AMEE AMIN | GENERAL ASSEMBLY DATA SCIENCE

PREDICTING HELPFULNESS OF AMAZON CUSTOMER REVIEWS

BACKGROUND

- Amazon displays top reviews to help customers with their buying decision process. Top reviews are from Top reviewers.

- A reviewer's rank is determined by the overall helpfulness of all their reviews, factoring in the number of reviews they have written.

- Review helpfulness plays an important part in determining rank. Writing thousands of reviews that customers don't find helpful won't move a reviewer up in the standings.

EXAMPLE PRODUCT – BUBBA'S PALEO GRANOLA

Top Customer Reviews

★★★★★ **Vanilla and Uber Chocolate. The taste is fresh and really good**

By L.A.G. on July 19, 2016

Flavor Name: Vanilla | Size: 1 | **Verified Purchase**

Tried cinn-ful Apple, Vanilla and Uber Chocolate. The taste is fresh and really good. I have to be careful not to eat the whole bag! Loved all 3 favors, cannot pick a favorite. Great for a snack or when craving something sweet. Did not expect to love the taste. Just reordered 6 bags of each. Happy that it is non gmo. Wish it was organic.

Comment | 3 people found this helpful. Was this review helpful to you? Yes No Report abuse

★★★★★ **especially something that was a little bit sweet. This is so delicious**

By SLD500 on April 5, 2016

Flavor Name: Uber Chocolate | Size: 1 | **Verified Purchase**

Arrived way ahead of schedule. I am new to the Paleo lifestyle and I missed the snacks...especially something that was a little bit sweet. This is so delicious!!! No guilt...just pure good food. I LOVE it!!! Thank you Bubba's Fine Foods!!!

Comment | 3 people found this helpful. Was this review helpful to you? Yes No Report abuse

PROBLEM

- When people don't vote, how should Amazon determine Helpfulness?
- New products are viewed less by customers, lower chance of review votes
- Problem: Build a model that predicts the helpfulness of a review

DATA

- Dataset consists of 568,454 fine food reviews from Amazon users left up to October 2012.
- Features include product id, user ratings, helpfulness score, timestamp, review summary, and review text.

	ProductId	UserId	ProfileName	HelpfulnessNumerator	HelpfulnessDenominator	Score	Time	Summary	Text
Id									
1	B001E4KFG0	A3SGXH7AUHU8GW	delmartian	1	1	5	1303862400	Good Quality Dog Food	I have bought several of the Vitality canned d...

HYPOTHESIS

- Reviews that mention whether the customer would buy the product again & score the product highly, will be more helpful to other customers.

METHOD

- Make Helpfulnesses a binary classification problem
- “Bag of words” model. Tf-idf vectorizer -> Logistic regression.
- K-means clustering. Select Top 10 words -> Logistic Regression

CLEANING THE DATA

- Punctuation
- Capitalization
- HTML tags

BINARIZE HELPFULNESS

- If (# of helpful votes) / (# of total votes) > 0.5
  - Review is helpful
- If (# of helpful votes) / (# of total votes) <= 0.5
  - Review is not helpful

EXPLORATORY ANALYSIS

- For reviews with > 10 votes
- If (# of helpful votes) / (# of total votes) > 0.5
  - Review is helpful
- If (# of helpful votes) / (# of total votes) <= 0.5
  - Review is not helpful

