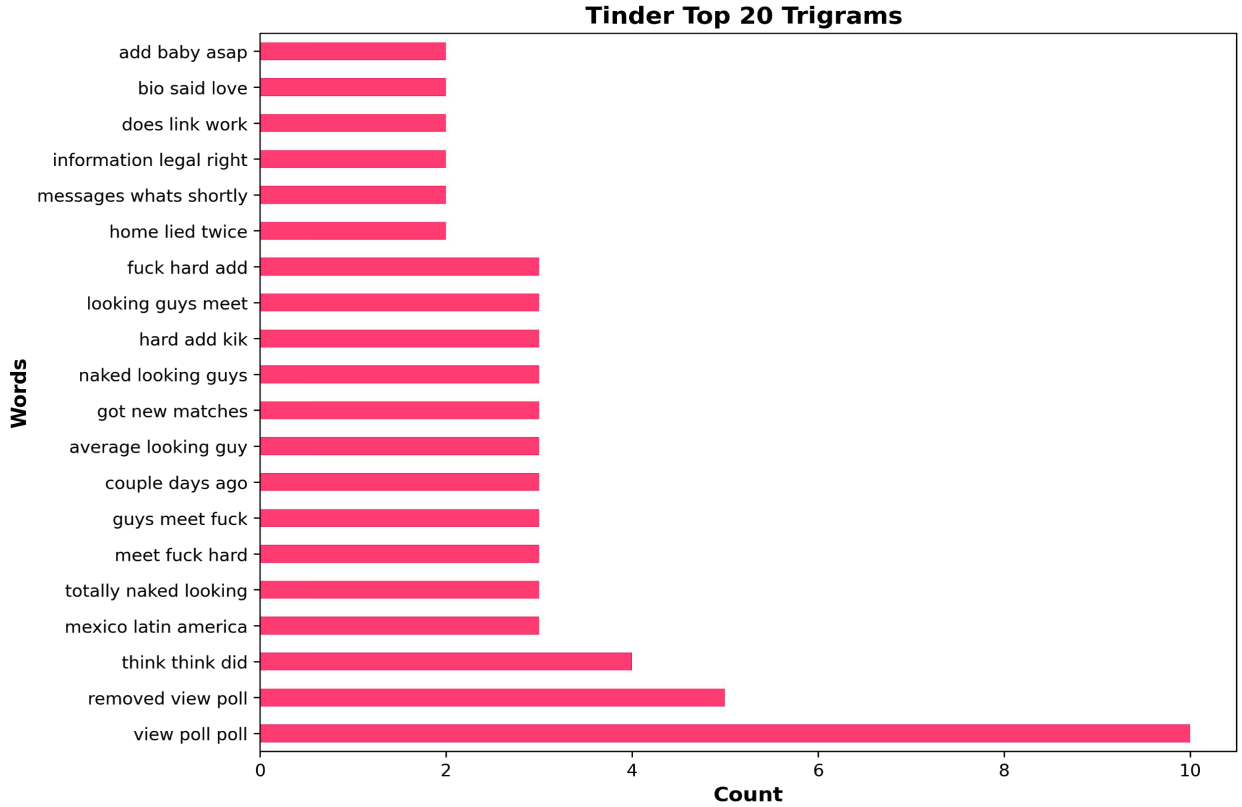


Avg. # of words per post-
Tinder: 22
Bumble: 52

Which subreddit is more “serious”?



Which subreddit is more “serious”?



Sensitivity/Specificity Scores

- Tinder : 87.1% of predictions were accurate (sensitivity).

- Bumble: 59.2% of predictions were accurate (specificity).

Conclusion: Swipe right on Bumble

1. Yes, there is a discernible difference between the two subreddits. Our model is able to predict with 73% accuracy.
2. I'd argue common patterns in tinder posts suggest less seriousness.

Further exploration

- Analyze images
- Differentiate between the polls that are posted
- Dig deeper into words/phrases that imply a long-term match like “marry” or “going on 6 years…”
- Look at geography or demographics
- Review comments