Would reddit Like My Comment?

Zach Heick

What is Reddit?

- Reddit is a social news aggregation website consisting of smaller discussion boards, called *subreddits*
- Subreddit topics consist of almost anything
- Users can submit and comment on *threads*

Comment Scoring System

- Each comment has an initial score of 1
- Users can upvote (+1) or downvote (-1) the comment
- Comments sorted from greatest to least (by default)

Comment Scoring System

Comments that you share an opinion with, funny

Hide blatant racist, sexist, and derogatory comments from being seen

Comment Scoring System

Comments that you share an opinion with, funny

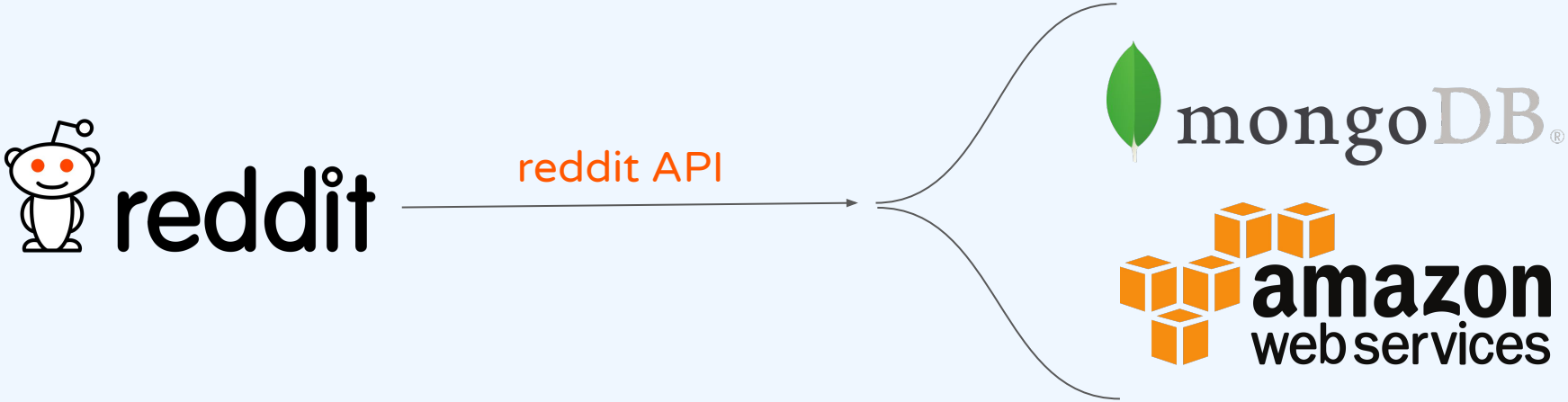
Genuine opinions that do not agree with the majority of users are downvoted in disagreement

"Thanks. Great political discussion."

Can we classify if reddit would like a comment?

Data Collection

- Collected **197,427** comments from multiple subreddits using Reddit API
- Politics, atheism, worldnews, science, and hip hop



Text Cleaning

- Removed markdown syntax, hyperlinks, and stop words
- Lemmatized cleaned words
- Used TFIDF on cleaned words to create sparse matrix

