Subreddit Classification

r/AmiTheAsshole vs r/relationships

r/AmiTheAsshole

- Subreddit for people to speak about their conflicts & get a second opinion.
- Inherently more confrontational
- Has voters use coded language

r/relationships

- Subreddit for people to speak about their interpersonal conflicts & get a second opinion.
- People generally have more to lose (since conflicts tend to be more personal)

Acronyms

What's with these acronyms? What do they mean?

- The acronym present in the top voted comment will trigger a bot to assign a "post flair" to each thread.

Acronym	Meaning
YTA	You're the Asshole
NTA	Not the A-hole
ESH	Everyone Sucks here
NAH	No A-holes here
INFO	Not Enough Info

-
- YTA or "You're the Asshole" is for scenarios where the OP is at fault in their situation.

- NTA or Not the Asshole is for scenarios where the OP is NOT to blame and the other party described in their scenario is to blame.

- ESH or Everyone Sucks here is for scenarios where both parties are to blame- both people involved in the scenario should be held responsible.

- NAH or No Assholes Here is for scenarios where neither party is to blame. All parties actions are justified. Nobody needs to be held accountable. Shit happens.

- INFO or Not Enough Info is for situations where the OP never clarifies details that would determine the true judgment.

Problem Statement

This project aims to predict posts belonging to r/AmiTheAsshole & r/relationships.

- Subsequently, pulling the features that drive the coefficients to understand how people use language when describing general conflict in comparison to interpersonal conflict.

Data

- Data for this project was retrieved from Reddit through looping requests of .json files.
- Dropping initial duplicates found that 23% of AITA data was duplicates, and 52% of relationships data was duplicates.

	title	selftext	subreddit
0	New Resources for Anyone Looking to Help Those...	[This recent vice article](https://www.vice.co...	AmItheAsshole
1	Monthly Open Forum November 2020	Welcome to the monthly open forum! This is the...	AmItheAsshole
2	AITA for moving out because my husband bought ...	For some background we are both 25 and live in...	AmItheAsshole
3	AITA For wanting to sue my fiancee's sister af...	My M28 fiancee's F29 sister recently moved in ...	AmItheAsshole
4	AITA for telling my brothers how it really fel...	I (25F) was adopted from China as a baby. My p...	AmItheAsshole
5	AITA for not telling my coworkers why I don't ...	I had a yearly review last week and my manager...	AmItheAsshole
6	AITA for canceling a date when I was told I ha...	This guy constantly asks me out on dates. "Can...	AmItheAsshole
7	AITA for kicking out my parents because their ...	Here's the situation...I (27M) have my mom and...	AmItheAsshole
8	AITA for denying my parents access to my room ...	So I know it sounds weird but let me explain.\...	AmItheAsshole
9	AITA for keeping my ex husband's last name aft...	I've been married once before; my ex and I wer...	AmItheAsshole

```
op.tail(10)
```