EXPLORATORY ANALYSIS

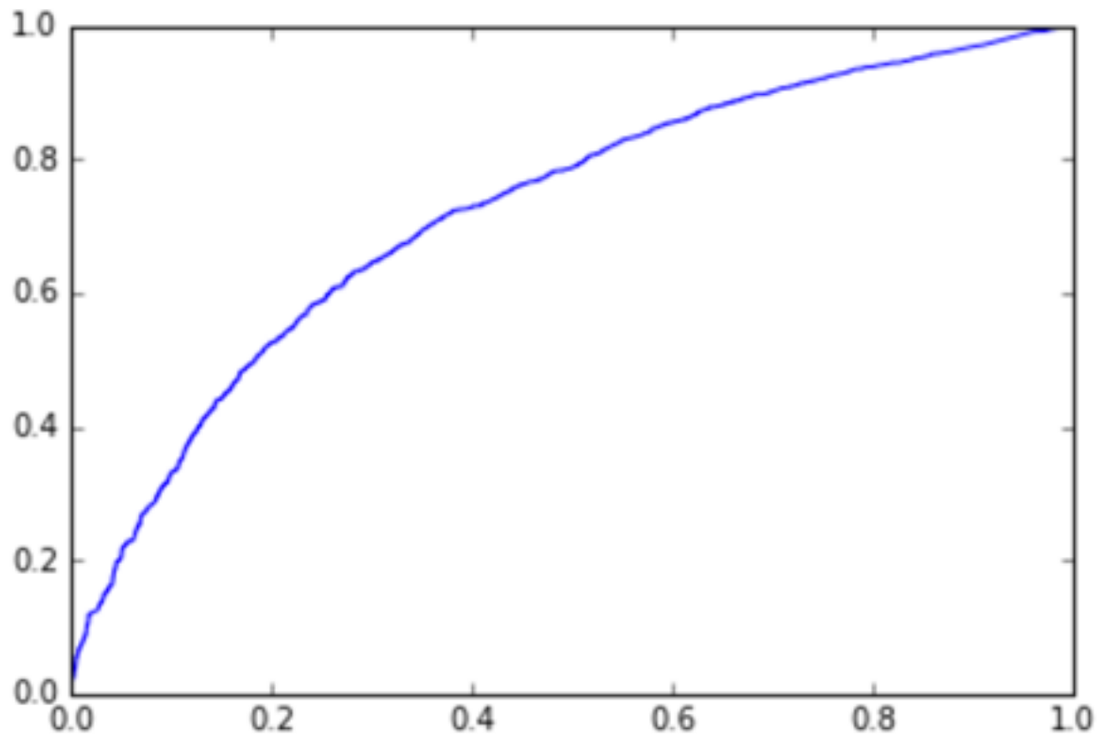
- Many more helpful reviews than not helpful, will bias the model
- Use stratified K-fold to compensate

	HelpfulnessNumerator	HelpfulnessDenominator	Score	Text
Helpful				
0	4100	4100	4100	4100
1	17363	17363	17363	17363

EXPLORATORY ANALYSIS

- Bag of words model: good AUC value, but difficult to interpret meaning from best coefficients

```
(u'eat', 1.1988757821801048),
(u'love', 1.327079240641496),
(u'used', 1.4182934021649025),
(u'time', 1.4314137846470354),
(u'amazon', 1.4488825074938934),
(u'try', 1.743612902002463 5),
(u'make', 1.8089264785709085),
(u'good', 1.9259597318553792),
(u'think', 2.2960891065828037),
(u'people', 2.4230429756217542)]
```



ROC / AUC SCORE: 0.72

K-MEANS CLUSTERING

- Hypothesis: There's a natural clustering to review vocabulary. I can use the most descriptive clusters to simplify the model.

```
model = KMeans(n_clusters=4, init='k-means++', max_iter=100, n_init=1,random_state=5)
```

