Identifying and Isolating Toxicity Using Reddit

Metis, New York

November 16th, 2018

Emma Huang

Toxicity on Reddit

Subreddits used:
- r/Incels
- r/TheRedPill
- r/LegalAdvice
- r/Confessions
- r/TwoXChromosomes
- r/Relationships

r/Incels

"Women are the ultimate cause of our suffering. They are the ones who have UNJUSTLY made our lives a living hell… We need to focus more on our hatred of women. Hatred is power."

"It is obvious that women are inferior, that is why men have always been in control of women."

r/TheRedPill

HOW TO GET LAID LIKE A WARLORD: 37 Rules of Approaching Model-Tier Girls
A complete guide to picking up 9s and 10s

3. A girl can ONLY get attracted to you if you have a stronger frame than her.
Think of this like a mathematical equation. There's no getting around it, no cheating it, no short-circuiting it. It is an immutable law.
Women want to submit to you. They want to submit to a strong man. But she can't submit to you if your frame is weaker than hers.
Physiologically, girls can't even become   for a guy who has a weaker frame than they do. It's literally impossible.

Word Vectors: 'lol'

All Posts

lmfao

lmao

hahah

Word Vectors: 'lol'

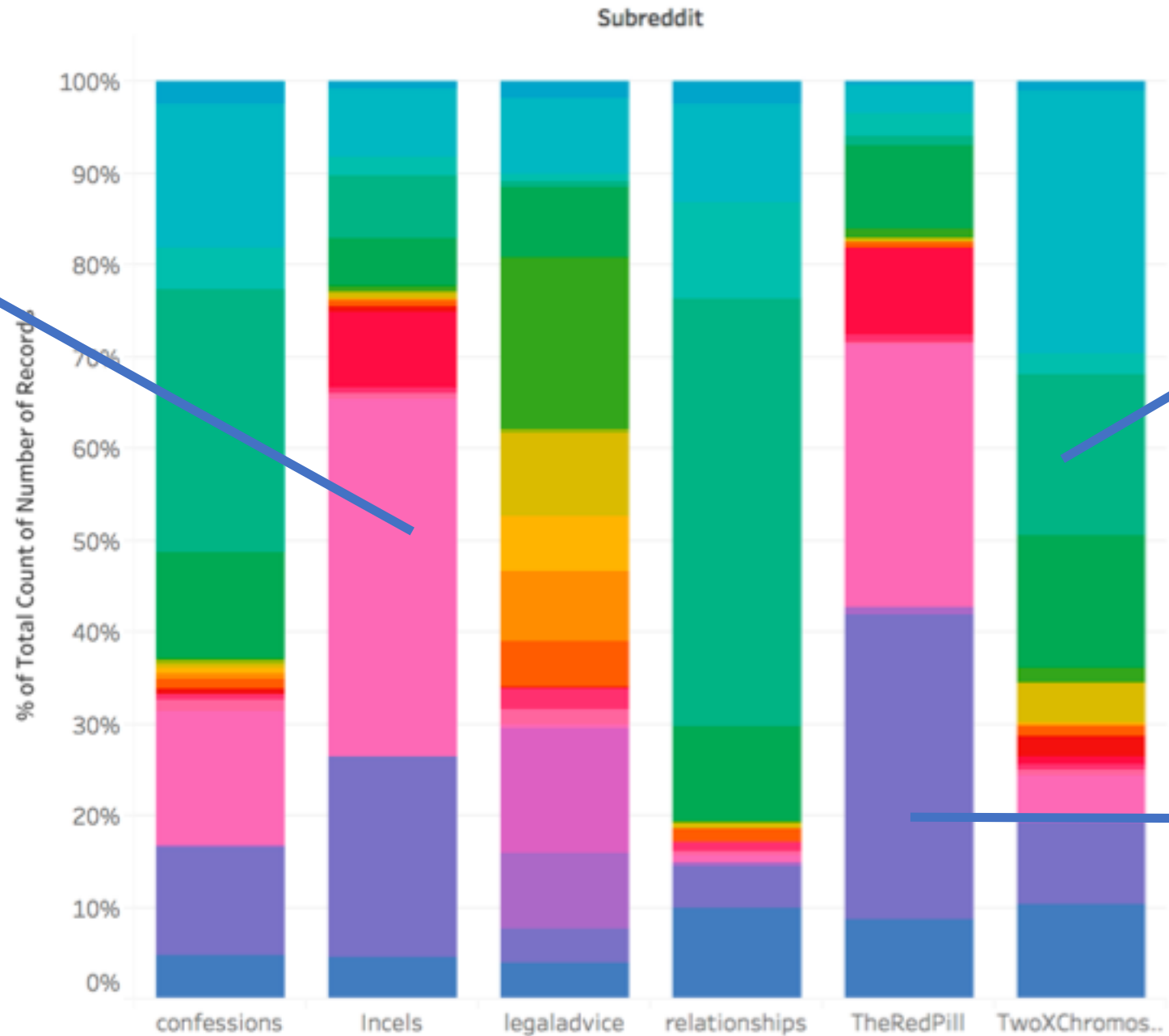| All Posts	r/Incels
lmfao	seriously
lmao	annoying
hahah	idiot

Is it possible to isolate a sexist topic?

- Preprocessing
  - Removed markdown code, slang, etc.
  - Created a custom tokenizer to lemmatize

Is it possible to isolate a sexist topic?

