

Final Review

DSE 220

Task

- Dataset – Amazon Product Reviews
- Given a review and its features (item ID, reviewer ID, price of item, category of item etc.), predict how many users will find the review to be helpful
- Format - Kaggle Competition
- Detailed instructions are available both on Kaggle and Class GitHub
- Deadlines
 - Kaggle competition: **11 June 2017 4:59 PM PST**
 - Submitting code to GitHub: **12 June 2017 11:59 PM PST**
 - Submitting report to Gradescope: **12 June 2017 11:59 PM PST**

Task

★★★★★ 5 Stars

By [Amazon Customer](#) on March 31, 2017

Size: 128GB | Color: Black | **Verified Purchase**

Great product, looks like brand new, and works perfect.

The packaging is lame, but the cellphone (what trully matters) is perfect, and it cost a hundred bucks less, so, its really convenient.
5 stars.

[Comment](#)

12 people found this helpful. Was this review helpful to you?

Yes

No

[Report abuse](#)

Task

```
{'categoryID': 0,  
  'categories': [['Clothing, Shoes & Jewelry', 'Wedding Party Gifts', 'Gifts for  
the Groomsmen'], ['Clothing, Shoes & Jewelry', 'Women'], ['Clothing, Shoes  
& Jewelry', 'Novelty, Costumes & More', 'Novelty']],  
  'itemID': 'I408260822',  
  'reviewerID': 'U441384838',  
  'rating': 3.0,  
  'reviewText': "it's better on a man's hand.I didn't find it very feminine against  
mine. It's a good price, came in a timely fashion but i prefer a thinner band",  
  'reviewHash': 'R345042616',  
  'reviewTime': '05 13, 2014',  
  'summary': 'Good price but...',  
  'price': 19.99,  
  'unixReviewTime': 1399939200,  
  'helpful': {'outOf': 2, 'nHelpful': 2}}
```

Given other fields, predict **'nHelpful'**

Task

- Scoring metric – Mean Absolute Error (MAE)
- Absolute difference between predicted value and actual nHelpful value
- Train/Test data is available on Kaggle
- Baseline.ipynb – A sample jupyter notebook that tells you how to parse data, put it into a Pandas Dataframe and create a submission file
- Grading details are provided in the PDF on GitHub

Kaggle

- Platform for predictive modelling and analytics competitions
- Acquired by Google in March 2017
- People upload their solutions and see their ranks in real time

Kaggle

- Sign up on Kaggle using your @ucsd.edu email id [**not @eng.ucsd**]
- You may choose any username you like so as to stay anonymous
- Your rank on public leaderboard is based on 50% of the test data
- Upon completion of competition, we evaluate you on the basis of private leaderboard – which is based on remaining 50% of test data
- Competition link - <https://inclass.kaggle.com/c/dse-220-final>