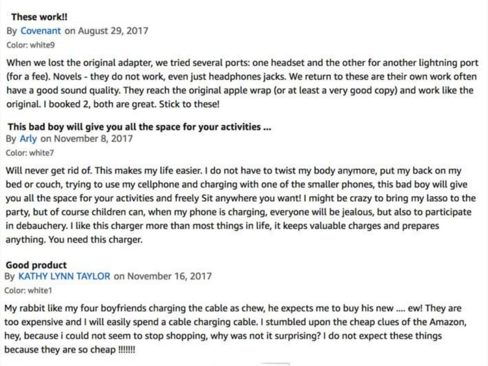
Predicting Amazon Product Ratings

By: Prince Wang, Benjamin Phang, He Jeff, Rustum Chhor

Our Question:

Can we create a model to accurately predict a 5-star scale rating from the review?



These work!!
By Covenant on August 29, 2017
Color: white9

When we lost the original adapter, we tried several ports: one headset and the other for another lightning port (for a fee). Novels - they do not work, even just headphones jacks. We return to these are their own work often have a good sound quality. They reach the original apple wrap (or at least a very good copy) and work like the original. I booked 2, both are great. Stick to these!

This bad boy will give you all the space for your activities ...
By Arly on November 8, 2017
Color: white7

Will never get rid of. This makes my life easier. I do not have to twist my body anymore, put my back on my bed or couch, trying to use my cellphone and charging with one of the smaller phones, this bad boy will give you all the space for your activities and freely Sit anywhere you want! I might be crazy to bring my lasso to the party, but of course children can, when my phone is charging, everyone will be jealous, but also to participate in debauchery. I like this charger more than most things in life, it keeps valuable charges and prepares anything. You need this charger.

Good product
By KATHY LYNN TAYLOR on November 16, 2017
Color: white1

My rabbit like my four boyfriends charging the cable as chew, he expects me to buy his new .... ew! They are too expensive and I will easily spend a cable charging cable. I stumbled upon the cheap clues of the Amazon, hey, because i could not seem to stop shopping, why was not it surprising? I do not expect these things because they are so cheap !!!!!!

Motivation

What can you get from Amazon Reviews?

1. Current product review systems depend heavily on the 5-star system but do not have a fail safe to see if ratings are accurate to their review.

2. Improve categorical assignment of satisfaction with a services or products that can be used in other areas which do not have a 5 star system.

3. Gain insight into customers' concerns without manuelly reading all reviews, which helps identify areas where a product can improve

Our Data

1 We began with a dataset of 413,839 entries. The columns were: Product Name, Brand Name, Price, Rating., Reviews, Review Votes

	Product Name	Brand Name	Price	Rating	Reviews	Review Votes
0	"CLEAR CLEAN ESN" Sprint EPIC 4G Galaxy SPH-D7...	Samsung	199.99	5	I feel so LUCKY to have found this used (phone...	1.0
1	"CLEAR CLEAN ESN" Sprint EPIC 4G Galaxy SPH-D7...	Samsung	199.99	4	nice phone, nice up grade from my pantach revu...	0.0
2	"CLEAR CLEAN ESN" Sprint EPIC 4G Galaxy SPH-D7...	Samsung	199.99	5	Very pleased	0.0
3	"CLEAR CLEAN ESN" Sprint EPIC 4G Galaxy SPH-D7...	Samsung	199.99	4	It works good but it goes slow sometimes but i...	0.0
4	"CLEAR CLEAN ESN" Sprint EPIC 4G Galaxy SPH-D7...	Samsung	199.99	4	Great phone to replace my lost phone. The only...	0.0
5	"CLEAR CLEAN ESN" Sprint EPIC 4G Galaxy SPH-D7...	Samsung	199.99	1	I already had a phone with problems... I know ...	1.0
6	"CLEAR CLEAN ESN" Sprint EPIC 4G Galaxy SPH-D7...	Samsung	199.99	2	The charging port was loose. I got that solder...	0.0
7	"CLEAR CLEAN ESN" Sprint EPIC 4G Galaxy SPH-D7...	Samsung	199.99	2	Phone looks good but wouldn't stay charged, ha...	0.0
8	"CLEAR CLEAN ESN" Sprint EPIC 4G Galaxy SPH-D7...	Samsung	199.99	5	I originally was using the Samsung S2 Galaxy f...	0.0
9	"CLEAR CLEAN ESN" Sprint EPIC 4G Galaxy SPH-D7...	Samsung	199.99	3	It's battery life is great. It's very responsi...	0.0

