Optimizing Tweets for Engagement on the Topic of HIV and AIDS

Sahar Manavi
Springboard
May 2021

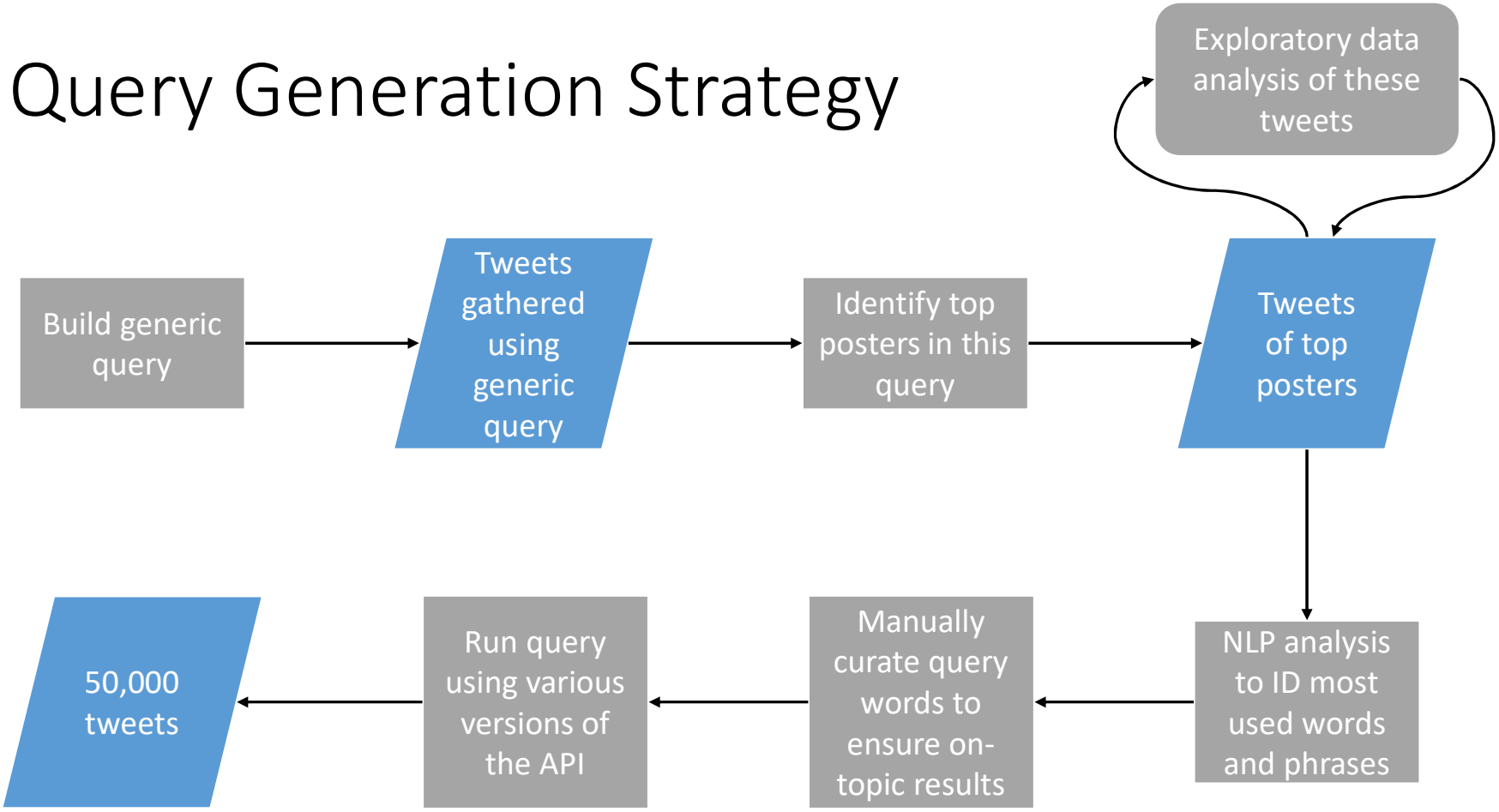
Goals for this project

- Identify phrases and hashtags
- That predict higher engagement (favorites, retweets, replies, and quote tweets)
- For tweets on the topic of HIV and AIDS.

