
Natural Language Processing

Classification of text between subreddits belonging to the same category.

Problem Statement

We hypothesis that by having two similar text-driven subreddits, NLP will have a difficult time spotting the difference between the 2 subreddits.

What is Reddit?



Reddit is an American social news aggregation, web content rating, and discussion website. It is built with several online forums of different communities of special interest. Some of them could be image focused, videos or text.

Selected Subreddits

r/tifu (Today I Fucked Up)

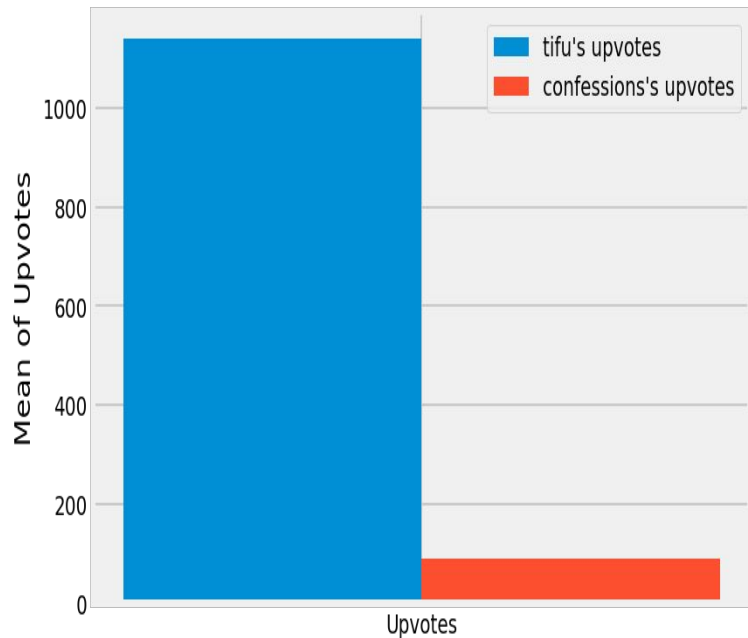
r/confessions

- Very text driven subreddits of personal recounts with very little image postings
 - Both have a very active community (r/tifu: 16.7m members, r/confessions: 786k members)
 - r/tifu tends to be more humorous while r/confession tends to be more serious
-

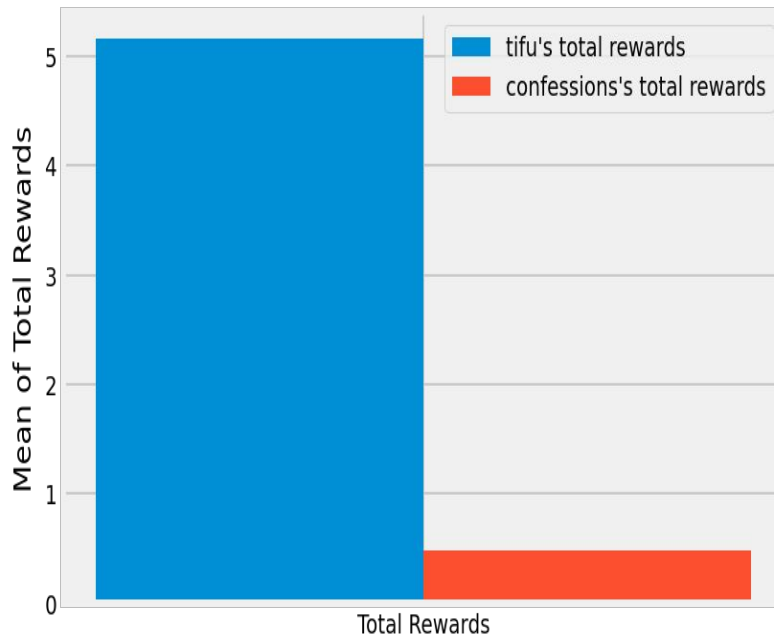
EDA & Model Selection

EDA

Tifu VS Confessions



Tifu VS Confessions



- 'Tifu' is more 5 times more popular than 'Confessions'