2 After cleaning NA values in the review and price columns, we were left with 395,725 rows of data to work with.

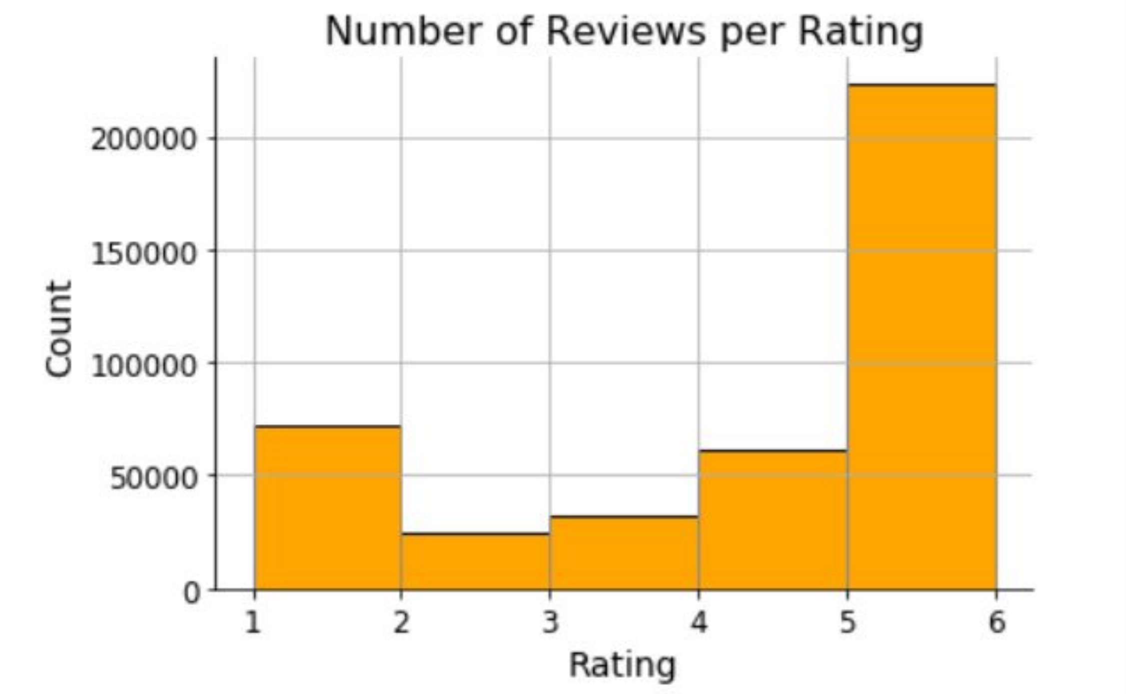
3 We cleaned punctuation using the C based translations function:

```
cleaned_data['Reviews'] = '|aa'.join(cleaned_data['Reviews'].tolist()).translate(translations).split('|aa')
```

4 We created Stop words using two methods: one using NLTK stop words, and another using our own defined stop words

Exploratory Data Analysis

What does our data look like?



68.9%

284,997 4-5 Star Reviews

-Heavily skewed

- the majority of customers are very satisfied with their product.

97,078 1-2 Star Reviews

23.4%

EDA (cont.)

```
In [29]: #word count of all comments
         len(transformed_comment)
         #previously before cleaning it was 82m words, now reduced to 52m

Out[29]: 52032256
```

Our Approach

Modeling:

Gather and Clean Data

Delete NA values and punctuation

Data-cleaning

Removing punctuation and removing stop words - came up with canonical method that involved selecting common words in both 5& 1 star ratings since they have no bearing on the outcome.

Stopword optimization

Running a rudimentary model and deciding on stopwords to remove based on accuracy of model.

Modelling

Used a multi-label classifier on a one vs rest classifier with logistic regression to decide the semantics of words. Started with basic true-false training but moved on to forced selection off the probabilities to ensure a prediction is output.

Further modelling

Trained said model on both individual words and sentences from ratings. Tested accuracy at 77.7%.

Results

Created database of words with one-star to five-star connotations and function that produces the model's output for any review input.

Insights:

Filtering out Stop Words

Generate a list of stop words and filter it out from the comments

Model Construction

Construct a Word2Vec model using the skip-gram method

Word2Vec

Using Word2Vec to figure out a proximity model of words that are closest to each other.

Optimize Word2Vec

Tune the model so it generates the best result which resembles each word most accurately

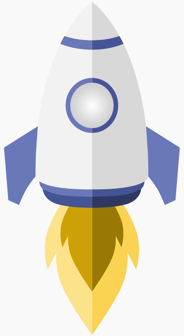
Results

interpreting the relations between words in a E-commerce context.

Rating Prediction Model

Step 1: Preprocessing

- Removal of punctuation
- Removal of stopwords
  - Lots of online libraries like NLTK but instead took a different approach more customized to our data.
  - Created function that collected the top n number of similar words in both five-star and one-star rating categories. The logic is that if these words commonly appear in both categories, they ultimately have little bearing on the outcome of the rating.



Step 2: Rudimentary modelling

- Since we're working with similar data input (all one string-based input) and five different categories as an output. Decided to use a OnevsRestClassifier with logistic regression as a means to predict the rating of different reviews.
- Used a 70/30 Train/Test split on data
- Initially trained model using true/false Boolean. For example, when the model trains on 5-star reviews, the dataframe it learns from would have all 5-star reviews as true, whilst rest are false. Model would then output percentage probability of true/false for each 5 categories to categorize new reviews.

