Petar Juric

Mohammad Alhabli

Phase 1 - Twitter Opioid Data Analysis

1. The problem

a. Social media is a realtime source of opioid use information that is not subject to disclosure concerns.

b. A fundamental issue in our society today is the increased opioid use in the United States. There is an equally proportional need to analyze health attitudes and online behavior to the utilization of non-medical opioids.

c. Using Twitter and their user content, it may be possible to garner the trends of opioid use and correlate the Twitter content to give insight to population health and mental wellbeing.

d. Slang will be a strong issue, where certain words have to be mapped to keywords.

   i. Example: dope meaning slang for “awesome,” a stupid person, to smear, and most useful for us, a drug taken illegally or recreationally, usually heroin

e. Many people using Twitter create content that will be irrelevant or create content that should be perceived as a joke. The analysis should be able to discern between the two

f. There are many bots on Twitter, so there will be some inaccuracy when dealing with Tweets from those users

2. Literature

a. Chan B, Lopez A, Sarkar U (2015) The Canary in the Coal Mine Tweets: Social Media Reveals Public Perceptions of Non-Medical Use of Opioids. PLoS ONE 10(8): e013507

b. Z. Prince, D. Jha, and R. Singh, “DUI: the drug use insights web server”, Bioinformatics, Vol. 37 (24), pp. 4895-4897, 2021

c. Ducharme J, Moore S. Opioid Use Disorder Assessment Tools and Drug Screening. Mo Med. 2019 Jul-Aug;116(4):318-324. PMID: 31527982; PMCID: PMC6699803.

3. Research, Design and Development Plan.

a. Using Python Twitter libraries such as Tweepy and The twitter API, we will crawl Twitter for tweets containing information on opioid consumption, selling, buying and effect statistics.

b. We may also be able to correlate data that we find with other projects, such as DUI (Drug Use Insights)

   i. DUI is a public and open-source web app allows users to search and explore addiction related terms that are comprised of semantically related topics

c. First, we must obtain Twitter information on a scale that meets our needs, while also adhering to developer rules given by Twitter

   i. 500,000 monthly tweets able to be captured a month

ii. 900 specific tweet information GET requests / 15 minutes per user

iii. 450 historical tweet lookups / 15 minutes per app

d. Tweet GET requests can be filtered to high detail

i. With Academic research access via Twitter's Developer Platform, we are able to programmatically access public tweets dating back to 2006

e. We will use real-time observation with live feeds on Twitter to gain current data

f. Tweets will be searched using Twitter's Python library and relevant data will be stored in our database using MongoDB

4. Evaluation Plan

a. We will be able to analyze data from that point on using Python

i. With Academic research access via Twitter's Developer Platform, we are able to programmatically access public tweets dating back to 2006

b. Correlating data can be difficult with just numbers, so visualization tools will be needed

i. Bar charts

ii. Correlation plots

iii. heatmaps

5. **Summary**
    a. With the ability to conduct these experiments, we will be able to see multiple trends with certain contexts. Questions to be considered:
        i. Are people using opioids more often and posting these experiences on twitter?
        ii. Are these experiences correlated to other data points seen elsewhere? (DUI)

6. **Demo of preliminary results**
    a. Use of Python Bot to scrape data
    b. Data is listed in slide

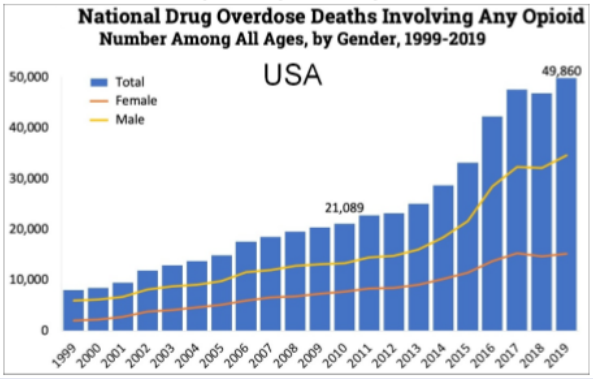
Twitter Opioid Analysis

Petar Juric

Mohammad Alhabli

The American Opioid Epidemic

- 70,000+ die from overdose each year
- 2 million opioid users in past year
- 10.1 million opioid prescription misusers



Opioid Statistics

- Opioid misuse and overdose deaths from opioids are serious health issues in the United States. Overdose deaths involving prescription and illicit opioids doubled from 2010 to 2016, with more than 42,000 deaths in 2016 [CDC 2016a].
- Provisional data show that there were more than 49,000 opioid overdose deaths in 2017 [CDC 2018a]. In October 2017, the President declared the opioid overdose epidemic to be a public health emergency.
- https://www.cdc.gov/niosh/docs/2019-101/pdfs/2019-101.pdf

Current Data Sources

Opioid Misuse Measure (COMM®)

