

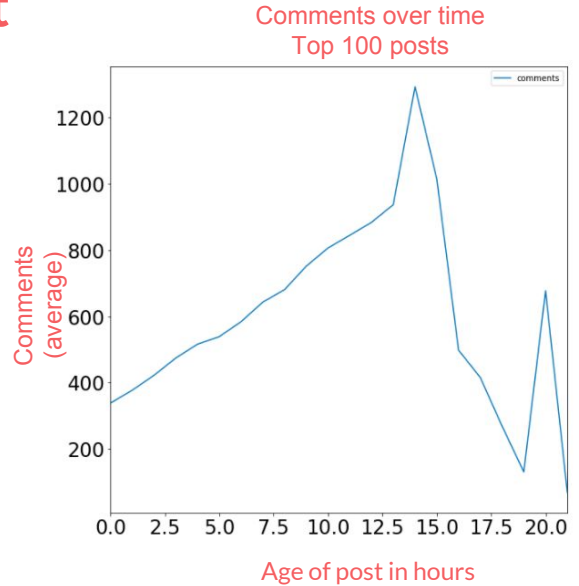


How To: Viral Reddit Posts

Reddit Post Engagement

What drives engagement?

- Time since post
- Subreddit/Content
- Author Popularity

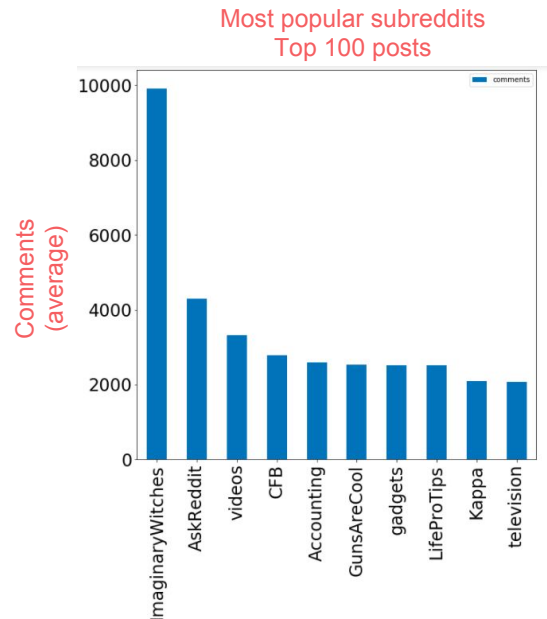


We know 'viral' is a function of time since post, but what else affects if a post goes viral? Reddit posts in the top 100, on the whole peak at 15 hours. Keep in mind this is an aggregate graph.

Reddit Post Engagement

What drives engagement?

- Time since post
- **Subreddit/Content**
- Author Popularity



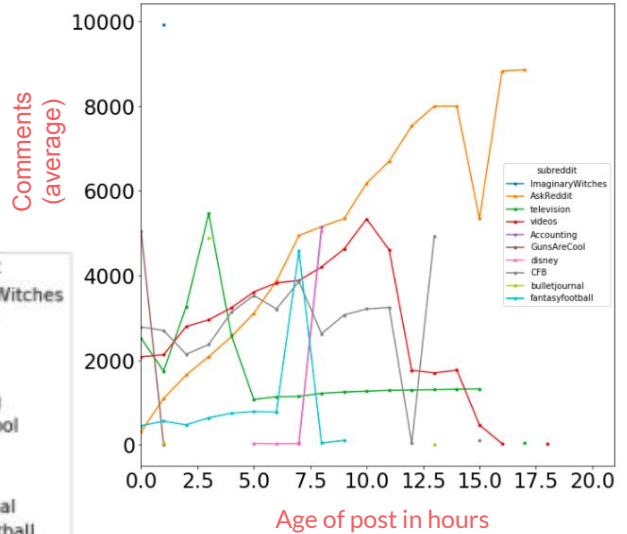
Reddit Post Engagement

What drives engagement?

