

# AskWomen AskMen Classification

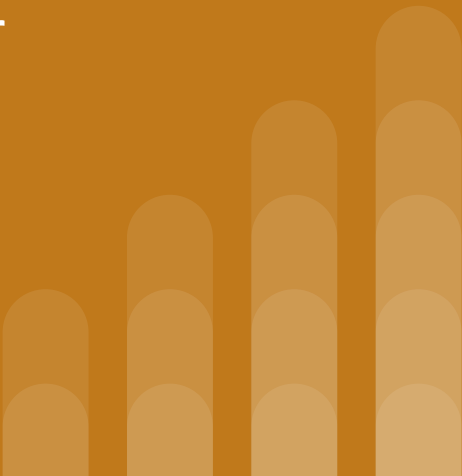
Subreddit comparison  
Sean Bjork





## **Problem Statement:**

To build an accurate (over 70%) predictive model for classifying Subreddit posts as from either AskMen or AskWomen.



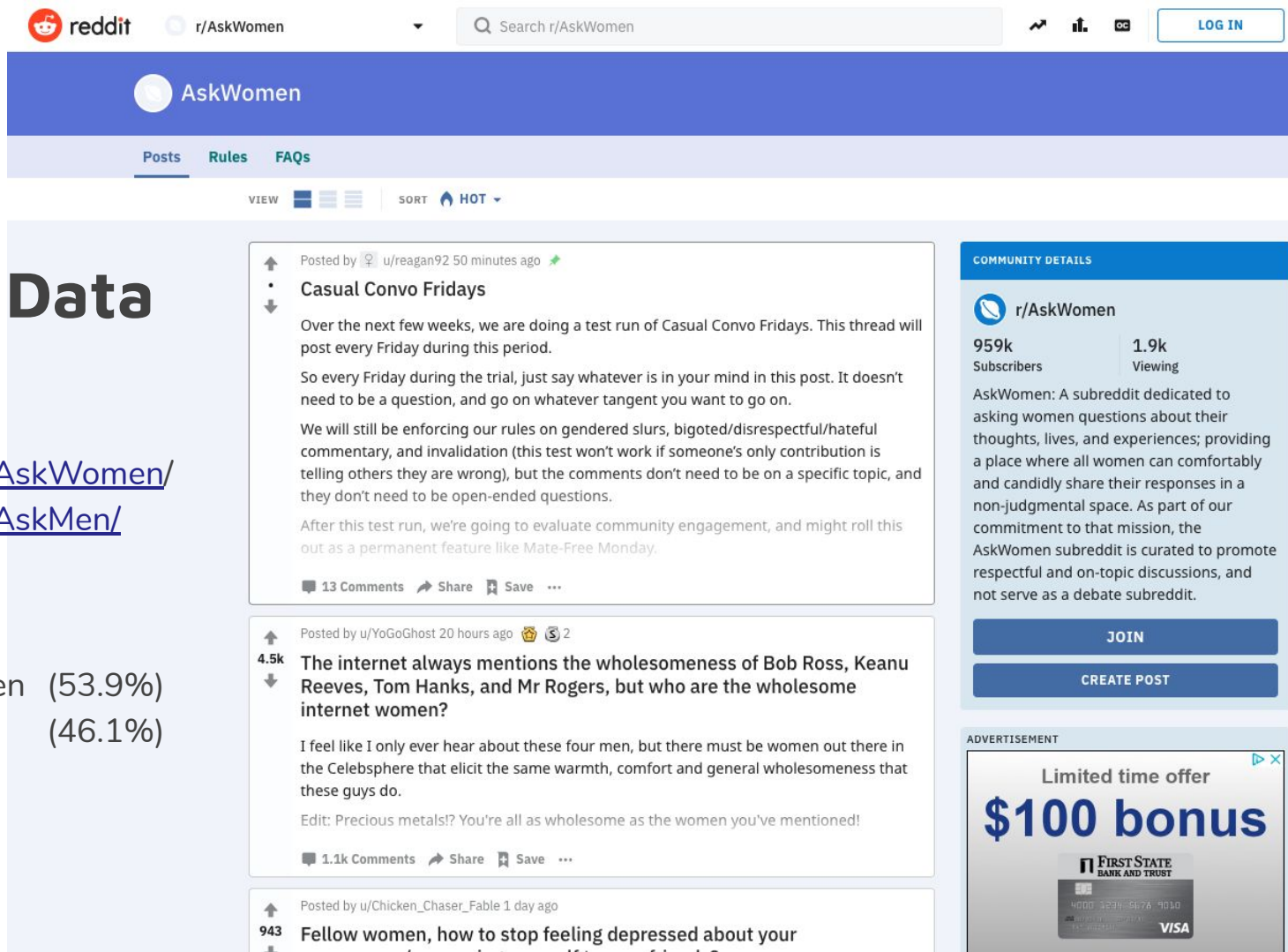
# About the Data

Collected questions from:

- <https://reddit.com/r/AskWomen/>
- <https://reddit.com/r/AskMen/>

Number of questions:

- 684 from AskWomen (53.9%)
- 585 from AskMen (46.1%)
  - 1269 total



The screenshot shows the Reddit interface for the r/AskWomen subreddit. At the top, the subreddit name "AskWomen" is displayed. Below it are tabs for "Posts", "Rules", and "FAQs". The "Posts" tab is active, showing a list of posts. The first post is titled "Casual Convo Fridays" by user u/reagan92, posted 50 minutes ago. It has 13 comments and is marked as a "Hot" post. The post content discusses a test run of "Casual Convo Fridays" where users can share thoughts and experiences. The second post is titled "The internet always mentions the wholesomeness of Bob Ross, Keanu Reeves, Tom Hanks, and Mr Rogers, but who are the wholesome internet women?" by user u/YoGoGhost, posted 20 hours ago. It has 4.5k upvotes, 1.1k comments, and is marked as a "Hot" post. The post content discusses the lack of representation of wholesome women in popular media. The third post is partially visible, titled "Fellow women, how to stop feeling depressed about your". On the right side of the page, the "COMMUNITY DETAILS" section shows that r/AskWomen has 959k subscribers and 1.9k members viewing. It also includes a "JOIN" button and a "CREATE POST" button. At the bottom right, there is an advertisement for First State Bank and Trust, offering a \$100 bonus.

reddit r/AskWomen Search r/AskWomen

AskWomen

Posts Rules FAQs

VIEW SORT HOT

Posted by u/reagan92 50 minutes ago

**Casual Convo Fridays**

Over the next few weeks, we are doing a test run of Casual Convo Fridays. This thread will post every Friday during this period.

So every Friday during the trial, just say whatever is in your mind in this post. It doesn't need to be a question, and go on whatever tangent you want to go on.

We will still be enforcing our rules on gendered slurs, bigoted/disrespectful/hateful commentary, and invalidation (this test won't work if someone's only contribution is telling others they are wrong), but the comments don't need to be on a specific topic, and they don't need to be open-ended questions.

After this test run, we're going to evaluate community engagement, and might roll this out as a permanent feature like Mate-Free Monday.

13 Comments Share Save

Posted by u/YoGoGhost 20 hours ago 2

**4.5k The internet always mentions the wholesomeness of Bob Ross, Keanu Reeves, Tom Hanks, and Mr Rogers, but who are the wholesome internet women?**

I feel like I only ever hear about these four men, but there must be women out there in the Celebsphere that elicit the same warmth, comfort and general wholesomeness that these guys do.

Edit: Precious metals!? You're all as wholesome as the women you've mentioned!

1.1k Comments Share Save

Posted by u/Chicken\_Chaser\_Fable 1 day ago

**943 Fellow women, how to stop feeling depressed about your**

COMMUNITY DETAILS

r/AskWomen

959k Subscribers 1.9k Viewing

AskWomen: A subreddit dedicated to asking women questions about their thoughts, lives, and experiences; providing a place where all women can comfortably and candidly share their responses in a non-judgmental space. As part of our commitment to that mission, the AskWomen subreddit is curated to promote respectful and on-topic discussions, and not serve as a debate subreddit.

JOIN

CREATE POST

ADVERTISEMENT

Limited time offer

**\$100 bonus**

FIRST STATE BANK AND TRUST

4000 3234 5476 9010

VISA



## Sample Questions with \*Signal\*

"What weird food craving did you have while pregnant? And what kind of food couldn't you stand anymore?"

- AskWomen

'What makes a strong woman in your opinion? As in how can you tell she is strong?'

- AskMen



# Early Process

## Cleaning

- Removed undesired special characters
- Replaced “SO” with “sigoth” (significant other)
- Made all letters lowercase

## Preprocessing

- Removed “stop words”
- Applied PorterStemmer
- Used vectorizers:
  - CountVectorizer
  - TFIDF

