Random Acts of Pizza

W207: Spring '17 Final Project
Erika Lawrence, Geoff Stirling, Jen Jen Chen, Leslie Teo

Project Overview

- Reddit community: Random Acts of Pizza

- Can ML help us predict who gets a free pizza?

- Data set of 5,671 text-based requests for cheesy goodness

Photo credit: thr.com, Hollywood Reporter, 2017

Solution Process & Tools

- The Data
  - Text
    - Request
    - Title
  - Metadata
    - Account: reddits, history
    - Post: age, votes

request_title
Request Colorado Springs Help Us Please
California, No cash and I could use ...
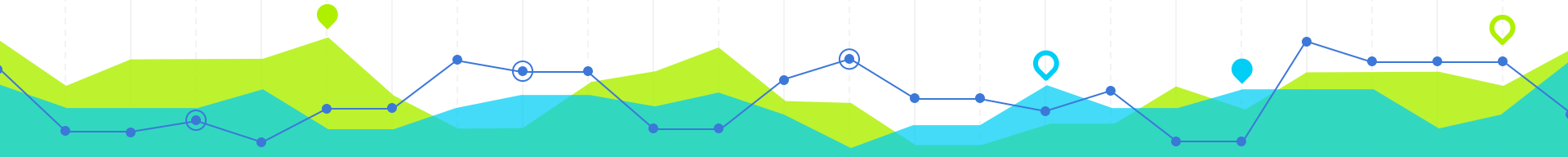Hungry couple in Dundee, Scotland wo...

requester_subreddits_at_request
[FoodstuffsAllAround, IAmA, RandomActsOfCookie...
[AskReddit, Guitar, Jazz, Music, NSFW_GIF, Psy...
[Albany, AskReddit, Brooklyn, Favors, ImGoingT...
[AskReddit, Favors, Jazz, boardgames, discgolf...

Solution Process & Tools

- Data cleaning/pre-processing
  - Text
    - Normalize/Denormalize
    - Vectorize
    - Clean
    - Lemmatize ('reprocessing' -> 'process')

Solution Process & Tools

- Metadata features:
    - Time/date of request
    - Up/down votes
- Sentiment Analysis
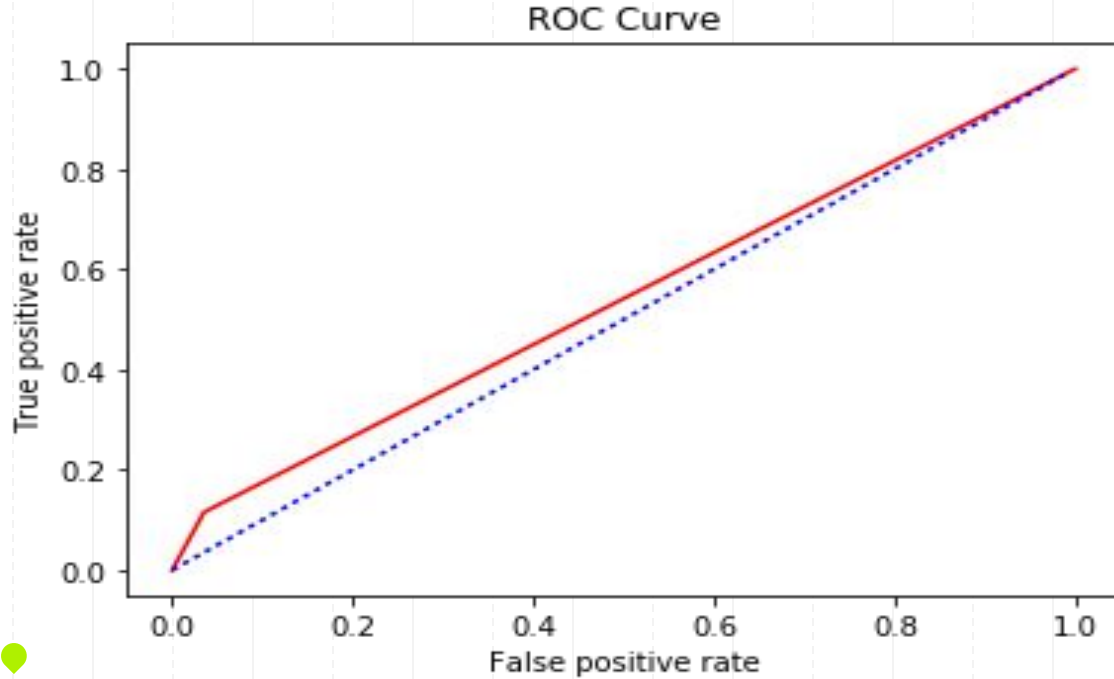   - vaderSentiment

