POSITIVE NEWS

Analysis of Headlines

By Preeya Sawadmanod

Agenda

01 Problem

02 Data Collection

03 Explore Data

04 Modeling

05 Recommendation

Problem

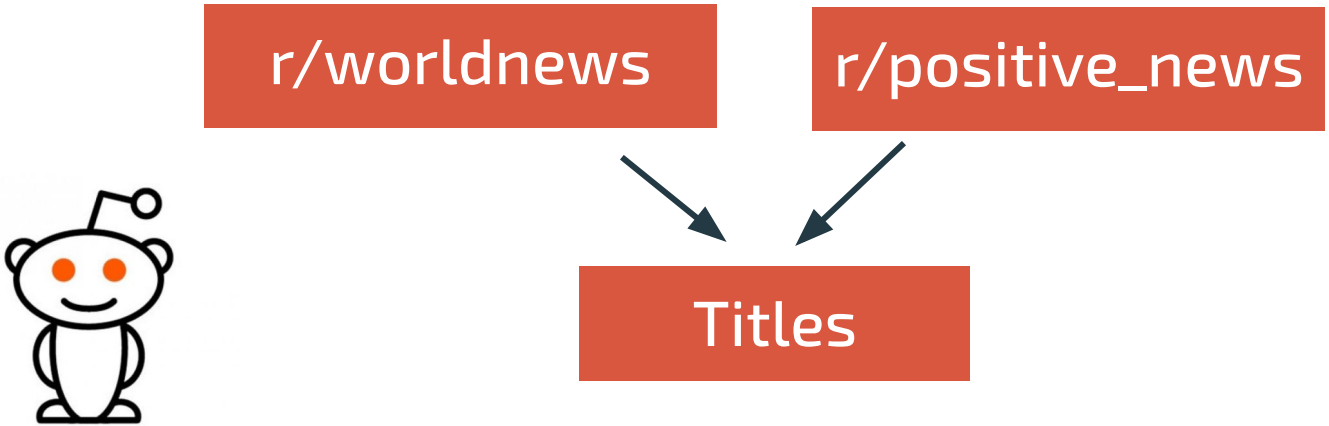
- **Great Headlines** to draw the reader's attention.
- Emphasize the **positivity** of news all over the world.
- Headlines should be **different** and **unique** compared to other news headlines.

Are there any distinct keywords in titles, that would make someone perceived news as positive?

Data Collection

Reddit is a **network of communities** based on people's interests.

Web Scraping:

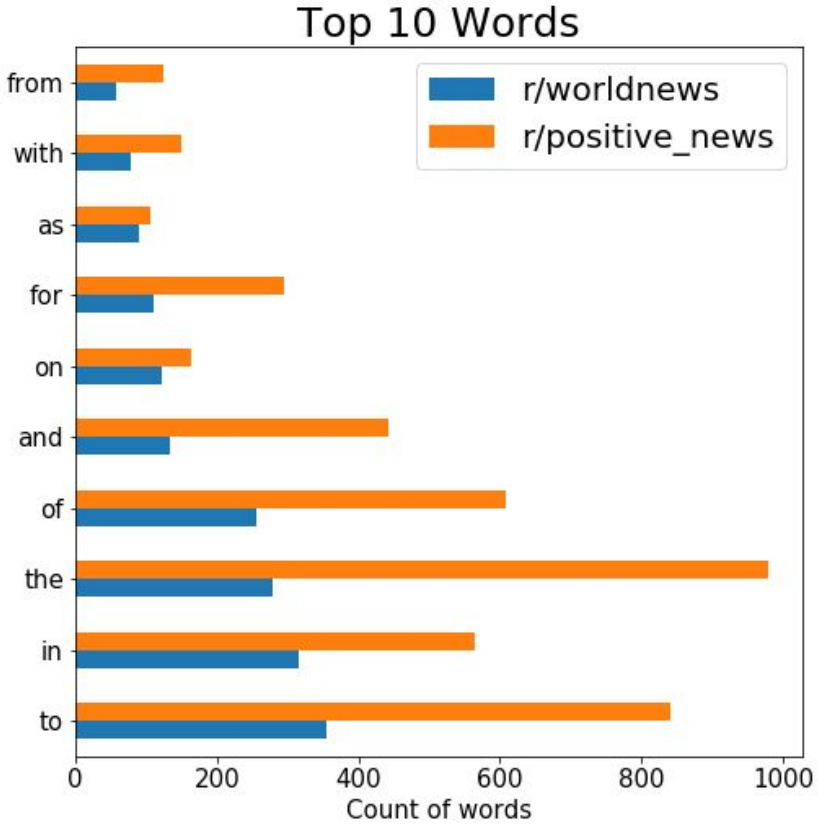


Explore Data

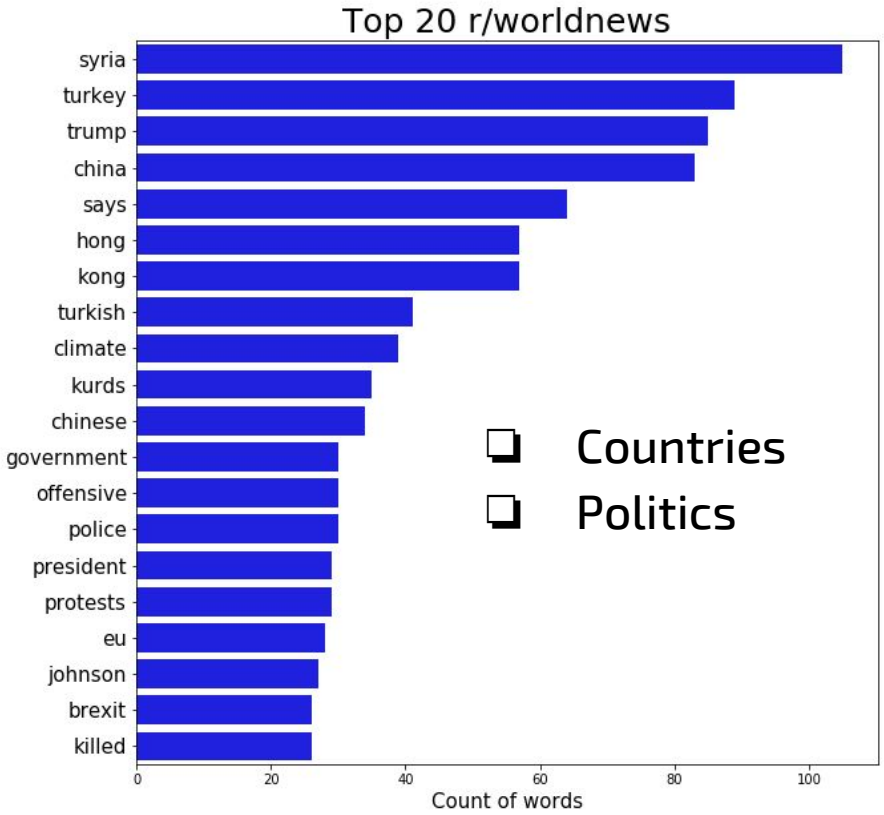
Cleaning:

- ❑ Removed Punctuation
- ❑ Removed Duplicates

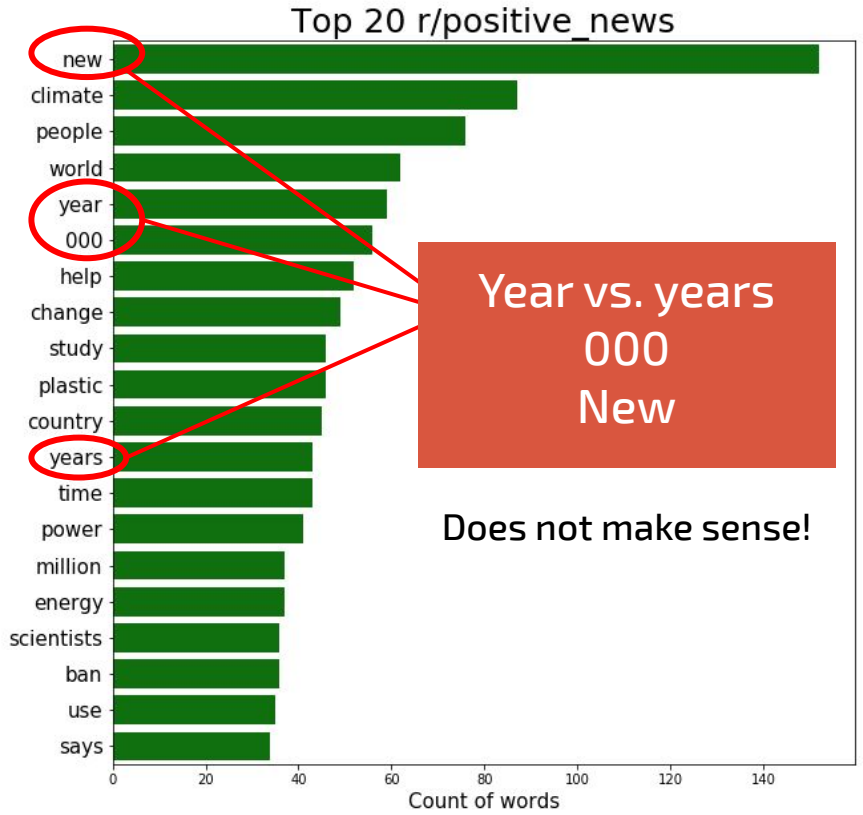
Remove stopwords



Explore Data



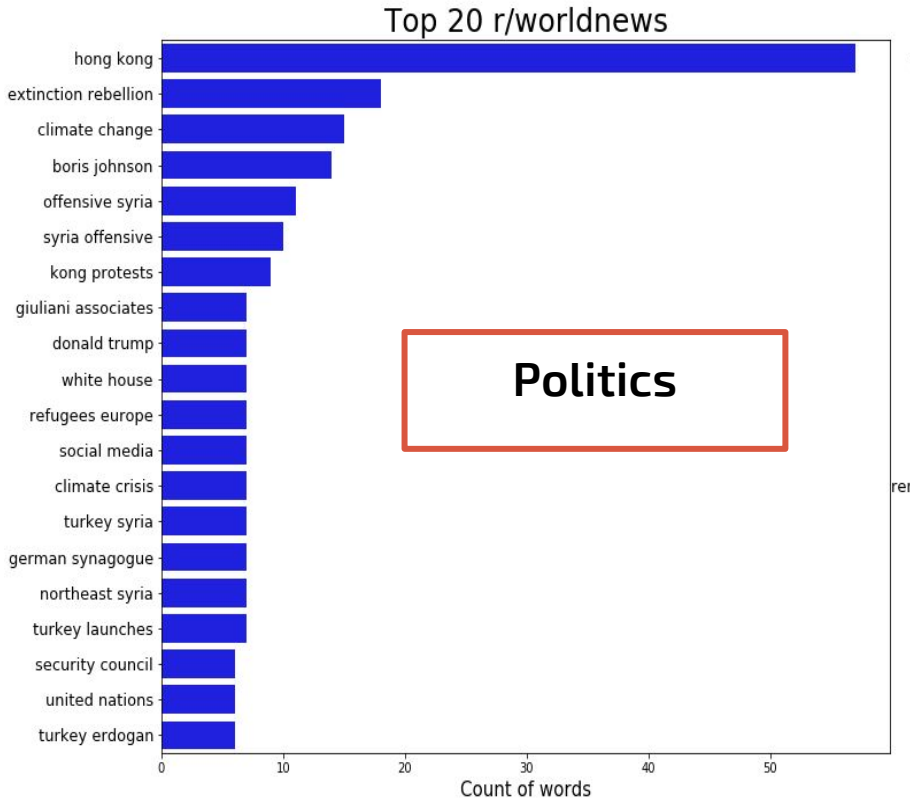
- Countries
- Politics

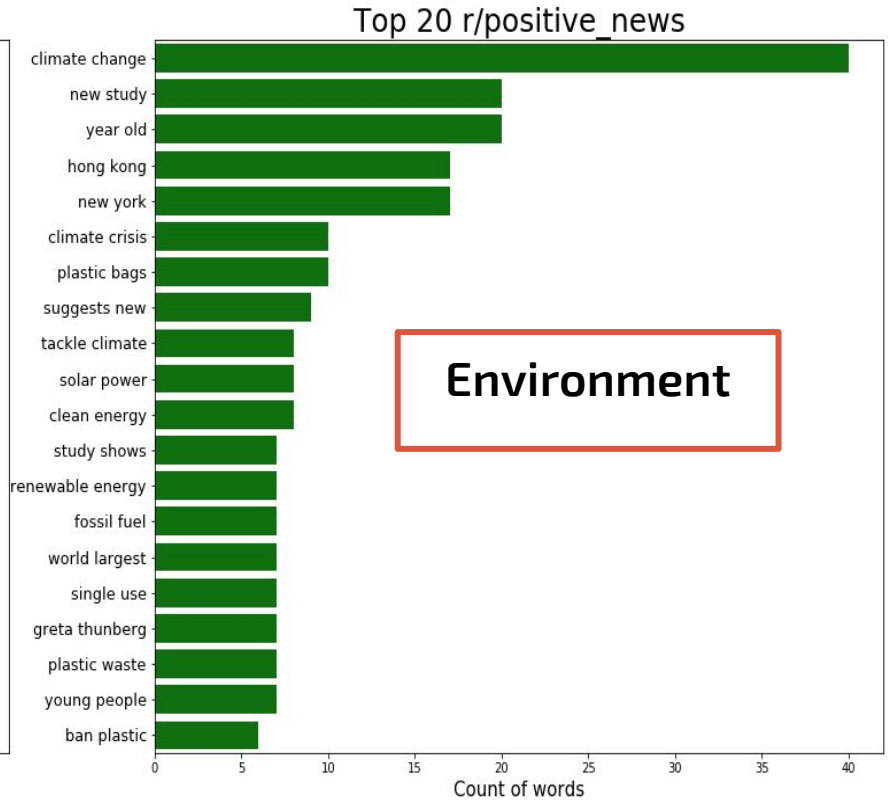


Year vs. years
000
New

Does not make sense!