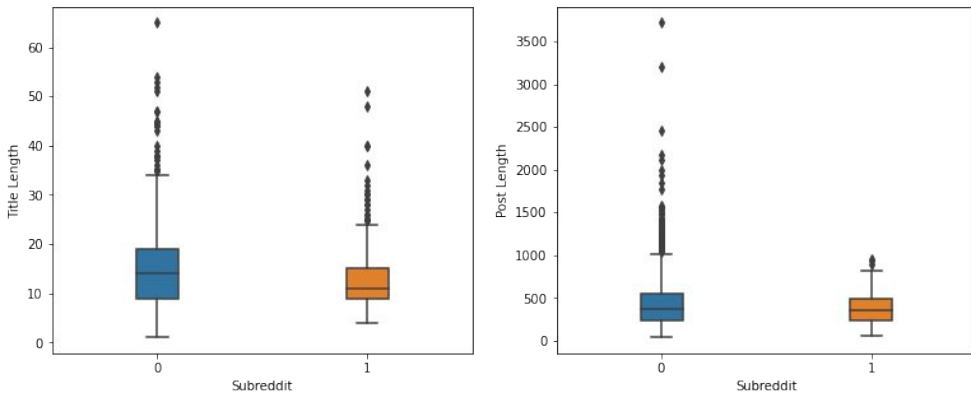|	title	selftext	subreddit
1982	I really care about this girl but she doesn't ...	So, this is about my ex gf/friend (we're both ...	relationships
1983	Anxiety from bf Snapchat score increasing	I (34f) have anxiety from past bad relationshi...	relationships
1984	I (18F) really don't think I'm good enough for...	Hi, first post here, sorry if I've done someth...	relationships
1985	I (23F) am scared to start a relationship with...	The guy that I am currently talking to and I h...	relationships
1986	How do I [29F] deal with my gift-giving anxiet...	First Christmas with my boyfriend of six month...	relationships
1987	I (20F) dont know if it's too early to settle ...	I've been with this guy for the past 3 years. ...	relationships
1988	My Roommate is Overstepping Boundaries	***UPDATE*** So I talked to my roommate and le...	relationships
1989	How do I (21F) let my friend (20F) know that I...	Basically what the title says. This girl (20F)...	relationships
1990	My [29M] girlfriend [25F] has given me an ulti...	Hello Reddit,\n\nMy girlfriend and I have been...	relationships
1991	Can somebody tell me if I am being paranoid or...	I've just had a big bust up with a woman I've ...	relationships

Word Cleaning

Stopwords

```
#Assigning stopwords
stop_words = set(stopwords.words('english'))
special_stops = {'aita', 'asshole', 'wibta', 'yta', 'nta', 'esh', 'nah', 'info', 'relationship', 'relationships',
                 'www', 'https', 'com' , 'reddit', 'husband', 'wife', 'girlfriend', 'boyfriend', 'bf' , 'gf', 'tl',
                 'dr', 'tldr'}
stop_words = stop_words.union(special_stops)
```

- Tokenized & Lemmatized
- Kept original data
- Selftext (Post Text) & Title

Word Count

- Similar medians between both subreddits
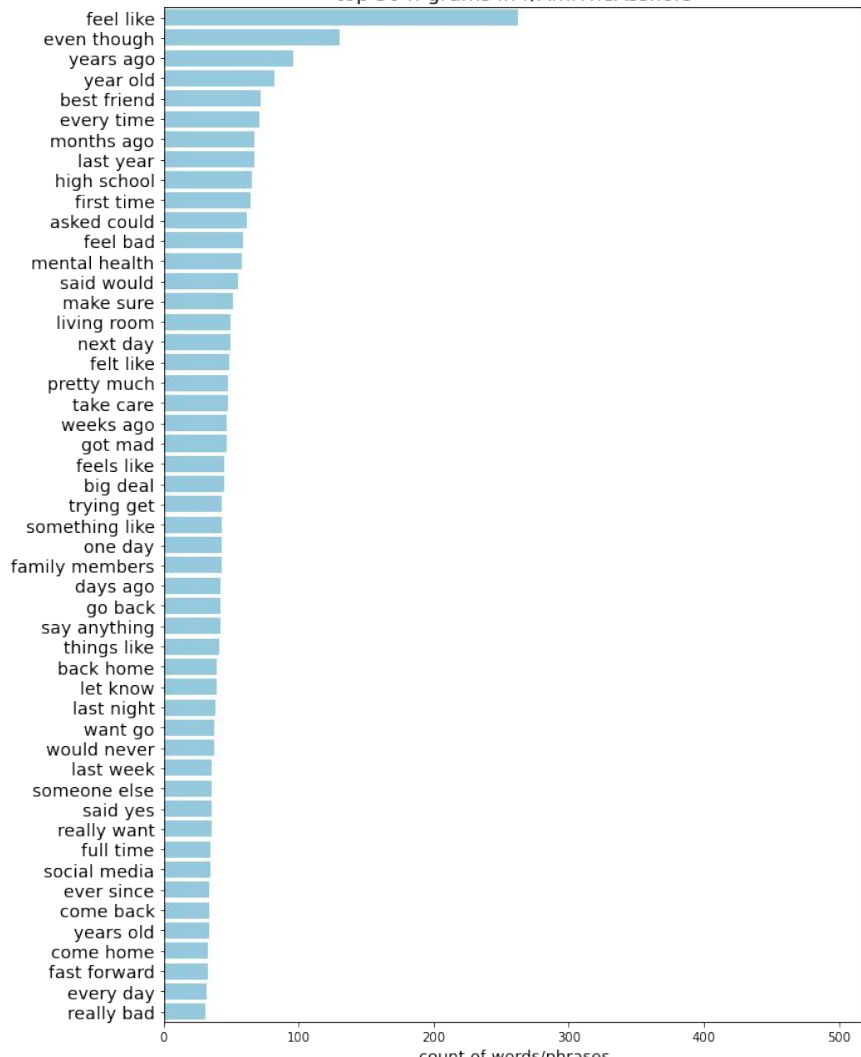- Where they differ is the large amount of outliers in r/relationships

