



r/reveal

uncover new Reddit
communities

Wes Barnett



Available in the
Chrome Web Store



330M+

active users

138K+

**active
communities**

often with prefix "r/"



source: redditinc.com



- **links**
- **pictures**
- **videos**
- **text**

138K+

**active
communities**

often with prefix "r/"



source: redditinc.com



138K+

**active
communities**

often with prefix "r/"



source: redditinc.com

How can I submit content to communities if I'm not aware of them?

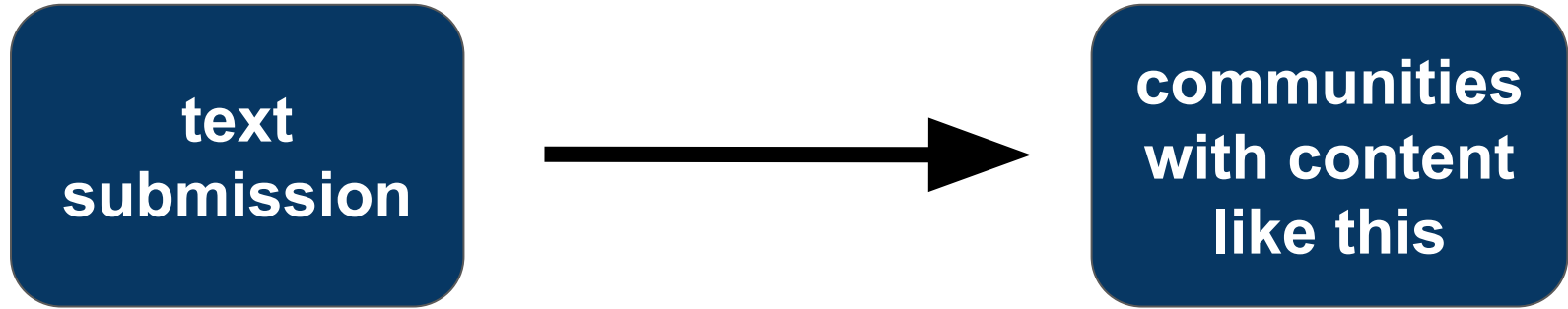
- Already aware of “popular” or “trending” communities
- Content producers keep communities active

750K+

**dead
communities**



How can I submit content to communities if I'm not aware of them?



The data: Reddit submissions

- 37 GB for just May 2018



The data: Reddit submissions

- 37 GB for just May 2018
- Only select a subset



The data: Reddit submissions

- 37 GB for just May 2018
- Only select a subset
- Train multiple models (3)
 - Split up models on popularity (high, med, low)
 - ~550-600 classes per model
 - 200k-700k samples per model



Multiclass classification model

- Support Vector Machine
 - Stochastic Gradient Descent



Multiclass classification model

- Support Vector Machine
 - Stochastic Gradient Descent
- Hashing vectorizer
 - 2^{18} features
 - Word frequencies in title and text



Multiclass classification model

- Support Vector Machine
 - Stochastic Gradient Descent
- Hashing vectorizer
 - 2^{18} features
 - Word frequencies in title and text
- Regularization parameter
 - Selected using validation set performance