Background and Reasoning

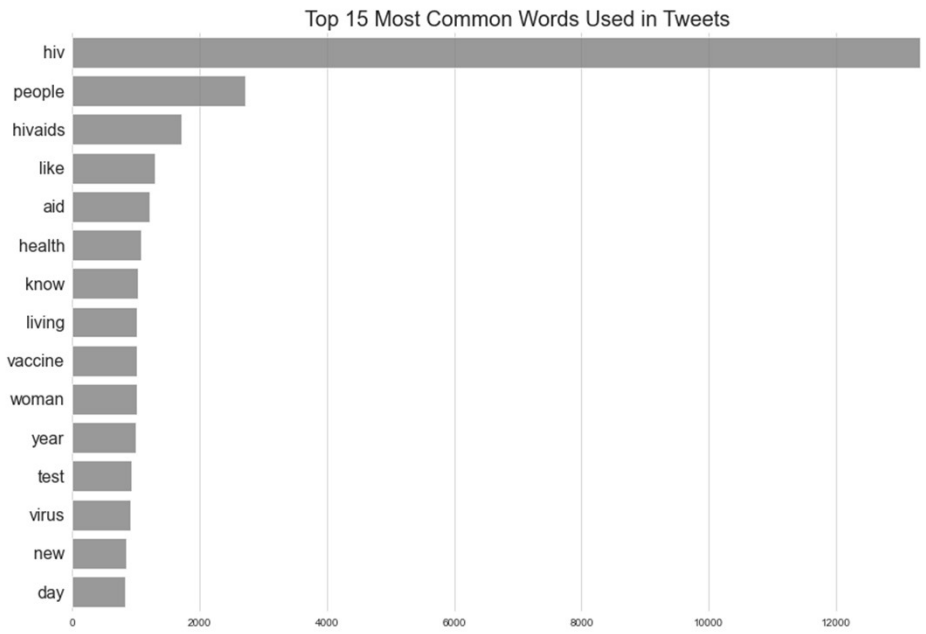
- The HIV/AIDS community on Twitter is relatively small and it's rare for tweets on the topic to have broad appeal outside of that community.
- However, there are still "celebrities" or more popular tweets and topics within the community.
- Donors looking to help organizations will be knowledgeable of and looking within the community to find groups to support.
- Therefore, establishing credibility and having broad reach within the community is important.