r/relationships

	count	mean	std	min	25%	50%	75%	max
subreddit	936.0	0.000000	0.000000	0.0	0.0	0.0	0.0	0.0
title_len	936.0	14.928419	8.243371	1.0	9.0	14.0	19.0	65.0
selftext_len	936.0	457.076923	338.426650	39.0	245.0	367.0	557.0	3717.0

r/AmiTheAsshole

	count	mean	std	min	25%	50%	75%	max
subreddit	1056.0	1.000000	0.000000	1.0	1.0	1.0	1.00	1.0
title_len	1056.0	12.133523	5.091781	4.0	9.0	11.0	15.00	51.0
selftext_len	1056.0	365.425189	155.679267	63.0	242.0	353.0	494.25	956.0

| top 50 n-grams in r/AmiTheAsshole	top 50 n-grams in r/relationships
feel like	feel like
even though	best friend
years ago	even though
year old	felt like
best friend	years ago
every time	feels like
months ago	get back
last year	first time
high school	mental health
first time	months ago
asked could	social media
feel bad	high school
mental health	long distance
said would	every time
make sure	someone else
living room	back together
next day	makes feel
felt like	two years
pretty much	started dating
take care	would like
weeks ago	let know
got mad	let go
feels like	every day
big deal	last year
trying get	long time
something like	things like
one day	year old
family members	pretty much
days ago	really want
go back	best friends
say anything	need advice
things like	got back
back home	want know
let know	make feel
last night	go back
want go	anyone else
would never	one day
last week	long term
someone else	next day
said yes	really know
really want	weeks ago
full time	want get
social media	feel bad
ever since	something like
come back	together years
years old	fast forward
come home	need help
fast forward	two months
every day	last night
really bad	made feel

