Subreddit Predictions

Objectives

- Goal
- Process
- Results
- Plans

Strategy

- Premium: $961,996.98 in 2016
- Engage
  - Ad-Free
  - Access to exclusive subreddit
- Match words to posts

Ref: www.feedough.com

Model

- Travel vs. Shoestring
- Baseline: 50.35 % of negative class
- TfidfVectorizer
  - Max_features: 300
- Logistic Regression
  - L1
  - Training: 75.57 %
  - Testing: 73.11 %

Features

- Random Forest Classifier
  - Training: 76.76%
  - Testing: 70.23 %
  - Cross_val: 70.23%

Weakest

currently	0.000002
australia	0.000003
nature	0.000018
island	0.000019
transport	0.000020
canada	0.000026
short	0.000027
spending	0.000030

Strongest

cheap	0.128739
budget	0.057682
cheapest	0.048647
itinerary	0.028247
cheaper	0.026075
time	0.025341
save	0.023941
day	0.022778

What next?

- Tune parameters in Random Forest Classifier
- VotingClassifier
- Other Models