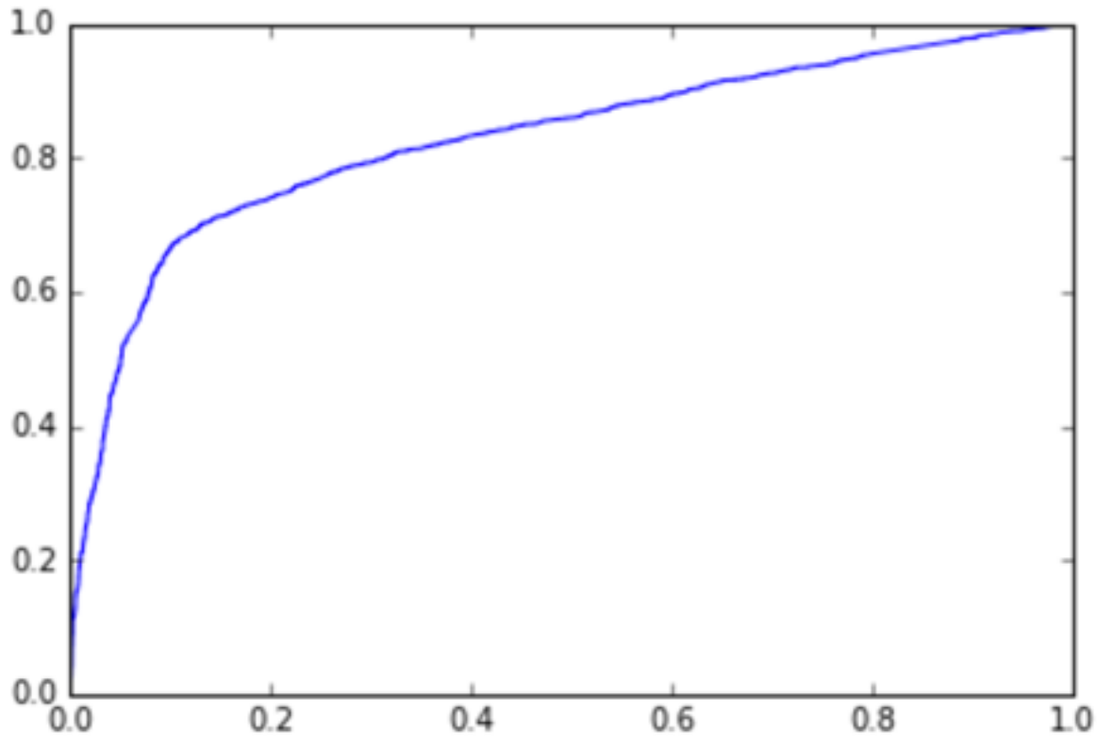
```
vectorizer = TfidfVectorizer(min_df = 0.05, max_df=0.95,
                             ngram_range=(1, 2),
                             stop_words='english')
```

- Select Top 10 words per cluster

LOGISTIC REGRESSION

- Features: Top cluster words + Product score

```
gs = grid_search.GridSearchCV(
    estimator=LogisticRegression(),
    param_grid={'C': [10**-i for i in range(-5, 5)], 'class_weight': [None, 'balanced']},
    cv=cross_validation.StratifiedKFold(df1.Helpful,n_folds=10),
    scoring='roc_auc'
)
```



ROC / AUC SCORE: 0.82

(Bag of words ROC / AUC: 0.72)

LOGISTIC REGRESSION

- Best coefficients. Can interpret better than bag of words.

```
(u'price', 1.2031181875427359),
(u'great', 1.3016982468873803),
(u'food', 1.4150144194030678),
(u'flavor', 1.5746769214376584),
(u'cup', 1.9186985579466715),
(u'br br', 2.196487516457194 6)]
```