Baseline

997 posts from TIFU, 992 from Confessions

49.87%

Models that we used

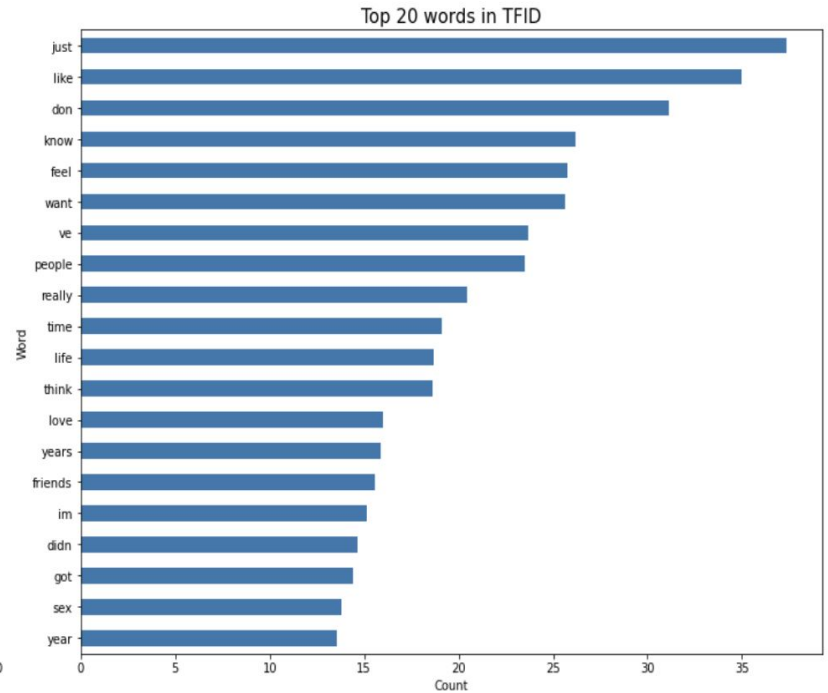
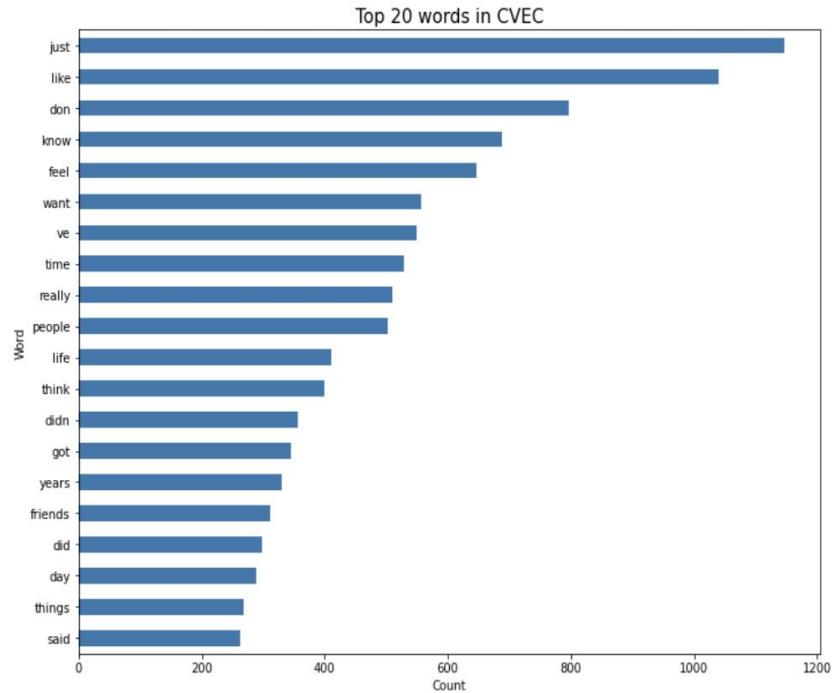
Text Feature Extraction:

- CVEC
- TFID

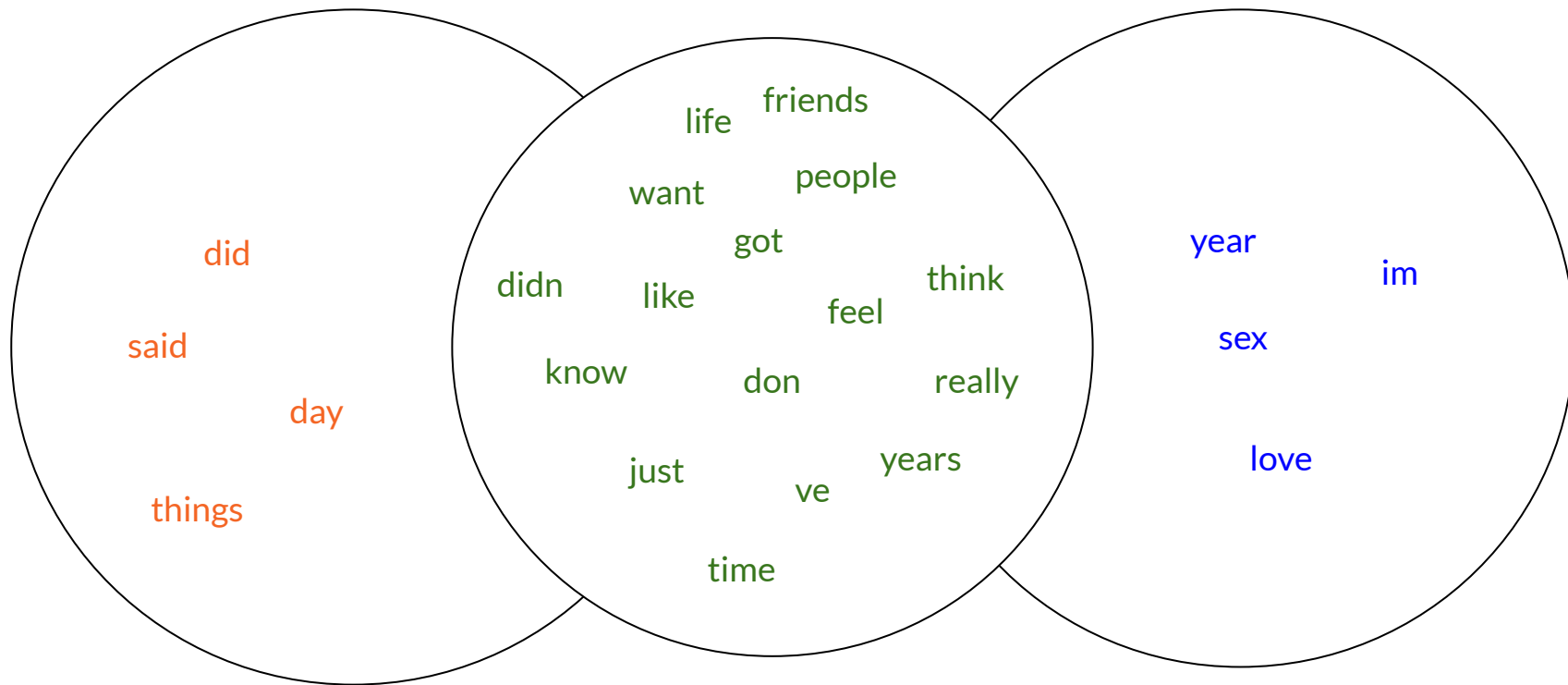
Classification Models:

- Naive Bayes
 - KNN (3, 5)
 - LogReg
-

Top 20 words appeared



Common words



Results and Recommendations

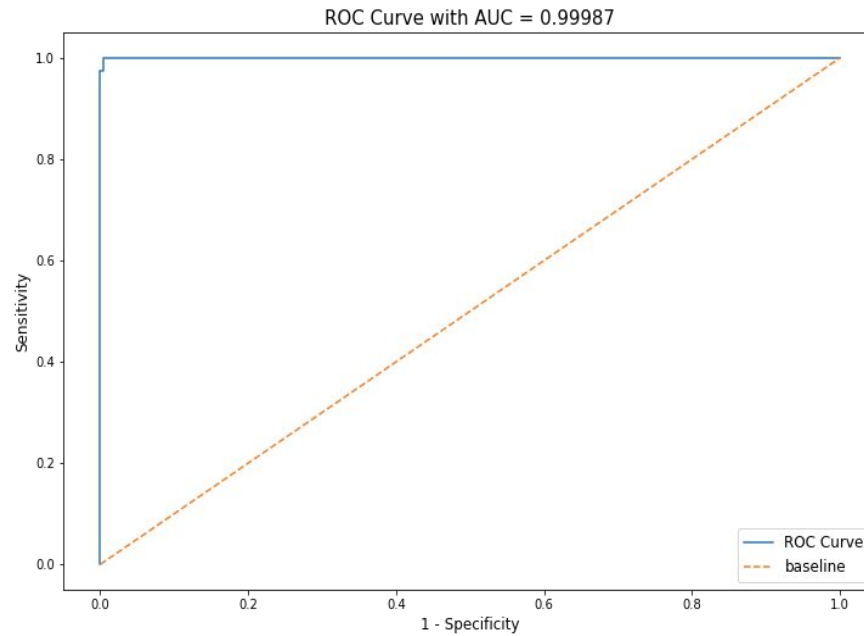
Accuracy

Using a combination of Count Vectorizer and Logistic Regression, model achieved a accuracy of **99.75%** on **397** points of unseen data.

Prediction Matrix

| Predictions / Actual Values | TIFU | Confessions |
|-----------------------------|------|-------------|
| TIFU | 198 | 0 |
| Confessions | 1 | 198 |

ROC AUC



Evaluating Success

Coefficients of Logistic Regression

| coef | word |
|----------|--------|
| 7.30307 | tifu |
| 1.84232 | tl |
| 0.576673 | dr |
| 0.556798 | decid |
| 0.541545 | went |
| 0.490661 | happen |
| 0.302628 | use |
| 0.236951 | work |
| 0.197744 | come |
| 0.17065 | ha |

| coef | word |
|------------|-------|
| -0.0957404 | year |
| -0.112575 | don |
| -0.120586 | know |
| -0.124124 | way |
| -0.155209 | told |
| -0.158582 | feel |
| -0.175814 | want |
| -0.191075 | veri |
| -0.207942 | thing |
| -0.330562 | life |

Posts Structure



r/tifu · Posted by u/cvosx 1 hour ago

TIFU by saucing my barista with hand san



This happened on Sunday morning, I'm only just getting over the eternal embarrassment, and to be honest, am baffled that something like this could happen to anyone let alone me.