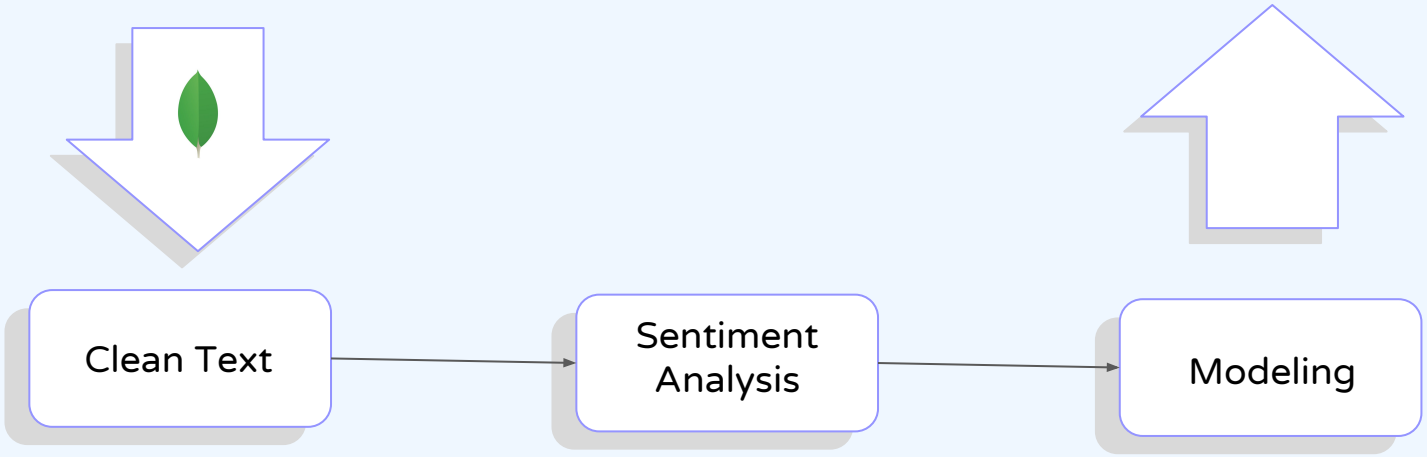
Sentiment Analysis

- Utilized TextBlob library for sentiment analysis
- Recorded subjectivity, polarity by sentence
- Comment word count and character count

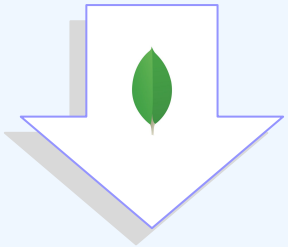
Modeling

- Binned and classified scores
  - **Class 0**: (-inf, 50], **Class 1**: (50, inf)
- Decision Tree algorithm to classify, chased **recall**

/r/worldnews
/r/science
/r/politics
/r/hiphopheads
/r/atheism

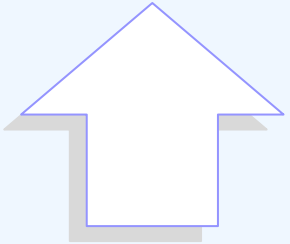


/r/worldnews
/r/science
/r/politics
/r/hiphopheads

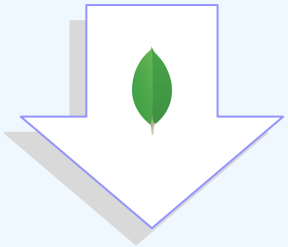


Clean Text → Sentiment Analysis → Modeling

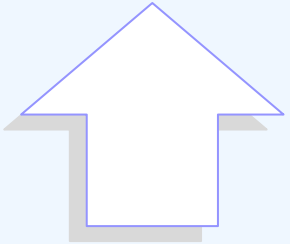
/r/atheism **model**

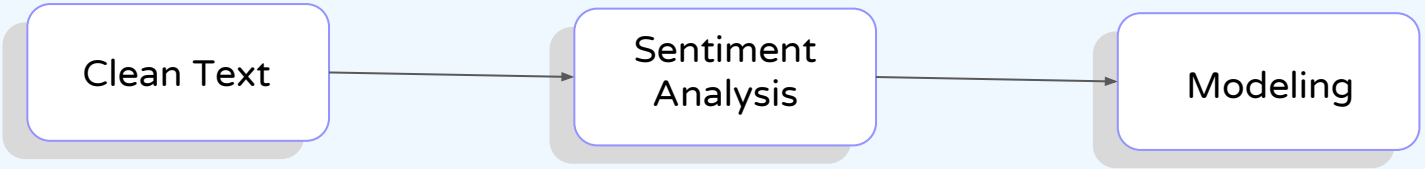


/r/worldnews
/r/science
/r/politics

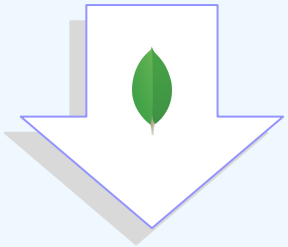


/r/atheism **model**
/r/hiphopheads **model**

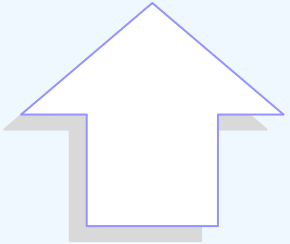




/r/worldnews
/r/science



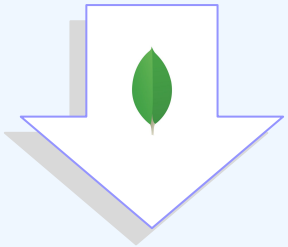
/r/atheism **model**
/r/hiphopheads **model**
/r/politics **model**

