

Project Proposal

Arjun Mahadkar
B00976832
amahadk1@binghamton.edu

Deepang Raval
B00924269
draval1@binghamton.edu

Rahul Verma
B00892091
rverma4@binghamton.edu

Sudeep Rawat
B00852066
srawat1@binghamton.edu

Yuraj Vartak
B00866245
yvartak1@binghamton.edu

1 RESEARCH QUESTIONS

We plan to answer these research questions for each of the social media platform we have collected the data from.

1.1 Twitter Tweets Data

For each named entity recognized from the tweets, what is the dominant sentiment towards it?

1.2 Subreddit Comments Data

What percentage of comments that are made related to the U.S. politics are toxic?

1.3 YouTube Comments Data

What is the structure of comments that are posted with the motive of scamming others?

2 ANALYSIS PRESENTED

2.1 Twitter Tweets Data

- (1) Number of tweets collected per context within the input date and time range.
- (2) Top hashtags for the provided date and time range.
- (3) Wordcloud of named entities from tweets.

2.2 Subreddit Comments Data

- (1) Number of subreddit comments received for the specified date and time range.
- (2) Wordcloud of named entities from subreddit comments.

2.3 YouTube Comments Data

- (1) Videos with most comments for the specified date and time range.
- (2) Authors with most comments for the specified date and time range.
- (3) Wordcloud of most frequent words in comments related to Spam/Scam.

3 TOOLS

- (1) Django
- (2) Chart.js
- (3) Matplotlib
- (4) Seaborn
- (5) Wordcloud