Explore Data

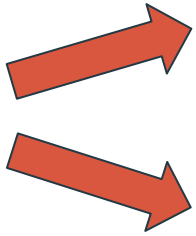




Modeling

Classification Models:

- ❑ Naive Bayes Model
- ❑ Logistic Regression Model



r/worldnews

r/positive_news

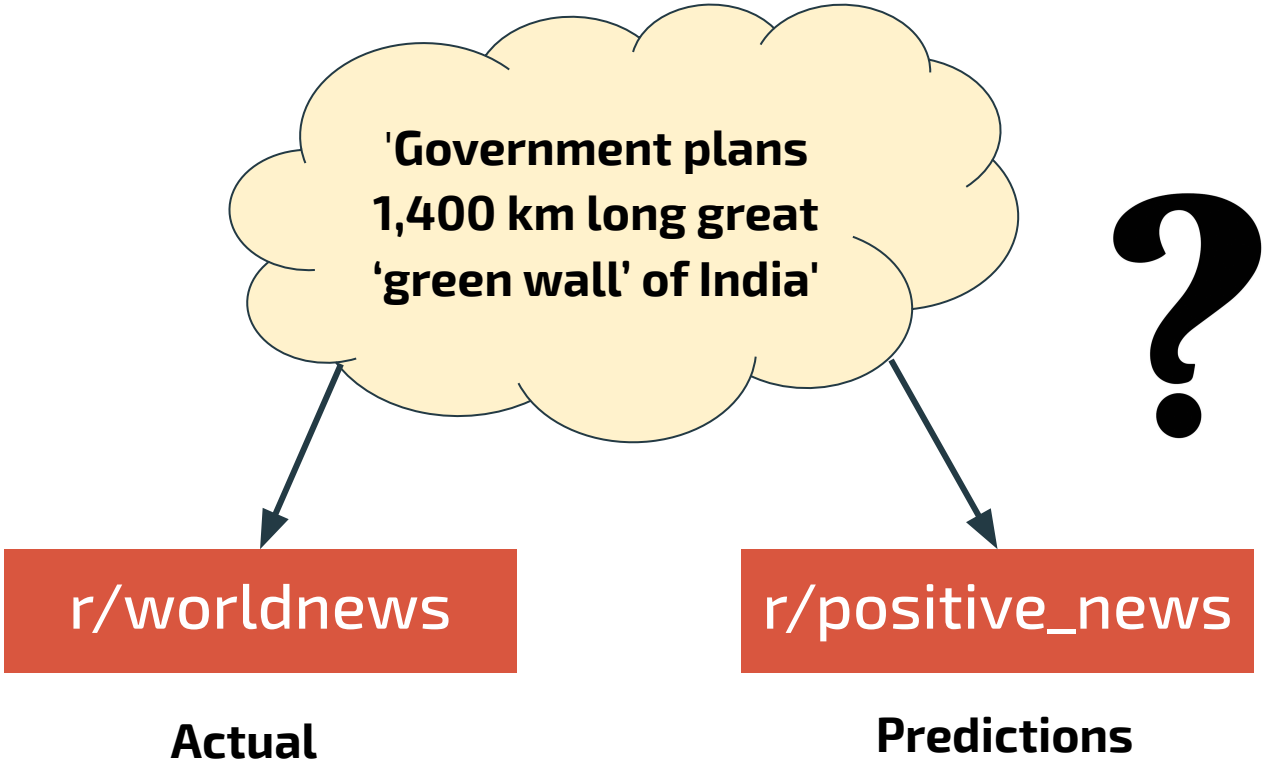
Best Model: Naive Bayes Model

- ❖ Accuracy score to measure success
- ❖ Misclassification are equally bad

86% Accuracy in Training set

96% Accuracy in Testing set

What do you think?



Recommendation

- **Overlap** between news titles
  - As shown in previous graphs
- **Improving** accuracy score by collecting more data
- There are different **keywords** in titles for positive news:
  - Climate change
  - Renewable energy
  - Fossil fuel