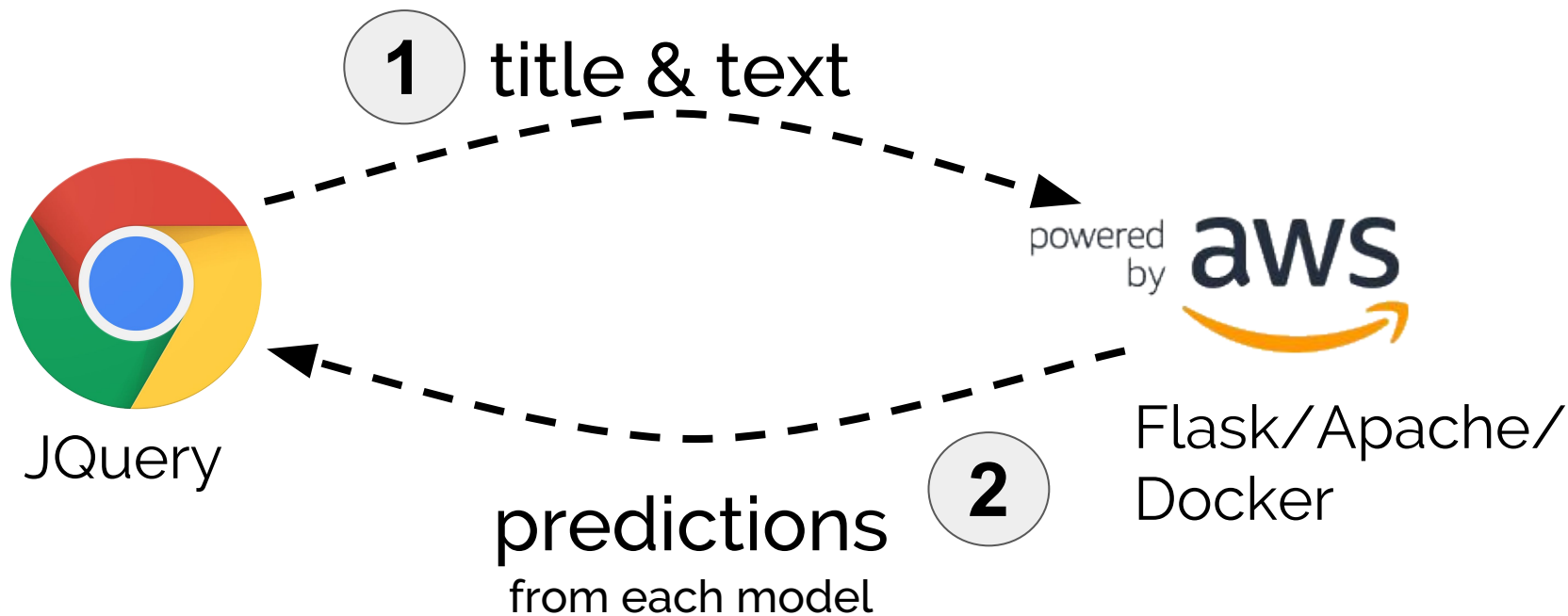
Model evaluation: Test set

Model	# classes	Mean accuracy
0	544	0.52
1	592	0.63
2	590	0.53

>250x
more accurate than
random chance
(< 0.002)



Model deployment



Demo

Demo



Wes Barnett



**PhD, Chemical and
Biomolecular Engineering**



**Postdoctoral Research
Scientist**









reddit

SUBMIT

submit to reddit

link

text

You are submitting a text-based post. Speak your mind. A title is required, but expanding further in the text field is not. Beginning your title with "vote up if" is violation of intergalactic law.

*title

text (optional)

[content policy](#) [formatting help](#)

communities with content like this

start typing above!





reddit

SUBMIT

submit to reddit

link

text

You are submitting a text-based post. Speak your mind. A title is required, but expanding further in the text field is not. Beginning your title with "vote up if" is violation of intergalactic law.

*title

Science fiction recommendations?

text (optional)

[content policy](#) [formatting help](#)

communities with content like this

[suggestmeabook](#) · [classicalmusic](#) · [writing](#) · [booksuggestions](#)
· [datascience](#) · [Stellaris](#) · [Bandnames](#) · [ApplyingToCollege](#) ·
[AskScienceDiscussion](#)





reddit

SUBMIT

submit to reddit

link

text

You are submitting a text-based post. Speak your mind. A title is required, but expanding further in the text field is not. Beginning your title with "vote up if" is violation of intergalactic law.

*title

Science fiction recommendations?

text (optional)

I just finished reading the Expanse series. Can some

[content policy](#) [formatting help](#)

communities with content like this loading...

[suggestmeabook](#) · [scifi](#) · [books](#) · [booksuggestions](#) ·
[dragonage](#) · [occult](#) · [TheExpanse](#)





reddit

SUBMIT

submit to reddit

link

text

You are submitting a text-based post. Speak your mind. A title is required, but expanding further in the text field is not. Beginning your title with "vote up if" is violation of intergalactic law.

