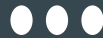


# Overcoming Opioids:

## Characterizing the Language of Reddit Users' Prior to Relapse

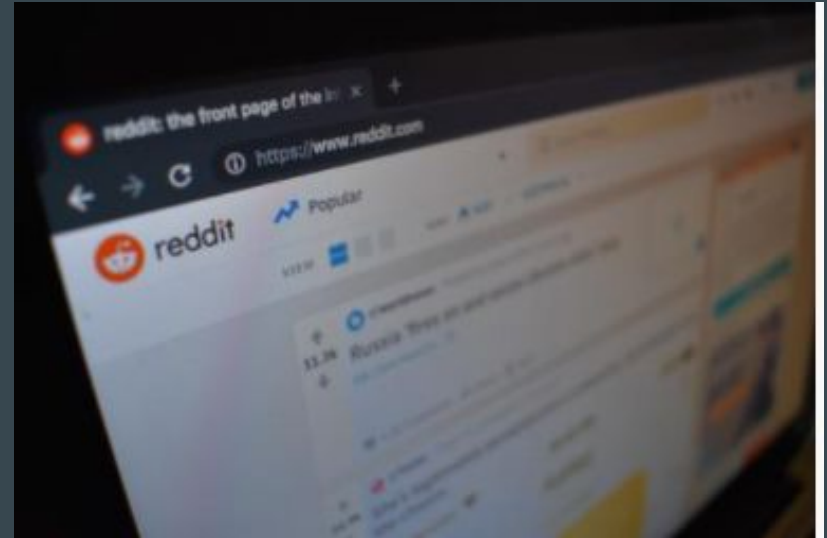


Stavya Arora  
Grade 12



# OUD And Recovery

Sixteen million people worldwide suffer from Opioid Use Disorder (OUD)



Dependence Heroin Detox  
Adverse Withdrawl Drugs Fatal Laws  
Epidemic Depression Abuse  
Usage **OPIOID CRISIS** esoprieno  
Communities Injections Usage  
Addiction Deaths Treatment Health  
Political Fatal Control Addicts  
Antidote Policing  
Drugs Doctors Political  
Overdose

Many of those affected by OUD turn to social media, particularly Reddit (a popular online-based post forum) to aid them in their journey to recovery

# Background

## THE 3 STAGES OF RELAPSE

### 1 EMOTIONAL RELAPSE

Emotions/behaviors may be leading to a relapse.



### 2 MENTAL RELAPSE

Individuals will start to feel tempted to use again.



### 3 PHYSICAL RELAPSE

Individuals give in to their urges and begin using again.



The first stage of relapse is emotional relapse:

- Occurs before an individual even considers using again
- Individual demonstrates a negative affect in their day-to-day language

# Significance of Work

Automated  
identification of  
recovery problems  
through the use of  
machine learning



Early detection and  
intervention by  
professionals



Tangible  
prevention of  
relapse and/or  
fatal overdoses



# Purpose

- Define specific means for **identifying Reddit users who have self-disclosed a relapse** in the subreddit r/OpiatesRecovery
- Demonstrate how the **tones differ** between a user who has self-disclosed a relapse from a user who has not mentioned a relapse in their post history on the same subreddit
- Characterize how an individual's **vocabulary and tone changes** in their Reddit posts as they move towards relapse

# Dataset Overview + Initial Cleaning

## DATASET OVERVIEW

- Retrieved using Reddit's public API from subreddit r/OpiatesRecovery
- Included all posts made by Reddit users from 2012-2021
- Used Python and Pandas (Python Data Analysis Library) to curate the dataset for the specific research question
- Executed primarily in Jupyter Notebooks

## INITIAL CLEANING

- Filtered out all [removed] or [deleted] authors and posts
- Created "fulltext" column - has all the text self-disclosed by user
- Ensured every user had 2 or more posts

```
In [3]: df = df[df.selftext != "[removed]"]
df = df[df.selftext != "[deleted]"]
df = df[df.author != "[removed]"]
df = df[df.author != "[deleted]"]
```

```
In [4]: df["fulltext"] = df["title"] + df["selftext"]
df["fulltext"] = df["fulltext"].astype(str)
```

```
In [5]: df = df.groupby("author").filter(lambda x: len(x) >= 2)
```

# Identifying Users Who Had Self-Disclosed a Relapse

## Built language matching system (regexes)

- Accounted for variations of the most common phrases used in the subreddit to disclose a relapse



↑ 20 ↓

[r/OpiatesRecovery](#) · Posted by [REDACTED] 4 days ago

### Im fkin screaming

So ive never been to my job sober, n im on day 9 sober and was building up quite a bit of anxiety about going back. Bought some kratom otw to work but wound up trying to toss and wash n basically threw it all up all over myself. Despite all that work has been great, nice to get out the house. Anyways im so fkin stoked! Found out my paycheck was 700\$ short, so should have that comin. Also got offered multiple audio engineer gigs for a major fest in my area and im elated 😊. Now i just wait for the inevitable shit storm to come thru cuz i know my life never goes this smoothly. Like i told myself i was going to pickup tmrw but i might for once in my life not

31 Comments   Share   Save   Hide   Report   100% Upvoted

Example Reddit post on r/OpiatesRecovery

# Setting Up Sentiment + Comparative Analyses

- **Treatment Group** - All the posts that discuss relapse in the subreddit
- **Control Group** - All posts from the subreddit made by users who have not self-disclosed a relapse in the past

