

# How to go Viral on Reddit



Prepared by Jess Chace for FiveThirtyEight aka DSI-CC4-NYC

# Reddit Stats

- 542 million monthly visitors (234 million are unique)
- 3rd most visited site in the U.S. (sixth in the world)
- In 2015, Reddit users submitted nearly 75 million posts and followed up with nearly three quarters of a billion comments

# Methodology

Create a Model to Predict which Features are strong predictors

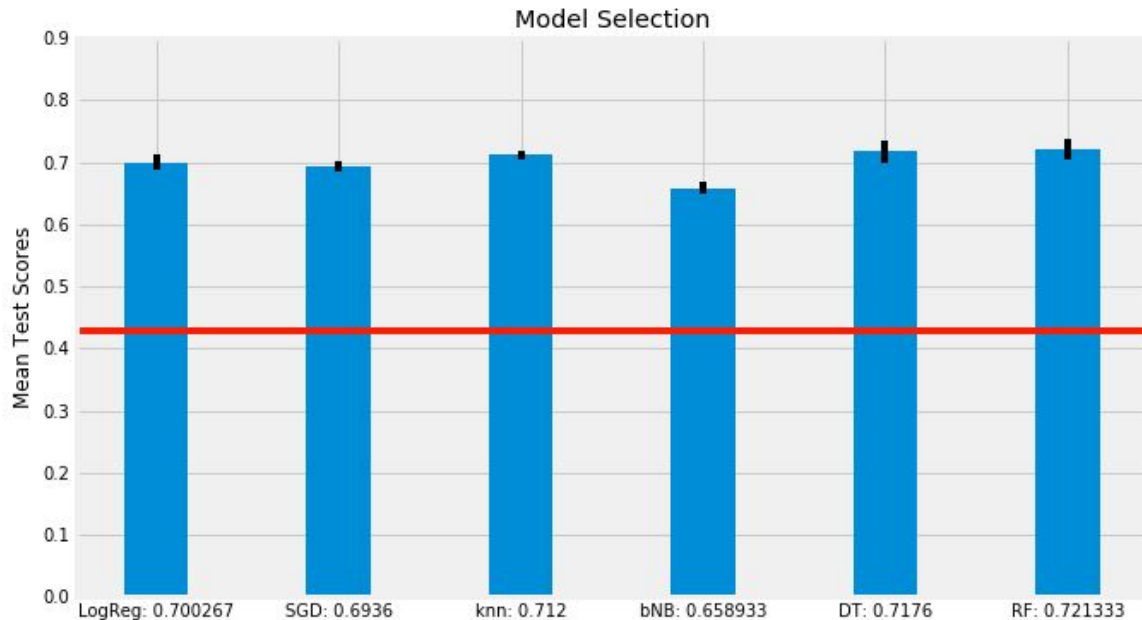
Examine those features for popularity, trends