*title

Science fiction recommendations?

text (optional)

I just finished reading the Expanse series. Can someone give me some recommendations

[content policy](#) [formatting help](#)

communities with content like this loading...

[suggestmeabook](#) · [booksuggestions](#) · [TheExpanse](#)





reddit

SUBMIT

submit to reddit

link

text

You are submitting a text-based post. Speak your mind. A title is required, but expanding further in the text field is not. Beginning your title with "vote up if" is violation of intergalactic law.

*title

Science fiction recommendations?

text (optional)

I just finished reading the Expanse series. Can someone give me some recommendations for another space opera?

[content policy](#) [formatting help](#)

communities with content like this

[suggestmeabook](#) · [scifi](#) · [booksuggestions](#) · [TheExpanse](#)



Wes Barnett



**PhD, Chemical and
Biomolecular Engineering**



**Postdoctoral Research
Scientist**





Demo 2

Demo 2





reddit

SUBMIT

submit to reddit

link

text

You are submitting a text-based post. Speak your mind. A title is required, but expanding further in the text field is not. Beginning your title with "vote up if" is violation of intergalactic law.

*title

text (optional)

[content policy](#) [formatting help](#)

communities with content like this

start typing above!





PERSONALFINANCE

comments



Investing

What do I need to watch out for when I exercise stock options? (self.personalfinance)

submitted 5 days ago * by [hammiec](#)

My options (ISOs) just started vesting for the publicly traded company I work for.

I joined the company before their IPO and so I have a strike price that is a fraction of the current price.

I also have 5 digits of student loans.

What do I need to watch out for to make sure I don't get screwed over by taxes or some other financial gotchas? I already know about the blackout periods.

TL;dr I have options that just vested and I want to exercise/sell. Also have student loans. What could screw me over?

6 comments share save hide [give gold](#) report crosspost

other communities with content like this

[investing](#) · [StudentLoans](#) · [options](#)