## Pre-processing Work:

- Removing punctuation
- Lowercase
- Stopwords - "is", "which", "at", etc.

title	selftext	id	fulltext	polarity-score	subjectivity-score
How long til my urine is clean?	I've been taking 10-20mg hydrocodone every day...	[REDACTED]	long til urin cleaniv take 1020mg hydrocodon d...	-0.050000	0.400000
Day 3 ct from 120mg pharma oxy	I waited to thank everyone for the support. Th...	[REDACTED]	day 3 ct 120mg pharma oxyi wait thank support ...	0.005556	0.511111

Returned polarity [-1,1] and subjectivity [0,1] of each post



# Calculated Confusion Matrix Determining Accuracy of Language Matching Systems

Count of Hand-Labeled Neg:39

Count of Hand-Labeled Pos:61

--

Prec, Recall, F1, Support

(0.8009337860780985, 0.7980243799915931, 0.7993876042656531, None)

Confusion

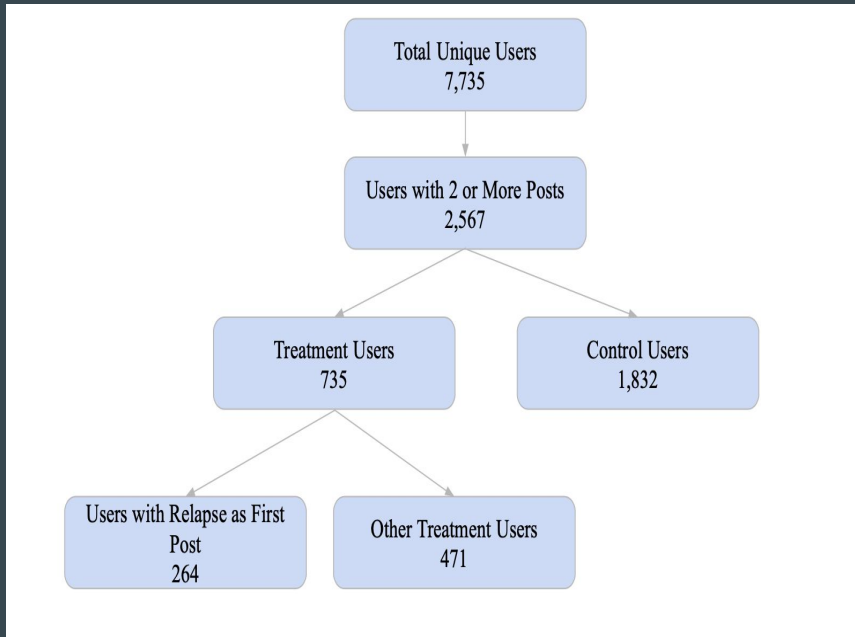
[[29 10]

[ 9 52]]

[29 10 9 52]

tn, fp, fn, tp

# General Results



- 735 users (9.5%) self-disclosed that they relapsed
- 35.92% of the users had their first post discuss a prior relapse

```
#second conclusion
conclusion2 = df_treat["author"].value_counts() / df_master["author"].value_counts() * 100
conclusion2 = conclusion2.dropna()
print (conclusion2)

50.000000
33.333333
Usernames 25.000000
blocked out 16.666667
for 50.000000
...
anonymity 25.000000
purposes 57.142857
17.647059
13.636364
9.090909
Name: author, Length: 735, dtype: float64

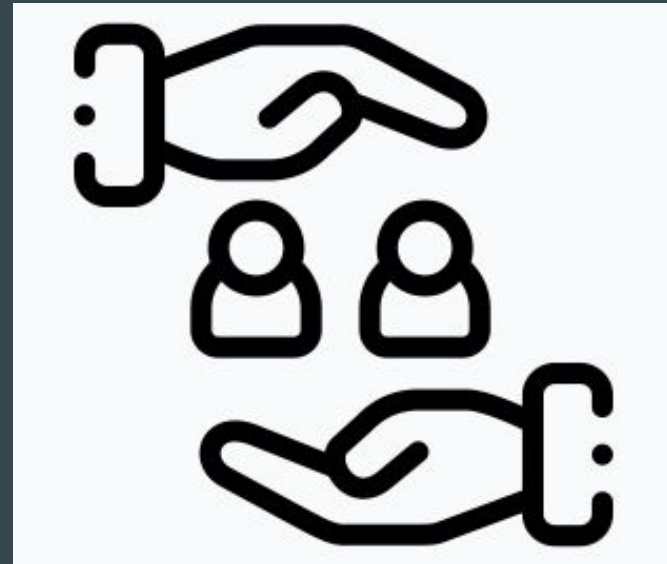
#second conclusion
conclusion2.mean()
39.18036707193818

#second conclusion
conclusion2.median()
33.33333333333333
```

Users discussed relapse in 39.18% of overall posts

# Conclusions

- The regular expressions created were accurate and robust to identify users who had self-disclosed a relapse, thus, **supporting the first hypothesis.**
- Results emphasize importance of online communities in recovery, in line with previous work



# Future Work

Quantify difference in sentiment of users who disclosed a relapse and users who didn't through a comparative statistical analysis

SHORT  
TERM



LONG  
TERM

Certain elements of robust regular expressions can be repurposed for many other addiction studies