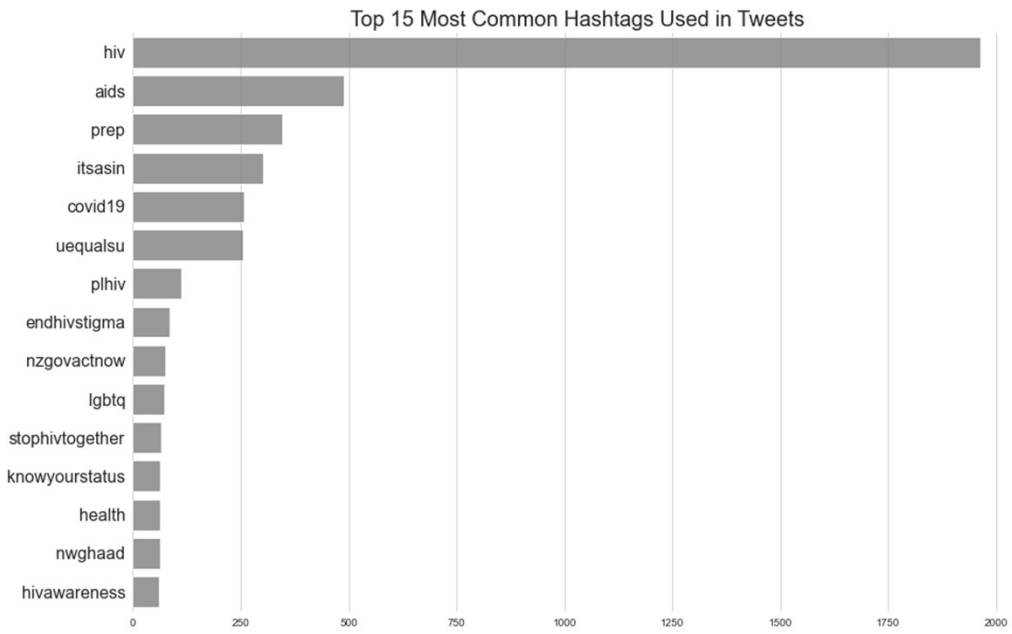
Query Generation Strategy



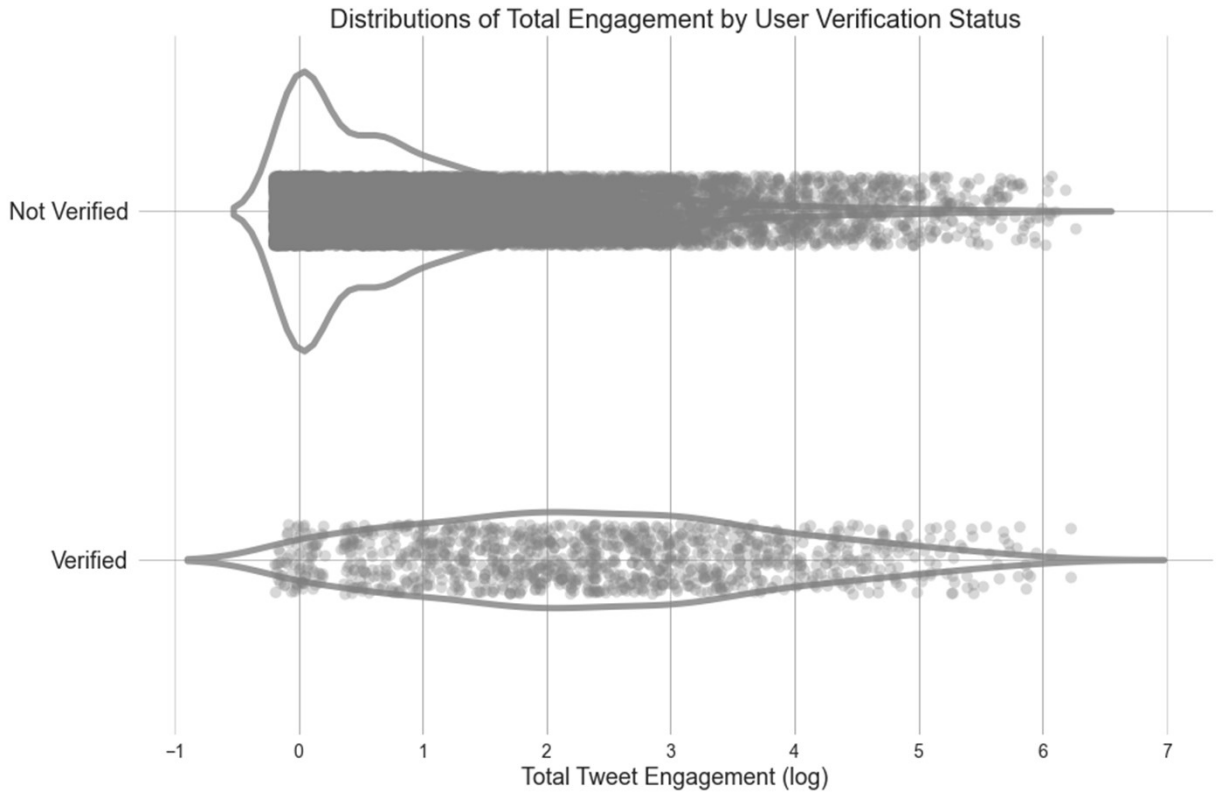