Mark each box that applies	Female	Male
Family history of substance abuse		
Alcohol	1	3
Illegal drugs	2	3
Rx drugs	4	4
Personal history of substance abuse		
Alcohol	3	3
Illegal drugs	4	4
Rx drugs	5	5
Age between 16—45 years	1	1
History of preadolescent sexual abuse	3	0
Psychological disease		
ADD, OCD, bipolar, schizophrenia	2	2
Depression	1	1
Scoring totals		

- 17 question assessment tool
- Identify opioid use and proneness to future use
- What issues do we get from these questions?
  - Influence involved
  - Ability to falsify data

Social Media Data Sources

- Real-time
  - No appointment needed
- Freedom to post*
  - usually no external influence involved
- More than just misuse. General involvement
  - Public opinions on drugs
- Problems with this data?
  - Semantics
  - No formal structure
  - Societal context
  - Community influence

DUI Resource

- Created contextual categories representing activities with substance use, recovery, acquisition of substances, ingestion, relapse and more
- Used random posts taken from drug use subreddits to identify themes and their representative terms

Semantics

- Drug use is covert and includes hidden meanings
  - Slang makes decoding statements difficult
- Social media posts are littered with misspellings

Oxycodone:

- Brand name: Oxycontin, Percocet, Percodan
- Oxycodone street names: Oxy, Hillbilly Heroin, Percs, O.C., Rims, Tires, Greenie

Hydrocodone:

- Brand name: Lortab, Norco, Vicodin
- Hydrocodone street names: Vike, Bananas, Fluff, Hydros

Morphine:

- Brand name: Kadian, Duramorph, MS Contin
- Morphine street names: Miss Emma, Monkey, White Stuff, M

Codeine:

- Brand name: Various
- Codeine street names: Lean, Captain Cody, Little C, Schoolboy

Fentanyl:

- Brand name: Actiq, Sublimaze
- Fentanyl street names: Apache, China Girl, Goodfella, China White, TNT, Friend, Dance Fever

Hydromorphone:

- Brand name: Dilaudid, Exalgo
- Hydromorphone street names: Smack, Dillies, Footballs, Juice

Methadone:

- Brand name: Dolophine, Methadose
- Methadone street names: Tootsie Roll, Red Rock, Mud, Dolls

Oxymorphone:

- Brand name: Opana
- Oxymorphone street names: Biscuits, Mrs. O, O Bomb, Octagon

Opioid Use Twitter Tool

- Analysis tool for twitter data to automatically categorize and subsequently analyze the following about opioids:
  - Health attitudes
  - Beliefs
  - Behaviors related to non medical use
  - Social Sentiment
  - Cultural Relations and History

Goals

- Extract relevant Tweets
  - Bulk over 2000 tweets
- Implement our work to keyword based search from DUI research
  - Parse with DUI, allowing substance-use term lists to be identified
- Analyze data collected
  - Compare results from DUI
  - Visualize different aspects from data (mentions in news, social sentiment, popularity, etc)

How our tool works

- parse Drug Use Insight data using Python script
- Saving parsed data to .csv
- Using saved data to analyze and run tests
  - Use LSA to find context and theme of text

Twitter Bot Prototype: Code

```
#define search twitter function
def search_twitter(query, tweet_fields, bearer_token = BEARER_TOKEN):
    headers = {"Authorization": "Bearer {}".format(bearer_token)}

    url = "https://api.twitter.com/2/tweets/search/recent?query={}&{}".format(
        query, tweet_fields
    )
    response = requests.request("GET", url, headers=headers)

    print(response.status_code)

    if response.status_code != 200:
        raise Exception(response.status_code, response.text)
    return response.json()


#search term
query = "skateboarding dog"
#twitter fields to be returned by api call
tweet_fields = "tweet.fields=text,author_id,created_at"

#twitter api call
json_response = search_twitter(query=query, tweet_fields=tweet_fields, bearer_token=BEARER_TOKEN)
#pretty printing
print(json.dumps(json_response, indent=4, sort_keys=True))
```

Sample Results


```
200
{
    "data": [
        {
            "author_id": "184756808",
            "created_at": "2022-10-02T05:51:42.000Z",
            "edit_history_tweet_ids": [
                "1576449829344923648"
            ],
            "id": "1576449829344923648",
            "text": "Yonge Street is for pedestrians all the way south from Yonge-Dundas Square. Crowds visible all the way. Skateboarding at Y-D Square and as you head south a mix of commis
        },
        {
            "author_id": "1557413159920705536",
            "created_at": "2022-10-02T03:50:16.000Z",
            "edit_history_tweet_ids": [
                "1576419269180805121"
            ],
            "id": "1576419269180805121",
            "text": "RT @cagomez: @RonFilipkowski Sounds like those PA elementary schools are on the same level as the the drinking, smoking, and skateboarding\u2026"
        },
        {
            "author_id": "26962434",
            "created_at": "2022-10-01T22:41:16.000Z",
            "edit_history_tweet_ids": [
                "1576341506457157633"
            ],
            "id": "1576341506457157633",
            "text": "@RonFilipkowski Sounds like those PA elementary schools are on the same level as the the drinking, smoking, and skateboarding dog sign https://t.co/Hh3sWxc1Pm"
        },
        {
            "author_id": "3328575166",
```