Modeling

Step 3: Stopword optimization

- Since our method is customized to our database. We decided to find a way to optimize the list of stopwords so we weren’t over/underfiltering.
- This involved running our model over 200 different scenarios from 0 stopwords removed to 199 stopwords removed and then testing its accuracy.
- The result was 13 stopword yielding the highest accuracy of ~77.7%. These stopwrods are: ['the', 'I', 'and', 'phone', 'a', 'it', 'to', 'is', 'for', 'this', 'of', 'in', 'was', 'that']

```
0
0.775290921725946
1
0.7761475772316634
2
0.7761602122686209
3
0.7761450502242719
4
0.77644576410386
5
0.7764533451260345
6
0.7767161538947501
7
0.7767161538947501
8
0.7766833027986607
9
0.7764937772442985
```

```
10
0.7769991787225977
11
0.7769991787225977
12
0.7769739086486828
13
0.7772316634026154
14
0.7772316634026154
15
0.7770749889443427
16
0.7770876239813002
17
0.7772316634026154
18
0.7754577042137848
```

Step 4: Refining the model
Removed the 13 stopwords from the last step from the data
Previous model using in-built predict function only output data that it’s sure about >50% threshold. So while accuracy score of guesses was high at ~85%. Lots of predictions were missing.
Changed model to train on direct values (1..5 ratings) versus previous true/false for each 1..5 category
Changed the model to output prediction probabilities instead of a single value so it’ll output the probability a review is for 1..5 ratings. Then it would select the maximum to produce a rating for each prediction. E.g. (‘bad’: 1:0.6, 2:0.1, 3:0.1, 4:0.1, 5:0.1) -> output: 1
Trained model on both sentences and individual words to check for differences. Marginal differences so stuck with training with sentences.

Conclusion

Our overall success.

Step 5: Creating databases and functions

- Collated all the data for each individual unique word. So we now have databases of words with 1-star all the way to 5-star ratings. This makes it easy to learn what features/words are associated with good/bad ratings.
- Creation of a predict function that allows a user to input any rating he likes and the model will output a prediction of that rating. Could be useful for testing out new ideas or understanding what users like/dislike in products.

```
In [60]: print(predict("this is a terrible phone"))
         1

In [61]: print(predict("I like this phone a lot"))
         5

In [64]: print(predict("this is okay. It's decent"))
         3

In [65]: print(predict('It works good but it goes slow sometimes. This is a great phone to replace my lost phone however the cha
         4

In [66]: print(predict('this sucks.'))
         1

In [67]: l go into settings to adjust. Other than that, it does the job until I am eligible to upgrade my phone again.Thaanks!'))
         4
```

	word	predictions
29	displeased	1
69	situation	1
108	suffered	1
181	aftermarketed	1
191	CHEAPLY	1
199	BADIm	1
269	Greatly	1
352	Hell	1
367	REFURBISHED	1

```
In [38]: five_w
```

2	1Mbps	5
3	howeverWith	5
5	dly	5
6	brighten	5
7	WiFilve	5
8	tipoMi	5
9	imposibble	5
10	nextbit	5
12	expectationsAt	5
14	TVMovies	5
15	Sounds	5

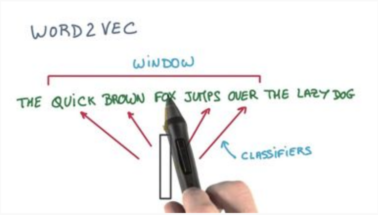
TensorFlow Word2Vec Model

How we used Tensorflow in our first model.

Model implemented through Tensorflow

Converted all comments into a list of words

Feed it to the Word2Vec Model



Reviews
feel lucky found used us used hard line someo...
nice nice grade pantach revue clean set easy s...
pleased
works good goes slow sometimes good love
great replace lost thing volume button work st...

Reviews
I feel so LUCKY to have found this used (phone...
nice phone, nice up grade from my pantach revu...
Very pleased
It works good but it goes slow sometimes but i...
Great phone to replace my lost phone. The only...
I already had a phone with problems... I know ...

```
'two',
'days',
'ordering',
'little',
'blemish',
'sidebut',
'cares',
```

```
['feel',
 'lucky',
 'found',
 'used',
 'us',
```

Conclusion

What insight this gave us.

- Word2Vec Result: Gaining insight through keywords
  - the word closest to “buy”: used, cheap, great, last(long?)
  - the word that is closest to “good” include “1080P”, “Crystal”, “Perfectly”, “big
    - sellers are able to observe which elements of the phone is most valued by the customers
    - which properties of the phones are in need of improvement

```
Nearest to easy: refined, assets, ring, ghana, weaker
Nearest to little: epic, octacore, browsed, UNK, est,
```

```
Nearest to screen: UNK, battery, camera, back, laser,
Nearest to works: work, UNK, great, worked, working,
```

Thanks for Watching

Any questions?