Final dataset

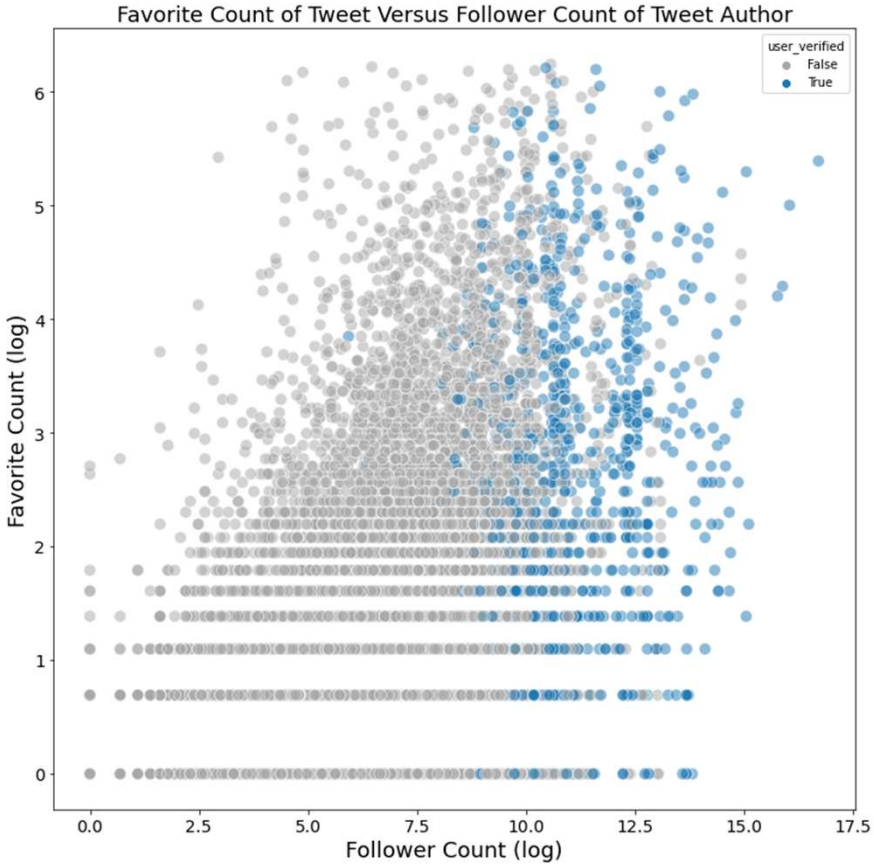
17,924 unique tweets from **12,674** unique users