I was at my local cafe, the one I've been going to all year. I should preface this by saying COVID hasn't really affected Australia, where I live, in the same way it has many other countries. Which means going to a cafe with friends, sitting down to eat and drink, and approaching the counter are all totally allowed and acceptable here.

So I'm sitting at this cafe with a few of my friends maybe 10m from the counter, and I get up and walk over to the barista to order my coffee and a sandwich. Once she's taken my order she gets to making the coffee right away just to the right of the cash register. I turn to walk away and out of the corner of my eye spotted the hand sanitiser right in front of the espresso machine. I'd already taken like 2 steps back to my table when it registered that "hey I should go and sanitize my hands."

So I turn 180 on the spot and stride back towards the counter. At this point my mates had obviously seen me walking back and the barista looked up as I started walking towards her. IE EVERYONE IS WATCHING ME DO THIS.

I look the barista in the eye and give her a smile while I pumped some hand sanitiser into my left hand. Turns out the little spout was facing her, presumably because she had used it moments before. At the risk of being too graphic, I don't know how else to phrase this, but I squirted it all over her chest.

She went bright red, I went bright red, and I think 3 of the people I was with peed their pants.

Safe to say I'm gonna need a new cafe.

TL;DR

Didn't look which direction the hand sanitiser was facing. Pumped it a bit too hard and sauced my barista.

r/confessions · Posted by u/PandaBunds 2 hours ago

I fart next to people who don't wear masks

I work at a community Rec center, where we run youth sports leagues. I don't know if it's comparable to retail, but people hate wearing their masks despite surging Covid numbers where I live (Utah). As part of my job, I have to baby sit some people, and by baby sit I mean telling them to put on their mask.

Usually I try to give them a bit in case they need to breathe for a sec, or get a drink of water. But if I ever have to tell someone twice to put on their mask, I always fart next to them while I'm talking. Always a silent fart, but in mind it's just a little "screw you"

False Positive

↑  r/confessions · Posted by u/cj_ruins 6 days ago

7

↓

I should be very, very dead.

This happened last week on new year's day into the second of January, so everything is still sort of sinking in and I get the passing thoughts that tell me "wow, that REALLY happened".

My friend and I were staying at my house and were already having a good time sober. This is where the "fun" begins. I hadn't seen him in a few months, so we were very excited about the hangout. We had promised to drink together (he never had, and thank God I limited him on how much he had, or things could have been much worse). I crack open the vodka. It is a 1.75 liter bottle of Popov vodka. There was JUST under half remaining. I poured his drink first and then mine. We sipped the first one a little but decided the real fun needed to wait until later when the least risk of being caught would be upon us.

Come 9:00, I start chugging the first one down. Thanks to me being a dumbass who doesn't properly measure drinks and just starts pouring them (I made mixed ones if it wasn't already obvious), I had actually put about THREE drinks worth into each mug of mixed lemon lime soda vodka mix. I have him one, and then another half, but I had about five and a half of these if my memory serves me right.

What happened after about 11:00 gets pretty hazy. However I do remember some details. I was having the best time. It started out in my room where I flopped everywhere around the floor. The moment my heart of hearts kicked in was when I couldn't feel my face or my head, even after falling and hitting it off of a solid wooden martial arts dummy made of mahogany. After this, we, as best we could went downstairs, and sat down for a few minutes before I, clouded by drunkenness, went into the liquor cabinet and snagged a Mickey of brandy, and chugged half of it. I HATE BRANDY. But I couldn't taste, and everything felt so good already that I just downed it. Thank goodness not all of it but still. That's when I decided it would be a splendid idea to go outside and run around in the freezing cold until the alcohol combined with motion sickness made me throw up. I normally freak the fuck out if I throw up (I have no clue why), but I just felt so good. I also tried to climb a tree, but that went to shit very quickly as I fell to the ground. Normally I would scale that fucker like it's a staircase but I just couldn't. There's this overwhelming, horrifying peace that sets over you as you lost control of your body. After we went inside, I chugged more brandy, and then passed THE FUCK OUT. I was sleeping on the bed maybe an hour...? If my memory serves me right, I passed out around 12:30 and was (thankfully) woken by my

Key Takeaways

- Accuracy relevant as metric as classes are well balanced
 - Despite posts being from the same category with little differentiation in terms of top words, model works well
 - Structure, as well as length of posts in one subreddit compared to the other allowed model to perform well
 - Applicable for use in classifying text where one class has a certain structure and key words compared to the other
-