- Preprocessing
  - Removed markdown code, slang, etc.
  - Created a custom tokenizer to lemmatize

- LDA select final topics(out of 20):
  - **"family"**: "mom" , "family", "dad", "sister", "parent"…
  - **"school"**: "school", "college", "class", "student", "university"…
  - **"work"**: "work", "job", "day", "week", "company", "period"…
  - **"dating"**: "feel", "like", "want", "relationship", "love"…
  - **"insecurity"**: "people", "like", "don't", "know", "think", "want"…
  - **"sex"**: "woman", "men", "sexual", "female", "sex"…
  - **"crude sex"**: "girl", "guy", "like", "f**k", "f*****g", "face", "good"…

Subreddit Topic Breakdown

Subreddit

"crude sex"

"dating"

"insecurity"

% of Total Count of Number of Records

100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%

confessions   Incels   legaladvice   relationships   TheRedPill   TwoXChromos…

LSTM Classifier

Target = "crude sex" and "insecurity"

Google Vector Embeddings

Optimized Accuracy/F1
- Accuracy: 0.896
- F1: 0.692

r/Confessions

"I've never had a boys phone number, never had a guy give me his jacket, never had a guy even talk to me like he would be interested in me. Honestly I feel like I'm going to **be lonely forever**... I feel like something's wrong with me **like I can't attract or talk to the opposite sex**... I just don't understand, does it have to do with **how ugly I am**? If so, how can I make myself prettier? I'm trying to lose weight, and it's slowly but surely working, but not enough ... I feel so lonely."

r/LegalAdvice

- "I've made a few memes out of some **annoying activists** at my, uhh... unique college. They've gone around in **anti sjw** circles on the internet and have been reposted by many people on many websites..."
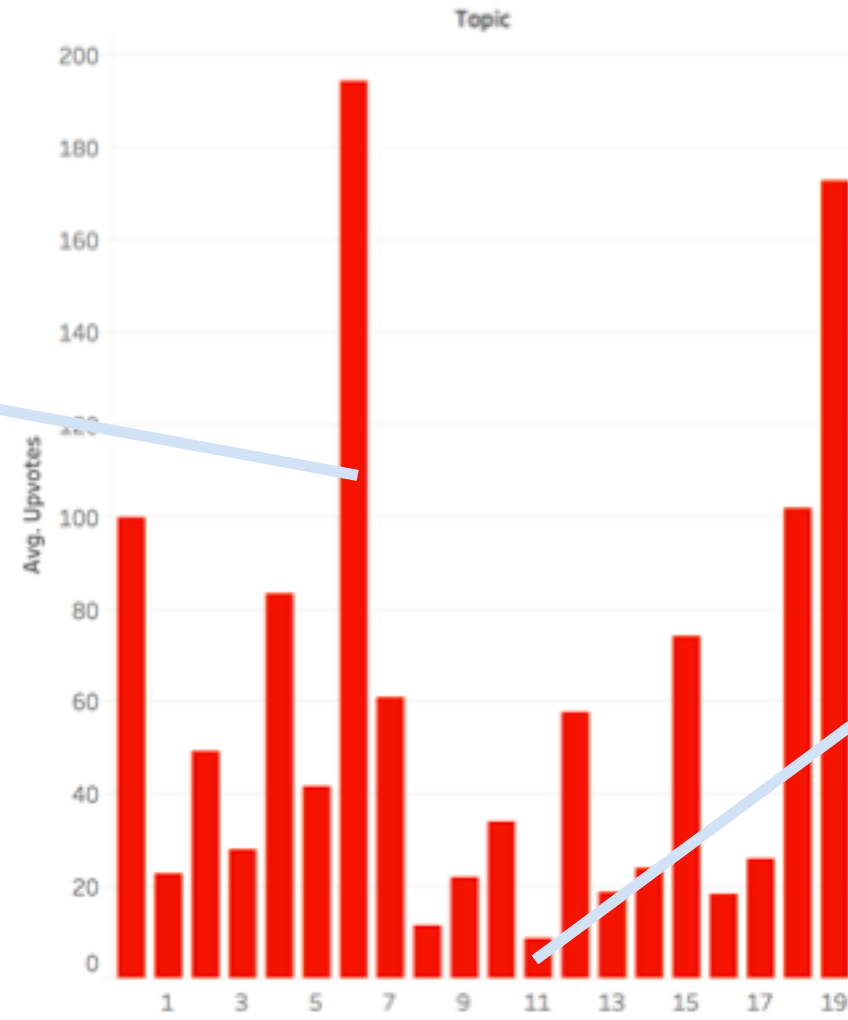
False Positive Comments

Classifications on Outside Text

Subreddit	Ratio of Positives
r/DataScience	0.072
r/Sex	0.068
r/The_Donald	0.549

r/Incels	r/TheRedPill	r/Relationships	r/TwoX	r/Confessions
male	male	male	male	male
alpha	dominated	acquaintance	dominated	younger
platonic	partner	acquaintances	warriors	met
femoid	romantic	bestfriend	persons	online
friendship	providing	males	characters	dated

Appendix 1: Word Vectors: 'woman'

Appendix 3: Topics

- Topic 0: family (parents)
- Topic 1: feelings
- Topic 2: friendships
- Topic 3: relationships/dating
- Topic 4: actions
- Topic 5: lawyer/court info
- Topic 6: school/college
- Topic 7: work
- Topic 8: driving ticket
- Topic 9: car accident
- Topic 10: house

- Topic 11: birth control
- Topic 12: sexual desire
- Topic 13: phone/texting
- Topic 14: family (partner/children)
- Topic 15: crude sex
- Topic 16: apartment
- Topic 17: salary
- Topic 18: self conscious
- Topic 19: other