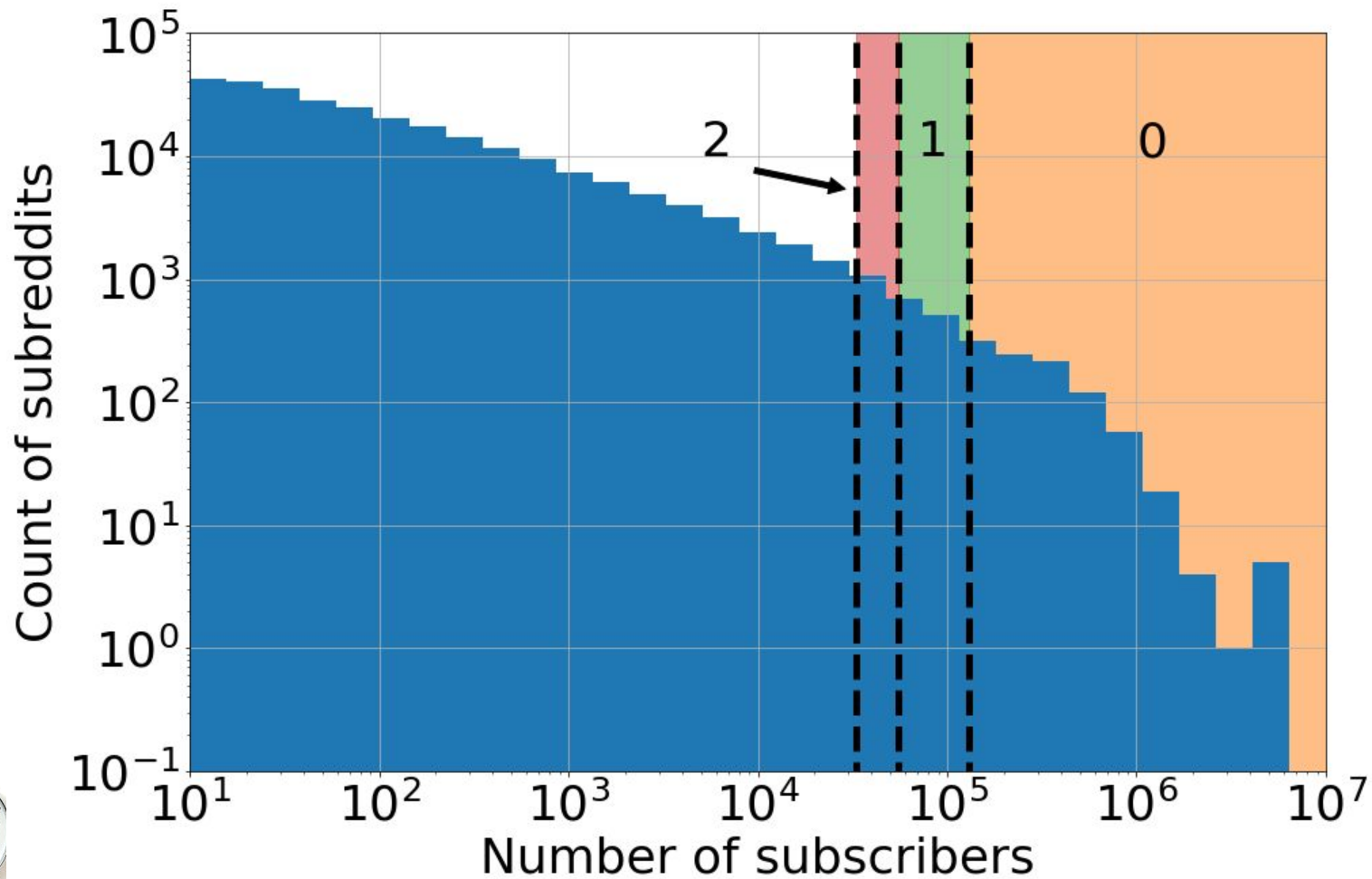




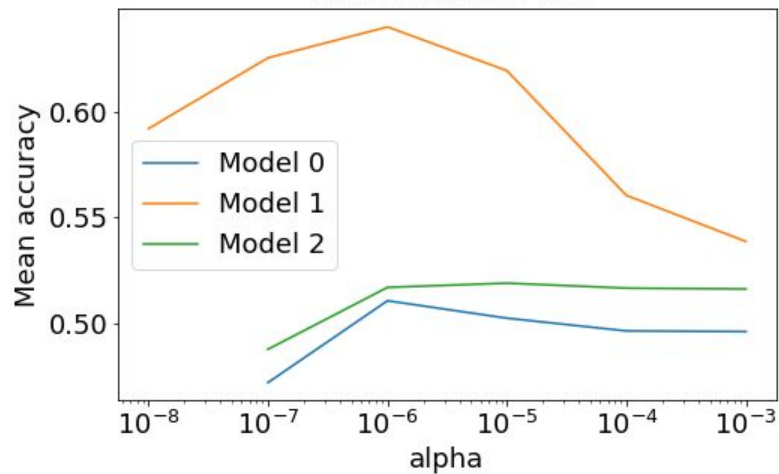
Future work

- Remove low quality submissions before training
- Use a more complex featurization
- Use content of web links in training
- Use more samples in training
- Train models on more communities
- Investigate how communities relate to one another in subject matter