Analyze Text in Titles, Posts, and Comments

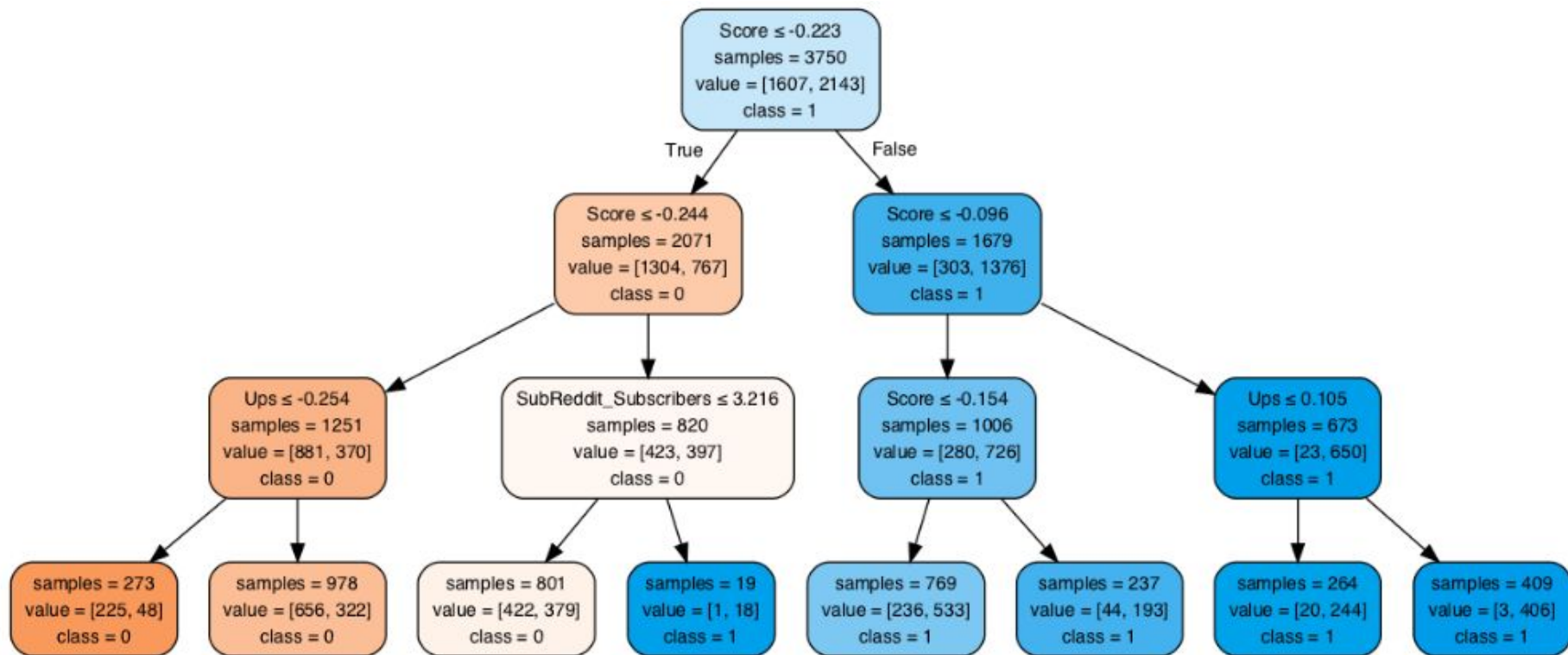
Write post based on findings

# Model Selection

- Target for HOT: 16
- Target for SUPER\_HOT: 44
- Used features: 'SubReddit\_Subscribers', 'Score', 'Ups', 'Num\_Crossposts', 'WLS', 'Created\_UTC', 'Pulled\_UTC', 'IsPinned', 'IsStickied'
- Pipelined and GridSearched across 6 different models to optimize parameters
- Sorted by highest test mean score across 5 cross-validations and across parameters
- Included highest standard deviation
- While models didn't perform as well as expected, they all beat the baseline accuracy score of: 0.427



# Decision Tree Classifier



# Pick a SubReddit

- Out of 5,000 posts, there were 1,950 unique SubReddits
- Top SubReddits have subscribers in the hundreds of millions
- Pick a SubReddit that aligns with your target audience

## SubReddit\_Subscribers

SubReddit	
funny	412170105
todayilearned	395679314
gaming	382102366
aww	361566598
pics	355289984
worldnews	338186535
AskReddit	328170223
gifs	307343070
videos	302774183
news	273454058

EarthPorn	260702814
mildlyinteresting	257815278
movies	246890619
Showerthoughts	240470545
Jokes	177702767
Art	154606843
OldSchoolCool	154313890
WritingPrompts	135494462
television	116745063
science	112509647

# Judge a Post by its Title

- 10,445 unique words exclusive of stop words in the 5,000 post titles
- Choose a title based on frequency of title words combined with target content

	Word	Count
0	new	149
1	just	148
2	like	119
3	time	87
4	day	82
5	oc	72
6	people	70
7	today	70
8	don	69
9	years	69
10	got	69

11	post	61
12	happy	60
13	old	58
14	year	58
15	trump	57
16	best	57
17	little	56
18	ve	56
19	think	52