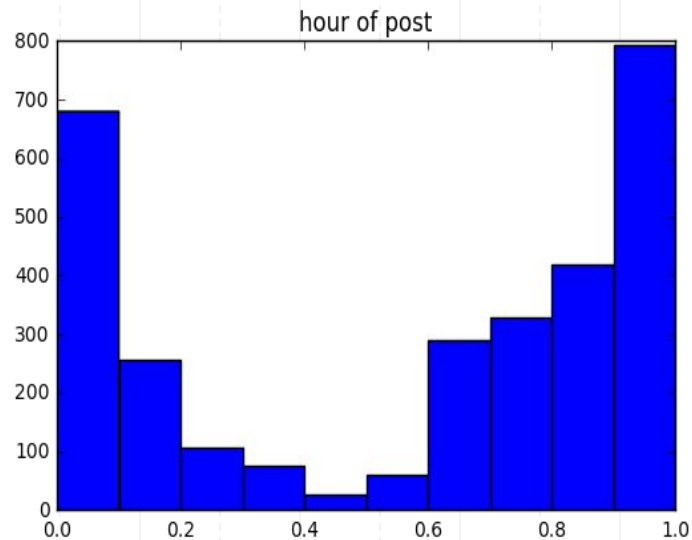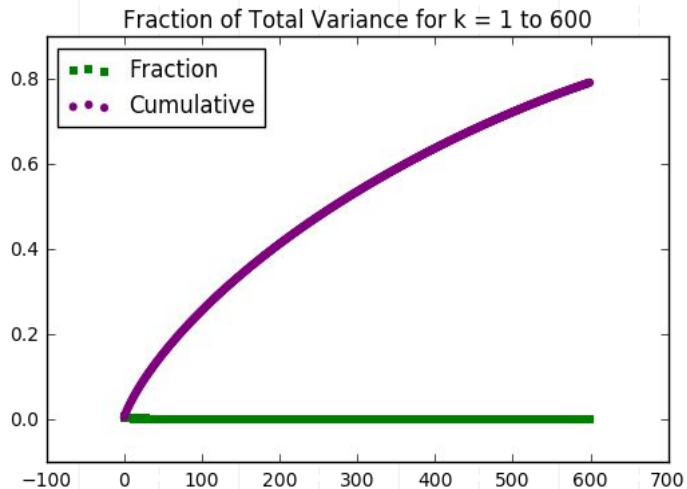# Solution Process & Tools

- Methods/Models:
  - Vectorize: bi/tri-gram, tf-idf
  - PCA
  - Logistic regression
  - Bernoulli NB
  - Random Forest
  - Ensemble
  - (+ SVM, Ada Forest, Gradient Boosting, and Decision Trees)

Results of our work: baseline 74%

ROC Curve

Results of our work: features

Results of our work: final

```
Random Forest with n=100
-------------------------------------------------------------------

Accuracy:  0.837
Area under curve of ROC:  0.721465696466
Classification report:

              precision    recall  f1-score   support

     False       0.84        0.96     0.90        740
      True       0.82        0.48     0.61        260

avg / total       0.83       0.84     0.82       1000
```
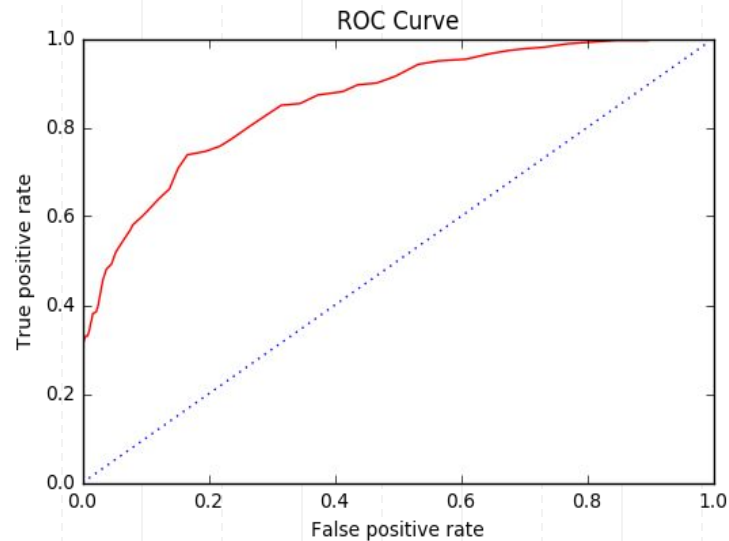

ROC Curve

What we learned for next time

- Expand time analysis

- Vader expansion

- Features are key. Start early.

- Never give up. Never surrender.

THANKS!

Any questions?