

Random Acts of Pizza

Pizza Hackers: Hung Pham, Tam Nguyen



Random Acts of Pizza

https://www.reddit.com/r/Random_Acts_Of_Pizza/



🍷 [Request] Barely can afford rent this month, really cravin' a pizza pie

submitted 14 hours ago * by betinob

Please help, I ate all my ramen. :(

[comment](#) [share](#)



🍷 [Request] 'Bout to lose it from exams [VA]

submitted 22 hours ago by newstarttn



awesome

Exams are next week and I'm hitting the books pretty hard. With Christmas coming up, I don't have any extra cash and I'm just having a hard time. Pizza would be appreciated but some nice words to get through exams would also help. I live in Virginia by the way.

[comment](#) [share](#)

Datasets

kaggle.com/c/random-acts-of-pizza

- December 8, 2010 - September 29, 2013
- Total: 4040 requests
- Received: 994 requests
- Not received: 3046 requests
- JSON

Important Attributes

Request text

request_title

request_text_edit_aware

Requester information

requester_account_age_in_days_at_request

requester_number_of_comments_at_request

requester_number_of_posts_at_request

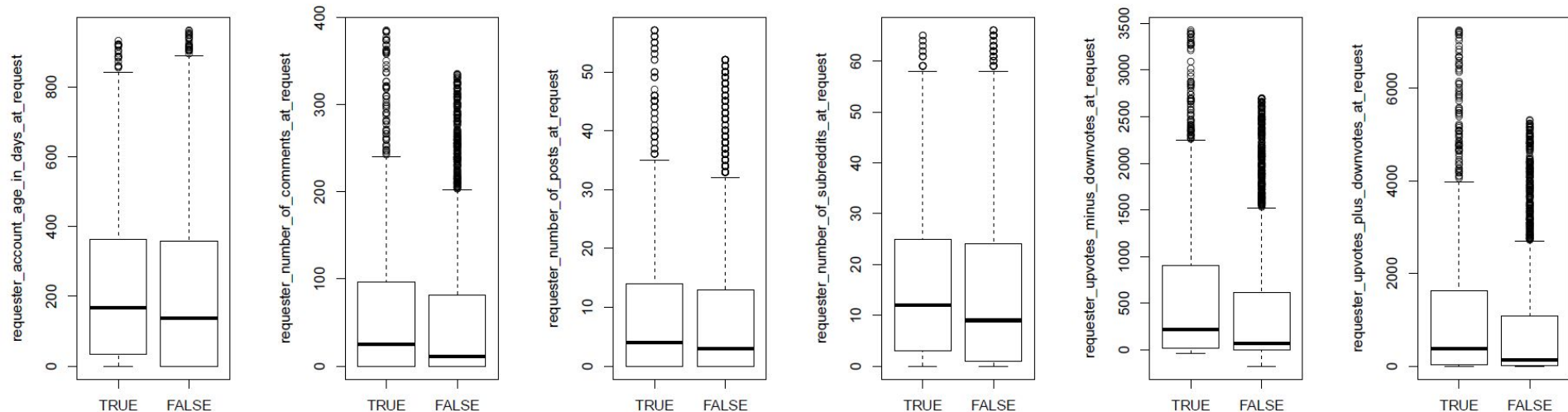
requester_subreddits_at_request

requester_upvotes_minus_downvotes_at_request

requester_upvotes_plus_downvotes_at_request

Requester Information

Predictive power



Boxplots of requester information attributes based on the success of requests

F-score

0%

- Logistic Regression
- 10-fold cross validation

		Actual	
		TRUE	FALSE
Predict	TRUE	0	0
	FALSE	92	312

Confusion matrix for a fold in cross validation