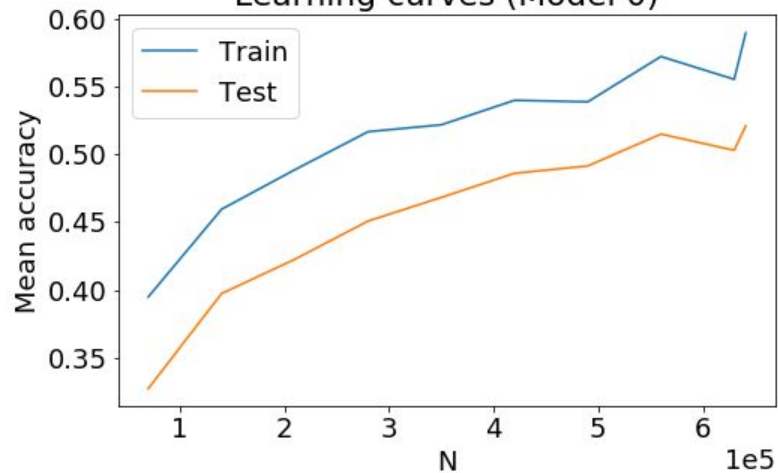




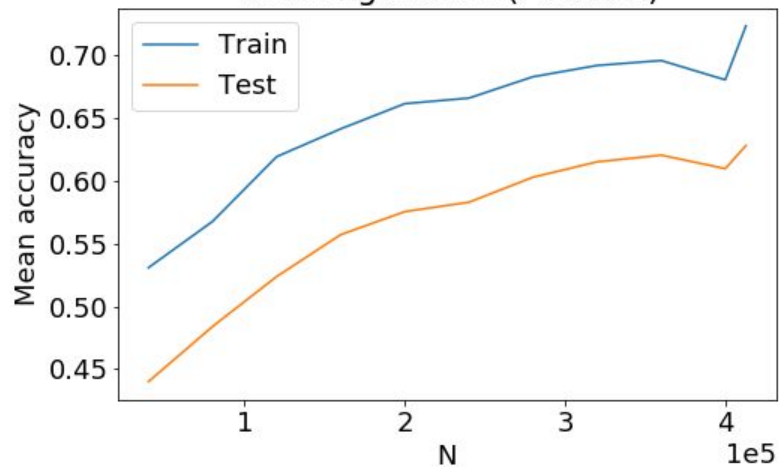
Validation curves



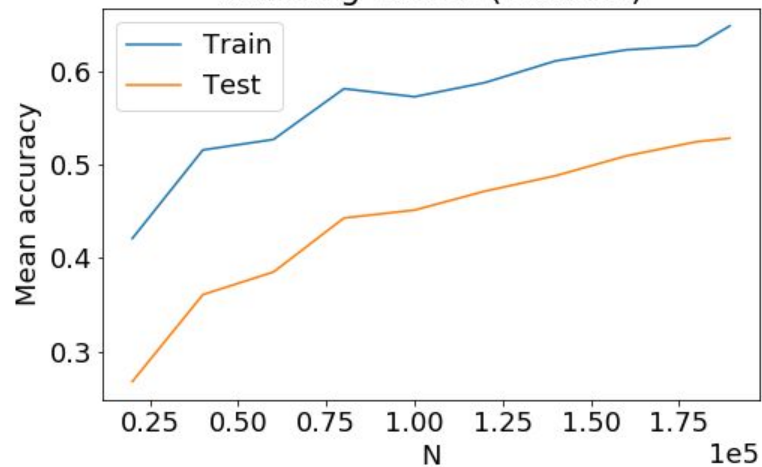
Learning curves (Model 0)



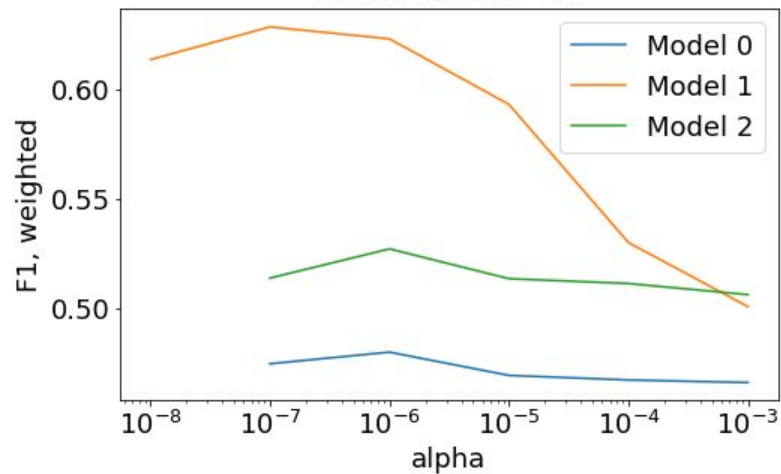
Learning curves (Model 1)