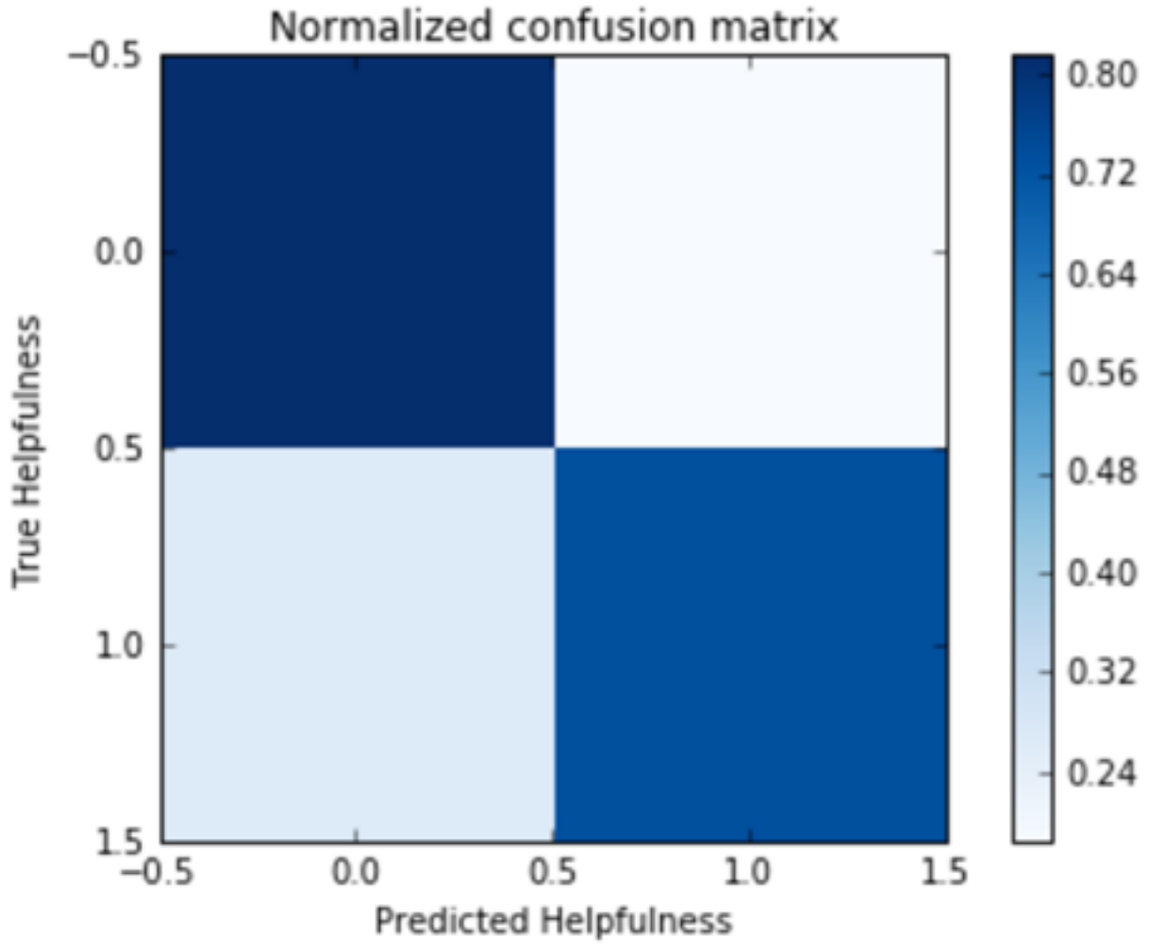
- My html parser isn't working :-/

```
(u'drink', -1.3446412638531735),
```

LOGISTIC REGRESSION

- Confusion Matrix
- Helpful reviews are being predicted as unhelpful.

Optimized against false positives



IMPLICATIONS

- Offer customers a guide or tips when writing reviews
  - "Describe the flavor"
  - "What's the value for the price you paid?"
- Need to understand the features, customer behavior, and domain better. More "unhelpful" reviews.
- Alternative methods: LDA for top "topical" words, Random Forest

QUESTIONS TO EXPLORE

- Is there natural clustering? Examine silhouette coefficient
- Binarize score - are reviews with score of 1 or 5 more helpful?
- Do customers tend to vote on reviews more when reviews are helpful?
- What should be defined as “unhelpful”?

ACKNOWLEDGEMENTS

- Thanks Sri!
- Thanks Dan!
- Thanks Kaggle!
- Thanks Amazon!