- Time since post
- **Subreddit/Content**
- Author Popularity



Subreddits with the Maximum Comments Over Time



Ask Reddit gains popularity over time

Fantasy Football spike

Video peaks at hour 10

College football talk

Imaginary Witches - well, we were collecting data over halloween weekend, after all.

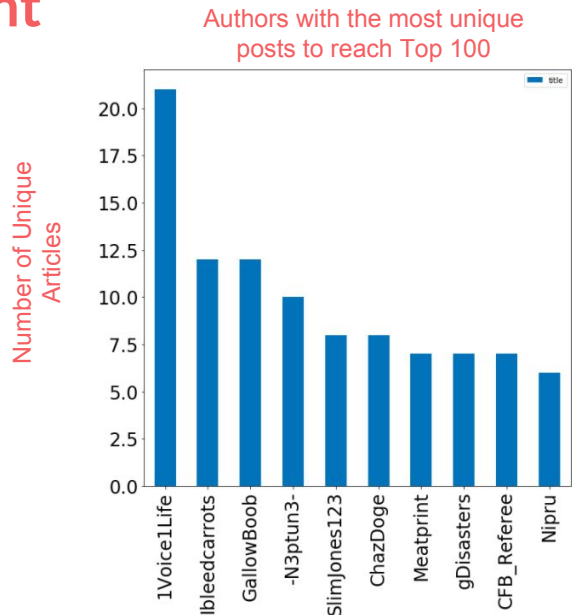
It's peak performance is a function of the immediacy of the topic.

Reddit Post Engagement

What drives engagement?

- Time since post
- Subreddit/Content
- **Author Popularity**

95% of authors only get 1 post in the top 100

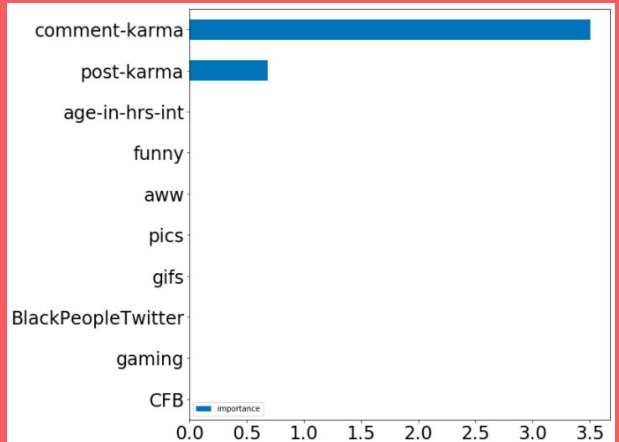


We'll be using Author's post-karma and comment karma as proxies for their involvement and popularity in the reddit community. After all, we only sampled for a weekend, and just because they didn't post 20 articles this weekend doesn't mean they don't have high influence in their circles.

Logistic Regression

With 3 folds for cross validation:

85.2% accurate
0.0 margin of error



We start with Logistic Regression because it's a simple and intuitive classifier. We use a linear model of variables to predict whether a post goes viral or not.

- Comment karma
- Post karma
- Subreddit
- Age (in hours)

To predict: Comments

Here we see subreddit dominates and it performs with admirable accuracy. However linear regression makes some strong assumptions about the nature of the data and its relationship with the outcome of interest. It assumes

- that the variables are normally distributed and
- that variables are linearly related to the outcome.
- It assumes in a more general sense that all variables that you put into it matter.

Feature Importance: the magnitude of the coefficient in the linear model scaled by the standard deviation of the data associated with that variable. (Scaled for magnitude). Something to take into consideration here: You must take the absolute value because the way the magnitude of the coefficient works is it can cancel itself out.

```
{('importance':abs(np.std(X,0)*lg.coef_[0])}, index=X.columns)
```

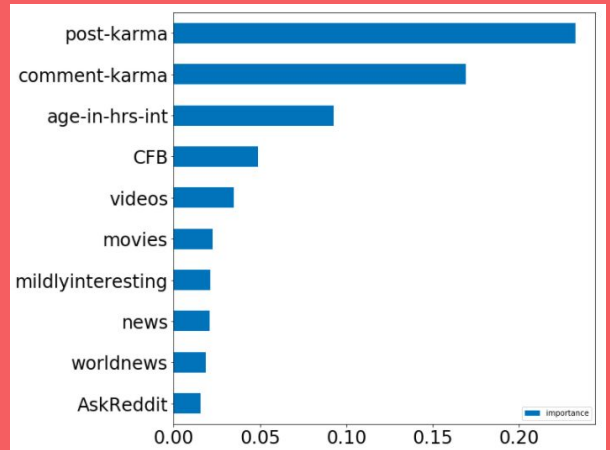
Let's use a model with fewer assumptions.