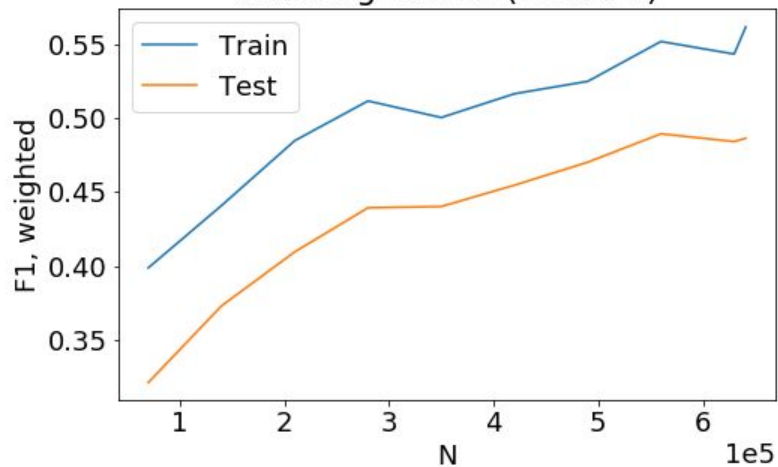
Learning curves (Model 2)



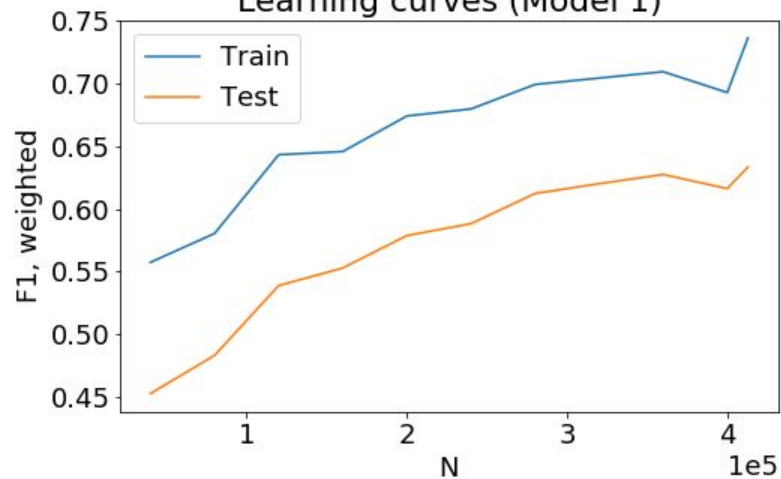
Validation curves



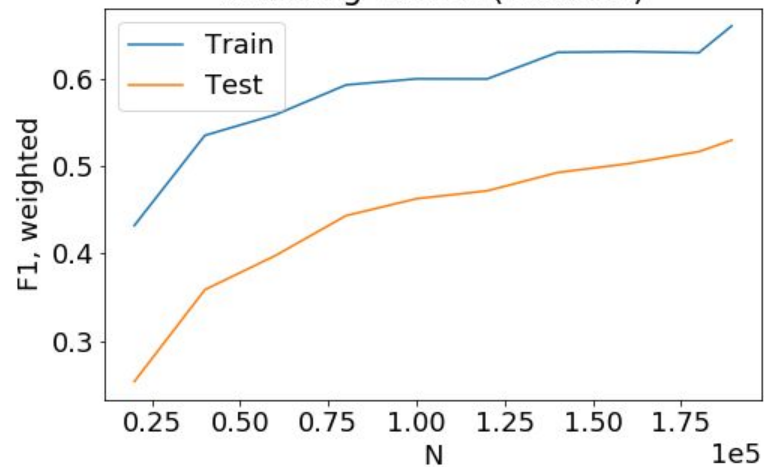
Learning curves (Model 0)



Learning curves (Model 1)



Learning curves (Model 2)



Model breakdowns

Model	Subscribers lower limit	Subscribers upper limit	Submissions (classes)	Subreddits (classes)
0	130,000		711,014	544
1	55,000	130,000	452,885	592
2	33,000	55,000	209,340	590



Average: weighted

Model	Precision	Recall	F1
0	0.54	0.52	0.49
1	0.71	0.63	0.63
2	0.63	0.53	0.53



Why not just use search?

- Search on Reddit is terrible
- Have to leave page where submitting content
- Users have to make effort to find new communities

