Data Science Project

Technology Addiction Twitter Study

by Edward Franke

EXECUTIVE SUMMARY:

The purpose for this project is to determine and track addiction to technology and how it changes over time. The data came from twitter feeds using twitter's api and code to determine the type of tweet related to addiction (if any being tracked). This code has been run twice (April of 2018 and July of 2019) with a net increase of technology addiction tweets of 1%. This needs to be tracked further better analyze the trend. Also, other sources of data should be examined.

Addiction	180419	190709
Total_addiction	3862	1991
illegal_drugs	778	417
prescription_drugs	717	309
shopping	89	30
sex	189	136
gambling	296	223
technology	363	236
alcohol	1202	460
smoking	71	51
Caffeine	157	129
Not Tech	2857	1418

IDEA: Code to capture tweets and classify them into nine categories of addiction.
CLIENT: Persons interested in monitoring addiction to technology (or other addictions).
REASON: Addiction to technology devices and solutions is a growing concern.
DATA: From Twitter feeds using Twitter's API to collect the data.
SOLUTION: Create code to analyze Twitter feeds for user's comments (tweets) about addiction topics.
DETAILS: See below.
DELIVERABLES: Code and a report outlining the discoveries.

BACKGROUND

A growing number of activists and technology professionals are concerned about people becoming addicted to technology and the negative impact on society it brings. This study was initially requested to determine whether 50%+ of the mentioned addictions are technology related. And if not, what percentage or ratio is technology related addictions compared to other non-technology addictions. It then has been updated to look at the trends of addiction, especially related to technology.

DATASETS and PROCESSING

Twitter's API was used on two occasions with the same search criteria and processing to determine technology addictions compared to other addictions and to determine the trend of technology addiction.

```
# Use the REST API for a static search of Twitter feeds
# primary addiction search

num_needed = 25000
tweet_list_f = []
last_id = -1 # id of last tweet seen
while len(tweet_list_f) < num_needed:
    try:
        new_tweets = api.search(q = '#%23addiction -filter:retweets', tweet_mode='extended', count = 100, max_id =
    except tweepy.TweepError as e:
        print("Error", e)
        break
    else:
        if not new_tweets:
            print("Could not find any more tweets!")
            break
        tweet_list_f.extend(new_tweets)
        last_id = new_tweets[-1].id

len(tweet_list_f)
```

```
Could not find any more tweets!
4994
```

Tweets were captured using search criteria of addiction and then 2 additional for phone down and texting driving which are considered symptoms of technology addiction. The data was extracted into the following categories from the following symptoms:
illegal_drugs = ["cocaine", "meth", "marijuana", "heroin", "drugs"]
perscription_drugs = ["pills", "opioid", "oxycontin", "painkiller", "painkillers"]
shopping = ["shop", "shopping", "buy", "buying"]
sex = ["sex", "porn", "pornography"]
gambling = ["gambling", "gamble", "bet", "betting", "casino", "horse", "horses"]
alcohol = ["alcohol", "alcoholism", "drink", "drinking", "party"]
smoking = ["smoking", "smoke", "cigarettes", "ecigarettes"]

technology = ["phone", "iphone", "digital", "technology", "facebook", "smartphone", "texting", "driving", "gaming", "internet", "video"]
caffeine = ["caffeine", "coffee", "tea", "red bull", "energy”]

DISCLAIMER

It is known that counting tweet information for a source of information is not the best method to gain information. Some outlets of addiction are excluded, and others could be mis-identified. For example, weed or pot can be used instead of marijuana, but this hasn't been accounted for. It is known that some areas of the country consider marijuana not to be an illegal drug but for the purpose of this analysis, it is considered as illegal. Only English words are used so foreign languages are not included.

RESULTS

When the code was first run in April of 2018, about 10% of the addiction was indicated as technology related. When it was run a second time in July of 2019, about 11% of the addiction was indicated as technology related. This is an increase of 1% over about 18 months timeframe.

	Addiction	Number
0	Total_addiction	3862
1	illegal_drugs	778
2	prescription_drugs	717
3	shopping	89
4	sex	189
5	gambling	296
6	technology	363
7	alcohol	1202
8	smoking	71
9	Caffeine	157
10	Not Tech	2857

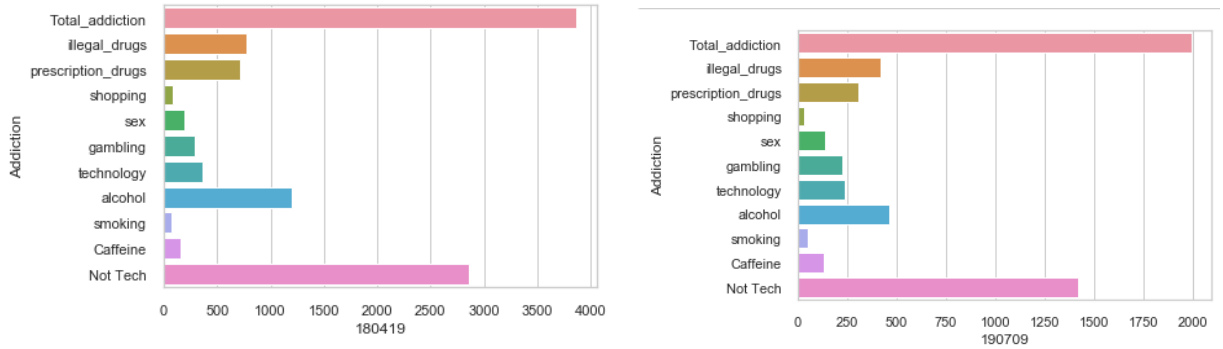
April 2018 results

	Addiction	Number
0	Total_addiction	1991
1	illegal_drugs	417
2	prescription_drugs	309
3	shopping	30
4	sex	136
5	gambling	223
6	technology	236
7	alcohol	460
8	smoking	51
9	Caffeine	129
10	Not Tech	1418

July 2019 results

Addiction	180419	190709
Total_addiction	3862	1991
illegal_drugs	778	417
prescription_drugs	717	309
shopping	89	30
sex	189	136
gambling	296	223
technology	363	236
alcohol	1202	460
smoking	71	51
Caffeine	157	129
Not Tech	2857	1418

Combined results



Barcharts showing the distribution of addiction types per time run.

The total number of tweets related to addiction dropped from 3862 to 1991 between the two dates the code was run. What factors contributed to this are unknown.

CONCLUSION

Further analysis is needed at different times to determine trends beyond just two datapoints. Also, other more credible sources of addiction data should be investigated.