Random Forest: Feature Importance

It matters who posts.

With 3 folds for cross
validation:

97.9% accurate
0.0 margin of error



Random Forest is really good because

- it can handle a lot of variables,
- it doesn't presume importance of variables
- Gives you a way to find out which ones are important
- Doesn't assume linear relationships

Feature importances are what they're called: it measures how important a given variable is to classifying correctly.

If I keep all the variables the same and just messed with the values of one variable, then the random messing will influence the outcome a lot or a little depending how influential that variable is.

Executive Summary

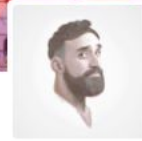
Reddit is a community

Popular posts have the following in common:

- Timeliness
- Active and well-regarded authors
- Commitment

Not as influential:

- Words in title



u/GallowBoob
@TheGallowBoob

18,372,511 Karma

I post things on reddit. Tinkered here a bit, ended up working in media, now it's hard to let go of an old habit. One day i'll unplug. One day... Follow me on Twitter & Instagram & Facebook : @TheGallowBoob

FOLLOW

Following this user will show all the posts they make to their profile on your front page.

More Options

Timeliness

- Current or relevant events reign supreme
- We scraped over halloween weekend with words that occur in 80% or less of the titles and Halloween was one of our most popular words in the titles; followed by costumes
- "Imaginary Witches" was one of our most sensational subreddits.

Active and well-regarded authors

- There are consistent viral posters and in a span of 3-ish days, one had 20 posts make it into the top 100 reddits DC page.
- In all our models

Commitment - Reddit is a community of volunteer moderators and content contributors. There are folks like GallowBoob who make it the tagline in their twitter profiles. The brand commitment goes a long way in terms of viral-ability.

Not as influential; outside of timeliness; is the words in the title. This left our model with the same accuracy score but higher variance.

Debut Idea:

TurkeyDay Stats

Option A: Football and College Football in particular have large, consistent followings

Option B: Post what data science can reveal about Thanksgiving

Option C: Post a cute picture

Build a brand

- Become a member ASAP
- Post and comment consistently
- Post about timely things
- Plan to 'go viral' perhaps around a holiday or major TV or sporting event.

Consider if 'viral' on Reddit is a part of 538's brand.

There are infinitely many subreddits and the point is to find what fits you rather than what fits everyone. Sometimes sticking to your brand is what will result in the best experience from Reddit.

Game Threads

User: CFB_Referee



[Game Thread] Penn State @ Ohio State Second Half (3:30PM ET) (self.CFB)

submitted 5 days ago by CFB_Referee [M]



Details

--!!-- Time | 3:30 PM Eastern - 2:30 PM Central - 1:30 PM Mountain - 12:30 PM Pacific Location | Ohio Stadium Watch | TV Odds | Spread: -6
- Over/Under N/A ** Resources** | Subreddit - Wiki ** Resources** | Subreddit - Wiki

- Keep trash talk **civil**
- **Get Team Flair!**
- **Get /r/CFB Logo Stickers here!**
- Game threads automatically sort comments by "**new**"
- **IRC chat** available at **#redditcfb**
- **See a problem?** Please click "**report**" under the comment.

LET'S TALK FOOTBALL!

[13318 Comments](#) [Share](#) [Report](#)

[Something with a nazi and
racism]

Titled: "Happy Halloween"

“Party On, Single Life”

User: oliverclothessoff



YouTube user demonstrating how Facebook listens to conversations to serve ads

User: AnimalCrust

Subreddit: Imaginary Witches