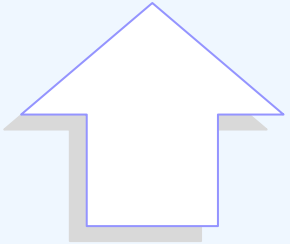


Clean Text → Sentiment Analysis → Modeling

/r/atheism **model**
/r/hiphopheads **model**
/r/politics **model**
/r/science **model**

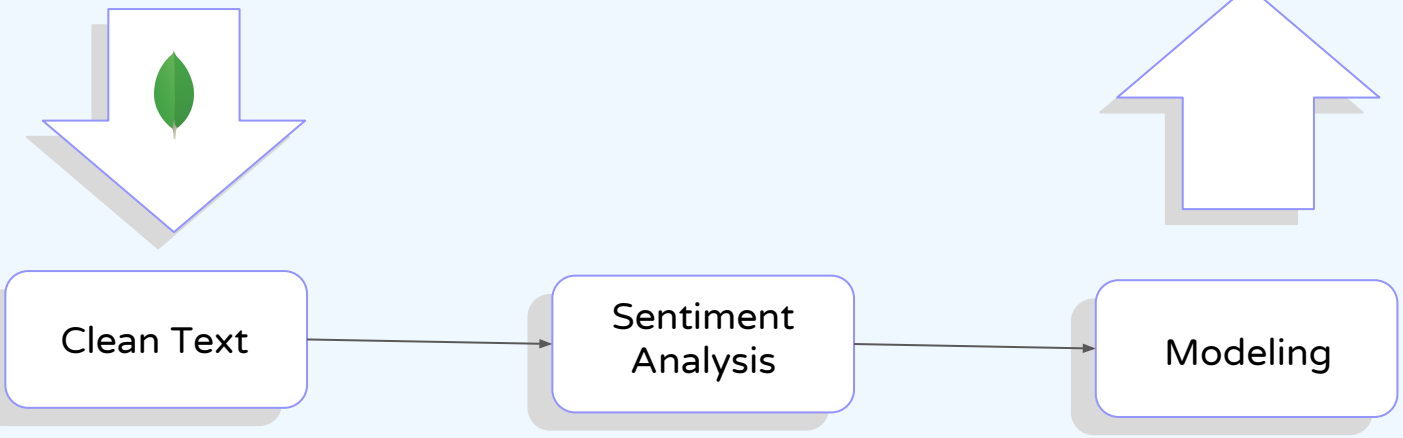
/r/worldnews

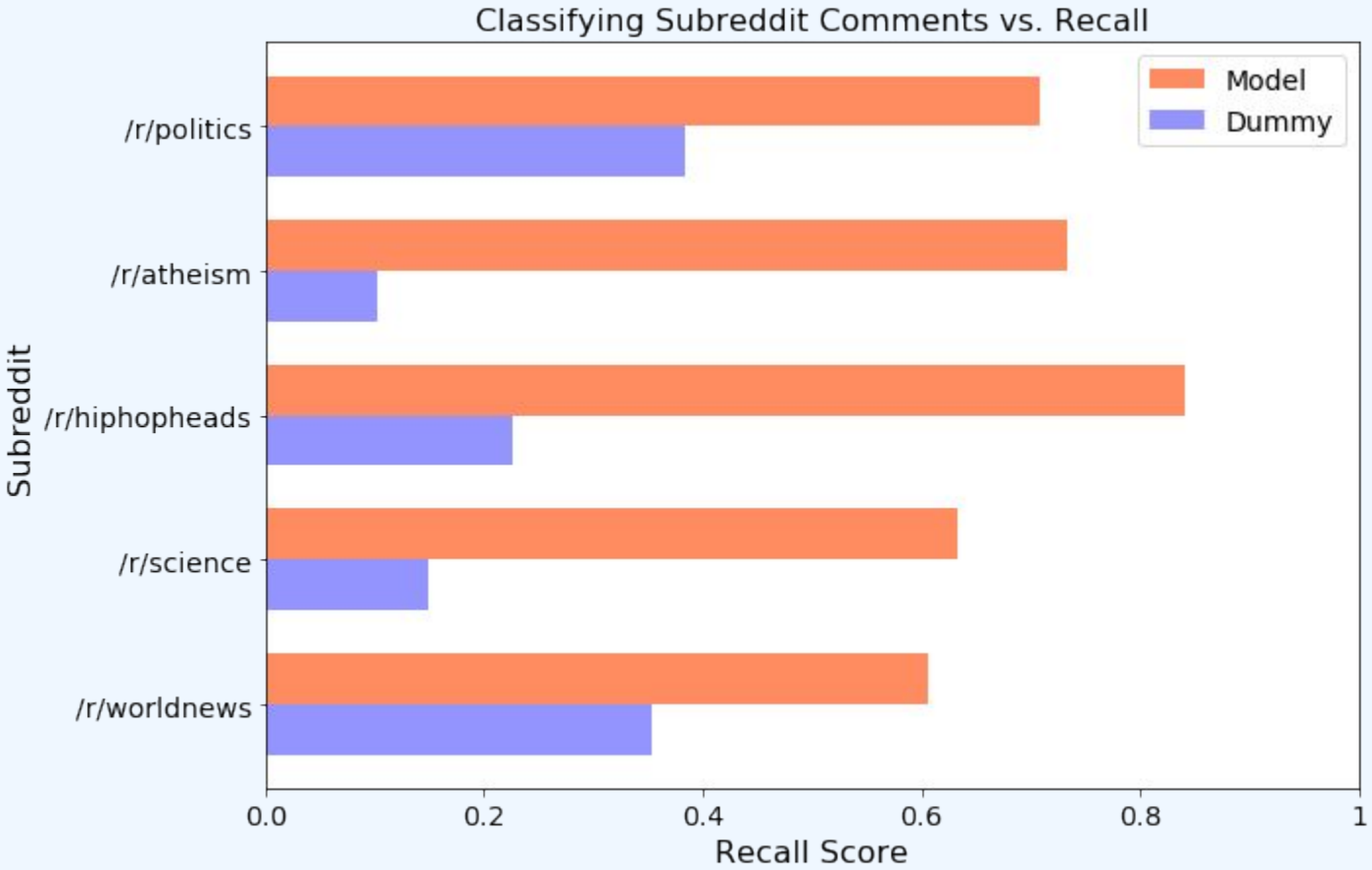




Clean Text → Sentiment Analysis → Modeling

/r/atheism **model**
/r/hiphopheads **model**
/r/politics **model**
/r/science **model**
/r/worldnews **model**




Classifying Subreddit Comments vs. Recall
Model
Dummy
/r/politics
/r/atheism
/r/hiphopheads
/r/science
/r/worldnews
Subreddit
0.0
0.2
0.4
0.6
0.8
1
Recall Score

Demo!

Appendix

Comment Scoring System

Steveweing 2 points 1 day ago

I just ran this through a linear regression model and have determined that data science will pay $0 in 2050.

Comment Scoring System

Steveweing 1 points 1 day ago

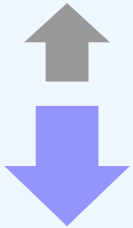
I just ran this through a linear regression model and have determined that data science will pay $0 in 2050.

Comment Scoring System

Steveweing 1 points 1 day ago

I just ran this through a linear regression model and have determined that data science will pay $0 in 2050.

Comment Scoring System



Steveweing 0 points 1 day ago

I just ran this through a linear regression model and have determined that data science will pay $0 in 2050.

Would reddit Like My Comment?

Simulates comment scores for specific Reddit subreddits depending on how many minutes the comment would be made after a thread was posted and the content of the comment itself.

/r/atheism

Thread was posted: 0 minutes ago.

Save

Clear Text Area | Clear Comments

Would reddit Like My Comment?

Simulates comment scores for specific Reddit subreddits depending on how many minutes the comment would be made after a thread was posted and the content of the comment itself.

/r/atheism

Thread was posted: 0 minutes ago.

Save

Clear Text Area | Clear Comments

Would reddit Like My Comment?

Simulates comment scores for specific Reddit subreddits depending on how many minutes the comment would be made after a thread was posted and the content of the comment itself.

/r/atheism

Thread was posted: 0 minutes ago.

Save

Clear Text Area | Clear Comments

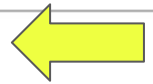
Would reddit Like My Comment?

Simulates comment scores for specific Reddit subreddits depending on how many minutes the comment would be made after a thread was posted and the content of the comment itself.

/r/atheism

Thread was posted: 0 minutes ago.

Save



Clear Text Area | Clear Comments

Wrapping up

- Created a class, **RCC (Reddit Comment Classifier)**, containing models for each subreddit
- Used RCC for web application