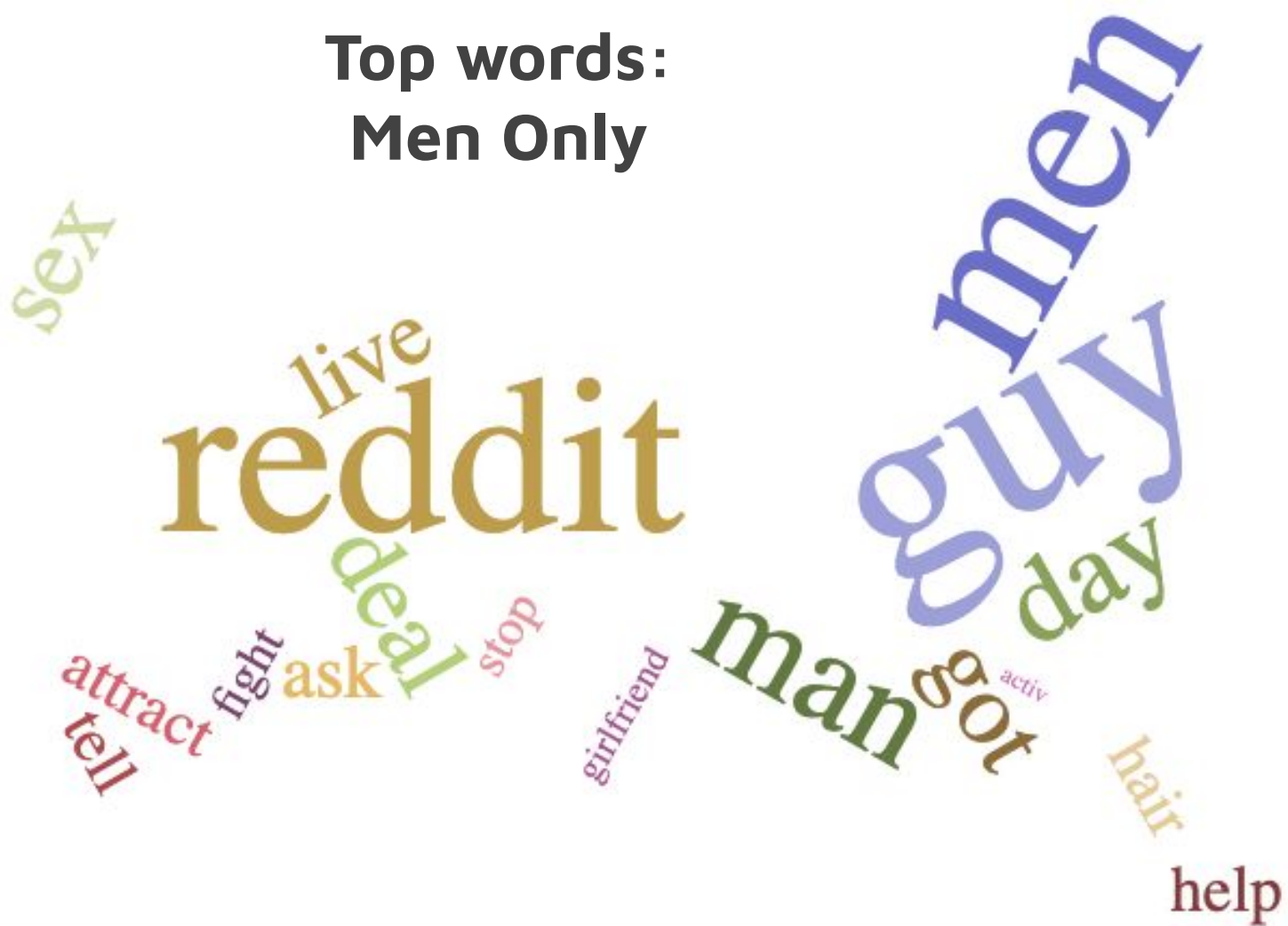
A word cloud featuring various terms related to relationships and social media. The words are arranged in a circular pattern, with some words being larger and more prominent than others. The colors of the words vary, including shades of blue, green, yellow, orange, red, and purple. The words include:

- Women
- like
- relationship
- date
- friend
- love
- time
- best
- way
- think
- make
- life
- thing
- people
- want
- feel
- someth
- use
- wear
- start
- partner
- love
- reddit
- time
- best
- way
- think
- make
- life
- thing
- people
- want
- feel
- someth
- use
- wear
- start
- partner
- love
- reddit

# Top words: Women Only



# Top words: Men Only







## Modeling

Logistic Regression  
Naive Bayes

with CountVectorizer and TFIDF

We used GridSearchCV to  
find best hyperparameters.

Best model:

Multinomial Naive Bayes

TFIDF & alpha = ~2.2

Best accuracy:     **66.4%**



# Conclusions / Limitations

- There signal coming from our predictive features is significant.
- We failed to achieve accuracy of greater than 70%.
- Our model may perform better with more training data.
- We recommend modeling with Multinomial Naive Bayes
  - TFIDF vectorizer.
- More data → Better model
- Tough to discern short questions with few / no distinguishing keywords
  - Both ask about “men” / “women”