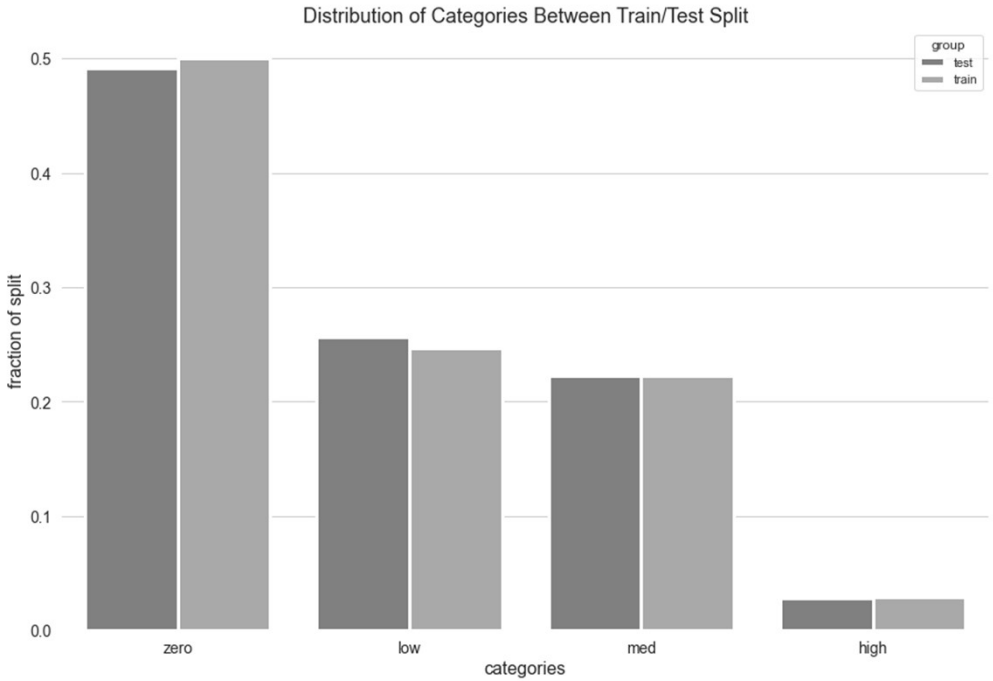


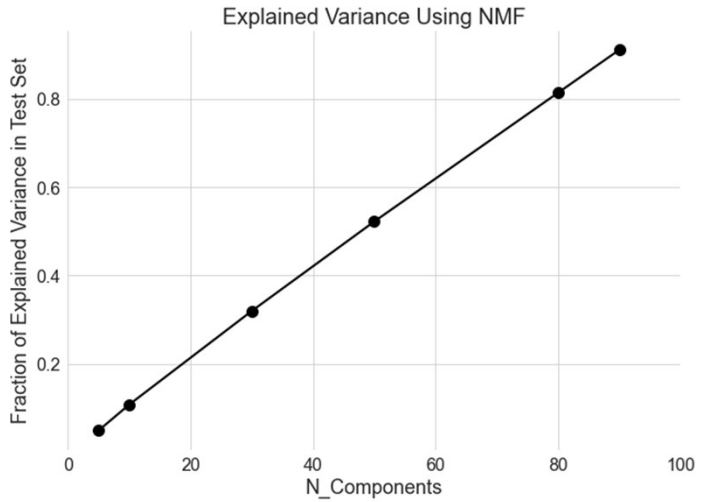
Follower Count and Verification Status



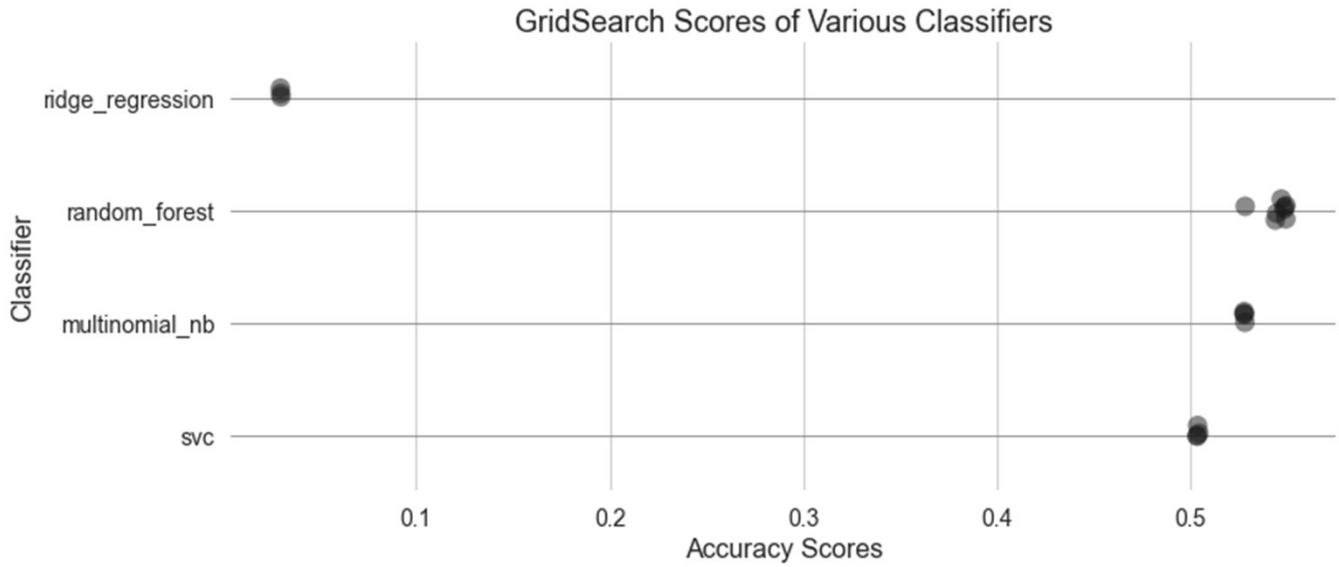


Data Pre-processing





Model Performance



Conclusion

- There does not appear to be an underlying pattern to the data.
- Possible explanations:
  - Engagement is mostly random and/or based on what the underlying Twitter algorithm chooses to prioritize.
  - A necessary aspect of my data collection (spreading it out over time) created too much topic randomness, since Twitter topics have their own lifecycle.

Future Directions

- Neural network/deep learning.
- Try resampling to even out classes.
- Change categories to be binary.
- Different features that aren’t all text dependent.