count of words/phrases — word/phrase count

EDA

- Relationships language geared toward relationships, featuring n-grams such as 'started dating', 'long distance'.
- AITA more confrontational with features like 'big deal', 'got mad', 'would never'.
- Caught a few more stop words and added them to the above
- There was a large overlap of n-grams in the top 50 features between both relationships and AITA
- Additionally, there was overlap between title text and post text for both subreddits
- Both texts were merged
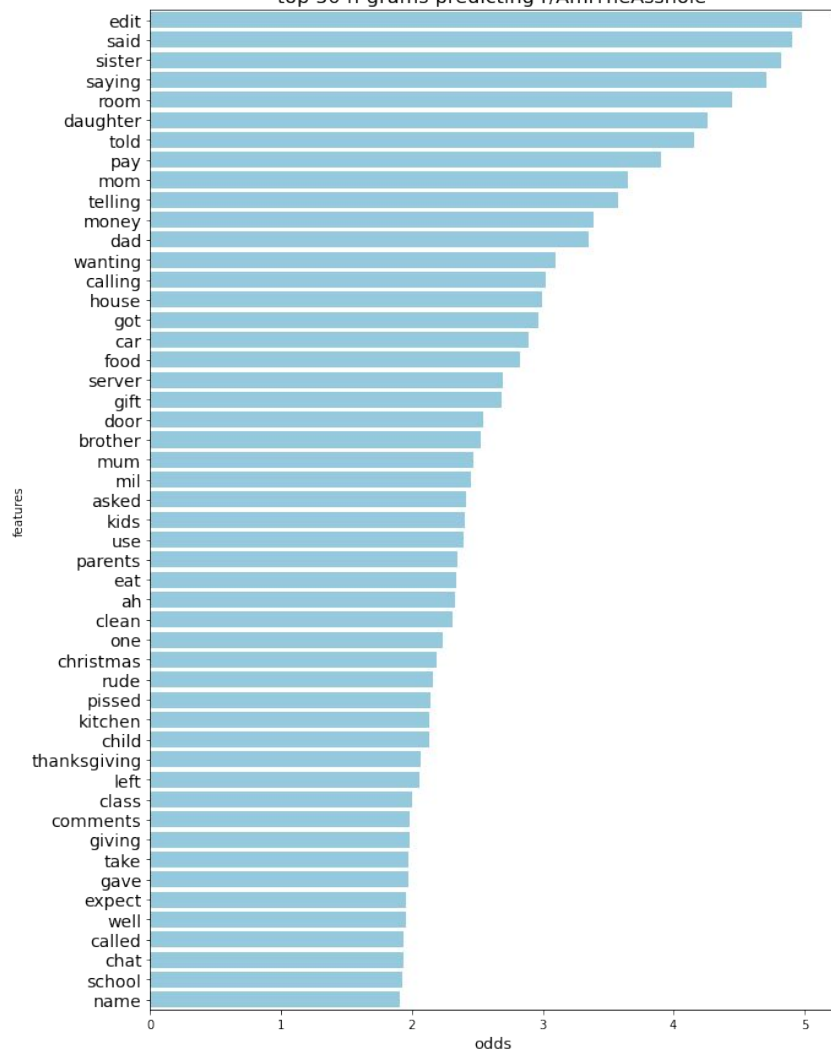

Overlap of AITA & Relationships

22   28   22

Relationships   AITA
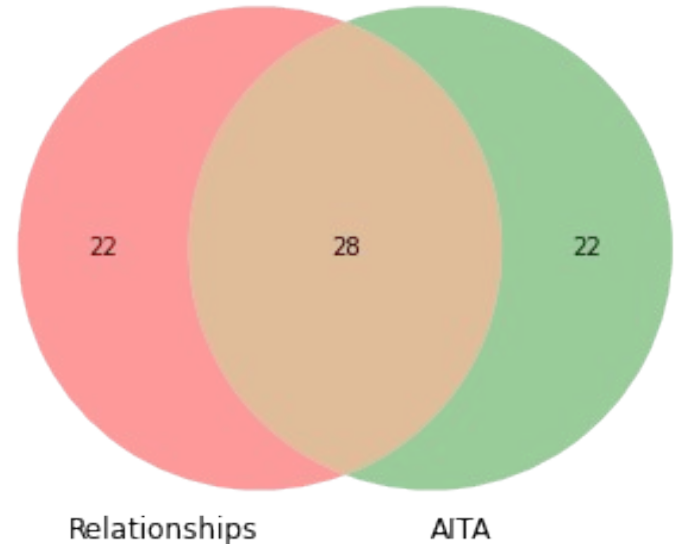
Model Building

- Rule of 2: Created a pipeline that input
    - 2 classifiers (TFIDF & Cvec)
    - 2 estimators(Naive Bayes & Logistic Regression)
- And input 2 sets of data.
    - Basic english stopwords & punctuation removal
    - Tokenized & Lemmatized

- Production Model Parameters:
    - Basic english stopwords & punctuation removal
    - TFIDF
    - Logistic Regression
        - 84.4%
        - Sensitivity: 86.25%(AITA)
        - Specificity: 82.52%(relationships)