Textual Content

Preprocessing

- Tokenization
- Stopwords
- **12930** distinct words

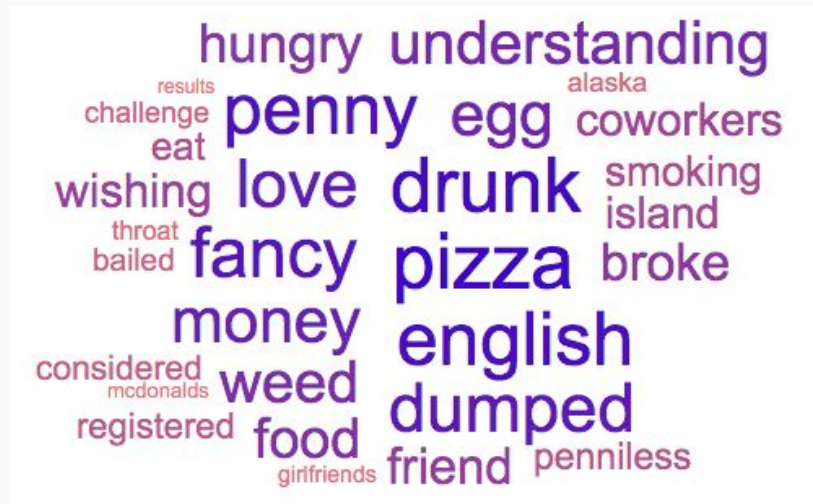
Textual Content

Contrast Analysis

$$\phi = \frac{p}{n} (p - n)$$



Top positive words



Top negative words

F-score

31%

- Bag of words
- Logistic Regression
- 10-fold cross validation

		Actual	
		TRUE	FALSE
Predict	TRUE	27	79
	FALSE	33	265

Confusion matrix for a fold in cross validation

Word2Vec clustering

Classification

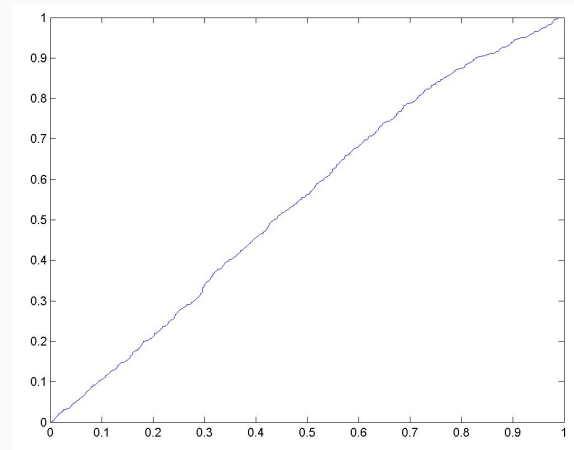
- Word2Vec
- K-means clustering
- Bag of cluster
- Support Vector Machine (SVM) classification

F-score

39.5%

- Bag of cluster
- SVM
- 10-fold cross validation

Receiver
operating
characteristic
AUC:
0.546137



Feature weight	5.182634	-4.404034	-3.256066	2.486054	-2.445433	2.181824
Top words per cluster	quebec	lentils	tinypic	ofallon	resources	graduates
	athens	buttered	ip	boise	government	seniors
	wilmington	packages	yelp	providence	disposable	physics
	antonio	raman	soundcloud	yup	walls	lads
	tempe	noodle	wordpress	lakewood	stamp	intern
	francisco	vegetables	gmail	fresno	agencies	nursing
	valley	salad	youtube	chapel	pantries	immigrant
	bellingham	boxed	png	ridge	snacks	apps
	dakota	pbj	tumblr	edmonton	toppings	veteran
	denton	dried	php	petersburg	electronic	sophomore

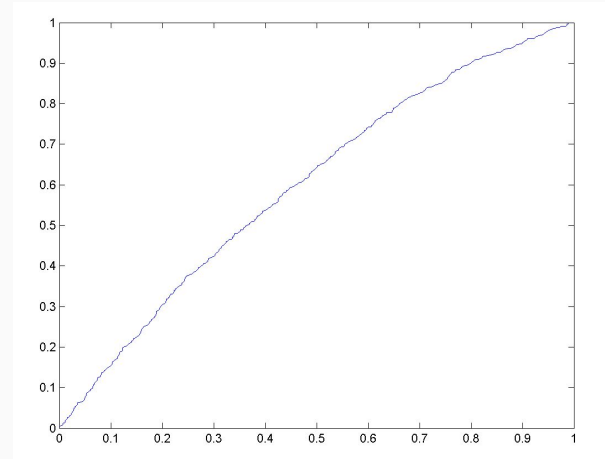
Cluster with high feature weight and their top words.

F-score

41%

- Average vector
- SVM
- 10-fold cross validation

Receiver
operating
characteristic
AUC:
0.599242



Finally

How do we get a free pizza?