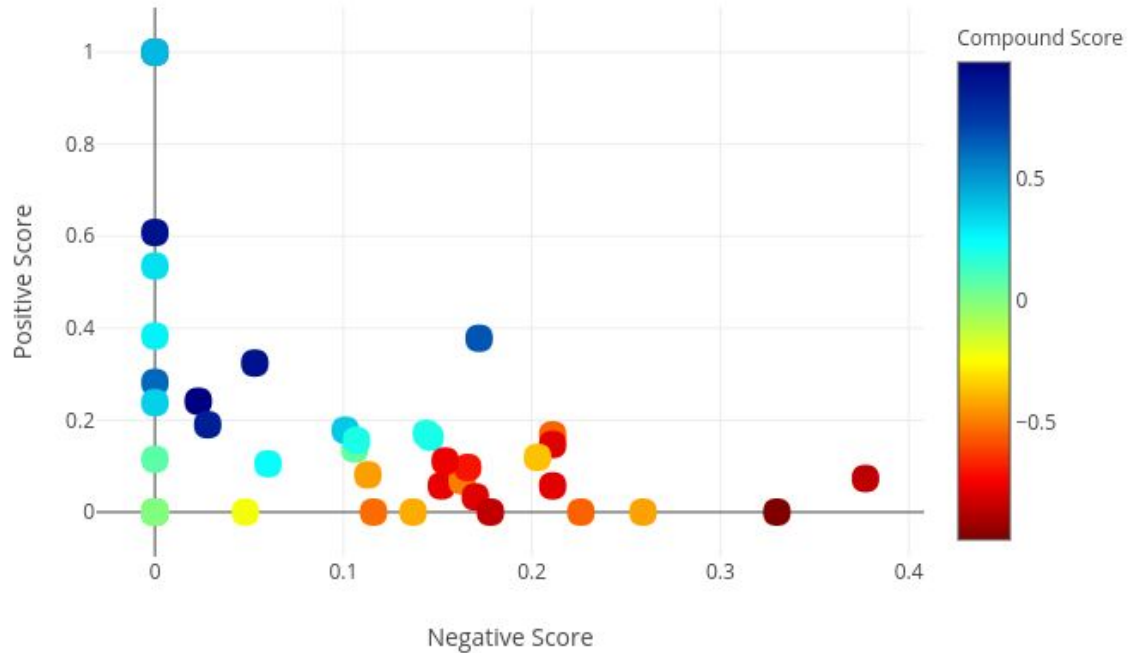
	Id	Title	SubReddit	SubReddit_Subscribers	Num_Comments
420	8nogl0	What celebrity has skeletons in their closet t...	AskReddit	19304088	20246
3009	8nmvqb	What food has made you wonder "how did our anc...	AskReddit	19304127	4731
4963	8nlhuo	What's a shallow reason you wouldn't date some...	AskReddit	19304150	4456
633	8npdh5	One Piece: Chapter 906	OnePiece	167348	4186
344	8nnoyh	My doctor charges me \$65 for being fat	Wellthatsucks	691651	3761
4016	8nmewn	Trudeau responds to Trump Tariffs: "We have to...	worldnews	18788144	3488
157	8nr4pb	What's a good movie everyone should watch but ...	AskReddit	19304083	3460
1098	8nneak	"ARAB ANDY" scares the shit out of a college c...	LivestreamFail	260629	3316
1682	8nneak	"ARAB ANDY" scares the shit out of a college c...	LivestreamFail	260629	3316
4569	8np3js	"YE" Album - First Thoughts Mega-Thread	Kanye	146302	2664
31	8nr18g	Two lions, two tigers and a jaguar have report...	news	16085493	2650
4158	8nnjqe	I texted my younger brother during graduation ...	pics	18699483	2543
4324	8nn95z	What's the creepiest video you've seen on the ...	AskReddit	19304135	2459
3765	8nqgr5	Match Thread: England vs Pakistan at Headingle...	Cricket	71824	2450
1447	8nmhhi	First Image of Shia Labeouf as Shia Labeouf's ...	movies	17635011	2227
435	8noo1r	Little cultural differences	PoliticalHumor	389478	2221

# Most Commented On

# Anticipate the Reactions

- Using Reddit's API wrapper PRAW, I pulled the top comment for the top 50 most commented on posts
- Then, using nltk's SentimentIntensityAnalyzer, I analyzed the top comment for the top 50 most commented on posts
  - SIA gives out four scores: pos, neg, neu, compound
- It's easy for things to devolve quickly on Reddit so I wanted to see whether there were posts that avoided that concern better than others
- Comments don't need to have only positive and good compound scores - rather just looking for really negative scores

Sentiment Analysis of Top Comments



# Write the Post

A few starter sentences using the Markov chain generator:

Women need to give me pity cash.

I will be replaced, and unfortunately, i don't think Ameer will be sticking this for four days.

I can't remember if it was pursuing me. I glanced up at her, and five days after my hospitalization it started to happen. I started working here almost two months ago. We had talked on the couch, took off my sunglasses, and checked my phone. Still no service.

Why would someone do this? There was just going to be literally upside down.

# Appendix A: Non-Canonical Packages

API:

```
import praw
```

Data Visualization:

```
from wordcloud import WordCloud, STOPWORDS
```

Natural Language Processing:

```
import nltk.data
```

```
from nltk.sentiment.vader import
```

```
SentimentIntensityAnalyzer
```

```
from nltk import sentiment
```

```
import markovify as mk
```