ROC Curve with AUC = 0.914
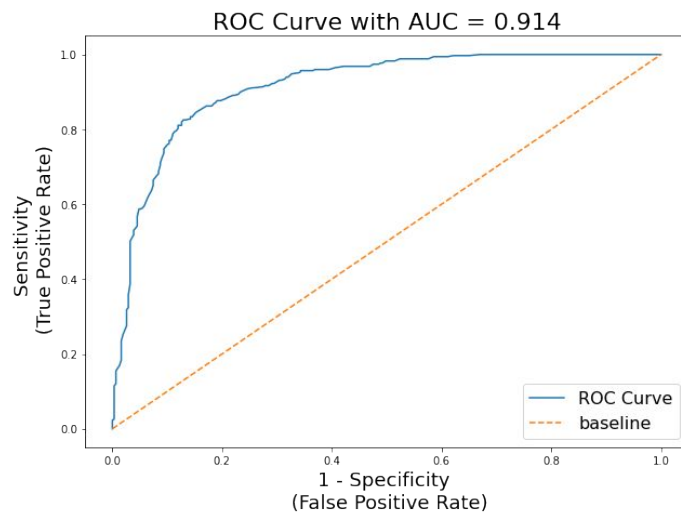
top 50 n-grams predicting r/AmiTheAsshole | top 50 n-grams predicting r/relationships
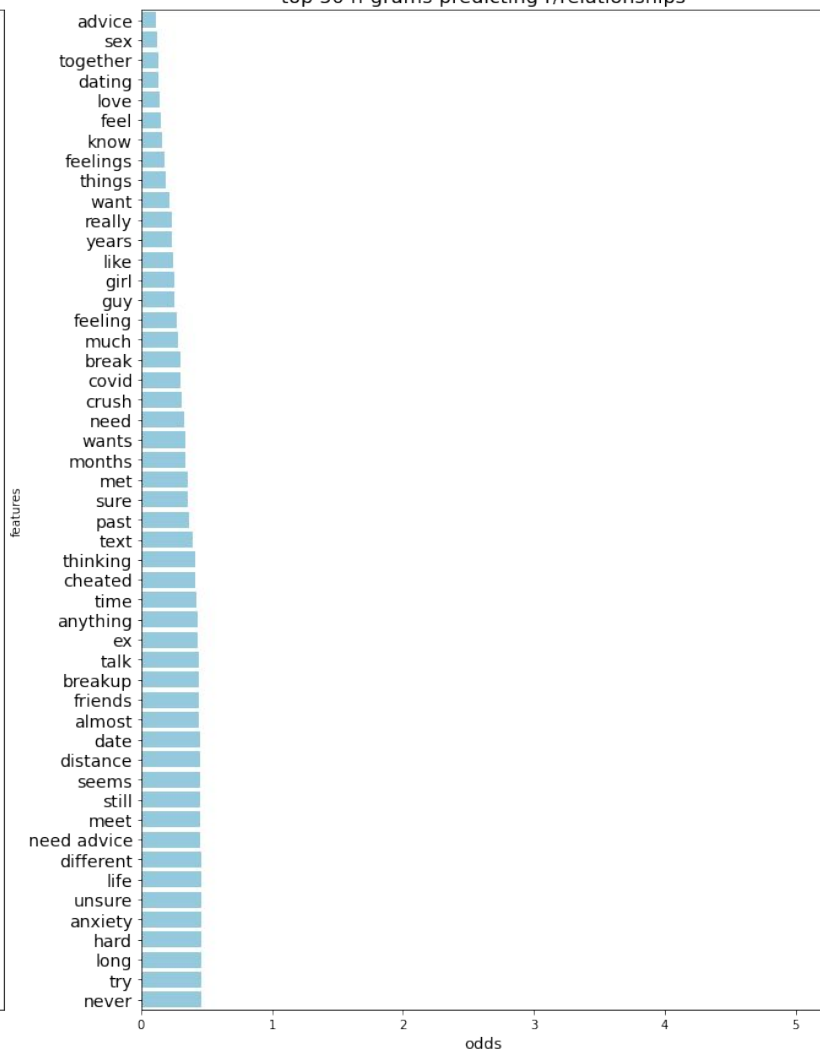
Final Observations

- While I set out to find the ways in which humans communicate.
- The machine picked the path of least resistance.
- E.g 'Edit' which r/AmiTheAsshole posters often use to add context to their conflict, or to justify their actions.

Thanks for listening!